



AISV

Associazione Italiana di Scienze della Voce



# AISV 2007

## 4° Convegno Nazionale

AISV - Associazione Italiana di Scienze della Voce

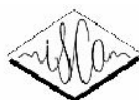
### “LA FONETICA SPERIMENTALE”

Metodo e Applicazioni

a cura di

Luciano Romito  
Vincenzo Galatà  
Rosita Lio

CD-rom incluso

Università della Calabria  
Arcavacata di Rende (CS) - Centro Congressi  
3-5 Dicembre 2007international speech  
communication association

UNIVERSITÀ DELLA CALABRIA



CAMPUS DI ARCAVACATA

Via P.Bucci  
87036 Arcavacata di Rende (CS)**L@bphon**Laboratorio di Fonetica  
Dipartimento di Linguistica – Cubo 20/A  
Università della Calabria



# **“LA FONETICA SPERIMENTALE”**

**Metodo e Applicazioni**

## **Atti del 4° Convegno Nazionale AISV 2007**

a cura di  
**Luciano Romito  
Vincenzo Galatà  
Rosita Lio**

**Università della Calabria  
Arcavacata di Rende (CS)  
Centro Congressi  
3-5 Dicembre 2007**



**international speech  
communication association**

UNIVERSITÀ DELLA CALABRIA



CENTRO DI ARCAVACATA

Via P.Bucci  
87036 Arcavacata di Rende (CS)

**L@bphon**

Laboratorio di Fonetica  
Dipartimento di Linguistica – Cubo 20/A  
Università della Calabria

“LA FONETICA SPERIMENTALE” - Metodo e Applicazioni

Atti del 4° Convegno Nazionale AISV 2007  
Centro Congressi, 3-5 Dicembre 2007  
Università della Calabria, Arcavacata di Rende

a cura di  
*Luciano Romito*  
*Vincenzo Galatà*  
*Rosita Lio*

ISBN: .....

Copyright © 2009 by  
Via...  
..... Padova



# INDICE

|                                                                                                                                 |            |
|---------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>INDICE .....</b>                                                                                                             | <b>I</b>   |
| <b>ORGANIZZAZIONE CONVEGNO .....</b>                                                                                            | <b>V</b>   |
| ORGANIZZAZIONE GENERALE: .....                                                                                                  | V          |
| ORGANIZZAZIONE LOCALE: .....                                                                                                    | V          |
| COMITATO SCIENTIFICO: .....                                                                                                     | V          |
| PROGRAMMA CONVEGNO: .....                                                                                                       | V          |
| CURA ATTI: .....                                                                                                                | V          |
| VALUTAZIONE ABSTRACTS: .....                                                                                                    | VI         |
| REVISIONE ARTICOLI: .....                                                                                                       | VI         |
| SEGRETERIA CONVEGNO AISV2007: .....                                                                                             | VI         |
| WEBMASTER: .....                                                                                                                | VI         |
| REDAZIONE ATTI E REALIZZAZIONE CD-ROM: .....                                                                                    | VI         |
| <b>“LA FONETICA SPERIMENTALE” METODO E APPLICAZIONI .....</b>                                                                   | <b>VII</b> |
| PREMESSA .....                                                                                                                  | VII        |
| TAVOLA ROTONDA .....                                                                                                            | VII        |
| CONCERTO DI VOCI .....                                                                                                          | XII        |
| SPAZIO “A(D)DOTTA UN DOTTORANDO!” .....                                                                                         | XII        |
| SPAZIO “DEMO” .....                                                                                                             | XIII       |
| PREMIO FRANCO FERRERO .....                                                                                                     | XIII       |
| RINGRAZIAMENTI .....                                                                                                            | XIV        |
| <b>FONETICA.....</b>                                                                                                            | <b>1</b>   |
| PROSPEZIONI SULLA STRUTTURA RITMICA DELL’ITALIANO BASATE SUL<br>CORPUS SEMISPONTANEO AVIP/API.....                              | 3          |
| <i>Chiara Bertini &amp; Pier Marco Bertinetto</i>                                                                               |            |
| SULLA REALTÀ ACUSTICA DELL’ACCENTO LESSICALE IN ITALIANO ED IN<br>SPAGNOLO: LA DURATA VOCALICA IN PRODUZIONE E PERCEZIONE ..... | 22         |
| <i>Iolanda Alfano, Renata Savy, Joaquim Llisterri</i>                                                                           |            |

|                                                                                                                                                                      |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| L'ACQUISIZIONE DEL TRATTO DI LUNGHEZZA CONSONANTICA IN APPRENDENTI L'ITALIANO L2.....                                                                                | 40  |
| <i>Patrizia Sorianello</i>                                                                                                                                           |     |
| UN CONTRIBUTO PER IL CONFRONTO TRA L'INTONAZIONE VENETA E QUELLA ANDALUSA .....                                                                                      | 62  |
| <i>Antonio Romano &amp; Renzo Miotti</i>                                                                                                                             |     |
| TEMPI DI ADATTAMENTO IN ELETTROPALATOLOGRAFIA: PRIMI DATI ARTICOLATORI, ACUSTICI E PERCETTIVI .....                                                                  | 77  |
| <i>Chiara Bertini, Irene Ricci</i>                                                                                                                                   |     |
| L'ORGANIZZAZIONE TEMPORALE DEI GESTI VOCALICI E CONSONANTICI NELLE CONSONANTI SCEMPIE E GEMINATE DELL'ITALIANO .....                                                 | 89  |
| <i>Claudio Zmarich, Barbara Gili Fivela, Pascal Perrier, Christophe Savariaux, Graziano Tisato</i>                                                                   |     |
| LÀ DOVE IL [ʃi] SUONA. UN'ANALISI PARAMETRICA DELLE SIBILANTI NELL'ITALIANO DI CARRARA. ....                                                                         | 105 |
| <i>Massimiliano Barbera, Marlen Barth, Diego Frassinelli, Gabriella Lapesa</i>                                                                                       |     |
| TRASCRIZIONE DI SEGNALI RUMOROSI.....                                                                                                                                | 126 |
| <i>Andrea Paoloni, Marco Medda, Adriano Soccorsi</i>                                                                                                                 |     |
| VOCALI ACCENTATE BREVI PRODOTTE CON INCREMENTO DI F0: COME E PERCHÉ? .....                                                                                           | 149 |
| <i>Arianna Uguzzoni</i>                                                                                                                                              |     |
| L'ETÀ DELLA VOCE .....                                                                                                                                               | 165 |
| <i>Antonella Giannini - Massimo Pettorino</i>                                                                                                                        |     |
| LA PREASPIRAZIONE NELLE LINGUE DEL NORD EUROPA E IN VARIETÀ DI ITALIANO: DATI DI BASE PER UN CONFRONTO TRA ISLANDESE E SENESE .....                                  | 179 |
| <i>P. Mairano, A. Romano, M. Stevens e J. Hajek</i>                                                                                                                  |     |
| L'INSIEME È LA SOMMA DELLE PARTI CONSIDERAZIONI PRELIMINARI SU DATI ARTICOLATORI DELLA LINGUA OTTENUTI TRAMITE ULTRASUONI ED ARTICULOGRAFO ELETTROMAGNETICO 3D ..... | 196 |
| <i>Mirko Grimaldi</i>                                                                                                                                                |     |
| IL FRANCESE COME LINGUA STRANIERA: L'ASSIMILAZIONE DEL TRATTO DI SONORITÀ.....                                                                                       | 215 |
| <i>Sonia d'Apolito, Barbara Gili Fivela</i>                                                                                                                          |     |
| SVILUPPO FONETICO E ACQUISIZIONE LESSICALE: L'EMERGERE DELLA FONOLOGIA DELL'ITALIANO IN UNA PROSPETTIVA INTERLINGUISTICA .....                                       | 242 |
| <i>Claudio Zmarich, Debora Stocco, Sara Giulivi, Mario Vayra</i>                                                                                                     |     |
| L'INTONAZIONE NELL'ITALIANO DELL'AREA LECCESE: PRIME OSSERVAZIONI DAL PUNTO DI VISTA AUTOSEGMENTALE-METRICO .....                                                    | 259 |
| <i>Antonio Stella, Barbara Gili Fivela</i>                                                                                                                           |     |
| EMOVO, DATABASE DI PARLATO EMOTIVO PER L'ITALIANO .....                                                                                                              | 293 |
| <i>Iacopo Iadarola</i>                                                                                                                                               |     |

|                                                                                   |            |
|-----------------------------------------------------------------------------------|------------|
| LA COMPOSIZIONE FONETICA DEL PRIMO VOCABOLARIO DEL BAMBINO                        | 324        |
| <i>Claudio Zmarich, Marco Dispaldro, Pasquale Rinaldi, Maria Cristina Caselli</i> |            |
| ETICHETTATURA FONETICA E TTS                                                      | 337        |
| <i>Pier Luigi Salza, Claudia Barolo, Beata Dobrzynska</i>                         |            |
| <b>DIALETTOLOGIA</b>                                                              | <b>348</b> |
| LE OCCLUSIVE PALATALI NEL DIALETTO DI SAN GIOVANNI IN FIORE (CS)                  | 349        |
| <i>Biagio Mele, Stephan Schmid</i>                                                |            |
| NUOVI DATI FONETICI E DIALETTOLOGICI AI MARGINI DELL' AREA                        |            |
| LAUSBERG: LE VARIETÀ DI ALIANO E ALIANELLO                                        | 372        |
| <i>Francesco Avolio &amp; Antonio Romano</i>                                      |            |
| IL PROGETTO BOMARZO                                                               | 405        |
| <i>Amedeo De Dominicis e Pamela Mattana</i>                                       |            |
| LA RETROFLESSIONE IN ALCUNI CENTRI DELLA PROVINCIA DI                             |            |
| CATANZARO: VERSO UNA MAPPA REGIONALE                                              | 412        |
| <i>Luciano Romito &amp; Paola Antonietta Scuticchio</i>                           |            |
| LA METAFONIA IN ALCUNI CENTRI DEL NORD CALABRIA: VERSO UNA                        |            |
| MAPPA REGIONALE                                                                   | 423        |
| <i>Luciano Romito &amp; Daniela Gagliardi</i>                                     |            |
| <b>FONOLOGIA</b>                                                                  | <b>437</b> |
| "GENERATIVO": DESIGNAZIONI E CONTENUTI DI UN TERMINE NEI MODELLI                  |            |
| FONOLOGICI DAL SECONDO NOVECENTO A OGGI                                           | 438        |
| <i>Manuela Lo Prejato</i>                                                         |            |
| <b>SIGNAL PROCESSING &amp; PATTERN RECOGNITION</b>                                | <b>450</b> |
| UN SISTEMA MULTIMODALE PER LA SEGMENTAZIONE DI TELEGIORNALI                       |            |
| BASATO SULL'INDIVIDUAZIONE AUTOMATICA DEGLI SPEAKER                               | 451        |
| <i>Leandro D'Anna, Gennaro Percannella, Carlo Sansone, Mario Vento</i>            |            |
| EFFETTO DEL RUMORE AMBIENTALE SULLA STIMA DELL'SNR NELLE                          |            |
| SOTTOBANDE DI INTERESSE DEGLI INDICI BIOMETRICI NELLE                             |            |
| COMPARAZIONI FONICHE                                                              | 469        |
| <i>Francesco Beritelli</i>                                                        |            |
| SVILUPPO DI UN SISTEMA DI <i>KEYWORD SPOTTING</i> PER L'INDICIZZAZIONE            |            |
| AUTOMATICA DEI DOCUMENTI AUDIO                                                    | 481        |
| <i>Graziano Tisato, Piero Così, Isabella Gagliardi</i>                            |            |
| SPERIMENTAZIONE DI SISTEMI SENZA ADDESTRAMENTO PER                                |            |
| L'ESTRAZIONE DI FONEMI VOCALICI                                                   | 501        |
| <i>Francesco Borchì, Monica Carfagni, Matteo Nunziati</i>                         |            |

|                                                                                                                                 |            |
|---------------------------------------------------------------------------------------------------------------------------------|------------|
| SVILUPPO DI UN SISTEMA DI RICONOSCIMENTO PER L'ARABO: PROBLEMI E SOLUZIONI .....                                                | 517        |
| <i>Piero Cosi, Mauro Nicolao, Giacomo Sommovilla, Graziano Tisato</i>                                                           |            |
| IMPLEMENTAZIONE DI UN ALGORITMO DI BSS IN FREQUENZA PER LA SEPARAZIONE DI SORGENTI ACUSTICHE .....                              | 531        |
| <i>Francesco Nesta, Maurizio Omologo, Piergiorgio Svaizer</i>                                                                   |            |
| <b>NATURAL LANGUAGE PROCESSING .....</b>                                                                                        | <b>543</b> |
| SPEECH AUTOMATION NEL CONTESTO DELLE ARCHITETTURE DI SERVIZIO DEI CONTACT CENTER .....                                          | 544        |
| <i>Alessandro Bruni, Gianpaolo Coro</i>                                                                                         |            |
| <b>ACUSTICA &amp; RESTAURO .....</b>                                                                                            | <b>559</b> |
| UN PROTOCOLLO DELLE PROCEDURE DI RESTAURO ALL'INTERNO DELL'ARCHIVIO SONORO CALABRESE.....                                       | 560        |
| <i>Luciano Romito, Vittorio Scullari</i>                                                                                        |            |
| <b>LOGOPEDIA, FONIATRIA &amp; AUDIOLOGIA.....</b>                                                                               | <b>574</b> |
| EVALUATION OF ELECTRO-ACOUSTIC PARAMETERS OF THE VOICE PRE AND POST ADENO-TONSILLECTOMY IN CHILDREN.....                        | 575        |
| <i>Nicola Lombardo, Eugenia Allegra, Alessandro La Boria, Aldo Garozzo</i>                                                      |            |
| CARATTERISTICHE DI PERSONALITA' E MODALITA' DI UTILIZZO DELLA VOCE NEI SOGGETTI CON PATOLOGIA VOCALE.....                       | 584        |
| <i>Nicola Angelillo, Brigida Di Costanzo, Assunta Fiorillo, Umberto Barillari</i>                                               |            |
| DEFICIT DELLE COMPETENZE SEMANTICO-PRAGMATICHE IN SOGGETTI NORMODUDENTI ED IPOACUSICI.....                                      | 597        |
| <i>Nicola Angelillo, Anna Vitiello, Brigida Di Costanzo, Umberto Barillari</i>                                                  |            |
| IL SISTEMA MDVP: SOGLIE DI NORMALITÀ A CONFRONTO.....                                                                           | 606        |
| <i>Antonio Mendicino, Vincenzo Galatà, Maria Tucci, Rosita Lio, Luciano Romito</i>                                              |            |
| <b>FONETICA FORENSE.....</b>                                                                                                    | <b>620</b> |
| VERSO UN FORMATO STANDARD NELLE INTERCETTAZIONI E UNA PROPOSTA PER L'ARCHIVIAZIONE E LA CONSERVAZIONE DELLE REGISTRAZIONI ..... | 621        |
| <i>Luciano Romito, Maria Tucci</i>                                                                                              |            |
| SPECIFICHE LINGUISTICHE DEL DATABASE UTILIZZATO PER LO SPEAKER RECOGNITION IN S.M.A.R.T. ....                                   | 632        |
| <i>Luciano Romito, Tommaso Bove, Stefano Delfino, Carla Rossi, Giovanna Jona Lasinio</i>                                        |            |

# ORGANIZZAZIONE CONVEGNO

## ORGANIZZAZIONE GENERALE:

Luciano Romito (Università della Calabria)

## ORGANIZZAZIONE LOCALE:

Rosita Lio (Università della Calabria)  
Vincenzo Galatà (Università della Calabria)  
Maria Tucci (Università della Calabria)  
Daniela Gagliardi (Università della Calabria)  
Paola Antonietta Scuticchio (Università della Calabria)

## COMITATO SCIENTIFICO:

Eleonora Bilotta (Università della Calabria)  
Sergio Canazza Targon (Università degli Studi di Udine)  
Ugo Cesari (Università Federico II, Napoli)  
Piero Così (Istituto di Scienze e Tecnologie della Cognizione, Padova - CNR)  
Francesco Cutugno (Università Federico II, Napoli)  
Amedeo De Dominicis (Università della Toscana)  
Vincenzo Di Nicola (Università di Bari)  
Mauro Falcone (Fondazione Ugo Bordoni)  
Barbara Gili-Fivela (Università degli Studi di Lecce)  
Mirko Grimaldi (Università degli Studi di Lecce)  
Giovanna Marotta (Università degli Studi di Pisa)  
Maurizio Omologo (ITC-irst)  
Andrea Paoloni (Fondazione Ugo Bordoni)  
Massimo Pettorino (Università L'Orientale Napoli)  
Massimo Privitera (Università della Calabria)  
Luciano Romito (Università della Calabria)  
Pier Luigi Salza (Loquendo S.p.A.)  
Renata Savy (Università degli Studi di Salerno)  
Stephan Schmid (Università di Zurigo)  
Graziano Tisato (Istituto di Scienze e Tecnologie della Cognizione, Padova)  
Claudio Zmarich (Istituto di Scienze e Tecnologie della Cognizione, Padova)

## PROGRAMMA CONVEGNO:

Luciano Romito (Università della Calabria)  
Antonio Mendicino (Università della Calabria)

## CURA ATTI:

Luciano Romito (Università della Calabria)  
Vincenzo Galatà (Università della Calabria)  
Rosita Lio (Università della Calabria)

## **VALUTAZIONE ABSTRACTS:**

Sergio Canazza Targon (Università degli Studi di Udine)  
Ugo Cesari (Università Federico II, Napoli)  
Francesco Cutugno (Università Federico II, Napoli)  
Amedeo De Dominicis (Università della Tuscia)  
Vincenzo Di Nicola (Università di Bari)  
Barbara Gili-Fivela (Università degli Studi di Lecce)  
Mirko Grimaldi (Università degli Studi di Lecce)  
Giovanna Marotta (Università degli Studi di Pisa)  
Maurizio Omologo (ITC-irst)  
Andrea Paoloni (Fondazione Ugo Bordoni)  
Massimo Pettorino (Università L'Orientale Napoli)  
Luciano Romito (Università della Calabria)  
Pier Luigi Salza (Loquendo S.p.A.)  
Renata Savy (Università degli Studi di Salerno)  
Stephan Schmid (Università di Zurigo)  
Graziano Tisato (Istituto di Scienze e Tecnologie della Cognizione, Padova )  
Claudio Zmarich (Istituto di Scienze e Tecnologie della Cognizione, Padova)

## **REVISIONE ARTICOLI:**

Francesco Cutugno (Università Federico II, Napoli)  
Amedeo De Dominicis (Università della Tuscia)  
Barbara Gili-Fivela (Università degli Studi di Lecce)  
Mirko Grimaldi (Università degli Studi di Lecce)  
Giovanna Marotta (Università degli Studi di Pisa)  
Maurizio Omologo (ITC-irst)  
Andrea Paoloni (Fondazione Ugo Bordoni)  
Luciano Romito (Università della Calabria)  
Renata Savy (Università degli Studi di Salerno)  
Stephan Schmid (Università di Zurigo)  
Claudio Zmarich (Istituto di Scienze e Tecnologie della Cognizione, Padova)

## **SEGRETERIA CONVEGNO AISV2007:**

Rosita Lio, Vincenzo Galatà  
L@bphon - Laboratorio di Fonetica - c/o Dipartimento di Linguistica  
Via P.Bucci - Cubo 20/A - 87036 Arcavacata di Rende (CS)  
Tel. (+39) 0984/494078 - Fax (+39) 0984/494114  
E-mail: [labfonuc@unical.it](mailto:labfonuc@unical.it) - WEB: <http://www.linguistica.unical.it/aisv2007/>

## **WEBMASTER:**

Vincenzo Galatà (Università della Calabria)

## **REDAZIONE ATTI E REALIZZAZIONE CD-ROM:**

Vincenzo Galatà (Università della Calabria)

# “LA FONETICA SPERIMENTALE” METODO E APPLICAZIONI

## PREMESSA

Il 4° convegno AISV ha avuto come momento aggregante la Fonetica Sperimentale e l'interdisciplinarietà dell'approccio metodologico tra tutte le discipline che si occupano di VOCE. Tale scelta aveva come scopo quello di stimolare un proficuo scambio di opinioni, un'attiva condivisione dei risultati raggiunti e anche di una possibile progettualità di intenti. L'uso di “Metodo” utilizzato al singolare rispetto ad “Applicazioni” al plurale, nel titolo della Tavola Rotonda, è stata una scelta dell'organizzazione. Un *unico* metodo utilizzato ed applicato per scopi anche molto differenti tra loro quali la sintesi vocale, la diagnosi di disturbi della voce, lo speech recognition, la prevenzione di eventuali patologie nascenti ecc... Allo stesso modo, lo studio di discipline quali la dialettologia, l'apprendimento di L2 o la linguistica forense, hanno trovato nel *metodo* un motivo di rinvolgimento.

## TAVOLA ROTONDA

La tavola rotonda dal titolo “*La Fonetica Sperimentale: Metodo e Applicazioni*” ha avuto come obiettivo quello di valorizzare e promuovere la Fonetica Sperimentale mettendo in evidenza l'interdisciplinarietà dell'approccio metodologico.

Vi hanno preso parte Emanuela Magno Caldognetto (ISTC-CNR, Padova) che, oltre ad essere moderatore della discussione, nel suo ruolo di linguista, ha avuto il compito di presentare la ricerca sullo studio e l'interpretazione delle emozioni sia dal punto di vista acustico sia sotto il profilo gestuale; Sergio Canazza Targon (Università degli Studi di Udine), con il compito più tecnico, ha parlato di restauro e di conservazione di corpora vocali; Nicola Lombardo (Clinica ORL, Università Magna Graecia – CZ) si è occupato della diagnosi e prevenzione di disturbi della voce attraverso l'analisi fonetica del parlato e, infine, Pierluigi Salza (Loquendo S.p.A. - Torino), è intervenuto sull'applicazione di tecniche basate sull'analisi fonetica del parlato, di speech recognition e di sintesi con finalità di mercato. Ciascuno dei relatori ha rappresentato il peso e l'importanza della Fonetica Sperimentale nel proprio ambito di ricerca/lavoro.

**Sergio Canazza Targon** ha aperto il confronto con un intervento di “Ri-mediazione dei documenti sonori” introducendo la platea alla molteplicità di trasformazioni che un documento sonoro subisce durante i suoi passaggi da un *medium* ad un altro (si pensi ad esempio al passaggio da analogico a digitale). Dopo una breve introduzione e contestualizzazione della tematica in ambito europeo, il relatore ha posto l'accento sull'interesse che molte comunità scientifiche nutrono per tale tematica presentando i, molti progetti finanziati da programmi europei differenti tra loro, ad esempio il 6° programma quadro e IST che promuovono la ricerca di base sulle scienze ingegneristiche, o progetti come Culture2000 che finanziano programmi culturali e altri ancora finanziati dal museo reale del Belgio. La “ri-mediazione” si basa sostanzialmente sulla corretta conservazione

dei documenti sonori e video. Il relatore ha illustrato alla platea una serie di passaggi relativi alle modalità di conservazione e ri-mediazione di documenti di tipo analogico in un medium di tipo digitale senza trascurare, durante questo processo, le diverse metodologie e tecniche di conservazione e restauro dei documenti sonori. Ha affermato e ribadito il suo punto di vista sulla definizione di “copia conservativa”, intesa come “il conservare l’unità documentale del documento originale garantendo la trasparenza massima del nuovo medium digitale in rapporto al documento originale”. A dimostrazione di quanto affermato, il relatore ha portato all’attenzione del pubblico alcuni esempi sull’insufficienza della semplice e sola digitalizzazione del segnale sonoro, mettendo in evidenza come risulti assolutamente necessario anche conservare tutte le informazioni relative alla custodia del supporto analogico (dove vengono riportate importantissime informazioni sulla registrazione o addirittura sul nastro stesso ecc.). Molto interessanti e rappresentativi sono risultati gli esempi sonori presentati con riferimento ai rumori comunemente riscontrabili in registrazioni provenienti da supporti analogici.

L’intervento di Sergio Canazza è risultato particolarmente interessante e la sua attinenza con il tema della tavola rotonda è stata evidenziata dall’intervento di Piero Così. Quest’ultimo, ricordando l’operato di Franco Ferrero con un accenno alle registrazioni da quest’ultimo effettuate su nastro nell’ambito degli studi elettroglottografici sin dagli anni ’70, ha riportato come la necessità di recuperare e catalogare tale materiale, faccia emergere l’importanza della Fonetica e del suo metodo nelle procedure di *restauro*: per operare qualsiasi tipo di restauro di materiale sonoro contenente registrazioni vocali sono necessarie robuste conoscenze legate alla voce per essere sicuri di non intaccare la sorgente vocale. La collaborazione tra musicologi e restauratori del suono è già ampiamente diffusa e si spera che lo stesso accada presto anche con gli studiosi di voce, dove la Fonetica, con il suo metodo, può sicuramente dare origine a nuove ed altrettanto fruttuose applicazioni.

**Nicola Lombardo** ha messo in evidenza l’utilità della fonetica nel campo della medicina e, nello specifico, nel campo dell’otorinolaringoiatria in cui egli stesso opera. Nell’affrontare il tema, il relatore ha illustrato un progetto di prevenzione dei disturbi della voce portato avanti all’interno dell’U.O. di Otorinolaringoiatria dell’Università Magna Grecia (Catanzaro) con il supporto di informatici e di bioingegneri.

Tale progetto, volto ad uno screening delle disfonie, così come ha sottolineato Lombardo nel suo intervento, si basa su una prassi clinica quotidiana operata dai medici, ovvero l’ascolto della voce dei propri pazienti per definirne la buona o la cattiva condizione. Il relatore ha spiegato a tal proposito che la chirurgia laringologica ha fatto enormi passi avanti nella ricerca e, allo stato attuale, un carcinoma delle corde vocali diagnosticato allo stadio iniziale ha una guarigione del 100%. Ciò conferma che il carcinoma non rappresenta più un problema medico, ma solo un problema di diagnosi precoce. In relazione a questo punto, Nicola Lombardo ha evidenziato tutta una serie di problemi etici e sociali che obbligano il medico all’adozione di una serie di mezzi e di interventi necessari a poter effettuare una diagnosi precoce. È proprio dal convergere di interessi comuni che si inserisce il progetto sullo screening delle disfonie. Il relatore ha sottolineato che quest’ultimo consente al paziente l’invio di campioni di voce direttamente on-line attraverso una procedura implementata sul server dell’Università (allo stato attuale la procedura non è ancora aperta al pubblico) permettendo, quindi, l’analisi delle produzioni inviate dal paziente (analisi di F0, Jitter, Shimmer ecc.). Nicola Lombardo ha più volte ribadito nel suo intervento che la presenza di alterazioni significative e costanti rilevate durante diverse settimane, e quindi non occasionali, costituisce l’indicazione per un dato paziente di approfondire le indagini cliniche con una laringoscopia o altro esame



approfondito, consentendo di diagnosticare precocemente eventuali carcinomi alle corde vocali.

Lombardo si è, inoltre, soffermato sul campo medico-legale, un altro settore in cui la fonetica viene in aiuto ai medici. La laringectomia porta inevitabilmente a temere il rischio di problematiche vocali o di perdita della voce (in passato, purtroppo, diversi pazienti hanno subito danni vocali a seguito di intervento di tiroidectomia). In buona parte del suo intervento, Nicola Lombardo ha trattato quanto, in ambito sanitario, sia molto sentito oggi il problema del risarcimento e delle denunce per contenziosi etichettati sempre più spesso come “mala sanità”. Il contenzioso e la litigiosità tra paziente e medico assume aspetti sempre più paradossali con pazienti che, a distanza di anni, a seguito di tiroidectomia, lamentano alterazioni vocali. La possibilità di registrare la voce e di verificarne la normalità consente, allo stato attuale, di documentare il problema identificando esattamente ciò che è imputabile alla negligenza o alla volontà del medico. È, infatti, possibile verificare e documentare, sin da subito, e con i mezzi che la fonetica mette a disposizione, la normalità della voce e quindi la riuscita dell'intervento. A ciò si aggiunga anche il fatto di poter fornire una risposta oggettiva e concreta attraverso esami spettro-acustici della voce. A tal proposito, il relatore ha saputo porre l'attenzione sul fatto che la fonetica rende tutto ciò possibile venendo in aiuto nel caso di contestazioni sollevate a distanza di anni. Quanto riportato da Lombardo rappresenta un aspetto medico-legale che la società di chirurgia e di otorinolaringoiatria, interessata a problematiche analoghe, ha inserito tra gli esami obbligatori da effettuare prima di un intervento di tiroidectomia oltre, naturalmente, ad una visita otorinolaringoiatrica, un esame della motilità vocale e, cosa non ancora obbligatoria, un esame spettro-caustico della voce e della registrazione vocale a cui dare un valore parametrico al fine di inquadrare il prodotto vocale. Lo studio elettroacustico della voce in ambito fonetico fornisce, quindi, un concreto aiuto anche sotto il profilo cautelativo nel caso di contenziosi legali.

**Pier Luigi Salza** della Loquendo S.p.A. ha invece posto l'attenzione sulla “etichettatura fonetica e TTS multilingue” come esempio di rapporto tra la Fonetica, con le sue finalità ed il suo modo di operare in modo sperimentale, e un'azienda attiva nel campo delle tecnologie vocali produttrice di software per la sintesi e il riconoscimento sia dello speaker che dello speech. L'etichettatura fonetica ha grande rilevanza nello sviluppo delle tecnologie della sintesi vocale, soprattutto nel caso in cui si fornisce la possibilità ad un sistema di TTS di *passare*, con la stessa voce, dalla sintesi di una lingua L1 a parole o spezzoni di frasi enunciati in una lingua straniera L2.

Dopo un'esauriente introduzione al sistema TTS e all'illustrazione delle problematiche più comunemente riscontrate riguardo i seguenti punti:

- la segmentazione automatica del database vocale registrato che costituisce il serbatoio di unità prelevate in *runtime* per generare il messaggio in sintesi;
- come parlare una lingua L2 con i foni della lingua L1 operante;
- l'uso dei codici utilizzati per la rappresentazione dei foni e la compatibilità fra i codici

l'intervento di Pier Luigi Salza ha messo in evidenza lo stretto legame tra gli algoritmi e il metodo della Fonetica Sperimentale. Pur essendo molte delle operazioni di generazione del corpus da cui attingere per la sintesi del messaggio affidate a programmi di tipo statistico automatico, basati su una grande quantità di testi trascritti, resta fondamentale il lavoro manuale di *rifinitura* operato da un consulente di madrelingua con robuste basi di fonetica. Non vi può, tuttavia, come sottolineato da Salza, essere sintesi senza etichettatura che, in questo caso, gioca un ruolo primario. Il relatore ha, infatti, spiegato che se il codice (inteso

come l'insieme: simbolo + tratti fonetico-articolatori associati) deve essere sia speaker-dependent sia language-independent, il livello di dettaglio dell'etichettatura deve essere sufficientemente alto mettendo in risalto che il codice deve consentire un livello di dettaglio e di accuratezza fonetica speaker-dependent per soddisfare: a) compatibilità con il segmentatore fonetico e b) non ambiguità nella rappresentazione dei foni tra lingue diverse. Una trascrizione con possibili ambiguità non è accettabile quando si deve poter *passare* da una lingua all'altra (mapping fonetico) e la "precisione" deve essere "assoluta" (deve cioè contenere scelte il più possibile giustificate foneticamente ma pragmaticamente adattate all'applicazione).

Pier Luigi Salza ha, successivamente, ribadito che la necessità di tenere conto del dettaglio fonetico nell'etichettatura a livello di speaker, mantenendo comunque una indipendenza dalla lingua, fa sì che il TTS si trovi a gestire livelli di trascrizione molto diversi:

- trascrizione di lingua, che si applica ad una prima trascrizione di parola (trascrizione larga o fonemica) e che deve garantire la compatibilità con ingressi fonetici esterni e consentire all'utente la costruzione di lessici fonetici;
- trascrizione di frase, che comprende regole di riscrittura sia all'interno sia ai confini di parola, incluse deaccentazioni (trascrizione stretta o fonetica – allofonica), riferita ad una tabella di lingua (set di fonemi) che deve garantire coerenza con le effettive realizzazioni fonetico articolatorie a livello di frase e garantire la non ambiguità tra le diverse lingue;
- trascrizione di voce, ovvero inserzione di allofoni speaker-dependent, anche diversi per ogni singola voce di una certa lingua (trascrizione fonetica ulteriormente stretta o allofonica idiosincratica), che fa riferimento ad una tabella di voce che deve garantire compatibilità con il segmentatore fonetico automatico voice-dependent.

Il relatore si è molto soffermato su un altro aspetto su cui impatta l'accuratezza dell'etichettatura fonetica: il Mapping fonetico su testi "mixed-language". La conversione testo-voce è intrinsecamente language-dependent, ma le applicazioni reali del TTS includono sempre più spesso testi "mixed-language" (parole o intere frasi in prestito da altre lingue). La pronuncia deve adattarsi alla lingua straniera mantenendo, preferibilmente, la stessa voce. La pronuncia approssimata è adatta per la lettura di parole o frasi in lingue straniere (L2) presenti nel testo (L1). L'architettura multilingue del TTS abilita così ciascuna voce a produrre frasi in tutte le lingue supportate dal sistema, sfruttando i mezzi messi a disposizione dalla fonetica.

**Emanuela Magno Caldognetto** ha relazionato sulla "comunicazione multimodale". La comunicazione multimodale è una tematica di ricerca che prevede un bilanciamento tra due tematiche assai ampie, trattate e sviluppate in decenni di ricerca, che sono la comunicazione verbale e la comunicazione non verbale: la prima trattata soprattutto da linguisti, fonetisti e psicologi cognitivisti; la seconda da psicologi cognitivi, sociolinguisti ecc.. La tematica unificata di questi due ambiti di ricerca può essere molto ambiziosa e non è, a tutt'oggi, completamente delineata. Si tratta tuttavia di una tematica che, secondo la relatrice, può favorire contatti con numerose altre discipline. La comunicazione multimodale può essere definita come "studio dei segnali (della loro struttura, dei processi di sincodificazione e di decodificazione multisensoriale) che trasmettono informazione nelle interazioni umane faccia-a-faccia" e, come tale, diventa una tematica unica e largamente interdisciplinare.

La relatrice si è poi soffermata anche sulla relazione tra sorgenti di segnali e sistemi di segnali nonché sulla relazione tra sistemi di segnali e tipi di informazione (secondo una classificazione cognitiva).

Considerando i segnali acustico-uditivi e ottico-visivi come veicolo di informazione, sussistono una serie di problematiche legate al fatto che questi possono trasmettere:

- informazioni diverse lungo lo stesso canale;
- informazioni eguali lungo canali diversi;
- informazioni diverse lungo canali diversi.

Lo studio della multimodalità comunicativa risulta assai complesso perché di solito non si tratta solo di individuare la giustapposizione di informazioni diverse generate da sorgenti specifiche e distinte, ma implica piuttosto l'individuazione di **strati** di informazioni diverse coprodotte. Qualunque sia la tipologia dell'informazione adottata, resta fondamentale, per l'elaborazione di una teoria della *Comunicazione Multimodale*, l'individuazione di:

- corrispondenza tra uno stimolo percepibile e un significato;
- relazioni temporali tra i segnali di più modalità che trasmettono uno stesso significato;
- specificità comunicativa di singoli canali e modalità.

Per mettere in evidenza come il metodo della fonetica possa avere molteplici applicazioni, la relatrice ha presentato alcune ricerche condotte presso l'ISTC di Padova riguardo il parlato e l'informazione trasmessa attraverso il canale visivo individuando, sostanzialmente, tre casi di bimodalità uditivo-visiva:

- la trasmissione di informazione fonologica;
- la trasmissione di emozioni nel parlato emotivo;
- la trasmissione di parlato e gestualità coverbale.

Per il primo caso, la Magno Caldognetto, ha presentato alcune ricerche con esperimenti percettivi effettuati inizialmente con ADOBE premiere ed in seguito con Elite nella sua versione più matura e recente di *Interface*. Tale tecnica mette in relazione tutta una serie di informazioni/rilevamento di dati con la possibilità di convertirli in parametri e soprattutto di relazionare segnale acustico e segnale visivo evidenziandone, da un lato le coerenze e le corrispondenze, dall'altro ciò che viene identificato con il termine di "Anisomorfia", ovvero un campo in cui inizio e fine del gesto articolatorio non coincidono con la segmentazione del segnale acustico.

Per il secondo caso di bimodalità, emozioni trasmesse da faccia e voce nel parlato emotivo, non è mancata una trattazione di quelle che sono le caratteristiche acustiche del parlato emotivo e delle espressioni facciali co-prodotte. Anche in tal caso le ricerche portate ad esempio sono risultate assai eloquenti:

- analisi della coproduzione di movimenti articolatori visibili e delle configurazioni delle labbra tipiche delle emozioni primarie (rabbia, paura, disgusto, gioia, tristezza, sorpresa) registrati sincronicamente con il segnale acustico corrispondente, eseguita rilevando gli spostamenti di markers collocati su punti ritenuti significativi;
- coproduzione di movimenti labiali e di pattern di movimenti facciali (fronte, sopracciglia ecc.) registrati sincronicamente col segnale acustico corrispondente.

Di notevole interesse risulta in questo contesto il fatto che le ipotesi formulate e i risultati ottenuti, siano stati ampiamente e proficuamente verificati attraverso la faccia parlante "Lucia" con una sintesi bimodale audiovisiva da testo partendo dai dati numerici.

Infine, ma di non minore importanza, è stato affrontato il terzo caso di bimodalità di coproduzione di gesto e parlato dove il tutto diventa ancor più complicato e ambizioso. Anche in questo caso gli esempi presentati riguardo il gruppo dell'ISTC di Padova.

## CONCERTO DI VOCI

Concerto di voci del Coro Polifonico UniCal diretto dal prof. Massimo Privitera, presso il “Teatro Piccolo” dell’Ateneo.

**Massimo Privitera** (Catania, 1956) ha svolto i propri studi universitari a Bologna, dove ha conseguito la laurea in DAMS, il diploma della Scuola di perfezionamento in musicologia ed il Dottorato di ricerca in musicologia (con una tesi su Orazio Vecchi). In seguito ha ottenuto una borsa di studio post-dottorato presso l’Università di Palermo. Ha insegnato per quindici anni Storia della musica presso i Conservatori statali e non statali, dove è stato anche Bibliotecario. Attualmente è professore associato di Storia della musica dall’antichità al rinascimento, presso l’Università della Calabria (Cosenza).

Ha pubblicato saggi sulla musica di Monteverdi, Marenzio, Vecchi, Gesualdo, Celano, Valentini, Tonelli, nonché sulla produzione di Erik Satie; e nel 2000 ha pubblicato una monografia su Arcangelo Corelli. Ha curato le edizioni moderne dei Madrigali di Frescobaldi (con Lorenzo Bianconi), delle Canzonette a sei voci di Vecchi (con Rossana Dalmonte), dei Madrigali di Achille Falcone. Ha curato la prima traduzione italiana degli scritti del musicologo americano Edward Lowinsky. Attende ad un volume sul madrigale italiano.

Alla principale attività di musicologo Massimo Privitera affianca quella di direttore di coro, compositore e vocalista. Ha fondato e diretto il Gruppo vocale Tens Clar, ha diretto il Madrigaletto dell’Università di Bologna, il coro femminile Kalicantus, il Laboratorio corale estivo di Perinaldo. Ha fondato e dirige il Coro polifonico dell’Università della Calabria.

## SPAZIO “A(D)DOTTA UN DOTTORANDO!”

Il quarto convegno AISV ha aperto le sue porte ai dottorandi di ogni tipo di scuola con progetti in corso sulla voce umana con un apposito spazio denominato “**A(d)dotta un Dottorando!**”. Si è trattato di un’iniziativa nuova ed unica nel suo genere mirata ad avvicinare i più giovani al mondo della ricerca.

Indipendentemente dalle comunicazioni scientifiche presenti durante il convegno, con l’iniziativa è stata favorita l’esposizione dei progetti di tesi (sotto forma di poster) per dare l’opportunità ai dottorandi di descrivere i loro progetti in corso di elaborazione, i risultati intermedi o il lavoro terminato ma ancora in fase di consolidamento ecc. In tale ottica, contrariamente a quanto comunemente avviene per i contributi presentati al convegno, non è stata attuata alcuna selezione e valutazione lasciando spazio a tutte le ricerche in itinere, o addirittura ai “progetti” di ricerca la possibilità di avere uno spazio poster. Il 4° convegno AISV ha così portato i giovani ricercatori a confrontarsi sulle loro attività, invitando e stimolando i soci più “anziani” ad intervenire con suggerimenti e pareri volti ad aiutare, indirizzare, consigliare e, perché no, a criticare alcune impostazioni.

L’iniziativa ha avuto, al suo esordio, un inaspettato successo. Sono stati complessivamente presentati 6 progetti di ricerca di dottorandi provenienti da tutta Italia di seguito riportati:

- Sviluppo di sistemi automatici per il riconoscimento del parlante, Matteo Nunziati - Università di Firenze;
- L’intonazione interrogativa nell’italiano regionale di Parma: una variabile sociolinguistica? Maria Chiara Felloni - Università di Pavia;
- Per un’analisi acustico-articolatoria e percettiva delle varietà salentine centrali, Angelica V. Costagliola - CRIL, Università del Salento;

- L'acquisizione fonetico-fonologica di L2 Inglese da parte di un gruppo di studenti universitari. Studio acustico e percettivo, Bianca Sisinni - CRIL, Università del Salento;
- Lingue (iso)accentuali e (iso)sillabiche: nuovi correlati e fattori di variazione, Paolo Mairano - Università di Torino;
- Correlazione tra impostazioni articolatorie, effetti acustici ed espressioni facciali: uno studio cross-linguistico-culturale europeo, Vincenzo Galatà - Università della Calabria.

## SPAZIO “DEMO”

Lo spazio demo rappresenta un'altra delle iniziative introdotte durante il 4° Convegno AISV. È stata, infatti, data la possibilità alle università ed ai responsabili di progetti di ricerca, di presentare e far conoscere tool o software senza scopi di lucro. Per tutta la durata del convegno sono state allestite delle postazioni dove le singole unità di ricerca hanno avuto modo di illustrare agli interessati i propri progetti. Nel dettaglio hanno preso parte:

- “InterVoice: un tool per l'analisi del segnale fonico in ambito forense”, Francesco Beritelli, Università di Catania;
- “Archivio sonoro calabrese”, Luciano Romito, Università della Calabria;
- “AGSpIT: Annotation Graph Spoken Italian Toolkit”, Francesco Cutugno, Università "Federico II" Napoli;
- “XGate – XML generator”, Francesco Cutugno, Università "Federico II" Napoli;
- “Banche dati lessicali sull'albanese d'Italia”, Francesco Altimari, Università della Calabria;
- “Istituto Regionale per la Comunità Occitana di Calabria”, Hans Peter Kunert, Comune di Guardia Piemontese;
- “Il Progetto Bomarzo”, Amedeo De Dominicis, Università della Tuscia (VT).

## PREMIO FRANCO FERRERO

Come da tradizione AISV, anche quest'anno, all'autore (studente o dottorando) del miglior articolo pubblicato negli Atti della Conferenza Annuale AISV del 2005 è stato assegnato il "Premio Franco Ferrero".

Per dare maggiore visibilità e importanza all'evento, la consegna dei premi è stata posticipata e celebrata durante la parte centrale del convegno prima della Tavola Rotonda.

Il “Premio Ferrero 2007” è stato consegnato dal prof. Roberto Guarasci, Direttore del Dip.to di Linguistica e Presidente del CoCoP. Il prof. Guarasci ha omaggiato i vincitori con una borsa di studio e con la consegna di una riproduzione fedele in scala di un'antica moneta Brutia riservata ad ospiti illustri dell'Ateneo.

I premiati di quest'anno, risultati vincitori del miglior articolo pubblicato negli Atti del 2° Convegno AISV 2005 tenutosi a FISCIANO (Salerno), sono stati: Mauro NICOLAO per la categoria *Tecnologie del Parlato* con il lavoro “Modellizzazione della prosodia e del timbro per la sintesi del parlato emotivo”, Mauro Nicolao, Carlo Drioli, Piero Così e Massimiliano BARBERA e Marlen BARTH per la categoria *Linguistica, Fonetica, Fonologia* con il lavoro “Qualità di voce e indebolimento consonantico: un caso di correlazione in scouse?”, Massimiliano Barbera e Marlen Barth.

## RINGRAZIAMENTI

Un ringraziamento particolare va a tutti i membri del comitato scientifico e a tutti coloro che ci hanno aiutato con il lavoro di revisione degli atti AISV2007 che in questo volume sono contenuti.

Non possiamo non ringraziare il prof. Guarasci per l'omaggio della medaglia *Brutia* ai vincitori del Premio Ferrero, medaglia consegnata durante la cerimonia di premiazione da lui stesso presieduta e celebrata.

Ringraziamo di cuore tutti gli studenti volontari che ci hanno aiutato con l'organizzazione di tutto il convegno.

Ringraziamo il Prof. Privitera e il coro da lui diretto.

*Luciano Romito*

*Vincenzo Galatà*

*Rosita Lio*

# **FONETICA**





# PROSPEZIONI SULLA STRUTTURA RITMICA DELL'ITALIANO BASATE SUL CORPUS SEMISPONTANEO AVIP/API

Chiara Bertini & Pier Marco Bertinetto

Scuola Normale Superiore (Pisa)

[c.bertini@sns.it](mailto:c.bertini@sns.it), [piermarco.bertinetto@sns.it](mailto:piermarco.bertinetto@sns.it)

## RIASSUNTO

In questo lavoro, presentiamo una ricerca condotta sui dati contenuti nella banca-dati AVIP/API, con specifico riguardo alla variante pisana dell'italiano. Trattandosi di materiali semispontanei, si è dovuto preliminarmente affrontare una serie di problemi per ciò che riguarda la selezione dei materiali, al fine di basare le osservazioni su porzioni di enunciato sufficientemente rappresentative.

Le analisi hanno fornito indicazioni interessanti e non sempre scontate. Se ne fornisce qui un primo elenco:

- la porzione finale dell'enunciato, dall'ultima sillaba accentata in poi, comporta una sensibile alterazione dell'andamento ritmico;
- un effetto relativamente modesto è invece legato alle momentanee sovrapposizioni dei due locutori;
- non si è notata alcuna differenza significativa per quanto riguarda il contrasto tra enunciati dichiarativi vs. interrogativi;
- è emerso un sostanziale contrasto fra "Giver" e "Follower", fortemente correlato, peraltro, con la diversa velocità di elocuzione (decisamente più lenti i "Givers");
- la velocità di elocuzione ha tangibili conseguenze sulla configurazione ritmica.

Nel condurre l'analisi, si sono sperimentati diversi algoritmi di calcolo, tra cui uno ideato dagli autori, denominato "Control/Compensation Index", il cui comportamento appare decisamente promettente.

## 1. INTRODUZIONE

Gli studi sull'indole ritmica delle lingue naturali hanno ricevuto nuovo impulso in questi ultimi anni, a partire dalla proposta di Ramus (Ramus *et alii*, 1999), presto ripresa da altri studiosi e segnatamente da Low (Low *et alii*, 2000) e Grabe & Low (Grabe & Low, 2002). Non sono ovviamente mancate le applicazioni all'italiano e ai suoi dialetti (Russo & Barry, 2004; Schmid, 2004; Romano, 2003). La discussione si è soprattutto concentrata sulla ricerca di un algoritmo capace di rappresentare in maniera efficace le proprietà ritmiche delle lingue naturali, traducendo in termini concreti l'intuizione di Pike (Pike, 1947) circa la polarizzazione tra lingue iso-accentuali e iso-sillabiche (syllable- vs. stress-timed). Benché gli studi successivi abbiano ampiamente chiarito che l'interpretazione letterale di tale dicotomia non poteva trovare conferma, è tuttora opinione diffusa che l'intuizione di partenza sia sostanzialmente corretta. Non sono del resto mancati i tentativi di reinterpretarla in chiave nuova, correlando un insieme di comportamenti fonetici con un insieme di proprietà fonologiche (Bertinetto, 1981; Dauer, 1983; Bertinetto, 1989). Tra i fattori rilevanti, sono stati individuati almeno i seguenti:

- riduzione vocalica vs. piena articolazione nelle sillabe atone;
- struttura sillabica complessa (con relativa incertezza dei confini sillabici) vs. semplice (con confini ben definiti);

- relativa flessibilità vs. rigidità nella collocazione dell'accento lessicale;
- densità di accenti secondari vs. relativa rarefazione;
- accelerazioni di velocità di eloquio ottenute (per lo più) mediante la compressione delle sillabe atone vs. compressione relativamente proporzionale tra sillabe toniche ed atone.

È stata successivamente avanzata l'ipotesi che questo fascio di tratti potesse trovare un punto di convergenza nel contrasto fra lingue "a controllo" e lingue "a compensazione" (Bertinetto & Vékás, 1991).<sup>1</sup> L'idea di base, già accennata in Bertinetto & Fowler (Bertinetto & Fowler, 1989), è la seguente. Si assume che le lingue "a controllo" siano basate su una struttura fonotattica semplice, che induce il parlante ad assumere un'impostazione articolatoria relativamente rigida, mentre il contrario avviene nelle lingue "a compensazione", fondate su una fonotassi complessa ed un'impostazione articolatoria più flessibile. Se l'idea è corretta, essa dovrebbe comportare una serie di conseguenze sul piano dei comportamenti articolatori e prosodici. Si dovrebbero in sostanza verificare le seguenti condizioni:

- Sul piano articolatorio, le lingue "a compensazione" dovrebbero comportare meccanismi di coarticolazione più spinti rispetto a quelli che agiscono nelle lingue "a controllo", il che si traduce in una maggior tendenza alla co-produzione dei gesti articolatori, con specifico riguardo alla sovrapposizione dei gesti consonantici e vocalici.

- Sul piano prosodico, le durate dei foni dovrebbero essere diversamente regolate: in particolare, nelle lingue "a controllo" i foni dovrebbero manifestare una certa dose di incomprimibilità anche al crescere della velocità di elocuzione. Ciò non significa, beninteso, che tali lingue non possano tollerare accelerazioni di eloquio, bensì che il grado di sovrapposizione tra gesti articolatori adiacenti non possa superare un certo limite, soprattutto per quanto riguarda i foni vocalici (si veda il punto precedente). Di conseguenza, questi ultimi dovrebbero tendere a conservare una corposa "visibilità" anche in sillaba atona. Per contro, nelle lingue "a compensazione" le vocali atone dovrebbero poter tollerare livelli anche molto elevati di coarticolazione, fino a dare l'impressione – nei casi estremi – di scomparire dallo spettrogramma in quanto porzione autonoma di segnale. Non va peraltro dimenticato che le porzioni indipendentemente misurabili dei foni, così come esse ci appaiono sullo spettrogramma, non sono che la parte visibile di un meccanismo estremamente complesso, fatto di continue sovrapposizioni tra gesti articolatori.

Se queste ipotesi colgono nel segno, ne consegue che l'impostazione adottata nelle recenti ricerche sul tema dell'isocronia contiene un forte elemento di criticità. Tanto nel modello di Ramus, quanto nel modello di Grabe & Low (nonché nelle varianti cui essi hanno dato luogo), viene analizzato il comportamento degli intervalli vocalici e consonantici senza riguardo ai segmenti che li compongono. Ciò impedisce di distinguere tra situazioni fonotattiche assai diverse, che potrebbero concretamente manifestarsi nelle lingue naturali. Per esempio: un medesimo intervallo potrebbe essere occupato da due consonanti oppure da quattro (e se abbiamo a che fare con il polacco, fino a non meno di sei). Trattare tale intervallo consonantico come un tutto indifferenziato significa perdere informazioni importanti, le quali potrebbero raccontarci molte cose circa le strategie articolatorie adottate dai parlanti e, di conseguenza, circa i riflessi che tali strategie possono avere sui comportamenti prosodici. Lo stesso dicasi per gli intervalli vocalici: una sequenza di vocali in iato non è la stessa cosa di una sequenza di vocali in sinalefe. Nel secondo caso,

---

<sup>1</sup> La terminologia è ispirata a Hoeqvist (Hoeqvist, 1983), anche se reinterpretata in accezione nuova.

i differenti gesti articolatori sono fusi entro un medesimo nucleo sillabico, mentre in caso di iato i nuclei sillabici restano distinti. Se dunque ammettiamo che tutto ciò possa contare nel dispiegarsi del ritmo delle lingue naturali, è importante trattare diversamente queste due situazioni.

Tenendo conto di ciò, è stato proposto dagli autori di questo lavoro il modello illustrato in Bertinetto & Bertini (2008), denominato ‘Control/Compensation Index’ (CCI). Esso si basa sul seguente algoritmo, dove  $m$  sta per ‘numero degli intervalli’ (vocalici o consonantici),  $d$  per ‘durata’,  $n$  per ‘numero di segmenti entro gli intervalli pertinenti’:

$$CCI = \frac{100}{m-1} \sum_{k=1}^{m-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right| \quad (1)$$

L’idea di prendere in considerazione il numero dei segmenti che costituiscono gli intervalli non è del tutto nuova. Essa è stata messa in pratica da Rouas & Farinas (Rouas & Farinas, 2004), sia pure limitatamente agli intervalli consonantici. Vi sono tuttavia importanti differenze tra la loro impostazione e quella del modello CCI, su cui ritorneremo nel § 3.

Ovviamente, la bontà di un metodo concepito per misurare le tendenze ritmiche delle lingue naturali dipende crucialmente da due elementi:

(i) la capacità di far emergere la correlazione tra gli orientamenti ritmici e il diverso assetto articolatorio adottato dalle lingue in rapporto alla propria struttura fonotattica;

(ii) la capacità di generare i raggruppamenti corretti delle lingue indiziate di appartenere prototipicamente all’uno o all’altro tipo ideale, comunque li si voglia denominare (la dicotomia “controllo vs. compensazione” ci appare, peraltro, assai più persuasiva di quella tradizionale, “syllable- vs. stress-timing”, notoriamente priva di riscontri effettivi).

Quanto al secondo aspetto, risultati molto incoraggianti possono già essere citati a partire da uno studio di Mairano & Romano (in preparazione), in cui sono state confrontate, attraverso l’applicazione dell’algoritmo CCI, otto diverse lingue, studiate a partire dalla lettura di brevi testi. Gli esiti della computazione hanno prodotto una separazione delle lingue analizzate più coerente rispetto a quella prodotta dagli altri modelli correntemente impiegati. Ciò dimostra che i presupposti su cui si fonda la nostra proposta sono alquanto solidi. La direzione intrapresa merita dunque di essere proseguita.

I dati che qui verranno illustrati, tuttavia, si riferiscono unicamente all’analisi condotta su materiali semi-spontanei di italiano pisano, estratti dal corpus AVIP/API.<sup>2</sup> Ci proponiamo beninteso di estendere in futuro l’analisi, confrontando i risultati ottenuti per l’italiano con quelli di altre lingue, studiate anch’esse a partire da materiali semi-spontanei, rigorosamente omogenei a quelli del nostro corpus. È infatti ben noto che lo stile di elocuzione (tipicamente: parlato letto vs. parlato spontaneo, diverse velocità di elocuzione etc.) produce effetti macroscopici a livello di comportamenti ritmici. Lo scopo di questa comunicazione non è comunque quello di confrontare l’italiano con altre lingue, bensì di verificare le diverse tendenze che possono emergere all’interno di un medesimo tipo di testo, confrontando locutori diversi e diversi ruoli discorsivi, sulla base di un corpus debitamente ampio. Quest’ultimo fatto permetterà, tra l’altro, di rispondere ad un’obiezione che si può muovere alle ricerche condotte nel filone inaugurato dall’avvento del modello di Ramus: ossia, la ridotta ampiezza della base di dati su cui si fondano molte delle classificazioni proposte. Il numero di intervalli vocalici e consonantici considerati è infatti

<sup>2</sup> <http://www.cirass.unina.it/ricerca/studi%20parlato/raccolta%20corpora/api/api.htm>

spesso piuttosto ristretto, con ovvie conseguenze sulla rappresentatività statistica, come mostra il seguente prospetto.

|                             |                         |
|-----------------------------|-------------------------|
| Barbosa, 2006               | 36 frasi, 1 locutore    |
| Dellwo, 2004                | 80 sillabe, 12 locutori |
| Low & Grabe, 2000           | 10 frasi, 10 locutori   |
| Mairano & Romano, 2006      | 15 frasi, 516 segmenti  |
| Ramus <i>et alii</i> , 1999 | 5 frasi, 4 locutori     |
| Rouas & Farinas, 2004       | 15 frasi, 10 locutori   |

Tabella 1: Ampiezza della base di dati per ogni lingua studiata in alcuni lavori di ricerca.

Occorre infine sottolineare che con il modello CCI non presumiamo affatto di cogliere in maniera esauriente le proprietà del ritmo delle lingue naturali. Il problema è assai più complesso, e a nostro avviso va affrontato su due livelli distinti e non necessariamente convergenti, come argomentiamo altrove (Bertinetto & Bertini, in preparazione), ossia:

(i) Livello-I, in cui si dispiega l'effetto della struttura fonotattica e che si fonda sull'accoppiamento di due oscillatori (vocalico e consonantico), secondo la suggestiva proposta di Goldstein (Goldstein *et alii*, 2007).

(ii) Livello-II, anch'esso imperniato sull'azione di due oscillatori accoppiati (accentuale e sillabico), secondo la proposta di O'Dell & Nieminen (O'Dell & Nieminen, 1999).

Da quanto detto, appare evidente che il modello CCI agisce unicamente al livello-I, ed è dunque a questo che ci limiteremo in questo lavoro. Trattandosi peraltro del livello in cui si dispiega l'effetto della fonotassi, esso costituisce una componente basilare ed inaggirabile del meccanismo ritmico delle lingue, pur senza esaurirne la complessità.

Per comodità del lettore, forniamo qui un breve riassunto dell'articolazione di questo lavoro. Nel § 2 sono riportati i criteri secondo cui sono stati selezionati i materiali a partire dal corpus di riferimento. Nel § 3 sono succintamente riassunte le principali differenze emerse nel confronto tra il modello CCI ed i modelli di Ramus (Ramus *et alii*, 1999), Grabe & Low (Grabe & Low, 2002), Dellwo (Dellwo, 2004), Rouas & Farinas (Rouas & Farinas, 2004). Il § 4 riporta l'esito delle analisi condotte sulle diverse variabili considerate nel nostro corpus: ruolo del locutore ("Giver" vs. "Follower"), velocità di elocuzione, tipo sintattico-intonativo (enunciati dichiarativi vs. interrogativi), sesso del locutore. Il § 5, infine, riassumerà le nostre conclusioni.

## 2. CRITERI DI SELEZIONE DEI MATERIALI

### 2.1 Materiali

La ricerca qui illustrata è stata condotta analizzando la variante pisana dell'italiano, contenuta nella banca-dati AVIP/API. Abbiamo preso in considerazione 5 coppie di parlanti (3 femmine e 7 maschi). Trattandosi di testi semispontanei elicitati con la tecnica del "map-task", è stato necessario stabilire preliminarmente dei criteri, per selezionare il materiale in modo tale da ottenere porzioni di parlato sufficientemente estese e allo stesso tempo uniformi in termini di "continuità" di eloquio. Ciò è indispensabile per estrarre indicazioni attendibili circa le tendenze ritmiche.

Per la selezione delle porzioni di enunciato utilizzabili abbiamo applicato nell'ordine i seguenti criteri:

I) sono stati scelti enunciati privi di esitazioni, pause, forme di assenso, esclamazioni, false partenze, fenomeni vocali non verbali, sequenze inintelligibili, routines discorsive (come le frasi fatte ricorrenti);

II) sono state ulteriormente eliminate le parti terminali di enunciato di tipo asseverativo (ad esempio: ..., *no?*), così come quelle introduttive (ad esempio: *cioè*, ...);

III) sono state selezionate sequenze che, in trascrizione ortografica, avessero almeno 9 sillabe ( $\geq 9$  sillabe) e, foneticamente, almeno 8 sillabe ( $\geq 8$  sillabe). La differenza è motivata dal fatto che non tutte le sillabe potenziali si realizzano effettivamente nel parlato, a causa di possibili fenomeni di fusione tra vocali adiacenti, riduzione di iato, ipoarticolazione etc.

Un problema particolarmente delicato è costituito dalla misurazione delle componenti segmentali dei dittonghi. In italiano, secondo l'opinione prevalente benché non unanime dei fonologi, i legamenti che compongono i dittonghi ascendenti e discendenti vanno considerati fonemi consonantici approssimanti (Bertinetto & Loporcaro, 2005). È dunque essenziale, per il funzionamento di un algoritmo come CCI calcolare esattamente le durate del legamento e della vocale nucleare.<sup>3</sup> Purtroppo, come ben sa chiunque si sia cimentato nella materia, la divisione tra legamento e vocale è impresa difficile e delicata; a maggior ragione in un corpus di parlato spontaneo, con fasi di elocuzione accelerata o addirittura ipoarticolate. È stato dunque necessario adottare una strategia operativa, con tutti i rischi del caso. Dopo una serie di analisi pilota, si è optato per la seguente soluzione:

(a) per i maschi, la durata del legamento è stata fatta coincidere convenzionalmente con le prime o, a seconda dei casi, con le ultime due vibrazioni laringali del dittongo;

(b) per le femmine, tale durata è stata fissata in 4 vibrazioni, tenuto conto della diversa frequenza intrinseca della voce femminile;

(c) quando l'intervallo corrispondente alle 2 o 4 pulsazioni superava un terzo della durata complessiva del dittongo (il che può accadere nel parlato veloce), si è convenzionalmente attribuita al legamento una durata pari ad 1/3 della durata complessiva.

La procedura adottata è sicuramente discutibile, ma ha almeno il vantaggio di non assegnare una durata fissa ad ogni legamento, dato che la durata delle pulsazioni laringali può variare da istante a istante. In tal modo, si è scongiurato il rischio di introdurre un livellamento artificiale della varianza.

Anche nel caso delle geminate si è adottato un criterio a base fonologica. L'intervallo consonantico corrispondente ad una geminata è stato pertanto interpretato come composto da due segmenti fonemati, la cui durata è stata convenzionalmente considerata pari alla metà dell'intero intervallo. Gli intervalli vocalici, per parte loro, consistevano per lo più di un singolo elemento, eccezion fatta per le sinalefi; le vocali in iato, in quanto riferibili a diversi nuclei sillabici, sono infatti state assegnate ad intervalli vocalici distinti. Beninteso, se in italiano vi fosse quantità vocalica pertinente (come in finnico o in giapponese), gli intervalli vocalici potrebbero anche constare di due elementi, in analogia a quanto accade con le geminate (e se avessimo a che fare con una lingua come l'estone, con vocali e consonanti superlunghe, dovremmo ammettere la possibilità di dividere addirittura per tre ogni fono vocalico o consonantico della lunghezza appropriata).

Le porzioni di enunciato selezionate per il nostro corpus sono state suddivise in enunciati interrogativi e dichiarativi. Tuttavia, gli enunciati in cui l'elemento interrogativo

---

<sup>3</sup> Ma lo stesso varrebbe, in realtà, per qualsiasi altro modello, a meno che non si voglia salomonicamente ignorare il problema della natura consonantica dei legamenti, come in effetti ci risulta che sia stato fatto nelle ricerche finora prodotte.

era unicamente riferibile all'espressione conclusiva preliminarmente eliminata (vedi il secondo criterio di selezione), sono stati etichettati come dichiarativi.

Per il calcolo della velocità di eloquio (in sillabe per secondo) è stata valutata la divisione sillabica della trascrizione fonetica, tenendo dunque conto delle sillabe effettivamente realizzate e considerando le vocali in iato come nuclei sillabici a sé stanti.

## *2.2 Verifiche preliminari*

In via preliminare, sono state effettuate alcune verifiche necessarie ad interpretare correttamente i risultati. In particolare, si è considerato il possibile effetto derivante da: (i) la sovrapposizione di voci dei due locutori; (ii) le diverse categorie sintattico-intonative degli enunciati.

### *2.2.1 Analisi delle sequenze con sovrapposizione di voci*

Le modalità con cui è stato registrato il parlato semi-spontaneo analizzato (i due parlanti disponevano di microfoni unidirezionali, ma si trovavano nel medesimo ambiente) possono comportare casi di sovrapposizione di voci, con possibili alterazioni della fluenza, magari impercettibili a prima vista.

Sono state dunque confrontate tre diverse selezioni dei materiali del corpus, così caratterizzate:

(a) eliminazione di tutti gli enunciati contenenti, anche se in minima parte, delle sovrapposizioni di voci;

(b) eliminazione delle sole porzioni di enunciato contenenti sovrapposizioni (quando però la sovrapposizione si colloca nella parte centrale, l'intero enunciato deve essere comunque eliminato, data la discontinuità che si verrebbe a creare giustapponendo parte iniziale e finale);

(c) come nel caso precedente, ma conservando gli enunciati in cui la sovrapposizione sia limitata alla parte finale, nell'ipotesi che in tal caso la distorsione ritmica sia di entità limitata.

I risultati, calcolati secondo l'indice CCI, sono riportati nella figura 1, avendo in ascissa il valore degli intervalli vocalici e in ordinata quello degli intervalli consonantici. La prossimità dei punti, anche in rapporto al "braccio" delle varianze, mostra che non ci sono significative differenze tra i tre tipi di elaborazione; ciò è confermato anche dalle indagini statistiche. È stato pertanto scelto di eliminare dalle successive analisi solo le porzioni di frase in cui si ha sovrapposizione, in modo da poter conservare il maggior numero possibile dei dati a disposizione.

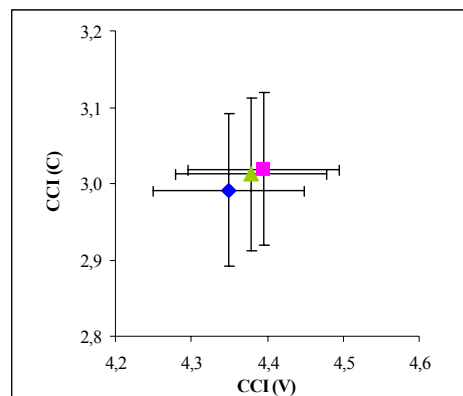


Figura 1: Confronto dell'indice CCI per diversi tipi di elaborazione del corpus, in rapporto alle sovrapposizioni di voci. Le "braccia" indicano la deviazione standard della media.

Quadrato rosa: caso (a); rombo blu: caso (b); triangolo verde: caso (c). Per una puntuale definizione dei tre tipi di selezione, si veda il testo.

### 2.2.2 Confronto tra diverse categorie di enunciato

Successivamente, si è considerato l'effetto congiunto della sovrapposizione di voci e dell'allungamento di fine enunciato. Sono stati eseguiti i calcoli relativi a tre insiemi di enunciati, così denominati;

- enunciati senza la porzione finale: per ogni enunciato, è stata eliminata la porzione che va dall'ultima sillaba accentata in poi;
- enunciati senza sovrapposizione di parlanti: da ogni enunciato è stata eliminata la porzione di frase con sovrapposizione (come nel caso (b) descritto nel paragrafo 2.2.1);
- enunciati senza ultima accentata né sovrapposizione di voci (in pratica, sommando le due condizioni precedenti).

Questi tre insiemi sono stati confrontati con il corpus inizialmente selezionato, indicato come "intero". Il calcolo dell'indice CCI riportato in figura 2 evidenzia, come atteso, il forte effetto della porzione di eloquio che segue l'ultima sillaba accentata, in netto contrasto con l'effetto prodotto dalla sovrapposizione di voci. La non significativa discrepanza tra gli enunciati del corpus "intero" e quelli privati delle sovrapposizioni conferma che le alterazioni causate dalla mescolanza di voci non hanno prodotto conseguenze rilevanti.

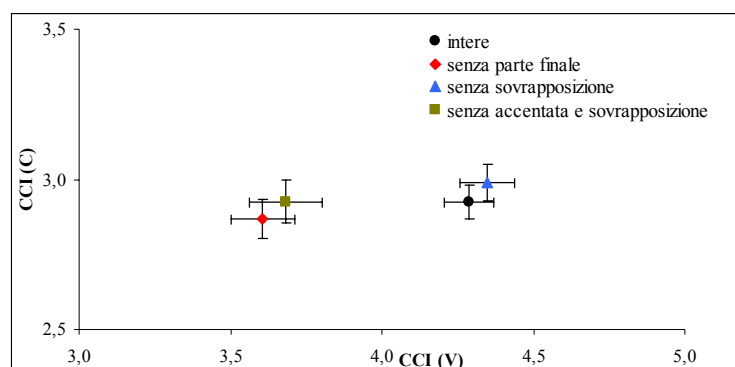


Figura 2: Confronto dell'indice CCI tra quattro diverse selezioni dei materiali: pallino nero: corpus "intero"; rombo rosso: enunciati privi della parte finale (dall'ultima sillaba accentata in poi); triangolo celeste: enunciati privi di sovrapposizione di voci; quadrato verde: somma delle due ultime condizioni.

In seguito ai risultati ottenuti in queste indagini preliminari, si è ritenuto di poter lavorare su tutti gli enunciati inizialmente selezionati, avendo unicamente cura di trascurare le porzioni finali degli stessi (dall'ultima sillaba accentata in poi). Ciò non significa, peraltro, che noi non attribuiamo alcun senso ritmico alla parte terminale degli enunciati, bensì che tale parte merita di essere analizzata indipendentemente, in quanto presenta caratteristiche proprie, con andamenti radicalmente diversi da quelli che riguardano il corpo dell'enunciato. A ciò si aggiunga che le caratteristiche della chiusa degli enunciati differiscono da lingua a lingua in maniera spesso capricciosamente idiosincratica, secondo le convenzioni adottate dalle diverse comunità di parlanti. Sappiamo, per esempio, che non tutte le lingue (o varietà di lingue) seguono la tendenza al rallentamento finale.

La tabella 2 riporta la quantità di dati utilizzati nelle analisi che seguiranno.

|                                      |                                                  |         |
|--------------------------------------|--------------------------------------------------|---------|
| FONEMI                               |                                                  | 8410    |
| INTERVALLI                           | vocalici                                         | 2991    |
|                                      | consonantici                                     | 2995    |
|                                      | Follower                                         | 1731    |
|                                      | Giver                                            | 4255    |
|                                      | dichiarative                                     | 5342    |
|                                      | interrogative                                    | 644     |
|                                      | maschio                                          | 4500    |
|                                      | femmina                                          | 1486    |
|                                      | TOTALI                                           | 5986    |
|                                      | N° medio di segmenti per intervalli vocalici     | 1.1     |
|                                      | N° medio di segmenti per intervalli consonantici | 1.4     |
| Durata media intervalli vocalici     |                                                  | 79.2 ms |
| Durata media intervalli consonantici |                                                  | 78.8 ms |
| SOGGETTI                             |                                                  | 10      |

Tabella 2: Riepilogo dei dati analizzati nel corpus "consolidato", ossia depurato degli elementi che possono introdurre significative deformazioni nell'andamento ritmico.



### 3. CONFRONTO TRA MODELLI

Va osservato che la formula CCI non prende in considerazione la nozione di ‘sillaba’. In Bertinetto & Bertini (in stampa) abbiamo fornito le ragioni teoriche di questa scelta, su cui qui non ritorniamo. In ogni caso, la nostra scelta coincide con quella adottata in altri ben noti modelli, come il modello PVI o quello di Ramus.

Il modello CCI assomiglia per certi aspetti a quello proposto da Rouas & Farinas (Rouas & Farinas, 2004), ma con importanti differenze. R&F introducono la nozione di ‘semi-sillaba’, definita come la sequenza che va da una vocale (esclusa) alla successiva (compresa). La parte consonantica di tale sequenza viene poi divisa per il numero di segmenti che la compongono, mentre gli intervalli vocalici sono mantenuti indivisi. Nel modello CCI, per contro, gli intervalli vocalici possono anche consistere di più segmenti (quando vi sia sinalefe o quando si abbia a che fare con una vocale di quantità fonologica lunga), anche se in molti casi essi consistono di un unico elemento. Inoltre, mentre il modello di R&F rientra nel novero degli algoritmi che si potrebbero definire “statici”, in quanto prescindono dall’effettiva successione dei segmenti che compongono la catena fonica, il modello CCI appartiene alla categoria degli algoritmi “dinamici”, il cui capostipite è il modello PVI di Grabe & Low. Ciò comporta una visione nettamente distinta delle cose, che potremmo metaforicamente riassumere così: nei modelli dinamici, si tiene conto delle variazioni locali nella durata degli elementi misurati, come se avessimo a che fare con una successione di fotogrammi; nei modelli statici si fornisce invece, per così dire, un fotogramma unico e riassuntivo.

In sostanza, CCI presenta le seguenti caratteristiche:

- i) sfrutta – a differenza del modello di R&F – le potenzialità dinamiche del modello PVI, a nostro avviso superiore al modello ramusiano nella sua capacità di seguire a passo a passo le effettive oscillazioni nella durata degli elementi adiacenti;
- ii) prende in debita considerazione – a differenza del modello PVI – il livello di complessità fonotattica, riflesso nel numero di elementi che compongono gli intervalli (e non solo quelli consonantici, come in R&F). Il fattore fonotattico è, in definitiva, alla radice del contrasto fra lingue ‘a controllo’ e ‘a compensazione’.

La figura 3 delinea la situazione ideale, quale si può predire sulla base di quanto ci è noto circa il comportamento delle lingue. Le lingue che chiamiamo “a controllo” dovrebbero cadere nella parte bianca del grafico, in quanto le fluttuazioni di durata delle vocali e delle consonanti tendono ad essere della medesima entità. Le lingue ‘a compensazione’ dovrebbero invece cadere nella parte grigia inferiore, perché le fluttuazioni vocaliche sono complessivamente maggiori di quelle consonantiche, a causa del maggior grado di coarticolazione (o sovrapposizione articolatoria) cui le vocali vanno soggette. La parte grigia superiore definisce, per contro, un’area in cui non ci si aspetta di trovare alcuna lingua: è infatti poco plausibile che le fluttuazioni di durata (ed i conseguenti livelli di co-articolazione/co-produzione) siano decisamente maggiori per le consonanti che non per le vocali.

È opportuno precisare che solo un modello come CCI permette di dare un’interpretazione altrettanto rigorosa dello spazio cartesiano, imperniato sulla retta bisettrice del piano. Negli altri modelli finora proposti, non potendosi sapere a quanti segmenti corrisponde ciascun intervallo, questo importante punto di riferimento viene a mancare, anche se la posizione in figura permette comunque di valutare quale, tra le due componenti (vocalica e consonantica, ma solo in riferimento ad intervalli indifferenziati), presenta complessivamente le oscillazioni maggiori.

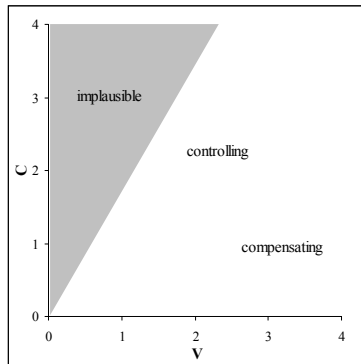


Figura 3: Rappresentazione schematica dei due prototipi ritmici fondamentali, secondo l'ipotesi "controllo/compensazione"..

La figura 4 mostra le posizioni ottenute a partire dai dati del nostro corpus per i seguenti modelli:

- CCI
- Ramus in due delle sue possibili versioni, con  $\Delta V$  o  $\%V$  come misura vocalica;
- PVI in due diverse versioni, con dati grezzi (rPVI) o normalizzati (nPVI; con normalizzazione riferita alle vocali);
- Varco in due delle sue possibili versioni, con VarcoV o  $\%V$  come misura vocalica;
- R&F, calcolato peraltro nella versione in cui, come misura vocalica, si ha la media della durata degli intervalli ( $D_v$ ).

Nella versione  $\Delta V$ , il modello ramusiano calcola la deviazione standard per gli intervalli sia vocalici sia consonantici. Nella versione  $\%V$ , la misura vocalica è invece data dalla percentuale di intervalli vocalici rispetto alla durata complessiva della sequenza. Quanto al modello PVI, esso si basa sulla somma delle differenze di durata tra intervalli adiacenti. Esso funziona praticamente come indicato dalla formula (1) sopra riportata, ma con la cruciale assenza della variabile  $n$ , che indica il numero di segmenti compresi in ciascun intervallo. Nella sua variante nPVI, ogni differenza tra intervalli consecutivi è normalizzata rispetto alla loro durata media, al fine di eliminare la variabilità dovuta alle variazioni della velocità di eloquio all'interno della sequenza considerata. Quanto al modello Varco, esso si basa sul rapporto tra la deviazione standard e il valor medio degli intervalli adiacenti di ogni porzione di enunciato. Infine, il modello di Rouas & Farinas (R&F) calcola la media della durata degli intervalli sia vocalici sia consonantici in ogni sequenza.

Per poter collocare su un medesimo grafico i risultati ottenuti dai diversi algoritmi, i valori vocalici dei due modelli ramusiani e di nPVI sono stati divisi per un fattore 10. Gli ovali della figura 4 indicano la dispersione spaziale dei 10 soggetti considerati nel corpus; per avere un'idea della quale (limitatamente all'applicazione del modello CCI) si può consultare la figura 5, da cui risulta peraltro che la distribuzione spaziale dei locutori è tutto sommato abbastanza omogenea, con una maggior fluttuazione vocalica piuttosto che consonantica. L'analisi qualitativa mostra la collocazione particolarmente periferica dei locutori D2G e BF rispetto agli altri. Peraltro, il valore elevato della deviazione standard dei suddetti locutori lascia intendere che la responsabilità della loro devianza sia soprattutto attribuibile al minor numero di dati presi in considerazione. Le analisi statistiche evidenziano inoltre significative differenze tra svariate coppie di locutori, con contrasti più o meno forti a seconda che vocali e consonanti siano considerate separatamente o

congiuntamente. La variabilità di questi risultati conferma l'idea già precedentemente avanzata, anche nelle ricerche sul ritmo, secondo cui lo studio di un solo locutore o di un numero ridotto di locutori alimenta ragionevoli dubbi circa la fondatezza delle conclusioni che si possono trarre, data la scarsa rappresentatività statistica dei dati.

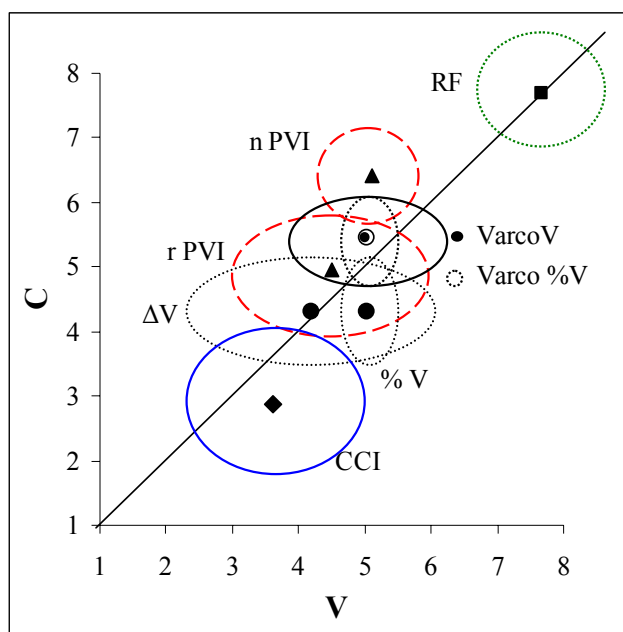


Figura 4: Rappresentazione spaziale dei modelli CCI, PVI, Ramus, Varco e R&F nelle possibili versioni. Gli ovali rappresentano la dispersione spaziale dei 10 soggetti considerati.

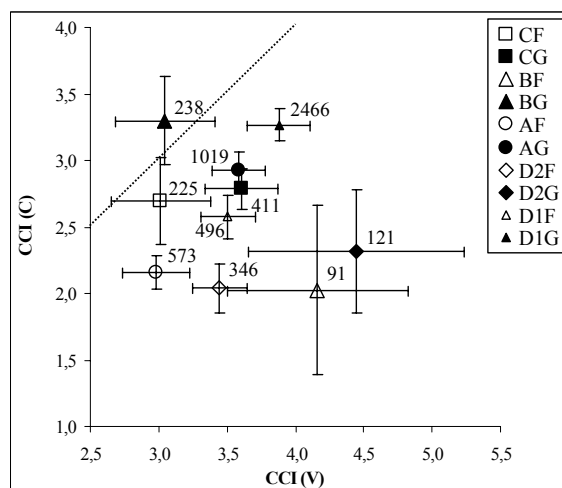


Figura 5: Confronto tra i 10 locutori secondo il metodo CCI, con indicazione del numero di intervalli analizzati per ognuno.

Pur con le cautele sopra segnalate quanto all'interpretazione della bisettrice, si possono avanzare alcune considerazioni in merito alla collocazione rispettiva dei punti generati dai diversi algoritmi. I modelli  $\Delta V$ , rPVI e R&F, che generano un punto collocato in pratica sulla bisettrice, suggeriscono che vocali e consonanti abbiano, nel complesso, un comportamento simile. Per contro, con nPVI e Varco (anche se in forma più lieve) l'oscillazione delle consonanti appare più forte di quella delle vocali: ma occorre dire che, nel primo modello, le misure di vocali e consonanti non sono rigorosamente omogenee, anche perché esso è specificamente concepito per ridurre la varianza tra le vocali. I modelli %V e CCI indicano invece la tendenza opposta, con fluttuazioni leggermente maggiori per le vocali che per le consonanti. Sulla base di quanto sappiamo dell'effettivo comportamento dell'italiano, quest'ultima situazione appare più consona ai dati reali. In italiano ci sono in effetti variazioni di durata relativamente apprezzabili tra vocali toniche ed atone, anche se esse sono assai più contenute di quanto spesso asserito (Bertinetto, 1981; Dell'Aglio *et alii*, 2002; Bertinetto & Loporcaro, 2005). Le variazioni sul piano della durata consonantica sono invece più contenute, e su di esse non può certo influire l'opposizione fonologica tra scempie e geminate, dato che queste ultime sono state considerate come una sequenza di due segmenti, in accordo con la maggior parte dei modelli fonologici. In sostanza, pur senza raggiungere la condizione delle lingue "a compensazione", l'italiano sembra parzialmente orientato in tale direzione, dato che non costituisce certo un esempio estremo di semplicità fonotattica.

Tenendo conto dei risultati prodotti dai vari modelli confrontati, ci sentiamo autorizzati ad asserire che, pur da questi risultati preliminari, il modello CCI sembra offrire i risultati più attendibili. In effetti, il nostro modello:

- è apparso non meno sensibile dei suoi migliori competitori nel rendere le fini differenze tra soggetti diversi (cf. figura 5);
- sembra offrire i risultati più affidabili in termini di posizionamento sul piano cartesiano.

Ovviamente, lo stadio successivo della verifica dovrà vertere sul confronto fra lingue diverse, rispetto alle quali il modello dovrà produrre raggruppamenti ragionevolmente interpretabili, sia nei rapporti reciproci tra le diverse lingue, sia soprattutto in relazione alle predizioni della teoria (cf. la figura 3). Il nostro prossimo compito consisterà dunque nel confrontare gli esiti ottenuti sull'italiano con quelli ottenibili su materiali analoghi, tratti da altre lingue. Stiamo attualmente lavorando su due corpora di dialoghi semispontanei, prodotti da parlanti tedeschi e brasiliani. Quanto al confronto con i materiali letti, ci proponiamo di studiare il comportamento dei nostri dieci parlanti nella lettura di frasi, a suo tempo registrate nella stessa occasione in cui è stato prodotto il corpus dialogato. Ci conforta non poco, comunque, l'esito ottenuto dal già citato lavoro di Mairano & Romano (in preparazione), che hanno confrontato otto lingue (tra cui la nostra) utilizzando l'algoritmo CCI, riscontrando comportamenti in linea con le nostre predizioni. In ogni caso, si può sin d'ora affermare che la copertura empirica offerta dai nostri materiali, in termini di stabilità statistica, è molto più ragguardevole di quella normalmente constatabile in questo genere di lavori, date le dimensioni del campione testato.

#### 4. CONFRONTO TRA VARIABILI

Uno degli scopi della nostra ricerca consisteva nel mettere a fuoco alcune variabili discorsive caratterizzanti il nostro corpus, per valutarne l'impatto in relazione all'andamento ritmico. Come nei casi già visti, i risultati sono rappresentati dalla proiezione di un punto sul piano cartesiano, avente come coordinate il valore dell'indice CCI calcolato

sulle vocali e sulle consonanti. I risultati verranno riportati sia graficamente sia in senso statistico, attraverso (M)ANOVA, t-test e test post-hoc. Possiamo fin d'ora asserire che i valori numerici estraibili dal nostro corpus possiedono una considerevole stabilità statistica, dato il numero di intervalli presi in esame.

La prima variabile considerata riguarda il contrasto tra enunciati dichiarativi e interrogativi, in rapporto all'eventuale influenza esercitata sulle tendenze ritmiche da questi due tipi sintattico-intonativi. La figura 6 mostra che non esiste una differenza significativa (statisticamente:  $p > 0.05$ ) per quanto riguarda tale contrasto, nonostante i punti sul piano siano separati.

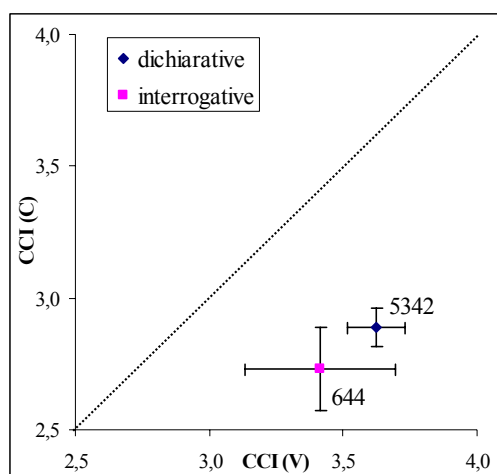


Figura 6: Confronto tra enunciati dichiarativi e interrogativi, con indicazione del numero di intervalli presi in considerazione.

Dalla medesima figura emerge, peraltro, anche il forte divario numerico tra i dati analizzati per i due gruppi a confronto, il che potrebbe aver influito sull'esito statistico. Per ovviare a ciò, è stata creata una selezione di frasi dichiarative, tratte in percentuali uguali dai dieci locutori (cf. la figura 7). In particolare, sono stati scelti enunciati tali che il valor medio della velocità di eloquio fosse analogo a quello degli enunciati interrogativi, e ciò al fine di neutralizzare la non trascurabile influenza del fattore velocità (come apparirà meglio in seguito). Sono state fatte più selezioni seguendo il criterio sopra indicato, ma in ciascun caso il risultato è rimasto invariato.

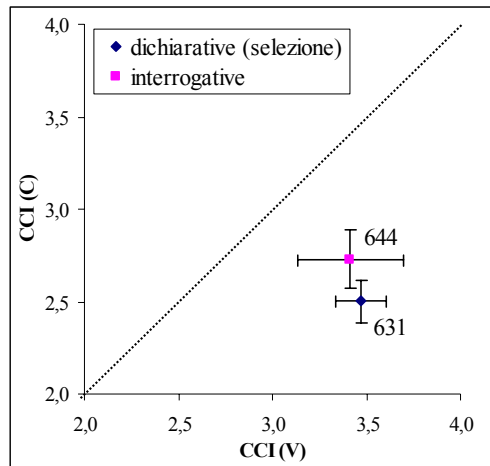


Figura 7: Confronto tra enunciati interrogativi e una selezione di enunciati dichiarativi, con numero di intervalli paragonabile e velocità di eloquio bilanciata.

Successivamente, abbiamo esaminato il contrasto tra ruolo di “Giver” e ruolo di “Follower” nei dialoghi del nostro corpus. Le registrazioni facenti parte della banca-dati AVIP/API sono state infatti realizzate con la tecnica del “map-task”, che obbliga i due locutori a svolgere un ruolo diverso nella conversazione: “Giver”, colui che dà le istruzioni sul percorso da seguire, e “Follower”, colui che deve mettere in pratica le istruzioni. Questo aspetto influisce inevitabilmente sulle strategie discorsive dei locutori e potrebbe di conseguenza influire anche sulle loro inclinazioni ritmiche. In effetti, dai nostri risultati emerge un sostanziale contrasto tra i due ruoli (cf. figura 8), confermato dalla differenza statisticamente significativa ( $F(1,234)$ ,  $p < 0.01$ ).

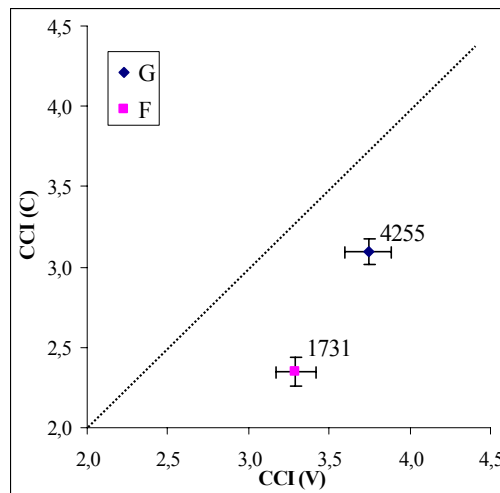


Figura 8: Confronto tra i due ruoli discorsivi (Giver vs. Follower), con relativo numero di intervalli analizzati.

Anche in questo caso, il divario tra il numero di intervalli analizzati per le due categorie a confronto potrebbe far strada all'idea che il risultato non sia veritiero. Come nel caso precedente, sono state dunque create diverse selezioni, in modo da rendere paragonabile la quantità di dati, ma il risultato è rimasto invariato. Si noti peraltro (e su ciò ritorneremo) che in questo caso non è stato possibile equilibrare la velocità di elocuzione, data la netta discrepanza che sussiste a tal riguardo tra i due ruoli discorsivi considerati.

Appare dunque evidente che tali ruoli influiscono sul comportamento ritmico dei parlanti. E poiché, nel nostro corpus di riferimento, ogni parlante assumeva un unico ruolo, è anche possibile che le differenze riscontrate abbiano a che fare con tendenze ritmiche individuali. Peraltro, osservando la figura 5 (dove i simboli pieni indicano i “Givers” e quelli vuoti i “Followers”), appare evidente come, al di là della diversità tra i singoli locutori, vi sia una ripartizione spaziale ben riconoscibile tra i due ruoli discorsivi.

Esiste tuttavia un'altra possibilità di interpretazione. Il contrasto tra i due ruoli appare infatti fortemente correlato con la velocità di elocuzione, dato che i “Givers” sono in media notevolmente più lenti dei “Followers” (6.4 vs. 7.5 sill/s). Abbiamo pertanto analizzato l'impatto del fattore velocità sulla configurazione ritmica, impatto già osservato in diversi studi sull'argomento (si vedano, per esempio, i lavori di Dellwo).

Ogni enunciato è stata catalogato a seconda della velocità di eloquio (lenta, media e veloce). Le tre classi, con valori medi equamente distribuiti, sono state così definite:

- v1: al di sotto di 6.99 sill/s (numero di intervalli vocalici e consonantici: 1301, 1308)
- v2: da 7 sill/s a 8.1 sill/s (numero di intervalli vocalici e consonantici: 660, 663)
- v3: al di sopra di 8.1 sill/s (numero di intervalli vocalici e consonantici: 514, 507)

Il grafico di figura 9 conferma quanto atteso: la diversa velocità di elocuzione ha tangibili conseguenze sull'andamento ritmico. All'aumentare della velocità, si nota una contrazione della fluttuazione nelle durate vocaliche e consonantiche, tutto sommato proporzionale anche se leggermente maggiore nel primo caso. Una prima indagine statistica indica significative differenze tra le velocità ( $F(2, 233)$ ,  $p < 0.01$ ); in particolare il test Games-Howell indica, tanto per le vocali quanto per le consonanti, una significativa ( $p < 0.01$ ) differenza tra velocità bassa e medio-alta, ma non tra alta e media.

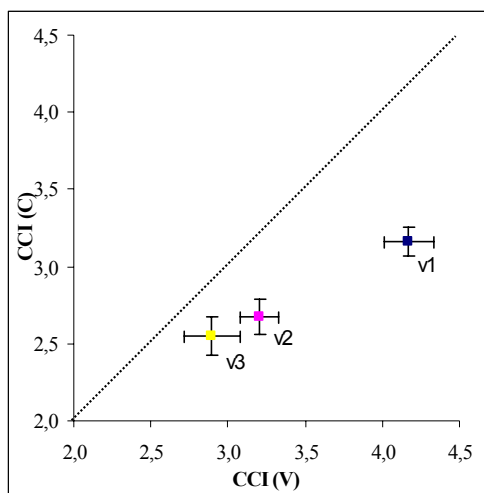


Figura 9: Effetto della velocità di eloquio (v1 = lenta, v2 = media, v3 = veloce).

Abbiamo inoltre fatto alcune prove per verificare quanto tale risultato dipendesse dalla scelta dei confini tra le classi di velocità. L'analisi ha mostrato che i risultati non sono sensibili a piccole variazioni di tali confini; per questo motivo, è stata mantenuta la sopraindicata suddivisione dello spazio delle velocità, fondata su una distribuzione equa dei rispettivi valori medi, malgrado il diverso numero di dati riferibili a ciascuna categoria. Ulteriori indagini, volte a creare nuovi e più fini raggruppamenti, sono attualmente in corso.

Alla luce di questi risultati, abbiamo rianalizzato il confronto tra "Follower" e "Giver", constatando che il suddetto contrasto è fortemente correlato con la diversa velocità di elocuzione (vedi figura 10). Le analisi statistiche hanno inoltre mostrato che il fattore "velocità" influisce maggiormente, ossia spiega una maggior quota di varianza, rispetto al fattore "ruolo discorsivo".

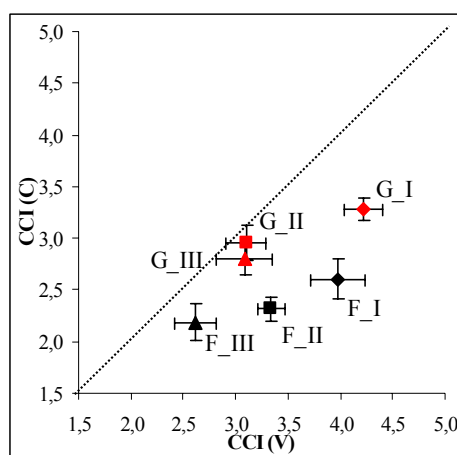


Figura 10: Confronto tra i due ruoli discorsivi (G vs. F) ripartiti secondo le velocità di elocuzione (I-III).

Sebbene risulti evidente che la variabile "velocità" è un fattore di ordine superiore rispetto alla variabile "ruolo discorsivo", resta da capire qual è l'effettivo influsso di quest'ultimo. Per dare una risposta a questa domanda, ci riserviamo di analizzare il comportamento di locutori che, nel corso del "map task", si scambiano di ruolo, come accade nei dialoghi del corpus CLIPS.<sup>4</sup>

Infine, abbiamo messo a fattore la variabile "sesso" (maschi vs. femmine), mediando sui valori riferibili ai due insiemi di locutori. Come previsto, le velocità medie differiscono: 6.3 vs. 8.0 sill/s per, rispettivamente, voce maschile e femminile. Peraltro, nonostante il grafico di figura 11 mostri come vi sia una certa differenza tra i due sessi, l'analisi statistica non rileva alcuna significatività ( $p > 0.05$ ).

<sup>4</sup> [www.clips.unina.it](http://www.clips.unina.it)



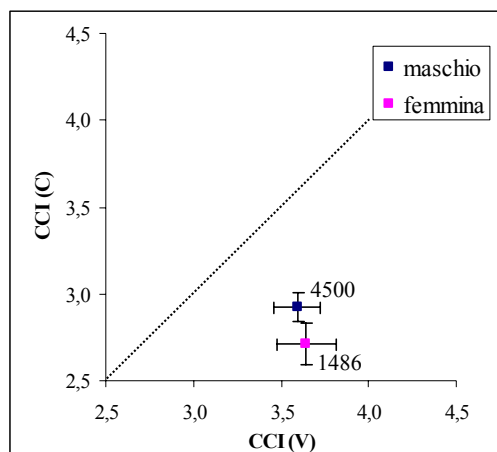


Figura 11: Confronto tra parlanti maschi e femmine, con relativo numero di intervalli analizzati.

Poiché esiste un divario notevole tra il numero di dati riferibili ai due insiemi, sono state create, in analogia con quanto fatto in precedenza, varie selezioni tra le produzioni maschili, tenendo conto del contributo di tutti i locutori e cercando per quanto possibile di equilibrare la velocità di eloquio. Il risultato è tuttavia rimasto statisticamente invariato.

## 5. CONCLUSIONI

Lo scopo di questo lavoro era duplice:

- (a) presentare un nuovo strumento per l'analisi degli andamenti ritmici del parlato;
- (b) applicarlo allo studio di alcune cruciali variabili discorsive individuabili nel corpus di riferimento, basato sulla varietà pisana dell'italiano registrata in condizioni semi-spontanee.

Quanto al primo punto, il confronto effettuato con altri algoritmi indica che il modello CCI sembra presentare alcuni indubbi vantaggi (cf. § 3). Ulteriori considerazioni in proposito, basate su accurati confronti statistici, si possono trovare in altri nostri studi (Bertinetto & Bertini 2008; Bertinetto & Bertini, in stampa).

L'analisi delle variabili discorsive ha mostrato (cf. § 4) che alcuni fattori influiscono fortemente sulla configurazione ritmica, mentre altri sembrano esercitare effetti piuttosto contenuti. In particolare, è emersa la forte influenza della velocità di eloquio e del ruolo del locutore (in quest'ordine gerarchico di importanza), a differenza di variabili come il tipo sintattico-intonativo (enunciati dichiarativi vs. interrogativi) o il sesso del parlante.

Questa ricerca ribadisce anche che l'utilizzo di un corpus debitamente ampio è assolutamente indispensabile per conferire sufficiente solidità statistica alle osservazioni. Ciò emerge in particolare dalla stabilità dei nostri esiti in tutti i casi in cui si è provato ad applicare il calcolo a selezioni diverse di enunciati, allo scopo di bilanciare le variabili in esame. Il che suggerisce, per converso, di considerare con il dovuto beneficio di inventario molti dati presenti in letteratura circa l'andamento ritmico delle lingue, quand'essi siano basati su un numero troppo limitato di dati o addirittura su un solo locutore.

## BIBLIOGRAFIA

- Barbosa, P. (2006), *Incursões em torno do ritmo da fala*, Campinas: Pontes.
- Bertinetto, P.M. (1981), Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici, Firenze: Accademia della Crusca.
- Bertinetto, P.M. (1989), Reflections on the dichotomy 'stress- vs. syllable-timing', *Revue de Phonétique Appliquée*, 91/93. 99-130.
- Bertinetto, P.M. & Bertini, C. (2008), On modeling the rhythm of natural languages: the 'Control/Compensation Index'. *Speech Prosody 2008*, Campinas, Brasil, May 6-9.
- Bertinetto, P.M. & Bertini, C. (in stampa), Modelización del ritmo y estructura silábica, con aplicación al italiano, in Sanchez Miret, F. (cur.), *Romanística sin complejos: Homenaje a Carmen Pensado*, Bern: Lang.
- Bertinetto, P.M. & Bertini, C. (in preparazione), Towards a unified predictive model of natural language rhythm, comunicazione presentata al convegno *Empirical Approaches to Language Rhythm*, Londra, Inghilterra, 28 marzo 2008.
- Bertinetto, P.M. & Fowler, C.A. (1989), On sensitivity to durational modifications in Italian and English, *Rivista di Linguistica*, 1. 69-94.
- Bertinetto, P.M. & Loporcaro, M. (2005), The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome, *Journal of the International Phonetics Association* 35: 131-151.
- Bertinetto, P.M. & Vékás, D. (1991), Controllo vs. compensazione: sui due tipi di isocronia, in *L'interfaccia tra fonologia e fonetica*, Magno Caldognetto, E. & Benincà, P. (eds.), Padova: Unipress.
- Dauer, R.M. (1983), Stress-timing and syllable-timing reanalyzed, *Journal of Phonetics*, 11. 51-62.
- Dellwo, V. (2004), Rhythm and speech rate: A variation coefficient for  $\Delta C$ , in *Language and language-processing*, Karnowski, P. & Szigeti, I. (ed.), Frankfurt am Main: Peter Lang. 231-241.
- Fowler, C.A. (1977), *Timing Control in Speech Production*, Indiana University Linguistics Club.
- Grabe, E. & Low, E.L. (2002), Durational variability in speech and the rhythm class hypothesis, *Papers in Laboratory Phonology* 7, 515-546.
- Hoeqvist, C.E. Jr (1983), The perceptual centre and rhythm categories, *Language and Speech*, 26, 4: 367-376.
- Low, E.L., Grabe, E. & Nolan, F. (2000), Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English, *Language and Speech*, 43 (4), 377-401.
- Mairano, P. & Romano, A. (2006), Lingue isosillabiche e isoaccentuali: misurazioni strumentali su campioni di italiano, francese, inglese e tedesco, in *Atti del III Convegno Nazionale AISV Scienze vocali e del linguaggio*, Trento, Italia, novembre- dicembre 28-1, CD-rom.

Mairano, P. & Romano, A. (in preparazione), A comparison of four rhythm metrics for six languages, *Workshop on Empirical Approaches to Speech Rhythm*, Londra, Inghilterra, 28 marzo 2008.

O'Dell, M. & Nieminen, T. (1999), Coupled oscillator model of speech rhythm, *Proceedings 14° Int. Cong. Phon. Sciences*, San Francisco, U.S.A., 2: 1075-1078.

Pike (1947), Pike, K.L. [1945], *The Intonation of American English*, Ann Arbor.

Ramus, F., Nespor, M. & Mehler, J. (1999), Correlates of linguistic rhythm in the speech signal, *Cognition*, 73. 265-292.

Romano A. (2003), Accento e intonazione in un'area di transizione del Salento centro-meridionale, in P. Radici Colace, G. Falcone & A. Zumbo (a cura di), *Storia politica e storia linguistica dell'Italia meridionale*, Atti del convegno internazionale di studi parlangeliani (Messina, 22-23 Maggio 2000), Messina-Napoli, Ed. Scientifiche Italiane, 169-181.

Rouas, J.L. & Farinas, J. (2004), Comparaison de méthodes de caractérisation du rythme des langues, *Workshop MIDL (Modélisation pour l'Identification Des Langues)*, Paris, France, novembre 29-30, 45-50.

Russo, M. & Barry, W.J. (2004), Interaction between segmental structure and rhythm. A look at Italian Dialects and Regional Standard Italian, *Folia Linguistica*, 38, 3-4, 277-296.

Schmid, S. (2004), Une approche phonétique de l'isochronie dans quelques dialectes italo-romans, in *Nouveaux départs en phonologie. Les conceptions sub- et suprasegmentales*, Meisenburg T. & Selig M. (curr.), Tübingen : Narr, 109-124.

# SULLA REALTÀ ACUSTICA DELL'ACCENTO LESSICALE IN ITALIANO ED IN SPAGNOLO: LA DURATA VOCALICA IN PRODUZIONE E PERCEZIONE

Iolanda Alfano\*, Renata Savy<sup>^</sup>, Joaquim Llisterri\*

\*Universitat Autònoma de Barcelona, <sup>^</sup>Università degli Studi di Salerno  
[Iolanda.Alfano@campus.uab.cat](mailto:Iolanda.Alfano@campus.uab.cat), [rsavy@unisa.it](mailto:rsavy@unisa.it), [Joaquim.Llisterri@uab.cat](mailto:Joaquim.Llisterri@uab.cat)

## 1. SOMMARIO

Il presente lavoro mira ad approfondire aspetti controversi relativi alle caratteristiche acustiche in produzione e in percezione dell'accento lessicale in italiano e in spagnolo, emersi in precedenti studi realizzati sulle due lingue con analoghe e pertanto comparabili metodologie sperimentali.

L'idea che muove i lavori cui facciamo riferimento (Llisterri *et al.*, 2005; Alfano, 2006; Alfano *et al.*, 2007) nasce dall'intento di valutare il peso dei correlati acustici dell'accento lessicale in percezione. Tali indagini sono state condotte partendo da esperimenti percettivi realizzati con stimoli (parole e non parole in isolamento) ottenuti in seguito alla manipolazione sistematica di durata e frequenza fondamentale.

A dispetto delle rilevanti proprietà strutturali condivise dai due sistemi linguistici, in italiano la durata risulta giocare un ruolo dominante (Alfano, 2006), mentre in spagnolo la  $f_0$  si delinea come il parametro responsabile della percezione dell'accento, se manipolato però insieme alla durata o all'intensità (Llisterri *et al.*, 2005).

I risultati mostrano, per ambedue le lingue, la natura complessa del processo di percezione, ma ancor più articolata si delinea l'analisi dei dati relativi allo stesso protocollo sperimentale impiegato in un lavoro linguisticamente incrociato, con stimoli in italiano e parlanti nativi di spagnolo (Alfano *et al.*, 2007).

Tale esperimento era stato messo a punto per capire se, ed eventualmente in che modo, la strategia percettiva dipende dalla lingua madre; i soggetti spagnoli paiono reagire all'ascolto di input in italiano diversamente da come fanno nella propria lingua. Mentre, infatti, in spagnolo si mostrano del tutto insensibili alla sola alterazione del fattore temporale, in italiano reagiscono agli stimoli con la durata alterata presentando evidenti difficoltà, sia nell'identificazione che nella discriminazione. Un'analisi del grado di coerenza delle risposte in relazione alla tipologia di modifica mostra, in conseguenza della manipolazione della durata, una situazione fortemente problematica. Data la peculiarità del comportamento dei soggetti in risposta all'alterazione di questo parametro, ci è parso necessario effettuare un'analisi delle caratteristiche acustiche di durata vocalica in italiano e in spagnolo.

Presentiamo, dunque, un'analisi contrastiva della durata vocalica nei *corpora* impiegati negli esperimenti rispetto a: a) posizione della vocale, b) sua tonicità/atonicità, c) profilo accentuale della parola. Tale indagine rileva differenze interessanti e talvolta inaspettate tra i due sistemi, che porgiamo altresì come chiave interpretativa della strategia percettiva impiegata in lingua straniera.

## 2. INTRODUZIONE

È ben noto che l'italiano e lo spagnolo hanno in comune proprietà strutturali di non poco peso, in primo luogo come diretta conseguenza dell'appartenenza di entrambe al

ceppo romanzo. In merito alle caratteristiche di strutture sillabiche ed accentuali, possiamo individuare diverse e forti analogie, tra le quali sottolineiamo che ambedue i sistemi sono:

- a sede accentuale libera in cui, oltre che una funzione delimitativa, compete all'accento anche una funzione distintiva, (Bertinetto, 1981a); l'accento è dotato di valore fonologico in quanto distingue significati (ad esempio in italiano *capito* vs. *capitò* ed in spagnolo *bebe* "beve" vs. *bebé* "bebè").
- Ad accento considerato dinamico: dal punto di vista dell'impressione percettiva, per ambedue le lingue la sillaba tonica è descritta come pronunciata con maggiore intensità rispetto alle atone, cioè con un presunto maggior sforzo nell'emissione dell'aria ed una maggior durata<sup>1</sup>, sebbene il peso attribuito ai correlati in percezione sia parzialmente distinto<sup>2</sup>.
- Con tendenza verso l'isosillabicità;
- con una netta prevalenza del profilo piano in termini di frequenza di occorrenza nei due lessici (italiano, 84%, Mancini e Voghera (1994), spagnolo, 79%, Quilis (1993)<sup>3</sup>;
- con struttura sillabica CV più frequente (italiano, 57,74%, Mancini e Voghera (1994), spagnolo, 59,86% Monroy-Casas (1980) e con una quasi totale coincidenza di tipi sillabici e della loro relativa distribuzione nel lessico<sup>4</sup>.

La portata delle caratteristiche strutturali condivise induce ad ipotizzare una forte somiglianza nei meccanismi di realizzazione dell'accento lessicale; le robuste affinità, delle quali abbiamo menzionato le principali, suggeriscono dunque l'ipotesi che l'accento si realizzi in maniera essenzialmente analoga tanto sul piano acustico quanto su quello percettivo.

Tuttavia, esperimenti percettivi realizzati sulle due lingue con soggetti nativi e non, con la stessa metodologia sperimentale, che riassumiamo brevemente di seguito, mostrano, invece, una situazione parzialmente diversa.

---

<sup>1</sup> Si vedano, tra gli altri, per l'italiano Bertinetto (1981a), per lo spagnolo Navarro Tomás (1974) e Quilis (1993).

<sup>2</sup> Mentre, infatti, in italiano il correlato ritenuto più stabile è stato unanimemente la durata, in spagnolo si trovano in letteratura opinioni discordanti, che attribuiscono una diversa rilevanza alla durata, alla frequenza fondamentale e all'intensità (per una rassegna sul tema, si veda Llisterri *et al.*, 2005).

<sup>3</sup> Segue in italiano il profilo proparossitono con il 10% e quello ossitono con il 6%, mentre in spagnolo, rispettivamente il primo raggiunge solo il 3% ed il secondo il 18% (in tale computo sono state escluse le parole monosillabiche).

<sup>4</sup> Confrontando i due sistemi, l'unica differenza rilevante risulta una maggior percentuale del tipo CVC in spagnolo rispetto all'italiano: rispettivamente 22% (Monroy-Casas, 1980) vs. 14% (Mancini e Voghera, 1994).

### 3. VERSANTE PERCETTIVO: IL DIVERSO RUOLO DEI CORRELATI IN L1 E IN L2

#### 3.1 Disegno sperimentale

I *corpora* impiegati in tali esperimenti consistono di trisillabi pronunciati in isolamento (parole e non parole), con struttura sillabica CV, suscettibili di portare accento sulla prima, sulla seconda e sulla terza sillaba (la tabella 1 mostra alcuni esempi<sup>5</sup>).

| italiano                        | spagnolo                      |
|---------------------------------|-------------------------------|
| parole                          |                               |
| [ˈse:mino] [seˈmi:no] [semiˈno] | [ˈlimite] [liˈmite] [limiˈte] |
| [ˈka:pito] [kaˈpi:to] [kapiˈto] | [ˈmeðiko] [meˈðiko] [meðiˈko] |
| non parole                      |                               |
| [ˈga:mino] [gaˈmi:no] [gamiˈno] | [ˈmaleðo] [maˈleðo] [maleˈðo] |
| [ˈna:vilo] [naˈvi:lo] [naviˈlo] | [ˈnuliβo] [nuˈliβo] [nuliˈβo] |

Tabella 1: esempi di item dei *corpora* impiegati in italiano (sinistra) ed in spagnolo (destra).

L'analisi acustica è stata effettuata con il programma *Praat* (Boersma e Weenink, 2005). Per ciascuna vocale sono stati calcolati:

- durata;
- frequenza fondamentale ad inizio, centro e fine vocale.

Per la preparazione degli stimoli sono stati sostituiti i valori originali con quelli medi trovati: è stata così creata una versione di ogni parola tramite l'algoritmo di risintesi del metodo PSOLA (*Pitch-Synchronous Overlap-Add*) incorporato nelle funzioni di *Praat*.

Gli stimoli manipolati proposti nei test sono stati come di seguito costruiti: alle parole parossitone sono stati sostituiti i valori di durata e di  $f_0$  delle corrispettive parossitone (che indichiamo con PP>P) ed alle parossitone i valori, precedentemente calcolati nell'analisi, delle ossitone (d'ora in poi P>O). Considerando, per esempio, la tripletta di non parole [ˈga:mino gaˈmi:no gamiˈno], a [ˈga:mino] sono stati sostituiti i valori di [gaˈmi:no] e a [gaˈmi:no] quelli di [gamiˈno].

Durata ed  $f_0$  sono state manipolate prima singolarmente e poi congiuntamente, ottenendo così stimoli con la sola durata modificata, altri con la sola frequenza ed altri ancora con entrambi i correlati alterati.

Il test prevedeva due prove di percezione, una di identificazione e l'altra di discriminazione. Nella prima si chiedeva a 30 soggetti di indicare quale sillaba percepivano come tonica; nella seconda si chiedeva loro di definire coppie di parole (di cui una parola era originale e l'altra aveva uno o più valori manipolati) come uguali o diverse rispetto al profilo accentuale. In particolare, nel compito di discriminazione, gli stimoli sono stati proposti in coppie, in numero bilanciato di originali e modificati, costituite da ogni stimolo manipolato combinato sia con il corrispettivo originale, sia con il corrispettivo stimolo del quale erano stati misurati e sostituiti i valori del/i correlato/i (ad es.: *gamino*

<sup>5</sup> Per consultare i *corpora* interi, si vedano le tabelle 3 e 4 poste in appendice. Mentre il *corpus* utilizzato nell'esperimento sullo spagnolo è costituito unicamente da triplette, quello sull'italiano è formato anche da coppie, data l'impossibilità di trovare triplette di trisillabi che rispettassero le caratteristiche sopra esposte.

[ga'mi:ɲo] modificato è stato proposto sia con *gamino* [ga'mi:ɲo] che con *gaminò* [gami'ɲo]). Ciascuno stimolo modificato è stato cioè accoppiato con la corrispettiva parola originale e, come controprova, anche con quella “target”, cioè quella di cui è stato calcolato il valore per la sostituzione. Per coerenza, l'uguaglianza nel primo caso implica la differenza nel secondo e viceversa: *gamino* [ga'mi:ɲo] modificato non può essere uguale sia a *gamino* [ga'mi:ɲo] che a *gaminò* [gami'ɲo]. Nell'interpretazione dei risultati si è dunque tenuto conto delle risposte relative ai due tipi di prove effettuate (identificazione e discriminazione), ma anche del grado di coerenza ottenuto nel test di discriminazione.

### 3.2 Risultati in L1

Confrontando gli esperimenti realizzati con soggetti nativi<sup>6</sup>, emergono varie differenze nelle strategie percettive impiegate da soggetti italiani e spagnoli: il ruolo svolto dai correlati acustici non risulta coincidente.

Presentiamo di seguito i risultati relativi all'identificazione delle parole; non riportiamo quelli che concernono le non parole e la prova di discriminazione, poiché mostrano le stesse tendenze.

L'alterazione della frequenza fondamentale in italiano non comporta un cambio percettivo: i soggetti non si mostrano sensibili alle modifiche di questo correlato, giacché continuano a percepire la prominente originale (figura 1a). Diversamente in spagnolo, pur non risultando tale manipolazione determinante, essa è in qualche modo percepita (figura 1b).

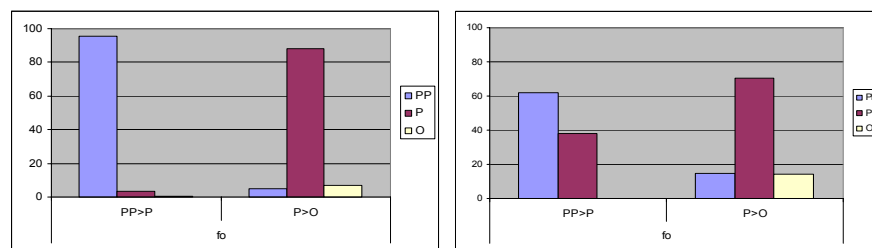


Figura 1: a) risultati della prova di identificazione (%) di parole con  $f_0$  manipolata in italiano; b) risultati della prova di identificazione (%) di parole con  $f_0$  manipolata in spagnolo. PP=proparossitone; P=parossitone; O=ossitone; PP>P=proparossitone con valori di parossitone; P>O=parossitone con valori di ossitone.

In risposta a stimoli con la durata manipolata invece, mentre i soggetti italiani percepiscono la modifica in circa il 60% dei casi (figura 2a), gli spagnoli risultano del tutto indifferenti a questo tipo di alterazione (figura 2b).

<sup>6</sup> Per l'italiano Alfano (2006) e per lo spagnolo Llisterri *et al.* (2005).

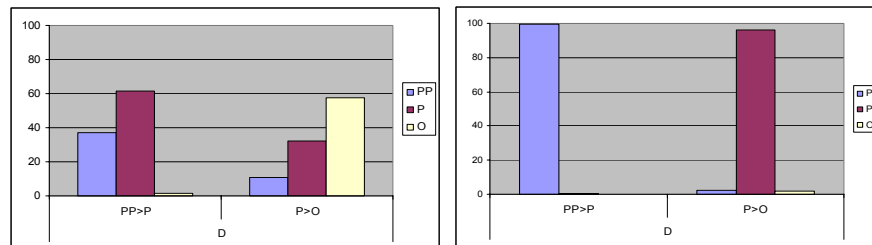


Figura 2: a) risultati della prova di identificazione (%) di parole con la durata manipolata in italiano; b) risultati della prova di identificazione (%) di parole con la durata manipolata in spagnolo. PP=proparossitone; P=parossitone; O=ossitone; PP>P=proparossitone con valori di parossitone; P>O=parossitone con valori di ossitone.

La manipolazione poi congiunta di ambedue i parametri provoca un cambio nella percezione della sede accentuale in modo considerevole in entrambe le lingue (si vedano per l'italiano la figura 3a e per lo spagnolo la 3b).

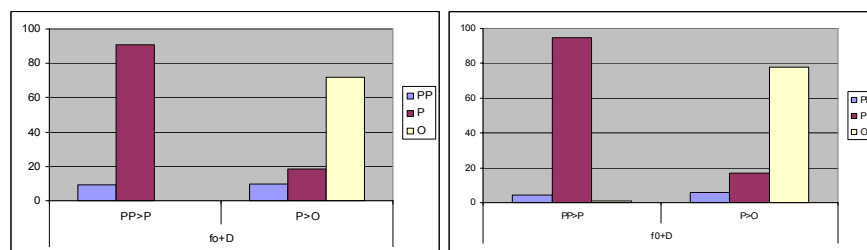


Figura 3: a) risultati della prova di identificazione (%) di parole con la durata e la  $f_0$  manipolate in italiano; b) risultati della prova di identificazione (%) di parole con la durata e la  $f_0$  manipolate in spagnolo. PP=proparossitone; P=parossitone; O=ossitone; PP>P=proparossitone con valori di parossitone; P>O=parossitone con valori di ossitone.

Il dato più interessante che emerge dal confronto riguarda il diverso comportamento dei soggetti italiani e spagnoli in risposta all'alterazione della durata. La sola manipolazione della frequenza risulta irrilevante in italiano e non determinante in spagnolo (figura 1a e 1b); la modifica congiunta di durata e frequenza provoca un cambio nella percezione della sede accentuale in percentuali molto simili (figura 3a e 3b).

Diversamente, mentre in italiano l'alterazione della sola durata comporta conseguenze sul piano percettivo, essa non viene affatto percepita in spagnolo (figura 2a e 2b).

### 3.3 Risultati in LS

Con il fine di capire se ed in che modo la strategia percettiva dipenda dalla lingua madre, è stato realizzato un esperimento analogo ai precedenti, seguendo lo stesso protocollo sperimentale, ma con stimoli in italiano e soggetti nativi di spagnolo (Alfano *et al.*, 2007).

Per evitare possibili meccanismi di interferenza tra le due lingue, sono stati esclusi dal *corpus* tutti gli stimoli che avevano una diretta somiglianza con parole spagnole<sup>7</sup>; la tabella 2 mostra gli stimoli impiegati nell'esperimento.

<sup>7</sup> Si confronti la tabella 2 con la 3 (§ Appendice).



| parole                                    |
|-------------------------------------------|
| [ˈse:mino] [seˈmi:no] [semiˈno]           |
| [ˈle:gami] [leˈga:mi] [roˈvino] [roviˈno] |
| [ˈka:pito] [kaˈpi:to] [kapiˈto]           |
| non parole                                |
| [ˈna:vilo] [naˈvi:lo] [naviˈlo]           |
| [ˈma:ligo] [maˈli:go] [maliˈgo]           |
| [ˈla:gano] [laˈga:no] [lagaˈno]           |

Tabella 2: *corpus* impiegato nell'analisi della percezione dell'accento lessicale in italiano LS (Alfano *et al.*, 2007).

In maniera del tutto analoga agli esperimenti realizzati sulla percezione in L1 precedentemente descritti, 30 parlanti nativi di spagnolo, di età compresa tra 19 e 22 anni, hanno svolto una prova di identificazione e una di discriminazione. Sono stati considerati tre gruppi di soggetti, in relazione al loro grado di conoscenza di italiano:

- gruppo a): aveva studiato italiano da 5 a 9 mesi;
- gruppo b): non aveva mai studiato italiano, ma aveva un alto grado di familiarità con questa lingua grazie ad esempio a viaggi in Italia o a relazioni con italiani;
- gruppo c): non aveva mai studiato italiano, né lo considerava familiare.

L'ipotesi iniziale era che i parlanti nativi di spagnolo (in particolar modo coloro che non avevano mai studiato l'italiano) impiegassero in LS strategie percettive fortemente condizionate da quelle impiegate in L1, riportando risultati dunque sostanzialmente diversi da quelli dei parlanti italiani. I risultati ottenuti, invece, svelano un comportamento quantomeno singolare.

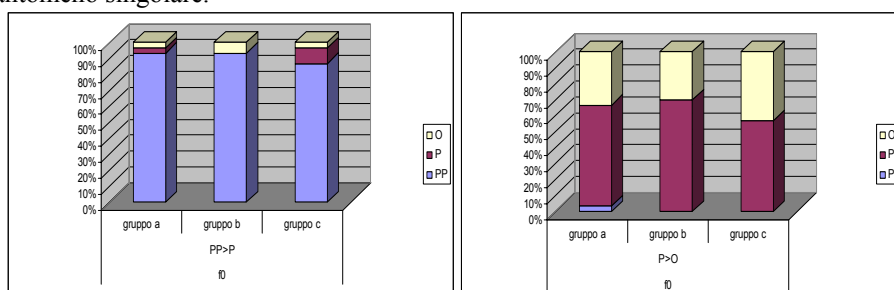


Figura 4: a) risultati della prova di identificazione (%) di PP>P (proparossitone con valori di parossitone) con  $f_0$  manipolata; b) risultati della prova di identificazione (%) di P>O (parossitone con valori di ossitone) con  $f_0$  manipolata. PP=proparossitone; P=parossitone; O=ossitone.

La manipolazione di  $f_0$  non provoca nessun cambio nella percezione della sede accentuale in maniera inequivocabile per la categoria PP>P (proparossitone con valori di parossitone, figura 4a) ed in percentuali meno elevate per P>O (parossitone con valori di ossitone, figura 4b).

L'alterazione della durata non implica un cambio in percezione, ma provoca un forte stato di confusione e di incoerenza nelle risposte; i soggetti di tutti e tre i gruppi si mostrano

incapaci di identificare gli stimoli con la durata modificata in italiano (figure 5a e 5b), laddove in spagnolo invece risultano indifferenti a questa modifica (figura 2b).

Come per i dati relativi alla percezione in L1, anche in questo caso non abbiamo presentato i risultati della prova di discriminazione, poiché convergono con quelli ottenuti nell'identificazione. Ci preme però sottolineare che in risposta agli stimoli con la durata manipolata, essi indicano chiaramente lo stesso problema che emerge nell'identificazione, mostrando un grado di coerenza quanto mai basso, che non si riscontra in nessun altro caso: come mostra il grafico in figura 5c, le PP>P vengono discriminate come diverse da una PP nel 66,7% dei casi (gruppo a), nell' 83,3% (gruppo b) e nell' 83,3% (gruppo c), ma, allo stesso tempo, diverse anche da una P nell'83,3% dei casi (gruppo a), nel 53,3% (gruppo b) e nel 63,3% (gruppo c); essenzialmente analoga risulta poi la confusione e l'incoerenza in risposta alle P>O (figura 5d).

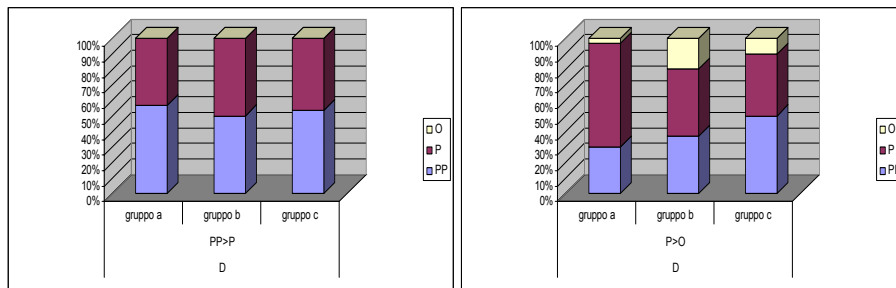


Figura 5: a) risultati della prova di identificazione (%) di PP>P (proparossitone con valori di parossitone) con la durata manipolata; b) risultati della prova di identificazione (%) di P>O (parossitone con valori di ossitone) con la durata manipolata; PP=proparossitone; P=parossitone; O=ossitone.

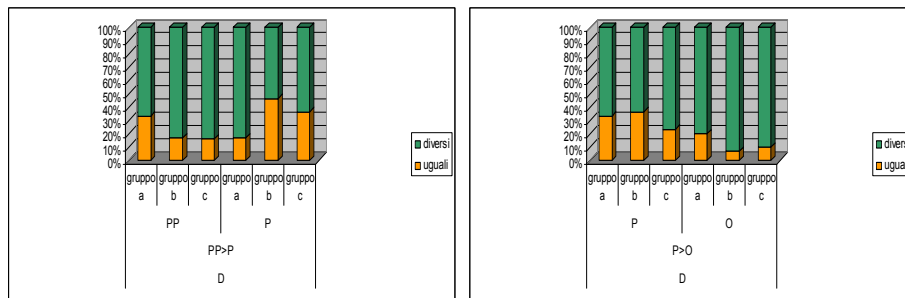


Figura 5: c) risultati della prova di discriminazione (%) di PP>P (proparossitone con valori di parossitone) con la durata manipolata; d) risultati della prova di discriminazione (%) di P>O (parossitone con valori di ossitone) con la durata manipolata; PP=proparossitone; P=parossitone; O=ossitone.

Infine, così come negli esperimenti percettivi in L1, in conseguenza poi della manipolazione congiunta di durata e frequenza, i soggetti percepiscono un cambio di prominenza in maniera chiara, ma in percentuali più alte nel caso della categoria PP>P rispetto a P>O (si osservino le figure 6a e 6b).

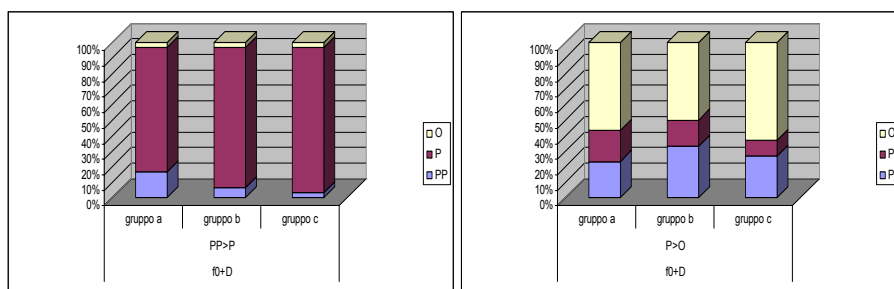


Figura 6: a) risultati della prova di identificazione (%) di PP>P (proparossitone con valori di parossitone) con  $f_0$  e durata manipolate; b) risultati della prova di identificazione (%) di P>O (parossitone con valori di ossitone) con  $f_0$  e durata manipolate. PP=proparossitone; P=parossitone; O=ossitone.

### 3.3.1 Percezione rispetto al grado di conoscenza della LS e differenza tra parole e non parole

Per considerare la variabile della conoscenza di italiano, come precedentemente accennato (§ 3.3), sono stati considerati tre gruppi da 10 soggetti.

L'ipotesi più ragionevole sembrava quella secondo cui la conoscenza della lingua straniera avvicina il comportamento di parlanti non nativi a quello dei nativi. Ci aspettavamo, cioè, che il gruppo a si differenziasse per un'attitudine più vicina a quella degli italiani nella loro L1 e che il gruppo c si comportasse in maniera non molto diversa da quanto avrebbe fatto nella propria L1.

Si osservino le figure 6a e 6b, relative all'identificazione di parole con ambedue i parametri manipolati. Confrontando il comportamento di ciascun gruppo, non emergono differenze molto grandi; i soggetti del gruppo c, però, percepiscono un cambio di pattern in percentuali più alte rispetto agli altri due: per PP>P, 93,4% rispetto a 80% del gruppo a e 90% del gruppo b; per P>O, 63,3% rispetto a 56,7% del gruppo a e 50% del gruppo b.

Si può avanzare l'ipotesi, in prima istanza, che il gruppo c si distingua in quanto, diversamente dai gruppi a e b, tende ad affidarsi unicamente alle caratteristiche acustiche degli stimoli ed a nessun tipo di conoscenza lessicale.

Non abbiamo presentato finora i risultati relativi alle non parole, poiché negli esperimenti condotti con soggetti nativi non si riscontrano differenze importanti con quelli relativi alle parole. Crediamo invece sia opportuno considerare in questo caso anche le non parole, poiché il comportamento dei soggetti spagnoli in italiano LS è parzialmente diverso in risposta ai due tipi di item. Si osservino i grafici che seguono (figura 7); nel caso delle P>O (parossitone con valori di ossitone), sulla destra di ogni grafico, mentre in risposta alle parole vi è una percentuale di casi in cui vengono percepite PP, questa confusione non avviene in risposta alle non parole.

Anche questo dato ci sembra suggerire l'ipotesi che nel caso delle non parole la percezione sia guidata prettamente dagli indici acustici e non influenzata da altri tipi di fattori.

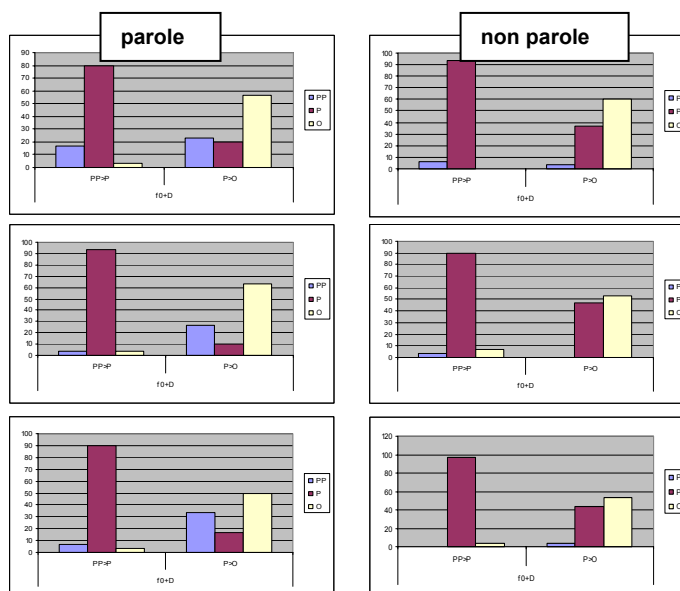


Figura 7: risultati dell'identificazione di parole (sinistra) e non parole (destra) del gruppo a (in alto), gruppo b (al centro) e gruppo c (in basso).

### 3.4 Risultati in spagnolo L1 vs. italiano LS

Confrontando globalmente il ruolo dei correlati in percezione in L1 rispetto a quanto avviene in italiano LS, la differenza più importante riguarda la durata: soggetti nativi di spagnolo, che ignorano le modifiche di durata nella propria L1, pur non percependo in italiano LS un cambio netto di profilo accentuale, risultano tutt'altro che indifferenti a questa manipolazione.

Pur consapevoli di un rapporto di non biunivocità tra il piano percettivo e quello acustico, abbiamo ritenuto necessario confrontare le caratteristiche acustiche dei *corpora* impiegati nelle due lingue, ritenendo che eventuali differenze in produzione potessero essere d'aiuto per la comprensione di quanto accade in percezione.

## 4. VERSANTE ACUSTICO: LA DURATA VOCALICA

Presentiamo, dunque, un'analisi della durata vocalica dei *corpora* impiegati negli esperimenti<sup>8</sup>.

Globalmente, la media della durata vocalica delle toniche risulta maggiore in italiano (165,9 ms) rispetto allo spagnolo (142,9 ms). Considerando poi ciascun pattern, emergono differenze importanti (si veda la figura 8):

<sup>8</sup> La tabella 5, posta in appendice, mostra un confronto di tutti i dati rispetto alla posizione della vocale, alla sua tonicità/atonicità ed al profilo accentuale della parola. L'analisi che si espone di seguito considera - tanto in italiano quanto in spagnolo - i valori medi di durata vocalica in *corpora* costituiti da parole pronunciate in isolamento: i dati ottenuti sono, quindi, naturalmente lontani dalla realtà del parlato connesso e sono da considerare, certamente, come misure relative e non come parametri assoluti.

- le toniche più lunghe in italiano sono quelle delle P, con una media di 175,6 ms, mentre in spagnolo quelle delle O, che durano mediamente 178,25 ms.
- le toniche di PP e P sono più lunghe in italiano che in spagnolo: le PP risultano, in media, rispettivamente di 163,67 vs. 117,61 ms e le P di 175,6 vs. 132,88 ms.
- le toniche di O risultano più lunghe in spagnolo (178,25 ms) che in italiano (156,82 ms).

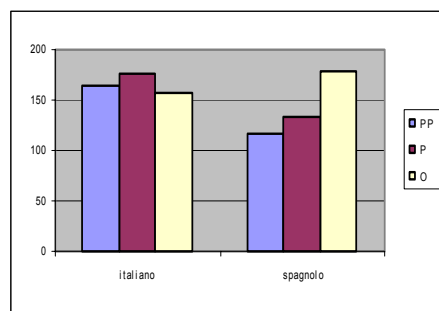


Figura 8: durata vocalica media (ms) in PP (proparossitone), P (parossitone) e O (ossitone) in italiano e spagnolo.

Analizzando la differenza tra toniche finali (cioè delle O) e non finali (considerando una media tra le toniche delle PP e delle P), mentre in italiano le finali (in media, 156,82 ms) sono leggermente più brevi delle non finali (170,1 ms), in spagnolo risultano decisamente più lunghe (178,25 vs 125,24 ms)<sup>9</sup>.

Guardando invece alle vocali atone, la media dell'italiano (102,9 ms) è molto simile a quella dello spagnolo (109,9 ms).

Anche distinguendo tra finali e non finali, la tendenza risulta la stessa: considerando la media delle vocali atone finali di PP e P, essa è in italiano 133,47 ms ed in spagnolo 146,79 ms; la media delle atone non finali di PP, P ed O risulta in italiano 87,01 ms ed in spagnolo 91,45 ms (figura 9).

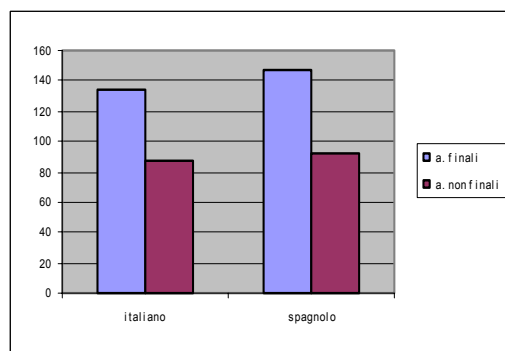


Figura 9: durata vocalica media (ms) delle atone finali (considerando la media tra PP e P) vs. non finali (valutando la media tra PP, P e O) in italiano e in spagnolo.

<sup>9</sup> La minor durata delle vocali toniche finali rispetto alle non finali in italiano è un dato ben noto e documentato in letteratura (si vedano, tra gli altri, Nespor & Vogel, 1986; Marotta, 1985).

La differenza poi che si dà in italiano tra pretoniche e post-toniche (89,59 vs 79,23 ms) risulta in spagnolo praticamente nulla (93,77 vs 92,38 ms), come mostra la figura 10.

Le post-toniche, dunque, sembrano tendenzialmente comprimersi in italiano, ma non in spagnolo, dato che suggerisce l'ipotesi di un'isosillabicità più stabile in spagnolo che in italiano<sup>10</sup>.

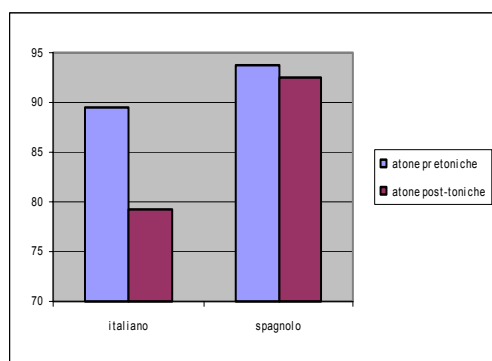


Figura 10: durata vocalica media (ms) delle atone pretoniche vs. post-toniche in italiano e in spagnolo.

Considerando poi la differenza tra toniche ed atone nelle due lingue (si osservi la figura 11), la tonicità risulta fortemente marcata in termini di differenze di durata per l'italiano, mentre in spagnolo ciò risulta vero per le O, ma non per le PP e le P, per le quali lo scarto tra toniche e atone è decisamente ridotto: rispettivamente, 117,61 vs. 118,45 ms (media delle atone delle PP) e 132,88 vs. 120,3 ms (media delle atone di P) (laddove esso risulta in italiano 163,67 vs. 100,56 ms e 175,60 vs. 116,65 ms).

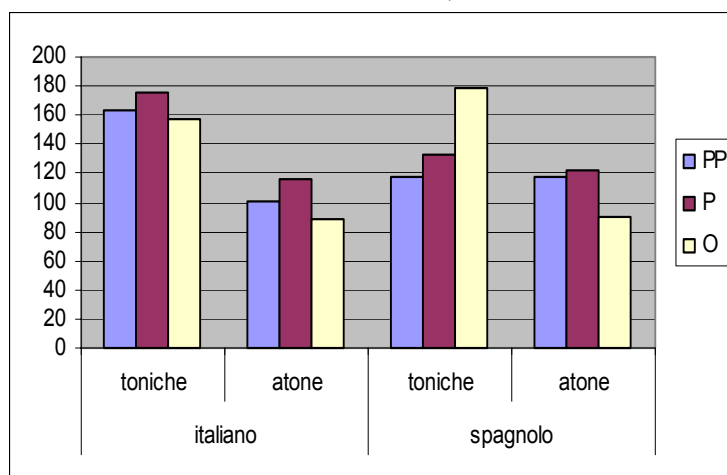


Figura 11: durata vocalica media (ms) delle toniche vs. atone di PP (proparossitone), P (parossitone) e O (ossitone) in italiano e in spagnolo.

<sup>10</sup> Si vedano, tra gli altri, per l'italiano Bertinetto (1981b, 1989, 1990) e Marotta (1985), per lo spagnolo Almeida (1997).

Distinguendo poi tra atone non finali e finali, confrontandole con le toniche (rispettivamente, figura 12 e 13), mentre le vocali toniche (considerando una media tra le toniche di PP, P ed O, pari a 165,9 ms) risultano in italiano più lunghe sia delle atone non finali (in media, 87,01 ms) che delle finali (in media, 133,47 ms), in spagnolo le toniche sono (valutando sempre una media tra i tre profili accentuali, pari a 142,9 ms) più lunghe delle atone non finali (91,45 ms), ma risultano più brevi delle finali (146,79).

Seppure, quindi, in italiano come in spagnolo le atone finali risultano in media più lunghe delle non finali, le toniche hanno, in ogni caso, in italiano una maggior durata rispetto alle atone. Diversamente, in spagnolo la media delle atone finali risulta addirittura maggiore della media delle toniche, confermando quanto già emerso riguardo a due aspetti:

- tonicità acusticamente marcata in maniera forte da variazioni di durata in italiano;
- relativa maggior stabilità della durata in spagnolo rispetto all'italiano.

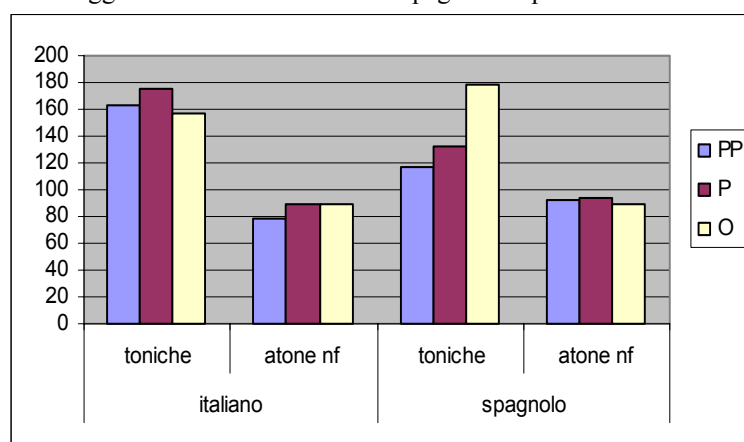


Figura 12: durata vocalica media (ms) delle toniche vs. atone non finali di PP (proparossitone), P (parossitone) e O (ossitone) in italiano e in spagnolo.

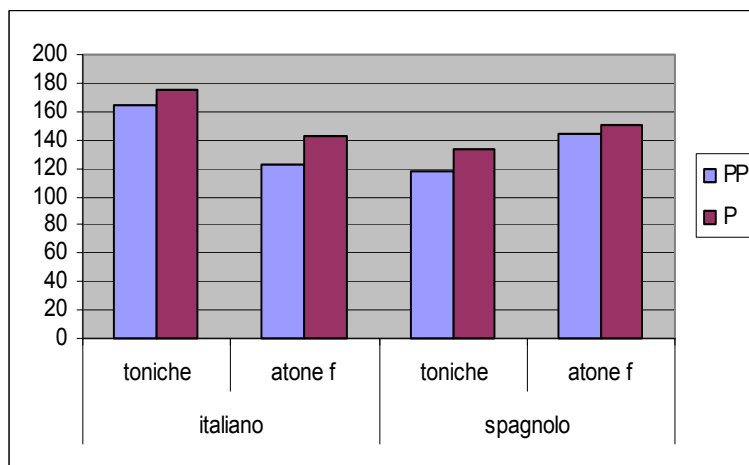


Figura 13: durata vocalica media (ms) delle toniche vs. atone finali di PP (proparossitone), P (parossitone) in italiano e in spagnolo.

Emergono dunque differenze tra i pattern accentuali delle due lingue; le figure che seguono (14a, 14b, 14c) rappresentano dei prototipi di proparossitone (PP), di parossitone (P) e di ossitone (O) costruiti secondo i valori di durata media calcolata nell'analisi acustica (rappresentata nelle figure sull'asse delle ordinate).

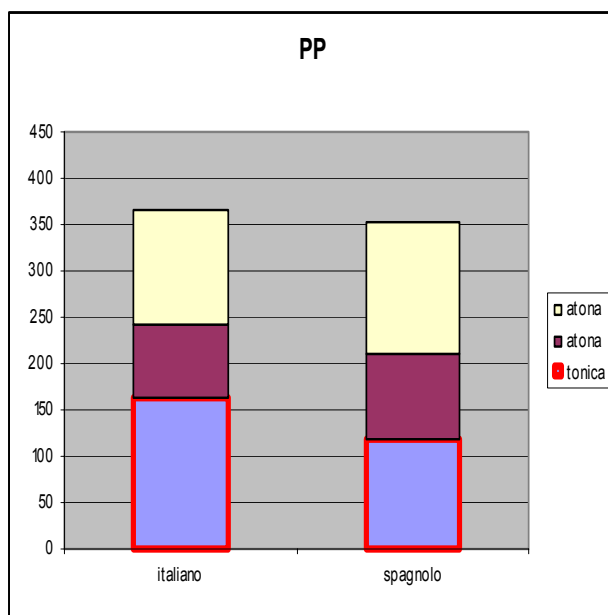


Figura 14a: prototipo con valori di durata vocalica media (ms) di una PP.

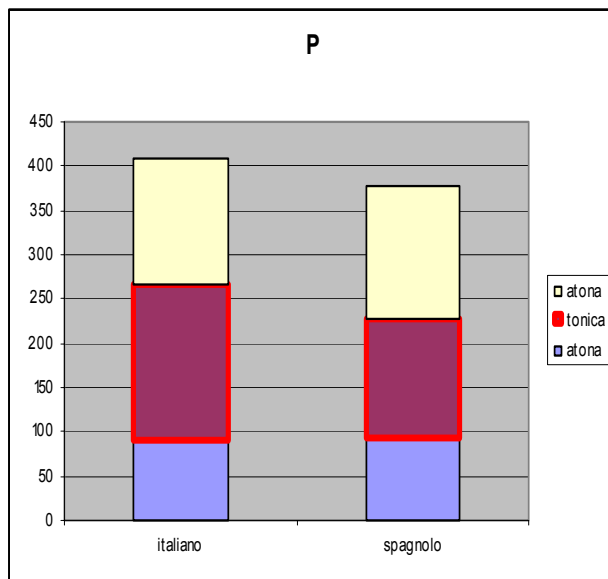


Figura 14b: prototipo con valori di durata vocalica media (ms) di una P.



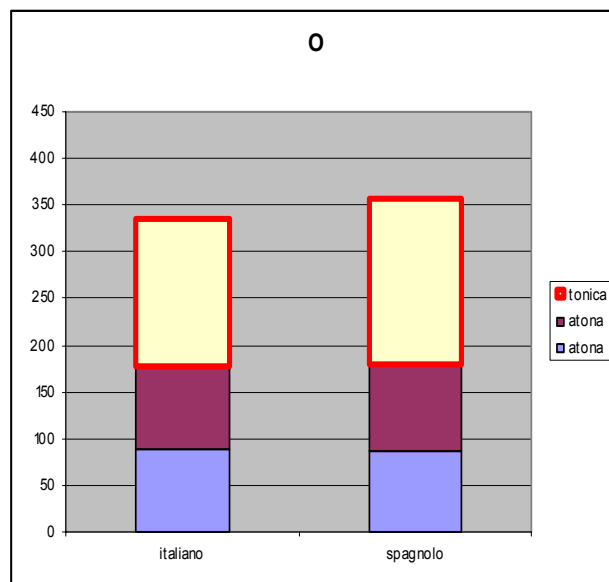


Figura 14c: prototipo con valori di durata vocalica media (ms) di una O.

Mettendo a confronto i valori medi di durata nelle due lingue si evince, dunque, che in italiano:

- sebbene le toniche delle O siano più brevi rispetto alle toniche di PP e P, per ciascuno schema accentuale le toniche risultano comunque più lunghe rispetto alle atone (si vedano le figure 14a, b e c);
- le toniche e le atone più lunghe sono quelle delle parossitone;
- quelle più brevi sono quelle delle ossitone (si osservino le figure 11, 12 e 13).

Diversamente in spagnolo:

- le toniche sono significativamente più lunghe delle atone solo per lo schema ossitono: non si riscontra una differenza marcata di durata tra toniche ed atone di proparossitone e parossitone (figura 11);
- le toniche con maggior durata sono quelle delle ossitone, che sono, altresì, le più lunghe dei due *corpora* (figura 11 e 14c).

La differenza più rilevante tra i due sistemi è che le vocali toniche in posizione finale risultano in italiano, in totale accordo con quanto indica la letteratura sul tema, il 7,8% più brevi di quelle in posizione non finale, laddove in spagnolo sono il 42,3% più lunghe (si veda la figura 14c).

Riteniamo dunque che la difficoltà nella percezione di stimoli italiani dei parlanti nativi di spagnolo si spieghi considerando le differenze acustiche tra i due sistemi linguistici: una P>O (parossitona con i valori di ossitona) in italiano subisce una diminuzione di durata, processo inverso rispetto a quanto accade in spagnolo. I soggetti spagnoli, quindi, si trovano a dover risolvere una sorta di conflitto tra le caratteristiche acustiche degli stimoli (una diminuzione di durata) e le aspettative legate alla loro L1, lo spagnolo (un aumento di durata).

## 5. CONCLUSIONI

Diversamente da quanto la forte somiglianza tra le due lingue facesse ipotizzare, si delineano differenze rilevanti sia in percezione che in produzione.

I risultati ottenuti negli esperimenti percettivi realizzati in L1, condotti, come abbiamo visto, con analoga metodologia, rivelano un diverso ruolo dei correlati acustici nelle due lingue: confermano la priorità della durata in italiano (Alfano, 2006) ed indicano cruciale un'interazione tra la variazione congiunta di  $f_0$  con almeno un altro parametro in spagnolo (Llisterri *et al.*, 2005).

L'analisi acustica dei valori di durata dei *corpora* impiegati negli esperimenti percettivi mostra importanti differenze: le vocali toniche in media più lunghe in italiano sono quelle delle P, seguite da quelle delle PP e infine delle O; in spagnolo, invece, le toniche delle O sono, con una differenza di 45,38 ms (considerando i valori medi), più lunghe di quelle di PP e di P. Dunque mentre per l'italiano si verifica un accorciamento delle vocali toniche finali rispetto alle non finali, lo spagnolo vede, nello stesso caso, un sensibile allungamento.

Ancora, le toniche non finali (guardando al valor medio quindi delle toniche di PP e P) si differenziano per una maggior durata rispetto alle adiacenti atone (finali e non) in italiano, ma non in spagnolo.

Quanto alla percezione degli ispanofoni in italiano LS, riteniamo che la diversità dei due sistemi sul piano acustico possa rendere conto delle difficoltà riscontrate; restano tuttavia da chiarire vari punti in merito alle strategie percettive impiegate, come abbiamo visto, in maniera diversa in L1 rispetto a quanto avviene in italiano LS.

Future analisi mireranno ad approfondire gli aspetti emersi in questo primo approccio, in particolare, attraverso l'esperimento speculare, ma con soggetti nativi di italiano e stimoli in spagnolo.

## BIBLIOGRAFIA

Alfano, I. (2006), La percezione dell'accento lessicale: un test sull'italiano a confronto con lo spagnolo, in Savy, R., Crocco, C. (a c. d.), *Atti del II AISV, Associazione Italiana di Scienze della Voce*, Salerno, 632-656.

Alfano, I., Llisterri, J. & Savy, R. (2007), The perception of Italian and Spanish lexical stress: A first cross-linguistic study, in *Proceedings of the 16th International Congress of Phonetic Sciences*. 6-10 August, 2007, Saarbrücken, Germany, 1793-1796. ISBN: 978-3-9811535-0-7 [CD-ROM].  
[http://liceu.uab.es/~joaquim/publicacions/Alfano\\_Llisterri\\_Savy\\_07\\_Perception\\_Stress\\_Spanish\\_Italian.pdf](http://liceu.uab.es/~joaquim/publicacions/Alfano_Llisterri_Savy_07_Perception_Stress_Spanish_Italian.pdf)

Almeida, M. (1997), Organización temporal del español: el principio de isocronía, *Revista de filología románica*, 14, 29-40.

Bertinetto, P. M. (1981a), *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*, Firenze: Accademia della Crusca.

Bertinetto, P. M. (1981b), Ancora sull'italiano come lingua ad isocronia sillabica, *Scritti linguistici in onore di G. B. Pellegrini*, Pisa: Pacini, 1073-1082.

Bertinetto, P. M. (1989), Reflections on the dichotomy 'stress' vs. 'syllable-timing', *Revue de Phonétique Appliquée*, 91/93, 99-130.

- Bertinetto, P. M., 1990. Coarticolazione e ritmo nelle lingue naturali. In *Atti del XVIII Convegno Nazionale dell'Associazione Italiana di Acustica*, L'Aquila, 1990.
- Boersma P. & Weenink, D. (2003), Praat: doing phonetics by computer (Version 4.0.4). <http://www.praat.org/>
- Llisterri, J., Machuca, M., de la Mota, C., Riera, M. & Ríos, A. (2005), La percepción del acento léxico en español, in *Filología y lingüística*. Madrid: CSIC–UNED–U. de Valladolid, 1, 271-297.
- Mancini, F. & Voghera, M. (1994), Lunghezza, tipi di sillabe e accento in italiano, in De Mauro, T. (a c. d.), *Come parlano gli Italiani*, Firenze: La Nuova Italia, 217-244.
- Marotta, G. (1985), *Modelli e misure ritmiche: la durata vocalica in italiano*, Bologna: Zanichelli.
- Monroy-Casas, R. (1980), *Aspectos fonéticos de las vocales españolas*, Madrid: Sociedad General Española de Librería.
- Navarro Tomás, T. (1974), *Manual de entonación española*, Madrid: Guadarrama.
- Nespor, M., and Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Quilis, A. (1993), *Tratado de fonología y fonética españolas*, Madrid: Editorial Gredos.

## 6. APPENDICE

| parole                          |
|---------------------------------|
| [ˈse:mino] [seˈmi:no] [semiˈno] |
| [ˈka:pito] [kaˈpi:to] [kapiˈto] |
| [ˈla:vati] [laˈva:ti]           |
| [ˈle:gami] [leˈga:mi]           |
| [ˈpa:gano] [paˈga:no]           |
| [laˈvo:ro] [lavoˈro]            |
| [roˈvi:no] [roviˈno]            |
| [reˈga:lo] [regaˈlo]            |
| non parole                      |
| [ˈga:mino] [gaˈmi:no] [gamiˈno] |
| [ˈnarvilo] [naˈvi:lo] [naviˈlo] |
| [ˈma:ligo] [maˈli:go] [maliˈgo] |
| [ˈla:gano] [laˈga:no] [lagaˈno] |
| [ˈni:vale] [niˈva:le]           |
| [ˈva:divo] [vaˈdi:vo]           |
| [ˈmi:lado] [miˈla:do]           |
| [diˈna:vo] [dinaˈvo]            |
| [gaˈdi:vo] [gadiˈvo]            |
| [viˈda:lo] [vidaˈlo]            |

Tabella 3: *corpus* impiegato in italiano (Alfano, 2006).

| parole                        |
|-------------------------------|
| [ˈlimite] [liˈmite] [limiˈte] |
| [ˈmeðiko] [meˈðiko] [meðiˈko] |
| [ˈnumero] [nuˈmero] [numeˈro] |
| [ˈbaliðo] [baˈliðo] [baliˈðo] |
| non parole                    |
| [ˈlaðeβo] [laˈðeβo] [laðeˈβo] |
| [ˈluʁiðo] [luˈʁiðo] [luʁiˈðo] |
| [ˈmaleðo] [maˈleðo] [maleˈðo] |
| [ˈnuliβo] [nuˈliβo] [nuliˈβo] |

Tabella 4: *corpus* impiegato in spagnolo (Listerri *et al.*, 2005).

|                    | italiano  | spagnolo |
|--------------------|-----------|----------|
|                    | <b>PP</b> |          |
| tonica non finale  | 163,67    | 132,87   |
| atona non finale   | 79,23     | 93,92    |
| atona finale       | 121,91    | 149,68   |
|                    | <b>P</b>  |          |
| atona non finale   | 89,94     | 93,92    |
| tonica non finale  | 175,6     | 132,87   |
| atona finale       | 143,37    | 149,68   |
|                    | <b>O</b>  |          |
| atona non finale   | 89,14     | 85,89    |
| atona non finale   | 89,25     | 93,62    |
| tonica finale      | 156,82    | 178,25   |
|                    |           |          |
| toniche            | 165,9     | 142,9    |
| toniche finali     | 156,82    | 178,25   |
| toniche non finali | 170,1     | 125,24   |
| toniche di PP      | 163,67    | 117,61   |
| toniche di P       | 175,6     | 132,88   |
| toniche di O       | 156,82    | 178,25   |
| atone              | 102,9     | 109,9    |
| atone finali       | 133,47    | 146,79   |
| atone non finali   | 87,01     | 91,45    |
| atone pretoniche   | 89,59     | 93,77    |
| atone post-toniche | 79,23     | 92,38    |

Tabella 5: valori medi della durata vocalica (ms) calcolata nei corpora impiegati negli esperimenti percettivi realizzati in italiano (Alfano, 2006) ed in spagnolo (Llisterri *et al.*, 2005).

# L'ACQUISIZIONE DEL TRATTO DI LUNGHEZZA CONSONANTICA IN APPRENDENTI L'ITALIANO L2

Patrizia Sorianello  
Università degli Studi di Bari  
*p.sorianello@lettere.uniba.it*

## SOMMARIO

Il presente studio intende verificare le strategie di acquisizione del tratto di lunghezza consonantica in tre apprendenti l'italiano di lingua madre tedesca. Il materiale considerato, conversazioni libere acquisite in modo longitudinale, è stato etichettato ed annotato su *text grids* formati da cinque livelli. I dati empirici sono stati ottenuti per mezzo di specifici *scripts* di interrogazione automatica.

I risultati raccolti mostrano che la produzione di consonanti geminate costituisce un'area di difficoltà per gli apprendenti tedescofonici. I soggetti manifestano non solo evidenti fenomeni di degeminazione, ma anche la formazione di categorie temporali mediane e persino l'iper-correzione del tratto di lunghezza. Non tutti i fonemi lunghi dell'italiano sembrano inclini allo scempiamento nella stessa misura, in linea con la marcatezza tipologica dei segmenti implicati.

Le risultanze sperimentali non disattendono i principali assunti della fonologia dell'interlingua: la formazione di una nuova categoria fonologica, qual è per l'appunto la classe dei fonemi geminati, rispetta i criteri di naturalezza e di marcatezza tipologica.

## 1. PREMESSA

Nella ormai copiosa letteratura relativa alla linguistica acquisizionale dell'italiano L2, la fonologia dell'interlingua, settore di indagine per molti aspetti ancora lacunoso, occupa un posto marginale<sup>1</sup>. Solo di recente, le caratteristiche fonologiche dell'italiano lingua seconda hanno attirato l'attenzione degli studiosi, questa volta con un interesse crescente difficilmente arrestabile; non a caso, una cospicua porzione di ricerche è stata condotta proprio negli ultimi anni<sup>2</sup>. Le fenomenologie indagate, ancora lungi dall'assurgere a carattere esaustivo, sono per lo più improntate sul versante della produzione che non della percezione<sup>3</sup>. All'interno di questo ambito l'analisi di processi segmentali è inoltre

---

<sup>1</sup> Questo filone tematico si apre in Italia nel 1988 con uno studio di Bernini condotto su 10 soggetti arabofoni. La strada è ormai aperta, ma l'interesse scientifico verso questa fenomenologia tutt'altro che vivo. Ci vorrà circa un decennio affinché le ricerche riprendano in questa direzione, sebbene ancora in modo poco sistematico e quasi episodico; a questo periodo risalgono gli studi svolti da Schmid (1994), Giannini e Costamagna (1997; 1998).

<sup>2</sup> Cfr. Costamagna (2003), Boula de Maureil *et al.* (2004), Celata (2004), Gamal (2007), Mori (2007), Sorianello (2008, in stampa).

<sup>3</sup> Nella letteratura internazionale sulla fonologia dell'interlingua, produzione e percezione non hanno ricevuto, nel corso degli anni, la medesima attenzione. Una rassegna ragionata delle ragioni che hanno favorito il livello della produzione, insieme alla disamina critica delle principali ricerche orientate alla percezione, è contenuta in Marotta (2008).

prevalente; gli studi sulla prosodia dell'italiano come L2 costituiscono, non troppo sorprendentemente, un'area fortemente trascurata<sup>4</sup>.

Lo studio che qui presentiamo si inserisce nel filone di ricerca volto a definire le caratteristiche fonetiche (segmentali e prosodiche) dell'italiano L2. La ricerca è stata realizzata in seno ad un progetto di ricerca nazionale (PRIN 2005/2007; prot. 2005109172\_003) facente capo a Giovanna Marotta (Unità di Ricerca dell'Università di Pisa) e coordinato da P.M. Bertinetto. La fenomenologia che sarà discussa in questa sede riguarda i meccanismi messi in atto nell'italiano interlingua per l'acquisizione del tratto di lunghezza consonantica. Le modalità di insorgenza di questa categoria fonologica risultano particolarmente evidenti nell'interlingua dei soggetti stranieri privi nel loro sistema linguistico di partenza (L1) di tale opposizione fonemica.

La lunghezza è un parametro prosodico che fa riferimento alla durata temporale di un suono. Una maggiore durata, oltre ad essere rilevabile acusticamente, è un robusto indice percettivo presente nella coscienza metalinguistica dei parlanti. In alcune lingue, la lunghezza, o quantità, di un segmento, vocalico o consonantico, può assumere valore funzionale. In merito a questo aspetto si distinguono:

a) lingue fonologicamente prive di opposizione di quantità vocalica o consonantica, ad es. lo spagnolo;

b) lingue in cui solo la quantità vocalica ha valore funzionale, ad es. l'olandese, *poot* (zampa) vs. *pot* (pentola);

c) lingue in cui solo la quantità consonantica ha natura fonologica, ad es, l'italiano (cfr. *ultra*)<sup>5</sup>;

d) lingue in cui sia la quantità vocalica che quella consonantica è distintiva, è questo il caso del libanese o del finnico; in quest'ultima l'opposizione è evidente anche a livello grafematico, ad es. *tuli* (fuoco) vs. *tuuli* (vento), *tule* (vieni, imperativo) vs. *tulle* (egli viene).

Dal punto di vista tipologico, l'opposizione di quantità consonantica non è frequente. Secondo i dati contenuti nel *database* UPSID (cfr. Maddieson, 1984), le lingue in cui la geminazione consonantica ha statuto fonemico sono nel complesso poco rappresentate, per alcuni fonemi persino rara.

Gli scritti specialistici condotti sull'argomento, pur essendo incentrati su lingue tipologicamente non omogenee, documentano un aspetto ricorrente: il tratto fonetico a cui è affidata la distinzione percettiva tra una consonante breve e una consonante lunga è la durata temporale. Ciò è vero nell'arabo libanese (cfr. Khattab, 1999), nel greco parlato a Cipro (cfr. Arvaniti, 1999), in alcune lingue indonesiane (cfr. Chon *et alii*, 1999), come pure nel finnico (cfr. Kunnari *et alii*, 2001) e nell'italiano (cfr., tra gli altri, Payne, 2005).

Nel confronto interlinguistico, lo scarto temporale tra consonanti geminate e scempie è comunque variabile, poiché risente del modo di articolazione del segmento implicato, del materiale linguistico esaminato (parlato letto o spontaneo) e, non per ultimo, della velocità elocutiva o del condizionamento prosodico. La distanza temporale tra classi di suoni lunghi

---

<sup>4</sup> Poche le eccezioni menzionabili, Giannini & Celata (2002-2003), Boula de Maureil *et al.* (2004), Gamal (2007) e il già citato Marotta (2008).

<sup>5</sup> In inglese come pure in tedesco, il tratto di lunghezza consonantica non è invece distintivo; in alcuni lessemi la lunghezza è tratto grafematico, ma non fonologico, ad es. ted. *Rolle*, ingl. *better*, *daddy*. In francese, la pronuncia geminata è presente solo in alcuni lessemi dotti, ad es. *villa*, in alcune forme verbali, come *courait/courrait*; le geminate grafiche sono di norma pronunciate scempie, ad es. *terre*, *personne*.

e brevi è sovente quantificata sotto forma di rapporto (*ratio*); l'indice numerico risultante dal rapporto aritmetico tra /C:/ e /C/ è tanto più alto quanto maggiore è la differenza temporale tra gli stessi suoni. Di norma, tale rapporto è superiore nel parlato letto, mentre si riduce nell'eloquio poco controllato<sup>6</sup>.

Diversamente, la geminazione sembrerebbe poco sensibile agli effetti della velocità di eloquio; la validità di questo assunto è tuttavia controversa. Nel cipriota la durata del fonema geminato, in condizioni elocutive sostenute, pur riducendosi, non si sovrappone mai alla durata della medesima consonante scempia realizzata a velocità normale (cfr. Arvaniti, 1999). In arabo libanese, per converso, la sovrapposizione temporale tra le due serie di fonemi (geminati e scempi) è presente soprattutto nel parlato spontaneo. Il rendimento funzionale dell'opposizione sarebbe in questo caso salvaguardata dall'informazione veicolata da altri indici non fonetici, *in primis* dal contesto semantico e sintattico di ricorrenza.

Il quadro descrittivo diventa ancora più complesso se consideriamo che l'incremento della velocità elocutiva sortisce effetti diversi dipendenti, non solo dalla lingua, ma anche dalla consonante implicata. Ad esempio, in un campione di parlato letto di lingua cipriota, il valore *ratio* tra /m:/ e /m/ rimane pressoché identico, a prescindere dallo *speech rate*, laddove il contrasto temporale tra /r:/ e /r/ è invece significativo<sup>7</sup>.

La durata di C scempie e geminate varia anche a seconda del modo di articolazione del segmento. Nel complesso, la classe di suoni in cui si osserva il più basso rapporto C:/ C è quella delle fricative; per contro la classe in cui il rapporto è sistematicamente più elevato è rappresentata dalle liquide, tra queste *in primis* la vibrante. In quest'ultimo caso, il notevole incremento temporale rilevato in diverse lingue nella realizzazione geminata di /r/ si deve sovente ad una diversa pronuncia del segmento, approssimante quando è scempio, vibrante se geminato.

Per quanto concerne l'italiano, il rapporto temporale tra /C:/ e /C/, in media pari a 1.9, è maggiore se il fonema è posto in posizione intonativa nucleare o se adiacente all'accento lessicale; è inoltre elevato per /l/, minimo per /f/ e /p/, (cfr. Payne, 2005)<sup>8</sup>. Diversamente, in uno studio condotto da Endo e Bertinetto (1999) su tre varietà di italiano regionale, piemontese, pisano e napoletano<sup>9</sup>, il rapporto, calcolato sulla base delle misure acustiche ivi riportate, risulta alto per /n:/ vs /n/, /l:/ vs. /l/, e, limitatamente alle varietà di italiano

<sup>6</sup> Ad esempio, nel libanese il rapporto rinvenuto è 2.50 in un campione di frasi lette, ma 1.82 nel parlato spontaneo (cfr. Khattab, 1999).

<sup>7</sup> A velocità normale, i rapporti rinvenuti sono per la precisione: /m:/ vs. /m/= 1.40, /n:/ vs. /n/= 1.88, /l:/ vs /l/= 1.88, /r:/ vs. /r/= 2.39; nell'eloquio veloce il rapporto tra la durata della C geminata e quella della scempia è invece /m:/ vs. /m/= 1.48, /n:/ vs. /n/= 1.78, /l:/ vs /l/= 1.77, /r:/ vs. /r/= 1.78. (cfr. Arvaniti, 1999).

<sup>8</sup> Questa ricerca è stata condotta su un campione di frasi lette da cinque locutori pisani; i fonemi considerati sono /p t b d f ʃ m l/.

<sup>9</sup> Il fine di questa ricerca era quello di valutare lo *status* delle consonanti rafforzate /ts dz ʃ n ʎ/, motivo per cui la durata di questi fonemi è stata posta a confronto con quella di /t t:/, /d d:/, /n n:/, /l l:/, /s s:/.



piemontese e pisano, in /s:/ vs. /s/, rispettivamente 2.4 e 2.5; il valore più basso si rinviene per /t:/ vs /t/<sup>10</sup>.

Sebbene la durata sia l'indice fonetico più robusto nell'opposizione fonematica tra /C/ e /C:/, altri tratti articolatori ed acustici contribuiscono a rafforzare la resa percettiva di una C geminata, come la durata della vocale che la precede, il luogo di articolazione del segmento medesimo, la sua posizione rispetto all'accento lessicale e persino la sua posizione all'interno dell'unità prosodica.

Per l'italiano L2 non si dispone di dati sperimentali relativi al contrasto temporale tra /C:/ e /C/<sup>11</sup>; di conseguenza, assumeremo quale riferimento, con le dovute cautele, i dati disponibili sull'italiano nativo.

## 2. L'INDAGINE

In questa sede saranno presentati e discussi i dati sperimentali relativi all'acquisizione del tratto di lunghezza consonantica da parte di tre apprendenti di lingua madre tedesca. Si tratta di giovani soggetti femminili studenti presso l'Università per Stranieri di Perugia la cui competenza dell'italiano è compresa tra B1 e C2. Il materiale analizzato consta di conversazioni guidate a carattere dialogico acquisite in modo longitudinale. Le sessioni di registrazione disponibili per apprendente oscillano tra quattro e sei; nel complesso la ricerca è stata condotta su 14 conversazioni<sup>12</sup>, per un totale di 580 unità prosodiche. Il materiale linguistico, già fornito di trascrizione ortografica, è stato digitalizzato a 22050 Khz e 16 bits.

Al fine di agevolare la successiva ispezione spettro-acustica, si è reso necessario procedere con la ripartizione dei dialoghi in unità prosodiche; ad ogni singola unità prosodica è stato assegnato un nome composto da otto sigle, univoco ed informativo, condizione indispensabile per la sottoclassificazione ordinata dell'archivio sonoro.

I *corpora* sono stati etichettati mediante un modello di codifica, creato all'uopo all'interno del suindicato progetto PRIN, denominato SAIL2, (*Sistema per l'Annotazione Morfo-fonologica dell'Italiano Lingua Seconda*). Si tratta del primo sistema di annotazione appositamente dedicato all'etichettatura dei fenomeni fonologici di campioni di italiano lingua seconda. La struttura di SAIL2, espressamente multilivello e flessibile, è stata elaborata su piattaforma *Praat*; cinque sono i livelli di etichettatura previsti: morfologico (*Parole*), fonologico (*Fonemi*), fonetico (*Sampa*) Prominenza prosodica (*Prominenza*) e intonativo (*ToBI*). Tutte le procedure di trascrizione fonetica sono state effettuate con l'alfabeto SAMPA (cfr. Wells, 1994), anche se sono state apportate diverse modifiche per

---

<sup>10</sup> Il dato rilevato per la fricativa /s/, in apparente controtendenza, si deve probabilmente alla pronuncia intervocalica sonora di /s/, al contrario di quanto avviene per la varietà di Napoli.

<sup>11</sup> Le modalità di insorgenza delle consonanti lunghe nell'interlingua di soggetti stranieri è stata in realtà già indagata da Giannini e Costamagna (1997; 1998). Il quadro sperimentale di questa ricerca, impostata uditivamente, ricco di implicazioni teoriche, non fornisce misurazioni di durata della consonante lunga, bensì valori percentuali relativi alla resa geminata di alcuni fonemi.

<sup>12</sup> Quanto da noi analizzato rappresenta una sottosezione di un ampio campione linguistico registrato dall'Università per Stranieri di Perugia. Corre l'obbligo ringraziare Stefania Giannini e Lidia Costamagna per aver permesso la fruizione di tale archivio sonoro da parte dell'intero *team* di ricerca Prin sopra menzionato.

meglio adattarlo alle caratteristiche linguistiche dell'italiano interlingua. L'estrazione dei dati è stata affidata a specifiche *routines* di interrogazione automatica formulate in linguaggio *Perl*<sup>13</sup>; a seconda delle esigenze della ricerca, l'interrogazione può coinvolgere anche più livelli di annotazione dando luogo ad informazioni multiple ed incrociate<sup>14</sup>.

In tedesco, lingua materna degli apprendenti esaminati, la quantità consonantica non è distintiva, a differenza di quanto avviene in italiano (cfr. Wiese, 1996)<sup>15</sup>. Di conseguenza, si può ragionevolmente presumere che l'acquisizione di questa categoria fonologica susciti non poche difficoltà. Per indagare le strategie messe in atto dagli apprendenti durante l'acquisizione della consonante lunga dal repertorio linguistico a nostra disposizione è stato estratto un campione di 584 fonemi geminati ricorrenti in posizione interna di parola. Sono stati sottoposti ad indagine i quindici fonemi dell'italiano per i quali vige opposizione distintiva di lunghezza, per la precisione /p t k b d g tʃ dʒ f v s m n l r/. Le consonanti intrinsecamente lunghe /ts dz ʃ ɲ ʎ/ sono state invece considerate a parte e non saranno discusse in questa sede.

La banca dati è stata ottenuta ponendo a confronto in modo incrociato le annotazioni indicizzate a livello fonologico, fonetico e non per ultimo morfologico. Ogni fonema geminato è stato infatti classificato per modo e luogo di articolazione, struttura sillabica, posizione dell'accento lessicale e per categoria morfosintattica del lessema che lo contiene. Ciò ha permesso di catalogare il fonema su più piani, di registrare il numero delle realizzazioni totali, la percentuale di quelle conformi al *target* e al contempo di quantificare il ventaglio delle sue produzioni allofoniche. Sul versante propriamente acustico, sono stati ricavati i dati relativi alla durata del fonema geminato e della vocale ad esso precedente; la durata delle consonanti è stata posta a confronto con un sottocampione di fonemi scempi estratti dal medesimo archivio.

### 3. I FONEMI GEMINATI

Nel *corpus* preso in esame ricorrono 584 fonemi geminati. Nell'interlingua dei soggetti esaminati, la geminazione è presente, sia pure con percentuali diverse, in tredici fonemi su quindici. Nel campione, infatti, non vi è alcuna ricorrenza dei fonemi /d:/ e /g:/, questa assenza non ha permesso di ricavare dati empirici relativi a questi suoni (cfr. anche Albano Leoni e Clemente, 2005)<sup>16</sup>. Tuttavia, i soggetti tedeschi non riescono a realizzare sempre, ove richiesto, dei fonemi geminati. Ciò che si verifica è un comportamento oscillante e variabile anche nella produzione elocutiva di uno stesso apprendente. Complessivamente, il fonema lungo /C:/ > [C:] è prodotto nel 54.1% delle ricorrenze totali; nel 28.8% i soggetti realizzano un segmento breve /C:/ > [C], nel 10.6% riscontriamo segmenti dalla durata intermedia (già Giannini e Costamagna, 1998), da noi indicizzati /C:/ > [C\_D], dove D sta per degeminazione e nel rimanente 6.5% altre realizzazioni. Per categoria intermedia è stato

<sup>13</sup> Lo *script* di estrazione produce in pochi secondi un *file* di testo contenente di volta in volta le informazioni richieste le quali sono già predisposte per un agevole inserimento in un *database* in formato *Excel*.

<sup>14</sup> Per l'esplicitazione delle norme di segmentazione e di etichettatura adottate in SAIL2, come pure per maggiori dettagli sulla struttura del sistema e sulle procedure di estrazione dei dati, rimandiamo a Marotta & Soriano (2008).

<sup>15</sup> Diversamente, nei dialetti tedeschi parlati in Svizzera vige un'opposizione fonologica tra occlusive sorde scempie e geminate (cfr. Kraehenmann & Jaeger, 2003)

<sup>16</sup> Questo dato è tuttavia in linea con la bassa incidenza statistica di /d:/ e /g:/ nell'italiano.

inteso un suono non comparabile, per durata e per giudizio percettivo, né al relativo fonema scempio, poiché più lungo, né al relativo fonema lungo, essendo più breve. La ricorrenza di questa categoria, pur essendo presente nei tre soggetti, si attesta come un fenomeno minoritario dalla bassa rilevanza statistica. La variante, infatti, non raggiunge, per nessun segmento, un cospicuo tasso di ricorrenza, essa si rinviene prevalentemente per /m/ (25%), e /k/ (20%); trascurabile è l'incidenza in /p/ (5%), /s/ (4.5%) e /b/ (2.9%).

Nel raggruppamento 'altre realizzazioni' abbiamo invece incluso tutte le varianti fonetiche che non coincidevano, ad una ispezione spettro-acustica, né con un segmento geminato, né con un segmento degeminato. In questo gruppo, sono state comprese ad esempio la pronuncia approssimante o fricativizzata, allofoni frequenti soprattutto per /r/ (37.5%), /v/ (33.3%) e per le occlusive /k/ (26.7%) e /b/ (23.5%). Si tratta comunque di articolazioni deboli, la cui durata è risultata di gran lunga inferiore a quella di una variante geminata. Diversamente, l'affricata /dʒ/, presenta un'alta variabilità allofonica, ben il 58.9%; la pronuncia di questo fonema, in realtà molto omogenea, corrisponde il più delle volte ad un suono sordo, ossia [tʃ].

Nel complesso, il quadro empirico fin qui delineato dimostra lo sforzo effettuato dall'apprendente per avvicinarsi al bersaglio articolatorio. L'assenza nel tedesco di consonanti lunghe rallenta, almeno negli stadi iniziali di interlingua, il processo acquisizionale del fonema lungo dell'italiano, caratterizzato da un tratto distintivo innovativo.

A fronte di questa premessa di validità generale, occorre rilevare che nell'interlingua dei soggetti non tutti i fonemi dell'italiano, per i quali vige opposizione di durata, risultano geminati con la medesima frequenza. Alcuni segmenti lunghi sono prodotti con percentuali nettamente più alte rispetto ad altri. Si vedano in merito i valori di fedele corrispondenza (% *Match*) o di mancata realizzazione del fonema geminato (% *Mismatch*) riportate per ciascun segmento nella Tab. 1 e nella Fig. 1.

| Fonema | Occorrenze<br>totali | % Match    |           | % Mismatch  |       |
|--------|----------------------|------------|-----------|-------------|-------|
|        |                      | /C:/ >[C:] | /C:/ >[C] | /C:/ >[C_D] | Altre |
| /p:/   | 20                   | 70         | 20        | 5           | 5     |
| /t:/   | 147                  | 71,4       | 8,2       | 13,6        | 6,8   |
| /k:/   | 15                   | 40         | 13,3      | 20          | 26,7  |
| /b:/   | 34                   | 17,6       | 55,9      | 2,9         | 23,5  |
| /d:/   | --                   | --         | --        | --          | --    |
| /g:/   | ---                  | --         | --        | --          | --    |
| /tʃ:/  | 24                   | 83,3       | 4,2       | 12,5        | --    |
| /dʒ:/  | 17                   | 23,5       | 5,9       | 11,8        | 58,9  |
| /f:/   | 25                   | 84         | 4         | 12          | --    |
| /s:/   | 89                   | 91         | 4,5       | 4,5         | --    |
| /v:/   | 6                    | --         | 66,7      | --          | 33,3  |
| /m:/   | 12                   | 25         | 50        | 25          | --    |
| /n:/   | 73                   | 42,5       | 43,8      | 13,7        | --    |
| /l:/   | 114                  | 21         | 68,4      | 10,5        | --    |
| /r:/   | 8                    | 12,5       | 50        | --          | 37,5  |

Tabella 1: occorrenze totali, percentuale di corretta realizzazione (% *Match*) o di mancata realizzazione del *target* (% *Mismatch*) per i fonemi elencati

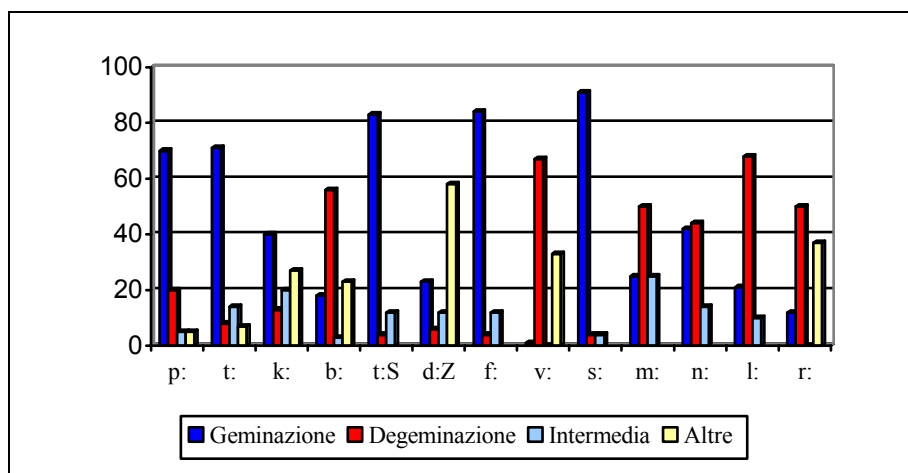


Figura 1: valori percentuali rinvenuti nei fonemi indicati per ciascuna resa allofonica<sup>17</sup>

Ad una prima analisi, la situazione che si delinea appare contraddittoria. Nell'interlingua dei soggetti tedeschi, il *target* geminato è per alcuni fonemi raggiunto con valori percentuali decisamente alti, è il caso di /s:/ (91%) o di /f:/ (84%), mentre per altri fonemi l'incidenza del processo è piuttosto contenuta, ad es. /l:/ (21%) oppure /r:/ (12%). La variante scempia, per contro, si registra con alte percentuali in /l:/ (68%), /v:/ (66.7%) e /b:/ (55.9%), esigua è la presenza di questo allofono per /f:/, /s:/, /t:/ e /t:/, gli stessi per i quali si ravvisa la migliore corrispondenza tra *target* fonologico e resa fonetica.

Un esame più approfondito mostra come l'attivazione del tratto di lunghezza consonantica, negli apprendenti ispezionati, sia fortemente subordinato alla qualità del fonema coinvolto. Il comportamento è speculare, da un lato si pongono le ostruenti sorde, la cui pronuncia è nel complesso più fedele al bersaglio fonologico, dall'altro si collocano le ostruenti sonore e l'intera serie delle sonoranti, più frequentemente articolati brevi.

Nella Fig. 2 riportiamo la realizzazione acustica, con relativa annotazione, di un sintagma prodotto dall'apprendente MB in cui si rileva la variante scempia del fonema /l:/ (*nella*), ms 42, insieme alla resa geminata di /t:/ (*settimana*), 118 ms; la divergenza è apprezzabile anche dal confronto tra la trascrizione riportata nel livello fonologico (*Fonemi*) e quella del sottostante livello fonetico (*Sampa*). Similmente, la Fig. 3, evidenzia lo scempiamento del fonema lungo /l:/ (*bellezza*), ms 37, parallelamente alla resa geminata dell'affricata /t:s/, ms 200.

<sup>17</sup> In trascrizione Sampa /tS/ equivale a /tʃ/, mentre /dZ/ a /dʒ/.

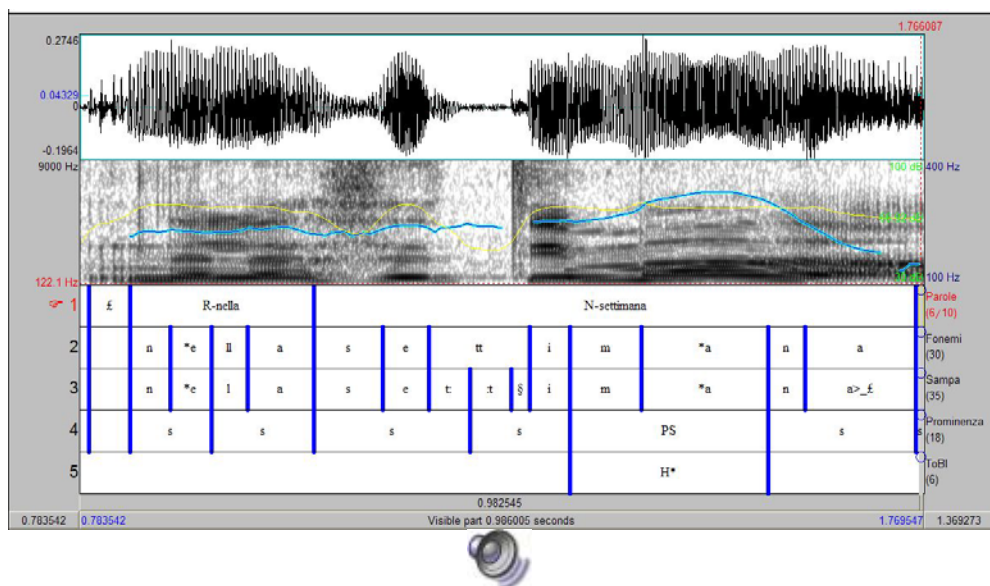


Figura 2: Forma d'onda, sonagramma a banda larga e annotazione in SAIL2 del sintagma *nella settimana*, realizzato dall'apprendente MB

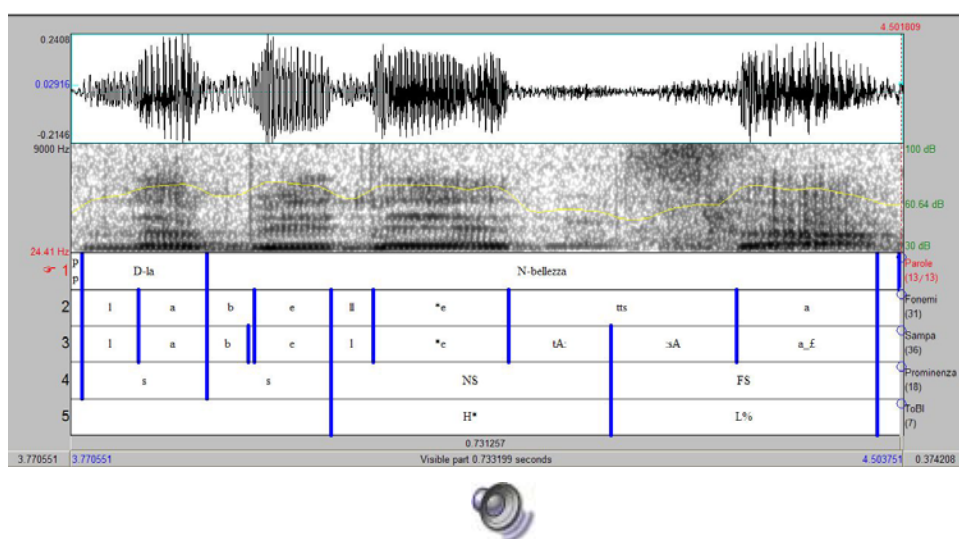


Figura 3: Forma d'onda, sonagramma a banda larga e annotazione in SAIL2 del sintagma *la bellezza*, realizzato dall'apprendente LM

#### 4. LA DURATA DELLA CONSONANTE

Ai fini dell'interpretazione del processo in esame, fondamentale si è rivelata l'analisi della durata temporale dei fonemi considerati. Un alto numero di C lunghe subisce uno spiccato processo di riduzione temporale. Il *target* geminato è molto spesso realizzato scempio. In questo caso, la durata del segmento, più breve, è approssimabile a quello del corrispettivo fonema scempio, come dimostra il confronto effettuato con un sottocampione di consonanti scempie (cfr. *ultra*). Dai dati calcolati per i tre soggetti è possibile stabilire un andamento temporale decrescente: la maggiore durata si rinviene per il segmento geminato (124 ms), un valore del tutto comparabile accomuna la durata del segmento degeminato a quella del fonema scempio (72 ms vs. 79 ms), mentre la categoria intermedia mostra una durata mediana, pari esattamente a 91 ms; si veda quanto riassunto nella Tab. 2.

| Processo     | Durata<br>(ms) | media | Deviazione<br>standard (ms) |
|--------------|----------------|-------|-----------------------------|
| /C:/ > [C:]  | 124            |       | 35                          |
| /C:/ > [C_D] | 91             |       | 8                           |
| /C:/ > [C]   | 72             |       | 21                          |
| /C/ > [C]    | <b>79</b>      |       | <b>24</b>                   |

Tabella 2: durata media e deviazione standard in ms per i processi indicati

Nella tab. 3 riportiamo la durata registrata per ciascun allofono di /C:/ ripartita questa volta per fonema.

| Fonema | /C:/ > [C:] | DS | /C:/ > [C] | DS | /C:/ > [C_D] | DS |
|--------|-------------|----|------------|----|--------------|----|
| /p:/   | 132         | 14 | 89         | 8  | 76           | -- |
| /t:/   | 143         | 17 | 77         | 18 | 101          | 8  |
| /k:/   | 162         | 18 | 100        | 18 | 98           | 6  |
| /b:/   | 126         | 9  | 65         | 12 | 79           | -- |
| /d:/   | --          | -- | --         | -- | --           | -- |
| /g:/   | --          | -- | --         | -- | --           | -- |
| /tʃ:/  | 159         | 21 | 106        | -- | 89           | 6  |
| /dʒ:/  | 140         | 23 | 85         | -- | 100          | 12 |
| /f:/   | 120         | 10 | 82         | -- | 91           | 5  |
| /s:/   | 138         | 15 | 87         | 3  | 91           | 6  |
| /v:/   | --          | -- | 48         | 9  | --           | -- |
| /m:/   | 154         | 5  | 57         | 16 | 92           | 6  |
| /n:/   | 137         | 17 | 63         | 11 | 90           | 7  |
| /l:/   | 146         | 28 | 54         | 16 | 84           | 21 |
| /r:/   | 68          | -- | 31         | 6  | --           | -- |

Tabella 3: durata media in ms e deviazione standard (DS) del fonema geminato e delle sue realizzazioni allofoniche

La durata temporale di /b:/, /t:/ e /p:/ è approssimabile a quella rinvenuta da Payne (2005), la durata di /f:/ è invece minore nel nostro campione, quella di /m:/ e /l:/ maggiore. Opportune cautele sono tuttavia necessarie, a causa della diversità del materiale linguistico esaminato, italiano nativo vs. italiano non nativo.

La durata delle consonanti geminate è stata posta a confronto con un campione di fonemi scempi, omorganici per modo e luogo di articolazione, estratti dal medesimo archivio sonoro. Si tratta di un insieme di 6767 fonemi scempi correttamente realizzati dagli apprendenti non nativi<sup>18</sup>. Le misurazioni di durata sono contenute nella Tab. 4.

Il segmento più lungo è l'affricata /tʃ/ = ms 131, seguito dalle fricative /s/ e /f/, rispettivamente 108 e 101 ms, e dalle occlusive sorde /k/ = 102 ms, /p/ = 93 ms e /t/ = 91 ms. I fonemi più brevi si rinvencono nella classe delle liquide, /l/ = 58 ms e /r/ = 39 ms. Da rilevare la maggiore durata di /k/ rispetto a /t/ e /p/ (già Vaggies *et alii*, 1978 per l'italiano letto). Nel nostro caso, non è tuttavia da escludere la presenza di un certo condizionamento fonetico imputabile alla lingua nativa degli apprendenti; in tedesco, lo ricordiamo, le occlusive sorde sono sovente realizzate aspirate.

| Fonema | Occorrenze | Durata media | DS   | Ratio C: / C |
|--------|------------|--------------|------|--------------|
| /p/    | 506        | 93           | 23   | 1,41         |
| /t/    | 707        | 91           | 27,9 | 1,57         |
| /k/    | 687        | 102          | 21,8 | 1,58         |
| /b/    | 76         | 70           | 19,2 | 1,8          |
| /d/    | 382        | 60           | 22,5 | --           |
| /g/    | 27         | 75           | 25   | --           |
| /tʃ/   | 181        | 131          | 29,5 | 1,21         |
| /dʒ/   | 25         | 71           | 16,7 | 1,97         |
| /f/    | 124        | 110          | 36   | 1,09         |
| /s/    | 693        | 108          | 36,2 | 1,27         |
| /v/    | 164        | 59           | 25,2 | --           |
| /m/    | 752        | 70           | 25,1 | 2,2          |
| /n/    | 1525       | 64           | 31,6 | 2,6          |
| /l/    | 758        | 58           | 27,1 | 2,51         |
| /r/    | 159        | 39           | 18,4 | 1,74         |

Tabella 4: numero di occorrenze, durata media, deviazione standard e rapporto C:/C per i fonemi riportati

Nell'interlingua dei soggetti tedeschi, il fonema /k/ manifesta sempre un VOT significativamente lungo, in media 47 ms nel fonema scempio, 69 in quello geminato<sup>19</sup>, che contribuisce ad incrementare la durata complessiva del fonema velare<sup>20</sup>.

<sup>18</sup> Il numero dei fonemi scempi ricorrenti nella nostra banca dati è in realtà molto più ampio. Ai fini dell'analisi sono stati considerati solo i segmenti 'ben realizzati', ovvero quelli articolatoriamente rispondenti al bersaglio fonologico; per fare un esempio su 106 ricorrenze del fonema /dʒ/, solo 25 si conformano fedelmente al *target*.

<sup>19</sup> Il VOT (*Voice Onset Time*), lo ricordiamo, è un parametro fisico assunto dalla letteratura internazionale per distinguere occlusive aspirate da occlusive non aspirate. Questo parametro fa riferimento all'intervallo temporale compreso tra il rilascio dell'occlusione e

Le durate rilevate per i fonemi scempi sono in linea con quelle riportate in altri studi; tendenzialmente, le consonanti sorde mostrano una durata maggiore delle sonore (già Lehiste, 1969, e per l'italiano Dell'Aglia, 2004). Il rapporto aritmetico calcolato tra la durata del fonema geminato e quella del fonema scempio, C:/C, (cfr. Tab. 4) è sostanzialmente in linea con i dati per l'italiano nativo. Indici elevati si ravvisano per /n/ (2.6), /l/ (2.51) e /m/ (2.2), in posizione diametralmente opposta si colloca il rapporto ottenuto per /f/ (1.09), /tʃ/ (1.21), /s/ (1.27).

La rappresentazione istografica contenuta nella Fig. 4 mostra, in modo comparativo, la durata registrata per ciascun allofono di /C:/ accanto a quella del corrispettivo fonema scempio. Dai dati si evince che la variante degeminata, /C:/> [C] sia talora persino più breve di quella del fonema scempio; ciò sembra provare come il fonema geminato subisca spesso un vero processo di ipoarticolazione.

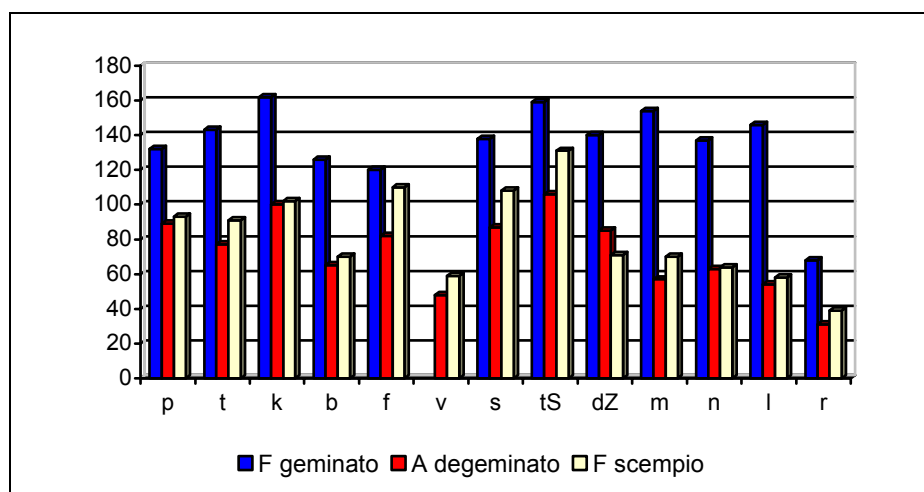


Figura 4: confronto tra la durata del fonema geminato correttamente realizzato, della sua realizzazione degeminata, e la durata del fonema scempio; F= fonema, A= allofono

l'attacco delle vibrazioni periodiche del fono successivo (cfr. Lisker e Abramson, 1964, Cho e Ladefoged, 1999); la sua durata è proporzionale al grado di posteriorità dell'occlusiva, essendo maggiore in /k/ e progressivamente minore in /t/ e /p/.

<sup>20</sup> La produzione fonetica delle occlusive sorde e sonore nell'interlingua di apprendenti tedeschi ha costituito di recente oggetto di specifica verifica empirica (cfr. Sorianello, 2008).



## 5. LA DURATA VOCALICA

L'opposizione di quantità consonantica ha una diretta conseguenza fonetica poiché rende predicibile, sebbene non distintiva, la durata della vocale precedente. Con riferimento all'italiano, la vocale accentata è infatti lunga in sillaba aperta non finale, breve in sillaba chiusa (cfr. Bertinetto, 1981, Marotta, 1985). Il mantenimento dell'opposizione temporale tra vocali toniche, poste rispettivamente prima di consonante geminata e prima di consonante scempia, è stato sottoposto a verifica sperimentale. Tramite uno *script* di estrazione automatica in formato *Perl*, abbiamo calcolato la durata delle vocali nei contesti indicati; il campione delle vocali è stato ripartito in timbri accentati e timbri non accentati<sup>21</sup>. Dai nostri dati tuttavia non emergono sostanziali differenze tra gli schemi sillabici posti a confronto, a prescindere dallo *status* accentuale, tonico o atono, della vocale. Questi sono i valori medi rinvenuti per i contesti osservati: 'VC.CV: ms 116 (DS: 30) vs. VC.CV: ms 125 (DS: 27); 'V:CV: ms 79 (DS: 15) vs. V:CV: ms 77 (DS: 13), si veda anche più esaurientemente quanto riportato nella Tab. 5;

| Apprendente | 'VC.CV   | 'V:CV    | VC.CV   | V:CV    |
|-------------|----------|----------|---------|---------|
| IE          | 123 (19) | 127 (29) | 81 (20) | 82 (10) |
| LM          | 118 (34) | 126 (23) | 87 (26) | 80 (12) |
| MB          | 108 (37) | 123 (28) | 68 (20) | 70 (16) |

Tabella 5: durata media in ms e deviazione standard (tra parentesi) della vocale tonica e atona nei contesti specificati

Nei tre soggetti, la vocale tonica mostra una durata lievemente superiore se la consonante successiva è scempia, lo scarto tuttavia non è statisticamente significativo; la vocale atona, come previsto, ha una durata simile tanto se seguita da un segmento consonantico breve, tanto se seguita da un segmento lungo.

Anche nell'italiano prodotto da apprendenti arabofoni marocchini, la correlazione tra la durata della vocale tonica e la struttura sillabica è molto debole, lo scarto temporale, di pochi ms, non è rilevante a livello statistico (cfr. Mori, 2007).

Quanto osservato sembra contravvenire la regola fonologica secondo la quale la vocale tonica in sillaba aperta detiene una durata maggiore rispetto alla stessa posta in sillaba chiusa<sup>22</sup>. In realtà, per quanto concerne l'italiano nativo, il comportamento della vocale è tutt'altro che sistematico. Più precisamente, l'opposizione temporale della vocale accentata tra sillaba aperta e chiusa sembrerebbe salvaguardata solo nel parlato stilisticamente molto controllato, come ad esempio nei testi elicitati mediante lettura. In Bertinetto (1981) la

<sup>21</sup> L'unità di riferimento per l'identificazione dell'accento è stata il lessema. In questa fase non sono stati considerati quei fenomeni di deaccentazione, tipici del parlato, che inducono sovente una ristrutturazione della struttura sillabico-accentuale del testo (cfr. Marotta, 1985). Questa scelta è stata dettata da specifiche esigenze tecniche imposte dal sistema di annotazione SAIL2.

<sup>22</sup> Questa tendenza è stata osservata anche in altre lingue. Ad es. in Madurese e in Toba Batak, lingue indonesiane, la durata della vocale posta prima di una consonante lunga risulta pressoché dimezzata rispetto alla durata della stessa ricorrente prima di una consonante scempia (cfr. Cohn *et al.*, 1999). Il risultato citato proviene dall'analisi di liste di frasi lette.

differenza è infatti significativa tanto nel campione delle parole isolate tanto in quello costituito da parole inserite in frasi cornice; i valori ottenuti per i tre segmenti della sequenza sono rispettivamente: 'V:CV= ms 142, 132, 87 (p. isolamento), 80 ms, 73, 43 (p. contesto), 'VC:V= 95 ms, 229, 71 (p. isolamento), 66, 116, 40 (p. contesto), (cfr. anche Farnetani e Kori, 1986).

Per converso, dati sperimentali ottenuti dall'analisi di parlato dialogico spontaneo (cfr. Albano Leoni *et alii* 1995, Landi & Savy, 1996, Dell'Aglio *et alii*, 2002, Sorianello & Calamai, 2005,) documentano come l'azione svolta dalla struttura sillabica sulla durata vocalica sia piuttosto debole. Il condizionamento diafasico sembra quindi rivelarsi una variabile alquanto robusta, capace di annullare, o quanto meno di ridurre, la distanza temporale tra vocale accentata in sillaba aperta e chiusa.

Pur tuttavia, un ulteriore elemento potrebbe concorrere alla definizione di questo scenario fonetico. Nello studio condotto da Endo e Bertinetto (1999) emerge come non tutte le consonanti, fra quelle considerate, mostrino la medesima propensione ad accorciare la vocale che le precede. Il comportamento è bilaterale: da un lato si pongono le ostruenti /t: d: s: ʃ ts/ le quali, essendo lunghe, chiudono la sillaba rendendo più breve la vocale precedente, dall'altro si collocano le sonoranti /n: ɲ l: ʎ/, le quali, non sembrano condizionare la durata della vocale posta immediatamente a sinistra.

Questo controllo è stato effettuato anche nel nostro materiale linguistico; i dati ottenuti collimano solo in parte con quelli degli autori menzionati, sebbene siano meno sistematici e, per alcune classi di suoni, persino contraddittori<sup>23</sup>. Gli esiti sono stati ricavati ponendo a confronto, per ciascuna consonante considerata<sup>24</sup>, la durata della vocale accentata nelle due sequenze 'VC.C e 'V.CV. Per quanto concerne le occlusive, la vocale tonica è più breve prima di /p:/, /t:/, /b:/, ma non prima di /k:/. Anche la fricativa /s:/ e l'affricata /tʃ:/ chiudono la sillaba, lo stesso non avviene però per /f:/. Più omogeneo appare il comportamento delle sonoranti /m: n: r:/; la vocale che precede il fonema geminato è sistematicamente più lunga rispetto alla stessa ricorrente prima del fonema scempio, l'unica eccezione è costituita da /l:/ la cui ricorrenza non modifica la durata vocalica<sup>25</sup>.

La valenza empirica degli esiti fin qui discussi, pur essendo informativa, non può a nostro avviso essere generalizzata; essendo stata ricavata da un campione di italiano interlingua, che come tale risente, a tutti i livelli, di una serie di condizionamenti imputabili alla L1 dei soggetti. Va altresì ricordato che l'eloquio degli apprendenti esaminati manifesta profonde alterazioni ritmico-prosodiche le quali, sovvertono l'equilibrio ritmico- temporale della lingua italiana.

<sup>23</sup> In merito a questo aspetto, i dati sull'italiano non sono sempre univoci; in Landi e Savy (1996) si registrano persino casi in cui il rapporto tra la durata della vocale e la struttura della sillaba è del tutto sovvertito, essendo la vocale più lunga in sillaba chiusa che non in quella aperta.

<sup>24</sup> In questa sede non è stato valutato il comportamento dei fonemi lunghi in posizione intervocalica, ossia /ts dz ʃ ɲ ʎ /.

<sup>25</sup> L'allungamento della vocale prima di C sonorante, *in primis* della vibrante, è un condizionamento contestuale peraltro ben noto in letteratura (cfr., fra gli altri, Peterson & Lehiste, 1960; Lehiste, 1969 e, per l'italiano, Farnetani & Kori, 1986).

## 6. DISCUSSIONE

Dai dati sperimentali emerge una situazione variabile in cui anche in medesimo intorno contestuale, il fonema lungo è talora conforme al *target*, altre volte realizzato breve. Complessivamente, l'incidenza statistica delle realizzazioni canoniche appare dominante (54,1%). Ciò prova che il processo di acquisizione di questa nuova categoria fonologica, C [+lungo], sia ben avviato, sebbene non ancora consolidato. A tale riguardo, è interessante osservare l'evoluzione diacronica cui è soggetta la formazione di tale competenza fonologica. Nel primo stadio dell'interlingua, la percentuale di realizzazioni coincidenti con il bersaglio fonologico è pari al 42,1%, mentre nell'ultima registrazione, effettuata a distanza di circa sei mesi dalla prima, la produzione del fonema geminato si attesta su valori percentuali più elevati, esattamente 56,6%. L'incremento rinvenuto a favore della resa geminata, si traduce, parallelamente, in un decremento della variante allofonica scempia e della categoria intermedia; del tutto inalterata è invece la situazione per quanto concerne le altre varianti (6.6%), si veda a riguardo quanto riportato nella Fig. 5.

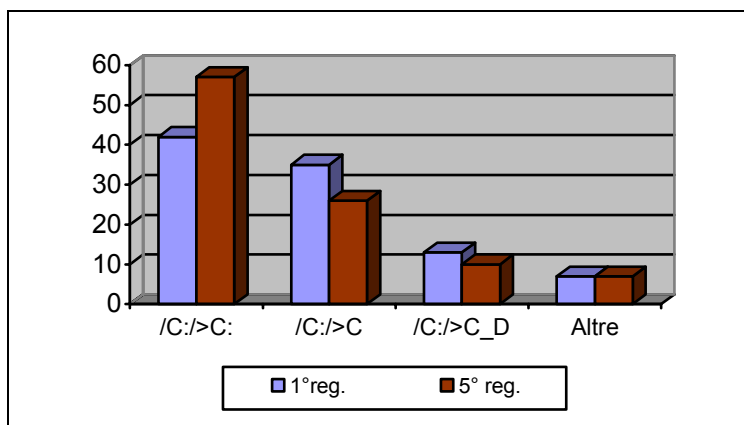


Figura 5: valori percentuali medi per resa allofonica relativi alla prima ed ultima sessione di registrazione

L'acquisizione di questa nuova competenza fonologica si prefigura come un processo graduale e lento, sostanzialmente continuo, sebbene i soggetti possono presentare nel loro percorso momenti di rallentamento acquisizionale e persino dei blocchi, determinando vere fasi di fossilizzazione<sup>26</sup>.

A fronte della validità di quanto detto, va comunque osservato come i tredici fonemi considerati non manifestino la medesima propensione alla riduzione temporale. Come già discusso in § 3 (cfr. Tab. 1 e Fig. 1), nei locutori tedeschi, l'attivazione del tratto di lunghezza consonantica è di gran lunga maggiore in alcuni fonemi, ad es. /s/ o /t/, che non in altri, come /l/ o /r/. Ad una prima analisi, la distribuzione del fenomeno appare poco sistematica; più chiaro ed eloquente è il quadro che risulta dal riordino dei valori percentuali di resa geminata secondo una scala progressivamente decrescente. Dalla rappresentazione istografica riportata nella Fig. 6 si evince che il *target* geminato è più

<sup>26</sup> Tale tendenza è peraltro ben documentata nella letteratura specialistica (cfr. Major, 2001).

frequentemente raggiunto nei suoni sordi, occlusivi, fricativi e affricati. Significativamente inferiore è, per contro, l'incidenza del fenomeno tra le consonanti sonore.

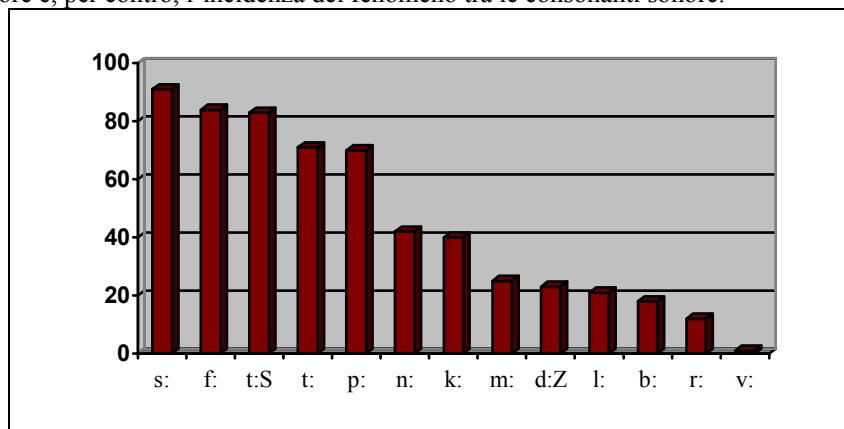


Figura 6: valori percentuali di geminazione in ordine decrescente

Quanto emerge è in linea con quanto postulato nelle classiche gerarchie di sonorità o di forza consonantica<sup>27</sup>. Alle consonanti ostruenti, e tra queste alle sorde, è sempre assegnato un indice di forza superiore rispetto alle altre classi di suoni naturali. In quanto forti, queste consonanti sono anche più inclini a subire un ulteriore rafforzamento, qual è appunto la geminazione. Inversamente, le consonanti liquide, suoni dotati di maggior grado di sonorità, sono segmenti tendenzialmente deboli.

Analiticamente, le fricative sorde /s:/ e /f:/, seguite dall'affricata /t:s/ e dalle occlusive sorde /p:/ e /t:/, sono pronunciate lunghe con percentuali che oscillano dall'91% al 70%. Diverso il comportamento delle occlusive /b:/ e /k:/; per /b:/ si rinviene una scarsa presenza di varianti lunghe (18%); questo fonema, nella quasi totalità dei casi ricorrente in una forma verbale (ad es. *abbiamo*, *dobbiamo*, *sarebbe*), mostra di norma una forte riduzione temporale, sovente accompagnata dall'approssimazione del gesto articolatorio. Più alta la percentuale di corretta realizzazione rinvenuta per /k:/ (40%); il comportamento di questa occlusiva si allontana da quello osservato per /p:/ e /t:/, a riprova della maggiore debolezza del luogo di articolazione velare<sup>28</sup>. Relativamente alla fricativa labiodentale sonora, si osserva la totale assenza della variante lunga; il fonema geminato /v:/ è difatti sempre scempio o, più spesso, approssimante<sup>29</sup>. Un discorso a parte va fatto per /d:z/, la percentuale di geminazione riscontrata per questa consonante è in realtà ben più alta, il più delle volte,

<sup>27</sup> Nelle gerarchie di scala consonantica, la forza di un suono è inversamente proporzionale al grado di sonorità detenuto; di conseguenza, suoni intrinsecamente sonori (vocali, approssimanti, liquide e nasali) sono anche i più deboli, mentre, per converso, i suoni dotati di scarsa sonorità (ad esempio, le ostruenti) sono forti. Cfr. Hooper (1972), Vennemann (1988), Kenstowicz (1994).

<sup>28</sup> Il luogo di articolazione velare risulta anche tipologicamente meno frequente (13.6%) rispetto a quello alveolare (15,3%) e bilabiale (14.3%), (cfr. Stefanuto & Vallée, 1999).

<sup>29</sup> Non va tuttavia ignorato che pure in italiano, la realizzazione intervocalica di /v/ è sovente approssimante (cfr. Rispoli & Savy, 1993). Non si può pertanto escludere che questo modello di pronuncia funga da *input* anche nell'acquisizione dell'italiano da parte di soggetti stranieri.

però, l'affricata risultante è [tʃ], segmento cioè sordo e non sonoro<sup>30</sup>; ai fini del computo abbiamo considerato solo le realizzazioni fonetiche conformi al fonema bersaglio. Va altresì rilevato che i fonemi /b:/, /k:/, /dʒ/ e /v:/, come pure /d:/ e /g:/, questi ultimi assenti nel nostro campione, hanno in italiano una scarsa frequenza statistica; tra la classe delle geminate equivalgono pertanto a suoni marcati per diffusione e distribuzione (cfr. Albano Leoni & Clemente, 2005)<sup>31</sup>.

Nell'italiano L2, la produzione del fonema lungo si prefigura quindi come un evento complesso di non facile o immediata acquisizione che, da un lato, si manifesta come un vero e proprio fallimento articolatorio, dando luogo a categorie fonetiche caratterizzate da un diverso grado di riduzione temporale, dall'altro genera non pochi ipercorrettismi, ad es. *befana* > *beffana*, *dopo* > *doppo*, *latino* > *lattino*. Nella Fig. 7 riportiamo, a scopo esemplificativo, l'ipercorrezione di /t/ > [t:], ms 223, nel lessema *latino*; si noti nello stesso enunciato la realizzazione canonica di /t:/ in *quattro*, ms 121, e lo scempiamento di /n:/ in *anni*, 50 ms.

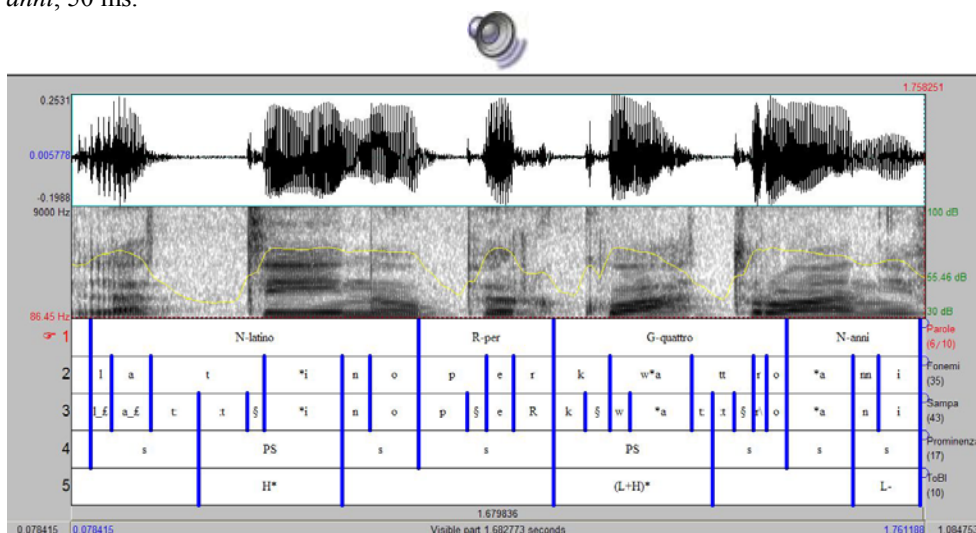


Figura 7: forma d'onda, sonagramma a banda larga e annotazione in SAIL2 dell'enunciato *latino per quattro anni*, realizzato dall'apprendente MB

Fondamentale, ancora una volta, è il fatto che l'ipercorrettismo, presente in tutti e tre i soggetti, a prescindere dallo schema accentuale, coinvolge in modo maggioritario solo alcune consonanti, tipicamente /s/ (31%), /f/, /t/ (18%) e /p/ (16%), (cfr. Fig. 8). Anche in

<sup>30</sup> Questa pronuncia non è affatto sorprendente, se consideriamo che tale affricata, anche nella sua versione scempia, assente nell'inventario fonologico della lingua tedesca, è quasi sempre sostituita con il corrispondente suono sordo, ad es. *oggi* > ['ɔtʃi], *Perugia* > [pe'ru:tʃa], (già Costamagna, 2003).

<sup>31</sup> Questo rilievo fonologico contrasta, almeno per /b/ e /dʒ/, con il quadro fonetico rilevabile in numerose varietà di italiano centro-meridionale, una vasta area geografica in cui la realizzazione intervocalica dei fonemi citati è sistematicamente lunga, ad es. *abito* > ['ab:ito], *agile* > ['ad:ʒile].

questo caso, si rinvenivano categorie mediane, annotate come C\_G, dove G sta per geminazione; tali varianti risultano prevalenti per /t/, /s/ /p/ e /k/.

Per quanto concerne il comportamento temporale, questi segmenti mostrano una durata approssimabile a quella del fonema lungo, in media 151 ms (DS: 25.5) per quelli geminati, ovvero /C/ > [C:], 119 ms (DS: 16) per la resa intermedia, ovvero /C/ > [C\_G].

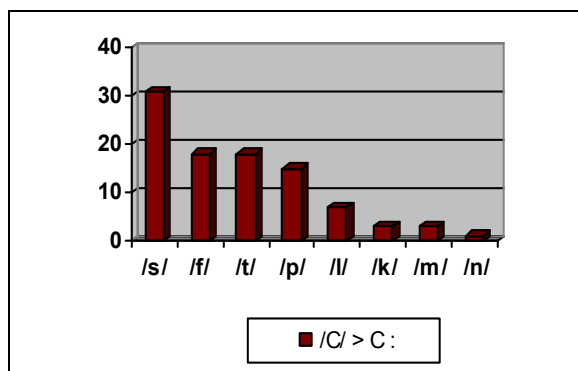


Figura 8: valori percentuali di ipercorrettismo /C/ > [C:]

Il quadro fin qui delineato rafforza l'ipotesi secondo cui l'attivazione del tratto di lunghezza, con o senza valore funzionale, sia avvertito come più naturale tra i suoni più forti, già dotati di maggiore durata intrinseca e, non a caso, segmenti tipologicamente più diffusi e quindi più naturali.

Le tendenze emerse in questa ricerca si prestano ad essere interpretate alla luce dei postulati teorici postulati da J. Flege (1987; 1991; 1995; 1999; 2003) nell'ambito del noto modello acquisizionale denominato *Speech Learning Model* (SLM). Si tratta del primo apparato teorico che ha rivalutato il ruolo della percezione nell'acquisizione di una seconda lingua, evidenziando l'esistenza di un rapporto implicazionale tra la percezione dei suoni e la loro produzione. Secondo Flege, la formazione di una categoria fonologica è possibile solo se gli apprendenti riescono a discriminare la differenza fonetica che esiste, per un determinato suono, tra la loro lingua nativa e la lingua seconda<sup>32</sup>. Nell'interlingua di soggetti non nativi, molti errori di produzione, secondo l'autore, sarebbero così causati da un fallimento percettivo. Ciò nonostante, in questa elaborazione percettiva non tutti i suoni si comportano allo stesso modo; essenziale è a tale riguardo la distinzione, postulata da Flege, tra suoni simili e suoni dissimili<sup>33</sup>. Maggiore è la distanza tra i suoni di L1 e quelli di L2, maggiore sarà anche la probabilità che tali differenze vengano percepite.

<sup>32</sup> Altri fattori, tuttavia, influenzano l'esito e la rapidità dei meccanismi di apprendimento. Un ruolo fondamentale è riconosciuto da Flege (1987; 2003) al fattore 'tempo', nel suo duplice aspetto: tempo di inizio del processo di apprendimento ed età di arrivo nel paese di immigrazione. Il livello di *native-likeness* aumenta in modo direttamente proporzionale alla precocità con cui ha inizio il processo medesimo.

<sup>33</sup> E' 'nuovo' un suono di L2 distante, sul versante acustico e percettivo, dal suono di L1 ad esso più vicino, tanto da richiedere, in fase di trascrizione fonetica, un nuovo simbolo o l'aggiunta di un diacritico IPA. Un suono 'simile', invece, condivide molti più indici spettro-acustici con un determinato suono di L1 ed è foneticamente trascritto allo stesso modo (cfr. Flege, 1987).

La classe dei fonemi consonantici lunghi dell'italiano rappresenta una categoria fonologica nuova per gli apprendenti tedeschi. In conformità alla *Markedness Differential Hypothesis* (cfr. Eckman, 1987), all'inizio prevalgono i comportamenti più naturali e meno marcati, ovvero C [- lungo], nel tempo però le categorie nuove avranno maggiori probabilità di essere apprese rispetto ai suoni simili, poiché i meccanismi di discriminazione sono più efficaci; la distinzione tra suoni dissimili risulta infatti dotata di maggiore salienza percettiva. L'acquisizione di un suono 'simile' è, per contro, spesso bloccata da un meccanismo di classificazione equivalente; in quest'ultimo caso la categoria, già presente in L1, sarà usata per processare percettivamente sia i suoni di L1 che i suoni di L2, determinando frequenti errori di sostituzione.

L'interpretazione dei nostri esiti empirici non può, tuttavia, prescindere dal riferimento ai principi di naturalezza e di marcatezza tipologica. Marcatezza e somiglianza tra suoni sono spesso assunti quali validi indici predittori del grado di difficoltà cui va incontro un soggetto nell'acquisizione di una lingua seconda; entrambi infatti rallentano il processo medesimo (cfr. Major, 2001). Un assunto della linguistica acquisizionale è che nell'apprendimento di una lingua seconda, i processi marcati siano acquisiti più lentamente (cfr. Eckman, 1987; Major, 2001)<sup>34</sup>. Ciò spiega perché, sebbene tutti i fonemi lunghi dell'italiano siano suoni nuovi per i tedescofoni esaminati, l'insorgenza del tratto di lunghezza non avvenga in eguale misura per tutte le consonanti. Il processo di geminazione avviene per stadi progressivi, ma ha inizio da alcune consonanti prototipiche, nello specifico le ostruenti sorde. Si tratta di suoni più stabili dal punto di vista articolatorio, che meglio rispondono ai criteri di massima resa percettiva ed acustica e che pertanto realizzano una condizione fonetica e articolatoria ottimale.

Posta questa condizione, non si può d'altra parte escludere che, nel processo di progressiva acquisizione del tratto di lunghezza consonantica, intervengano anche altri fattori condizionanti. Dai nostri dati si evince infatti come il processo di geminazione sia, in qualche misura, correlato anche alla natura morfologica del lessema contenente i fonemi lunghi. Il tasso di conformità al *target* fonologico è difatti più alto nei Nomi e nei Verbi; in maniera del tutto speculare lo scempiamento della consonante lunga, contravvenendo quanto riscontrato nell'italiano L2 da Giannini e Costamagna (1997; 1998), è maggioritario nelle parti del discorso funzionali, come la Preposizione o la Congiunzione, classi lessicali chiuse, limitate per numero di entrate, frequentemente soggette a fenomeni di ipoarticolazione e di riduzione segmentale. In questa direzione abbiamo volto il nostro interesse, al fine di pervenire ad una conoscenza integrata, non solo fonologica, del processo in esame, (cfr. Soriano, in stampa).

---

<sup>34</sup> I principi linguistici universali e naturali hanno un ruolo centrale anche nel modello acquisizionale elaborato da Major (2001), denominato OPM (*Ontogeny Phylogeny Model*). In questo sistema teorico, l'interlingua è il risultato dell'azione congiunta di tre componenti: L1 (sistema linguistico di partenza), L2 (sistema linguistico di arrivo) e U, ovvero gli universali.

## BIBLIOGRAFIA

Albano Leoni, F. Cutugno, F. & Savy, R. (1995), The vowel system of Italian connected speech, in Elenius, K. & Branderud, P. (eds.), *Proceedings of the 13<sup>th</sup> International Congress of Phonetic Sciences*, Stockholm: Stockholm University, vol. 4, pp. 396-399.

Albano Leoni, F. & Clemente C. (2005), Numeri, fonemi e foni, in De Mauro, T. & Chiari, I. (acd), *Parole e numeri*, Roma: Aracne, 27-44.

Arvaniti, A. (1999), Effects of speaking rate on the timing of single and geminate sonorants, in Ohala J. J., Hasegawa Y., Ohala M., Granville, D. & Bailey, A. C. (eds.), *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, University of California, San Francisco, USA, 3 voll., 599-602.

Bernini, G. (1988), Questioni di fonologia nell'italiano lingua seconda, in Giacalone Ramat A., (acd.), *L'italiano tra le altre lingue: strategie di acquisizione*, Bologna: Il Mulino, 77-90.

Bertinetto, P.M. (1981), *Strutture prosodiche dell'italiano*, Firenze: Accademia della Crusca.

Boula de Mareüil, P., Marotta, G. & Adda-Decker, M. (2004), Contribution of prosody to the perception of Spanish/Italian accents, in Bel, B. & Marlien, I. (eds.), *Speech Prosody 2004*, Nara (Giappone), 681-684.

Celata, C. (2004), *Acquisizione e mutamento di categorie fonologiche*, Milano: Franco Angeli.

Cho, T. & Ladefoged, P. (1999), Variation and universals in VOT: evidence from 18 languages, *Journal of Phonetics* 27, 207-221.

Cohn, A. C., Ham, W. H & Podesva, R. J. (1999), The phonetic realization of singleton-geminate contrasts in three languages of Indonesia, in Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D. & Bailey, A. C (eds.), *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, University of California, San Francisco, USA, 3 voll., 587-590.

Costamagna, L. (2003), *Affricates in Italian as L2: the role of psycho-attitudinal parameters*, in Costamagna, L. & Giannini, S. (acd), (2003), 95-129.

Costamagna, L. & Giannini, S. (acd), (2003), *La fonologia dell'interlingua. Principi e metodi di analisi*, Milano: Franco Angeli.

Dell'Aglio, M. (2004), Spunti sulla durata consonantica in pisano, in Mori, L. & Stefani, M. (acd), *Costituzione, gestione e restauro di corpora vocali*, Atti delle XIV Giornate del Gruppo di Fonetica Sperimentale, Viterbo, Roma: Esagrafica, 159-164.

Dell'Aglio, M., Bertinetto, P.M. & Agonigi, M. (2002), *Le durate dei foni vocalici in rapporto al contesto nel parlato di locutori pisani: primi risultati*, in Regnicoli, A. (acd), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia*, Atti delle XII Giornate di Studio del Gruppo di Fonetica Sperimentale, Macerata 2001, Roma: Il Calamo, 53-58.



Eckman, F. R. (1987), Markedness and the contrastive hypothesis, in Ioup, G. & Weinberger, S. H (eds.), *Interlanguage phonology*, Cambridge: Newbury House Publishers, 55-69.

Endo, R. & Bertinetto, P.M. (1999), Caratteristiche prosodiche delle così dette 'rafforzate' italiane, in Delmonte, R. & Bristot, A. (acd), *Aspetti computazionali in fonetica, linguistica e didattica delle lingue: modelli e algoritmi*, Roma: Esagrafica, 243-255.

Farnetani, E. & S. Kori (1986), Effects of syllable and word structure on segmental duration in spoken Italian, *Speech Communication* 5, 17-34.

Flege, J. E. (1987), The production of new and similar phones in a foreign language. Evidence for the effect of equivalence classification, *Journal. of Phonetics* 15, 47-65.

Flege, J. E. (1995), Second-language speech learning: findings and problems, in Strange, W. (ed.), *Speech perception and linguistic experience. Issues in cross-languages research*, Timonium: York Press, 233-277.

Flege, J. E. (1997), The Role of phonetic category formation in second-language speech learning", *New Sounds 97, Proceedings of the Third International Symposium on the Acquisition of Second Language Speech*, University of Klagenfurt, 8-11 September 1997: 79-88.

Flege, J. E. (1999), The relation between L2 production and perception, in Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D. & Bailey, A. C (eds.), *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, University of California, San Francisco, USA, 1273-1276.

Flege, J. E. (2003), *Assessing constraints on second-language segmental production and perception*, in Schiller, N. A. & Meyer, A. S. (eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, Berlin: Mouton de Gruyter, 319-355.

Gamal, D. (2007), Sul ritmo in italiano L2: l'eventuale isocronia, in Giordani, V., Bruseghini, V. & Cosi, P. (acd), *Scienze vocali e del linguaggio*, Atti del 3° Convegno nazionale AISV (Associazione Italiana di Scienze della Voce), Povo di Trento, Ravenna: EDK Editore, CD-Rom, 101-118.

Giacalone Ramat, A., (acd) (2003), *Verso l'italiano. Percorsi e strategie di acquisizione*, Roma: Carocci.

Giannini, S. (1997), Le strategie di acquisizione della fonologia. Preliminari ad una ricerca sull'italiano L2, in Ambrosini, R., Bologna, M. P., Motta, F. & Orlandi, Ch. (acd), *Scritta a aim n-ogaim. Studi in memoria di E. Campanile*, Pisa: Pacini, 605-633.

Giannini, S. & Celata, C. (2002-2003), Semplicità e semplificazione percettiva. Qualche dato sperimentale dallo studio dell'acquisizione, *Studi e saggi linguistici, Atti del Convegno di Studi in memoria di Tristano Bolelli*, Pisa: ETS, 121-145.

Giannini, S. & Costamagna, L. (1997), Language learning strategies in interlanguage phonology of Italian L2: a case history: the acquisition of consonant length, in Leather, J. &

- James, A. (eds), *New sounds 97, Proceedings of the 3° International Symposium on the acquisition of second language speech*, 96-103.
- Giannini, S. & Costamagna, L. (1998), Acquisizione di categorie fonologiche e diffusione lessicale del mutamento linguistico: affinità strutturali, *Archivio Glottologico Italiano* 85, 150-187.
- Hassan, Z. M. (2003), Temporal compensation between vowel and consonant in Swedish and Arabic in sequences of CV:C and CVC: and the word overall duration, *Phonum*, 9, 45-48.
- Hooper, J. B. (1976), *An introduction to natural generative phonology*, New York: Academic Press.
- Kenstowicz, M. (1994), *Phonology in generative grammar*, Oxford: Blackwell.
- Khatab, G. (2007), A phonetic study of gemination in Lebanese Arabic, in Trouvain, J. & Barry, W. J. (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 6-10 august 2007, 153-158.
- Kraehenmann, A. & Jaeger M. (2003), Phrase-initial geminate stops: articulatory evidence for phonological representation, in Solé, M. J., Recasens, D. & Romero, J. (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona (Spain) 3-9 August 2003, 2725-2728.
- Kunnari, S., Nakai, S. & Vihman, M. M. (2001), Cross-linguistic evidence for the acquisition of geminates, *Psychology of language and communication*, 5, 13-24.
- Landi, R. & Savy, R. (1996), Durata media, struttura sillabica e velocità di eloquio nel parlato connesso, in Peretti, A. & Simonetti, P. (acd), *Atti del XIV Convegno nazionale dell'Associazione Italiana di Acustica (AIA)*, Padova: Arti Grafiche Padovane, 65-70.
- Lehiste, I. (1969), *Suprasegmentals*, Cambridge (Ma): MIT Press.
- Lisker, L. & Abramson, A. S. (1964), A cross-language study of voicing in initial stops: acoustical measurements, *Word*, 20, 384-422.
- Maddieson, I. (1984), *Patterns of sounds*, Cambridge, CUP.
- Maddieson, I. (1997), Phonetic universals, in Hardcastle, W. J. & Laver, J. (eds.), *The Handbook of Phonetic Sciences*, Cambridge (Ma): Blackwell, 619-639.
- Major, R. C. (2001), *Foreign accent: the ontogeny and phylogeny of second language phonology*, London: Lawrence Publishers.
- Marotta, G. (1985), *Modelli e misure ritmiche*, Bologna: Zanichelli.
- Marotta, G. (2008), Sulla percezione dell'accento straniero, in Lazzeroni, R., Banfi, E., Bernini, G., Chini, M. & Marotta, G., (acd), *Diacronica et synchronica. Studi in onore di Anna Giacalone*, Pisa: ETS.

- Marotta, G. & Sorianello, P. (2008), SAIL2: un sistema integrato di annotazione morfo-fonologica per l'italiano L2, in Andorno, C. (acd), *Corpora di italiano L2: tecnologie, metodi, spunti teorici*, Perugia: Guerra.
- Mori, L. (2007), *Fonetica dell'italiano L2. Un'indagine sperimentale sulla variazione nell'interlingua dei marocchini*, Roma: Carocci.
- Nespor, M. (1993), *Fonologia*, Bologna: Il Mulino.
- Payne, E.M. (2005), Phonetic variation in Italian consonant gemination, *Journal of the International Phonetic Association* 35 (2), 153-189.
- Peterson, G. E. & Lehiste, I. (1960), Duration of syllable nuclei in English, *Journal of Acoustical Society of America*, 32, 693-703.
- Rispoli, G. & Savy, R. (1993), Alcune considerazioni spettroacustiche sulla [v] italiana, in *Atti del XI Convegno Nazionale dell'Associazione Acustica Italiana (AIA)*, Padova, 91-94.
- Schmid, S. (1994), *L'italiano degli spagnoli. Interlingue di immigrati nella Svizzera tedesca*, Milano: Franco Angeli.
- Sorianello, P. (2008), Persistenza di categorie fonologiche nell'italiano L2, in Costamagna, L. & Marotta, G. (acd), *Processi fonetici e categorie fonologiche nell'acquisizione dell'italiano*, Pisa: Pacini, 85-110.
- Sorianello, P. (in stampa), Categorie morfo-fonologiche nell'italiano interlingua, *Studi e Saggi linguistici*.
- Sorianello, P. & Calamai, S. (2005), *Il sistema vocalico romano*, in Albano Leoni, F. & Giordano, R. (acd), *Italiano parlato. Analisi di un dialogo*, Napoli: Liguori, 25-70.
- Stefanuto, M. & Vallée, N. (1999), Consonant systems: from universal trends to ontogenesis, in Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D. & Bailey, A. C (eds.), *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, University of California, San Francisco, USA, 1973-1976.
- Vaggies, K., Ferrero, F. E., Magno Caldognetto, E. & Lavagnoli, C. (1978), Some acoustic characteristics of Italian consonants, *Journal of Italian Linguistics* 3-4, 69-85.
- Vennemann, T. (1988), *Preference laws for syllable structure and the explanation of sound change*, Berlin: Mouton de Gruyter.
- Wells, J.C. (1994), Computer-coding the IPA: a proposed extension of SAMPA, disponibile presso il sito <http://www.phon.ucl.ac.uk/home/sampa/home.htm>
- Wiese, R. (1996), *The phonology of German*, Oxford: Clarendon Press.

# UN CONTRIBUTO PER IL CONFRONTO TRA L'INTONAZIONE VENETA E QUELLA ANDALUSA<sup>1</sup>

Antonio Romano<sup>1</sup> & Renzo Miotti<sup>2</sup>

<sup>1</sup>Dip. di Scienze del Linguaggio - Lab. Fon. Sperimentale "A. Genre" – Università degli Studi Torino

<sup>2</sup>Dip. di Romanistica – Università degli Studi di Verona

*antonio.romano@unito.it, renzo.miotti@univr.it*

## 1. INTRODUZIONE

Com'è noto da numerose fonti (per una rassegna bibliografica v. Romano, 2001) è nell'intonazione delle frasi interrogative totali che sono spesso presenti le caratteristiche regionali più marcate, con differenze che appaiono piuttosto evidenti nel parlato spontaneo.

In particolare, nel panorama linguistico italiano, alcune varietà venete sono note per essere facilmente riconoscibili per la presenza di alcuni schemi ascendenti – che altrove sarebbero associati a domande – anche in enunciati conclusivi con valore dichiarativo.

Un altro luogo comune vorrebbe le domande formulate da alcuni parlanti di questa stessa regione come vagamente riconducibili ad alcuni schemi tipici di varietà iberoromanze.

In questo studio pilota, condotto applicando la metodologia *AMPER*, si confrontano i dati ottenuti per la varietà veneta (liventina) di Motta di Livenza (TV) con quelli della varietà iberica di Malaga. Grazie alle osservazioni svolte sulle realizzazioni presenti in due corpora di riferimento, abbiamo potuto osservare come, negli andamenti melodici delle interrogative nelle diverse varietà, si evidenzia un Contorno Terminale di Modalità (*CTM*) ascendente con caratteristiche simili (ma ancora tali da permettere di discriminare le diverse varietà una volta sopprese le informazioni lessicali e segmentali), associate a particolari condizioni di organizzazione temporale.

## 2. MATERIALI E METODO

Il confronto tra le caratteristiche prosodiche delle parlate delle due regioni studiate è basato su un campione di realizzazioni raccolte per due varietà: la varietà andalusa di Malaga e una varietà di italiano regionale (it. reg.) veneto (quella liventina di Motta di Livenza, provincia di Treviso, Veneto orientale). La scelta delle due varietà è avvenuta sulla base della disponibilità degli informatori registrati, due parlanti di età compresa tra i 20 e 40 anni, e grazie all'esperienza dell'autore RM in questi due sub-domini linguistici.

I corpora di riferimento, realizzati appositamente per questo studio, sono basati su alcune sezioni dei questionari adottati in *AMPER-ESP* e *AMPER-ITA* (v. *AMPER - Géolinguistique*, 2005 e siti *web*; le frasi usate per l'italiano sono state calcate su una selezione delle 63 del corpus *AMPER-ESP*; v. Fernández Planas, 2005; Martínez Celdrán & Fernández Planas, 2006).

---

<sup>1</sup> Questo articolo presenta il risultato di una collaborazione tra i due autori. Sebbene concepita congiuntamente, la sua redazione finale (a cura di AR) si presenta strutturata in sezioni alle quali i due autori hanno dato distinti contributi: RM è particolarmente responsabile dei paragrafi 2 e 5.1 mentre AR ha curato tutte le analisi e la presentazione dei confronti proposti nei paragrafi 4, 5.2. Le altre parti sono frutto di una redazione congiunta.

Presentiamo qui di seguito la lista delle frasi corrispondenti dei due questionari sfruttati per questo lavoro, per lo spagnolo (castigliano) e per l'italiano, separandole per mezzo del codice *AMPER* che definisce convenzionalmente la struttura accentuale della frase:

|                                                      |     |                                                   |
|------------------------------------------------------|-----|---------------------------------------------------|
| 1. <i>La guitarra se toca con pánico</i>             | twp | <i>La chitarra si suona con panico</i>            |
| 2. <i>La guitarra se toca con paciencia</i>          | twt | <i>La chitarra si suona con pazienza</i>          |
| 3. <i>La guitarra se toca con obsesión</i>           | twk | <i>La chitarra si suona con serietà</i>           |
| 4. <i>La guitarra se toca con pánico práctico</i>    | twv | <i>La chitarra si suona con panico pratico</i>    |
| 5. <i>La guitarra se toca con pánico finito</i>      | twf | <i>La chitarra si suona con panico costante</i>   |
| 6. <i>La guitarra se toca con paciencia práctica</i> | twz | <i>La chitarra si suona con pazienza pratica</i>  |
| 7. <i>La guitarra se toca con paciencia finita</i>   | twS | <i>La chitarra si suona con pazienza costante</i> |
| 8. <i>La guitarra se toca con obsesión práctica</i>  | twj | <i>La chitarra si suona con serietà pratica</i>   |
| 9. <i>La guitarra se toca con obsesión finita</i>    | twx | <i>La chitarra si suona con serietà costante</i>  |
| 10. <i>La cítara se toca con pánico</i>              | pwp | <i>Il piffero si suona con panico</i>             |
| 11. <i>La cítara se toca con paciencia</i>           | pwt | <i>Il piffero si suona con pazienza</i>           |
| 12. <i>La cítara se toca con obsesión</i>            | pwk | <i>Il piffero si suona con serietà</i>            |
| 13. <i>La cítara se toca con pánico práctico</i>     | pwv | <i>Il piffero si suona con panico pratico</i>     |
| 14. <i>La cítara se toca con pánico finito</i>       | pwf | <i>Il piffero si suona con panico costante</i>    |
| 15. <i>La cítara se toca con paciencia práctica</i>  | pwz | <i>Il piffero si suona con pazienza pratica</i>   |
| 16. <i>La cítara se toca con paciencia finita</i>    | pws | <i>Il piffero si suona con pazienza costante</i>  |
| 17. <i>La cítara se toca con obsesión práctica</i>   | pwj | <i>Il piffero si suona con serietà pratica</i>    |
| 18. <i>La cítara se toca con obsesión finita</i>     | pwX | <i>Il piffero si suona con serietà costante</i>   |
| 19. <i>El saxofón se toca con pánico</i>             | kwp | <i>Il carillon si suona con panico</i>            |
| 20. <i>El saxofón se toca con paciencia</i>          | kwt | <i>Il carillon si suona con pazienza</i>          |
| 21. <i>El saxofón se toca con obsesión</i>           | kwk | <i>Il carillon si suona con serietà</i>           |
| 22. <i>El saxofón se toca con pánico práctico</i>    | kww | <i>Il carillon si suona con panico pratico</i>    |
| 23. <i>El saxofón se toca con pánico finito</i>      | kwf | <i>Il carillon si suona con panico costante</i>   |
| 24. <i>El saxofón se toca con paciencia práctica</i> | kwz | <i>Il carillon si suona con pazienza pratica</i>  |
| 25. <i>El saxofón se toca con paciencia finita</i>   | kws | <i>Il carillon si suona con pazienza costante</i> |
| 26. <i>El saxofón se toca con obsesión práctica</i>  | kwj | <i>Il carillon si suona con serietà pratica</i>   |
| 27. <i>El saxofón se toca con obsesión finita</i>    | kwX | <i>Il carillon si suona con serietà costante</i>  |

A partire dall'analisi di queste 27 frasi nelle due modalità e delle 3 ripetizioni elicitate a due locutori rappresentativi delle due località, sono stati ottenuti grafici dell'evoluzione di  $f_0$  e delle sequenze di durate dei segmenti vocalici che hanno permesso di eseguire un certo numero di confronti.

In un primo momento sono state valutate le caratteristiche che rendono distinte le due modalità Dichiarativa (*D*) e Interrogativa (*I*) all'interno della stessa parlata, al variare della struttura della frase stessa (SN-V-SP) in funzione della posizione degli accenti lessicali e con il ricorso all'uso di espansioni aggettivali. In un secondo momento si è passati invece a valutare congiuntamente i risultati ottenuti, confrontandoli con quelli di studi precedenti e col ricorso a metodi di modellizzazione diversi.

### 3. RISULTATI PER L'ANDALUSO

I risultati delle analisi condotte con le procedure *AMPER* per la varietà di Malaga sono stati confrontati con quelli di varietà simili studiate nell'ambito di *AMPER-AND-OR* (cfr. Pamies, 2007, Pamies, Amorós & O'Neill, 2007) e *AMPER-AND-OCC* (Congosto Martín, 2005, 2007a&b; Grasso, 2007). Il confronto ha preso in considerazione in particolare gli andamenti multiparametrici osservati per le varietà andaluse di Almeria, Granada e Siviglia, e per quelle *extremeña* di Cáceres e Don Benito (Badajoz).

Nelle produzioni del parlante da noi preso in considerazione per la varietà di Malaga si conferma la tendenza riscontrata negli studi citati (in particolare Grasso, 2007:13) a una desonorizzazione delle vocali inaccentate, con profili melodici contrassegnati dalla realizzazione di rilievi accentuali ritardati rispetto alla sede accentuale (che interessano quindi le brevi vocali postaccentuali) con andamenti differenti nelle due modalità.

Come si può verificare in generale, a colpo d'occhio, nella selezione di grafici proposta nelle Figg. 1 e 2, nella *D* uno stesso profilo melodico locale ascendente (con escursioni di 5 o 6 semitoni) sembra ripetersi nella realizzazione delle configurazioni accentuali prenucleari, mentre nella *I* una maggiore gerarchizzazione (il primo picco si trova di solito 5 o 6 semitoni al di sopra dei valori raggiunti in corrispondenza della posizione in cui *D* presenta il secondo rilievo) favorisce scarti melodici più significativi (12 o 13 semitoni) nella realizzazione del primo movimento ascendente ed escursioni ridotte o nulle per i rilievi successivi a eccezione dell'ultimo, sul quale si manifesta il profilo ascendente del *CTM* (con un salto di 12 semitoni, distribuito su due sillabe soltanto con accelerazione costante), il quale è invece discendente in *D*, in un intervallo di 6-8 semitoni percorso nell'arco di quattro sillabe.

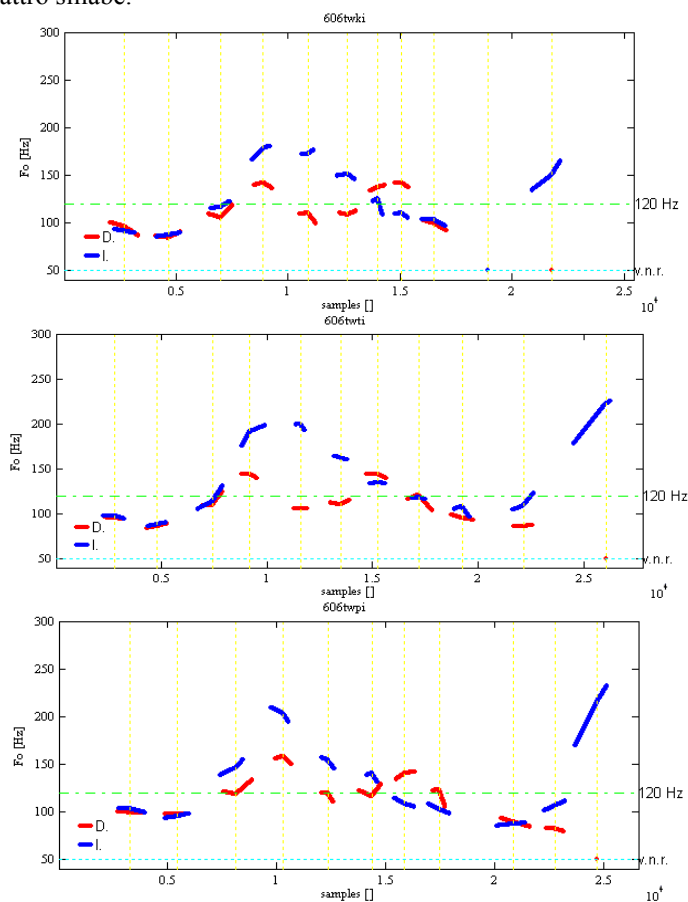


Figura 1. Profili melodici medi di enunciati dichiarativi (D, in rosso) e interrogativi (I, in blu) ottenuti a partire da frasi senza espansioni del tipo "La guitarra se toca con...": dall'alto in basso "... obsesión", "... paciencia", "... pánico".

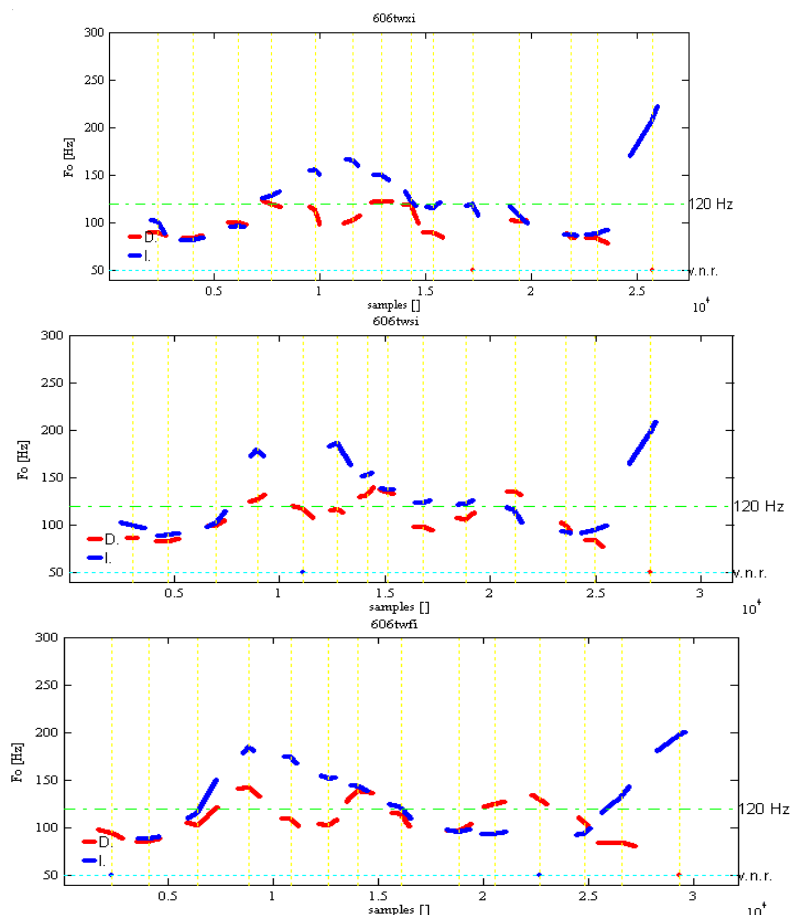
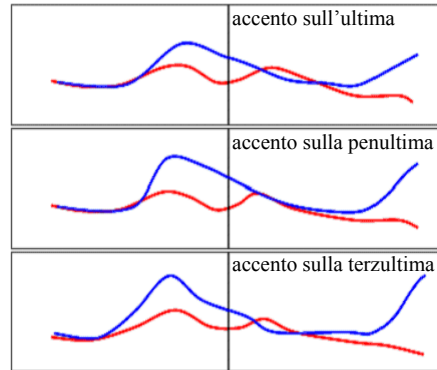


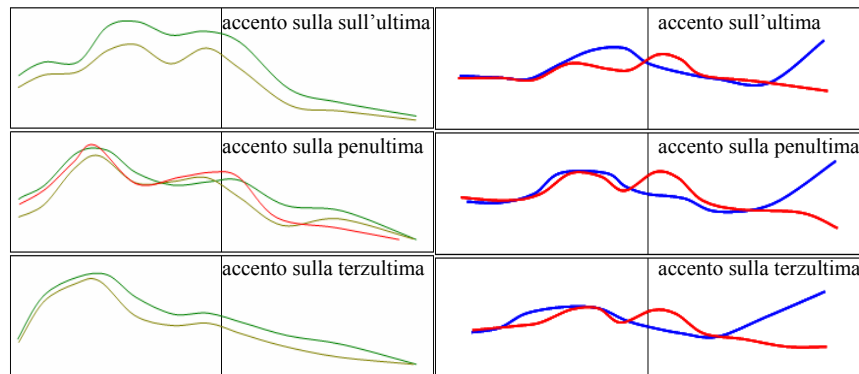
Figura 2. Profili melodici medi di enunciati dichiarativi (D, in rosso) e interrogativi (I, in blu) ottenuti a partire da frasi con espansioni del tipo "La guitarra se toca con...": dall'alto in basso "... obsesión finita", "... paciencia finita", "... pánico finito".

Il confronto con le altre varietà andaluse e *extremeña* è avvenuto prendendo in considerazione gli andamenti descritti per i dati relativi alle varietà di Almería, Granada e Siviglia (cfr. Pamies *et alii*, 2007; Congosto Martín, 2005, v. Fig. 3b), per Cáceres (v. Fig. 3b) e per Don Benito (Badajoz) (cfr. Congosto Martín, 2007a&b; Grasso, 2007; v. Fig. 3c) confermando tendenze simili<sup>2</sup>.

<sup>2</sup> In particolare, sulla scorta di una ricca bibliografia, Grasso (2007) discute alcuni interessanti aspetti del *continuum* intonativo che si è stabilito in queste regioni in virtù dell'affermarsi di modelli diversi legati alla diffusione di varietà di prestigio come quella del castigliano madrilenno (veicolata tra l'altro dai media) o quella del castigliano di Siviglia. Questi modelli sembrano influenzare anche gli andamenti presenti nei nostri dati.

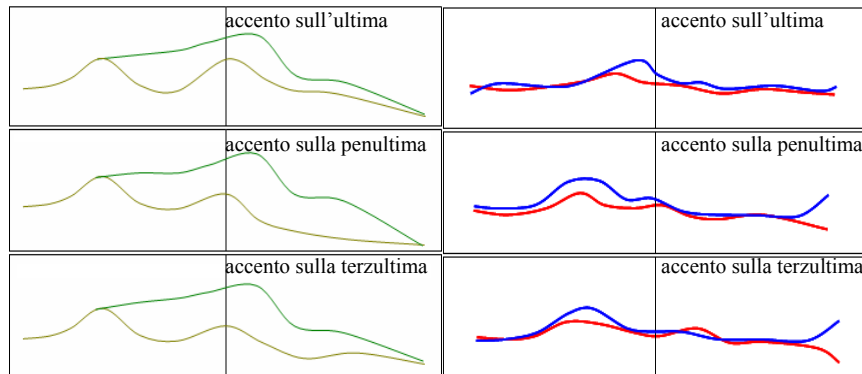


a. Malaga



b. Almería e Granada

c. Cáceres



d. Don Benito (parlanti conservativi)

e. Don Benito (parlanti innovativi)

Figura 3. Confronto tra i profili intonativi (*D* e *I*) da noi ottenuti per il locutore di Malaga (a.) con quelli presentati per i *tonemas declarativos* di Almería e Granada (b., adattati da Pamies *et alii*, 2007a&b), con quelli ottenuti col nostro stesso metodo da Grasso (2007) per Cáceres (c.) e con quelli relativi a dati raccolti per la varietà di Don Benito (d., adattati da Congosto Martín, 2007a&b; e., da Grasso, 2007).



I contorni dichiarativi da noi osservati (v. quelli schematizzati in Fig. 3a) risultano molto simili a quelli di queste varietà (praticamente identici a quelli di Cáceres di Grasso, 2007). Le nostre *I* sembrano differenziarsi, invece, soprattutto per il fatto che il profilo melodico sovrasta generalmente i due picchi della *D* (forse per una particolare enfasi presente nei nostri dati) e si conclude con un'ascesa il cui punto d'inizio sembra poco influenzato dal cambiamento di posizione dell'accento nucleare (è questo forse l'aspetto più interessante, che necessita di ulteriori verifiche, con un maggiore controllo delle scelte segmentali). La varietà di Don Benito, invece, rappresentata da locutori giovani nei dati di Grasso (2007), si manifesta un andamento ascendente finale molto meno pronunciato, con un'opposizione di modalità affidata maggiormente ai picchi prenucleari. Questa caratteristica tipica (già osservata in dati aragonesi da Contini *et alii*, 1995) appare come portatrice esclusiva dell'opposizione di modalità nel caso dei parlanti più conservativi a cui si riferisce Congosto Martín (2007a&b): la risalita finale sparisce infatti completamente (cfr. anche Alcoba & Murillo, 1998).

Sulla scorta di queste osservazioni e di queste verifiche abbiamo quindi riassunto le proprietà intonative più sistematiche presenti in queste produzioni in base a schemi prototipici come quelli riprodotti in Fig. 4.

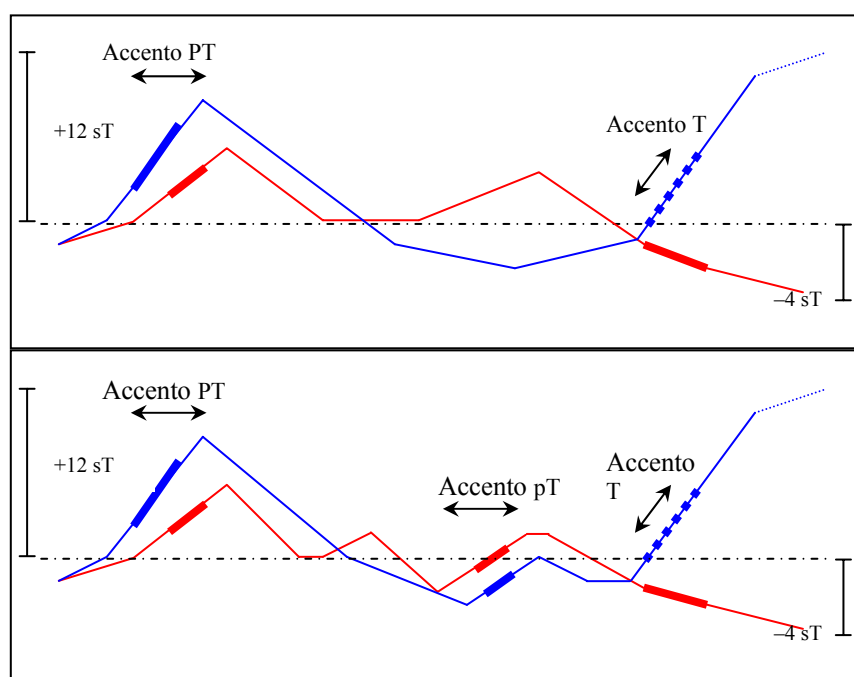


Figura 4. Schemi prototipici di variazione melodica negli enunciati dichiarativi (D, in rosso) e interrogativi (I, in blu) del locutore di Malaga (PT = accento di protonia; pT = ultimo accento di protonia (immediatamente pretonico), T = accento di tonia). Frasi senza espansioni (in alto) e con espansioni (in basso).

#### 4. RISULTATI PER IL VENETO

Per la varietà liventina di Motta di Livenza (v. Figg. 5 e 6), abbiamo invece riscontrato alcune delle caratteristiche tratteggiate per l'italiano padovano da Magno Caldognetto *et alii* (1978) e da Endo & Bertinetto (1997) e la maggior parte di quelle intuite e rappresentate schematicamente, sulla base di un'attenta descrizione diatopica dell'area, in Canepari (1985, 1999). Al contrario, nessun elemento di contaminazione o di gradiente areale sembra interessare la varietà in questione, per i fenomeni indagati, con le pur vicine varietà friulane (cfr. Miotti, 2002; D'Agostin & Romano, 2007).

Come si può verificare anche a partire dalla selezione di grafici proposta nelle Figg. 5 e 6, si sono presentati straordinariamente regolari gli schemi accentuali di *D* in protonia (con valori bassi sulla vocale accentata, tra un leggero innalzamento preaccentuale e i valori più alti sulla postaccentuale (di circa 6 semitoni), cui segue poi una discesa di 5 semitoni se la vocale seguente è ancora inaccentata; cfr. Canepari, 1999:401).

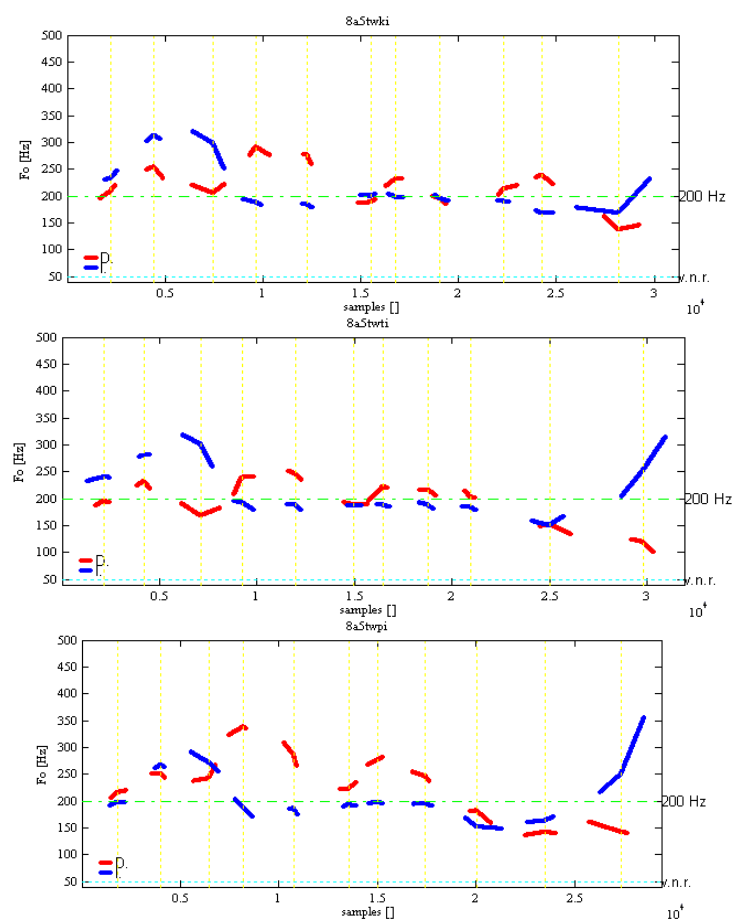


Figura 5. Profili melodici medi di enunciati dichiarativi (D, in rosso) e interrogativi (I, in blu) ottenuti a partire da frasi senza espansioni del tipo "La chitarra si suona con...": dall'alto in basso "... serietà", "... pazienza", "... panico".

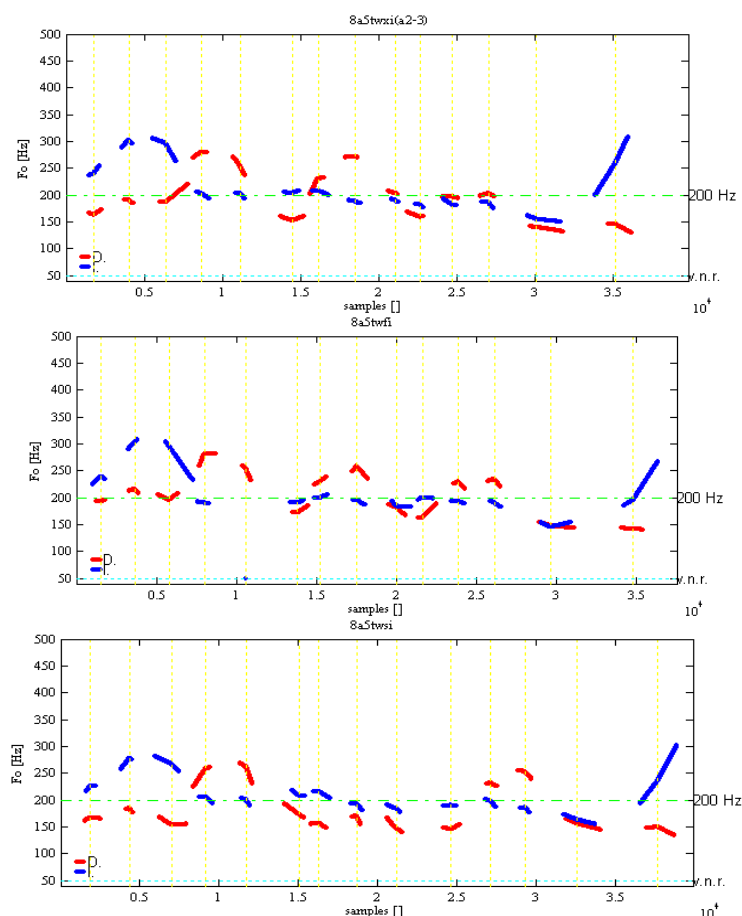


Figura 6. Profili melodici medi di enunciati dichiarativi (D, in rosso) e interrogativi (I, in blu) ottenuti a partire da frasi con espansioni del tipo "La chitarra si suona con...": dall'alto in basso "... serietà costante", "... pazienza costante", "... panico costante".

L'incostante anticipazione del primo rilievo melodico prenucleare di *I*, tra la preaccentuale e la vocale accentata, sembra in parte legato a una maggiore sospensione del tema presente in molte realizzazioni del corpus<sup>3</sup>. La compressione del *range* melodico nella resa degli schemi accentuali seguenti concentra i movimenti melodici intorno alla frequenza media del parlante ( $F_m$ ), con oscillazioni trascurabili. Lo schema della modalità *D* si chiude poi con una tonia discendente con un livello della postaccentuale che inizialmente conferma, o corregge su valori leggermente più alti (0-1 semitono), quello raggiunto della

<sup>3</sup> Fa eccezione proprio la prima realizzazione nella modalità *D* della frase *twx* (*La chitarra si suona con serietà costante*): in quel caso l'intonazione adottata dalla locutrice, più risolutiva (e tendenzialmente euritmica per via di un accento aggiuntivo sul verbo), ha portato all'abbassamento dei profili locali sui primi due accenti e a un sollevamento generalizzato finale della protonia.

vocale accentata finale (con uno schema simile a quello dato per Belluno in Canepari, 1999:402).

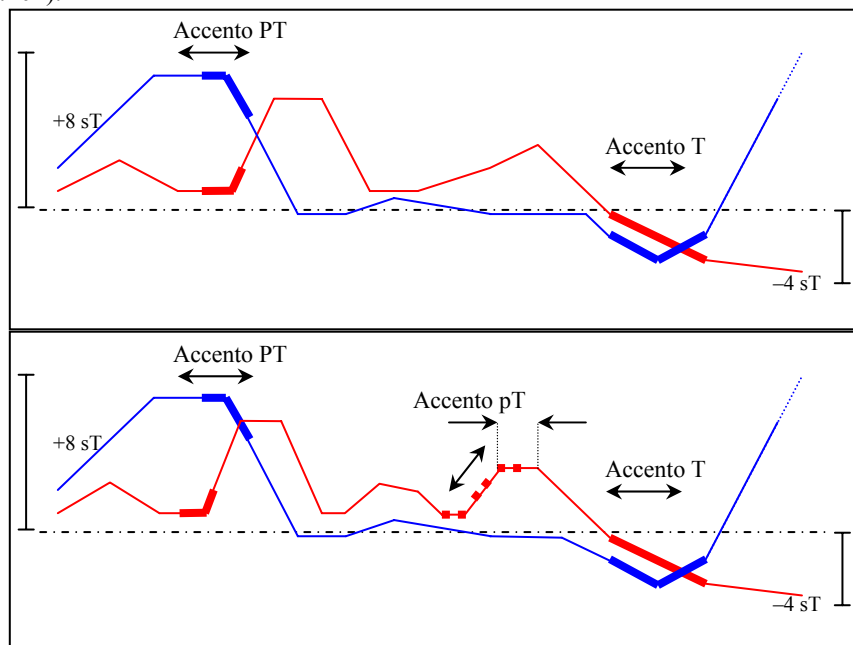


Figura 7. Schemi prototipici di variazione melodica negli enunciati dichiarativi (D, in rosso) e interrogativi (I, in blu) del locutore di Motta di Livenza (PT = accento di protonia; T = accento di tonia; pT = movimento melodico pretonico). Frasi senza espansioni (in alto) e con espansioni (in basso).

Il CTM della modalità *I* sembra invece seguire abbastanza fedelmente quello veneto generico (descritto da Canepari, 1999:402) o quello padovano di Endo & Bertinetto (1997:41): dopo una preaccentuale media (intorno alla *Fm*) si presenta una vocale nucleare più bassa di 4 o 5 semitoni da cui la curva sale rapidamente, con un'accelerazione crescente e un'escursione di 9-10 semitoni.

Anche per questa varietà, in base all'osservazione di queste caratteristiche, abbiamo provato a riassumere le proprietà intonative più sistematiche in un paio di schemi prototipici come quelli riprodotti in Fig. 7.

In particolare, nel caso di frasi con espansione, si noterà la presenza di un paio di fatti interessanti che riguardano i movimenti melodici pT e la realizzazione degli accenti divenuti interni: 1) nella frase *I* la presenza di materiale lessicale caratterizzato da accento di parola all'interno della frase, tra le due caratteristiche prominenze periferiche, non modifica in maniera rilevante lo schema melodico della frase; 2) nella frase *D* la realizzazione dell'accento divenuto pT si manifesta con una deformazione del profilo interaccentuale e la manifestazione di due rilievi melodici (smussati) piuttosto indipendenti dalla posizione dell'accento di parola (verificato accuratamente su materiali più adatti, questo fenomeno potrebbe avere interessanti ricadute teoriche).

## 5. CONFRONTO TRA LE VARIETÀ

Parallelamente all'osservazione di queste proprietà, in secondo luogo, è stato effettuato un confronto stringente tra le due parlate fin qui descritte separatamente, sulla base di alcuni metodi sperimentali.

### 5.1. Confronto tra varietà: tonogrammi

Un metodo a cui siamo ricorsi per verificare le condizioni di esecuzione del confronto è quello di valutare le caratteristiche intonative salienti presenti nella tonia prendendo in considerazione le rappresentazioni derivanti da due tipi di astrazioni diversi: uno tra quelli usati nell'ambito di *AMPER* (illustrato nei paragrafi precedenti) e uno derivante dalla stilizzazione mediante tonogrammi (v. Fig. 8), basata sul ricorso al metodo fonetico naturale di Canepari (metodo uditivo rigoroso, che stabilisce tre tonalità fondamentali: alta, media e bassa; cfr. Canepari 2006<sup>2</sup>). La definizione dei due prototipi, ottenuta separatamente dai due autori, ha permesso di giungere a risultati simili (cfr. Fig. 4 e 7).

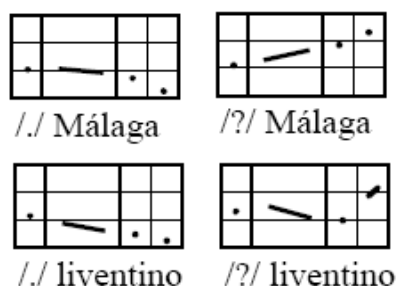


Figura 8. Tonogrammi delle tonie *D* (/./) e *I* (/?/) per la varietà andalusa di Malaga e quella liventina di Motta di Livenza.

### 5.2. Confronto tra varietà: distanze prosodiche

Dalla sovrapposizione delle curve normalizzate (e dalla distinta descrizione riportata succintamente più in alto) è emerso un diverso sfruttamento della dinamica melodica da parte dei due parlanti esaminati, con movimenti più accentuati per la modalità *D* della varietà liventina, ma con un'escursione finale più marcata ( $12 \pm 1$  semitoni in  $250 \pm 20$  ms) per la varietà andalusa rispetto a quella veneta ( $10 \pm 1$  semitoni in  $320 \pm 20$  ms). Come prevedibile, le curve intonative osservate sono infine nettamente differenziate per l'allineamento di questi movimenti con la posizione degli accenti lessicali e per la configurazione che assumono localmente: mentre nella varietà iberica il picco melodico è solitamente situato sulla vocale immediatamente postaccentuale (almeno nella modalità *D*), nella varietà liventina può essere spesso ritardato anche sulla seconda postaccentuale. Inoltre, nella modalità *I*, si nota una netta differenza nell'allineamento del minimo locale (il *turning point* di Bertinetto & Magno Caldognetto, 1993) del *CTM* che nelle produzioni venete è allineato con l'ultima vocale accentata da cui inizia, molto lentamente, l'ascesa finale (che assume i caratteri di una brusca risalita solo sulla vocale inaccentata finale).

mentre è piuttosto insensibile alla posizione dell'accento per Malaga (con profilo spesso totalmente ascendente sulla vocale accentata)<sup>4</sup>.

Un altro metodo che abbiamo adottato per valutare più obiettivamente questi dati è stato quello di sottoporli a una misura di distanza prosodica come quella applicata ai dati di Romano (1999\*2001), simile a quella usata da Hermes (1998)<sup>5</sup>. Applicandola ai soli valori di  $f_0$ , abbiamo verificato preliminarmente che una misura d'*intercorrelazione* produce valori che oscillano tra il 90 e il 94% quando si tratta di ripetizioni della stessa frase, nella stessa varietà e da parte dello stesso parlante<sup>6</sup>. Nel nostro caso abbiamo poi verificato che l'*intercorrelazione* tra i nostri dati di liventino e di *malagueño* oscilla tra l'84 e l'86,5% per le frasi a modalità *D* (quindi con una distanza sensibile ma minima), invece nel caso delle *I* i valori ottenuti sono persino negativi (tra -18 e -17%): per quanto simili possano sembrare, le due curve sono assolutamente diverse.

## 6. ALTRE OSSERVAZIONI E CONCLUSIONI

Con questo studio preliminare, basato sulle produzioni di due soli locutori, abbiamo osservato alcune caratteristiche intonative caratterizzanti di due varietà romanze, una dello spazio linguistico italoromanzo (veneto) e una di quello iberoromanzo (andaluso).

Le proprietà principali delle modalità osservate, riassunte schematicamente e contrastivamente, hanno permesso di osservare una grossolana somiglianza che non trova conferma in importanti dettagli nella strutturazione intonativa anche di semplici frasi.

In particolare, come abbiamo potuto verificare anche per mezzo di test d'ascolto informali sulla prosodia sintetica (in assenza d'informazioni segmentali e lessicali), le due varietà differiscono infine in maniera considerevole soprattutto per la loro organizzazione temporale. Tralasciando l'ovvia differenziazione che riguarda le proprietà temporali legate al consonantismo, osserviamo invece il dato interessante che si presenta per il vocalismo: mentre nel liventino la vocale accentata conserva sempre proprietà locali di maggior lunghezza (tranne che in finale di domanda dove può essere eguagliata dalla finale), nella varietà andalusa studiata sono prevalentemente solo le durate delle vocali accentate finali che presentano allungamenti più significativi rispetto alle vocali inaccentate.

In base a queste osservazioni e a quelle derivanti da valutazioni e verifiche più oggettive (come quella basata su una misura di distanza prosodica), ci sentiamo di confermare solo in minima parte l'intuizione generale che accosta la prosodia delle parlate venete a quella di varietà iberiche: soltanto la notevole ascesa finale della modalità *I* (che però – come visto – è più concentrata in liventino) è elemento condiviso. Quest'osservazione conferma però

---

<sup>4</sup> Anche questo interessante fenomeno, verificato su materiali più adatti, potrebbe sollevare importanti riflessioni teoriche visto che suggerirebbe, almeno per queste varietà, la potenziale autonomia di due livelli di strutturazione, uno accentuale (locale, in questo caso svincolato dal piano di realizzazione melodico) e uno intonativo (più propriamente melodico).

<sup>5</sup> Questa idea, già applicata in modo sperimentale negli studi menzionati, è stata recentemente riproposta da Albert Rilliard come possibile metodo di confronto tra i dati della *BD-AMPER*.

<sup>6</sup> La definizione d'*intercorrelazione* si basa su una *misura della deviazione globale tra le curve di  $f_0$  di diverse realizzazioni di enunciati comparabili* (Romano 1999\*2001:226-235). La formula usata è la seguente:

$\rho_{xy} = (1/N) * (\sum_{j=1}^N (x_j - \mu_x)(y_j - \mu_y)) / (\sigma_x * \sigma_y)$  (con  $-100\% \leq \rho_{xy} \leq 100\%$ ).

anche che in termini di sensibilità discriminatoria sulla distanza prosodica tra le varietà, il CTM è sede privilegiata.

## BIBLIOGRAFIA

- AMPER, v. Géolinguistique (2005). Projet AMPER - Atlas Multimédia Prosodique de l'Espace Roman, *Géolinguistique*, hors série 3 (éd. J.P. Lai).
- ALCOBA S. & MURILLO J. (1998). "Intonation in Spanish". In D. Hirst & A. Di Cristo (eds.), *Intonation Systems: a Survey of Twenty Languages*, Cambridge: Cambridge Univ. Press, 152-166.
- BERTINETTO P.M. & MAGNO CALDOGNETTO E. (1993). "Ritmo e intonazione". In A.A. Sobrero (a cura di), *Introduzione all'italiano contemporaneo. Le strutture*, 2, Roma-Bari: Laterza, 141-192.
- CANEPARI L. (1985). *L'intonazione. Linguistica e paralinguistica*. Napoli: Liguori.
- CANEPARI L. (1999). *Manuale di Pronuncia Italiana*. Bologna: Zanichelli.
- CANEPARI L. (2006<sup>[2]</sup>). *A Handbook of Phonetics*. München: Lincom.
- CONGOSTO MARTÍN Y. (2005). "Una primera aproximación a la prosodia del habla urbana de Sevilla". *Estudios de Fonética Experimental*, XIV, 225-246.
- CONGOSTO MARTÍN Y. (2007a). "Primeras aportaciones a la descripción prosódica del extremeño: interrogativas absolutas sin expansión en el habla de Don Benito". *Actas del III Congreso Int. de Fonética Experimental* (Santiago de Compostela, Oct. 2005), Santiago: Xunta de Galicia, 205-221.
- CONGOSTO MARTÍN Y. (2007b). "Interrogativas absolutas / declarativas en las hablas extremeñas. Estudio entonativo contrastivo". In J. Dorta (ed.), *La prosodia en el ámbito lingüístico románico (Acti delle "III jornadas científicas del proyecto AMPER"*, La Laguna - Tenerife, Isole Canarie, 24-25 Ott. 2006), Santa Cruz de Tenerife: La página, 271-297.
- DEVÍS HERRAIZ E. (2007). "La prosodia nell'interferenza tra L1 e L2: il caso delle interrogative polari tra veneti e catalani". *Estudios de Fonética Experimental*, XVI, 119-146.
- D'AGOSTIN R. & ROMANO A. (2007). "Analisi dell'intonazione delle varietà friulana e italiana parlate a Codroipo (Udine)". In J. Dorta (ed.), *La prosodia en el ámbito lingüístico románico (Acti delle "III jornadas científicas del proyecto AMPER"*, La Laguna - Tenerife, Isole Canarie, 24-25 Ott. 2006), Santa Cruz de Tenerife: La página, 101-121.
- ENDO R. & BERTINETTO P.M. (1997). "Aspetti dell'intonazione in alcune varietà dell'italiano". In F. Cutugno (ed.), *Fonetica e fonologia degli stili dell'italiano parlato. Atti delle VII Giornate di Studio del GFS* (Napoli, 1996), Roma, Esagrafica, 1997, 27-49 (v. anche *Quad. del Lab. di Ling.*, Scuola Norm. Sup., Pisa, 10/1996).
- FERNÁNDEZ PLANAS A.M. (2005). "Aspectos generales acerca del proyecto internacionales "AMPER" en España", *Estudios de Fonética Experimental*, XIV, 13-27.
- FRANCHON-CABRERA C., CONTINI M. & RHARDISSE A. (1998). "Analyse comparée de l'intonation en castillan et en aragonais". In G. Ruffino (ed.), *Actes du*

- XXI Congrès de Linguistique et Philologie Romane* (Palermo, set. 1995), Palermo: CSLFS, 1998, 9 pp.
- GRASSO A. (2007). "Analisi segmentale e soprasegmentale delle parlate dell'Estremadura (Spagna)". *Tesi di Laurea* (rel. A. Romano), Fac. di Lingue e Lett. Str. dell'Univ. degli Studi di Torino, inédita.
- MAGNO CALDOGNETTO E., FERRERO F., LAVAGNOLI C., VAGGES K. (1978). "F0 contours of statements, yes-no questions and wh-questions of two regional varieties of Italian". *Journal of Italian Linguistics*, 3, 57-68.
- MARTÍNEZ CELDRÁN E. & FERNÁNDEZ PLANAS A.M. (2006). "Hacia una geoprosodia de las lenguas íbero-romances en la "Web""", *Letras de Hoje*, 41/2, 9-22.
- MIOTTI R (2002). "Friulian", *Journal of the International Phonetic Association*, 32, 237-247.
- PAMIES A. (2007). "Observaciones sobre la estructura melódica en enunciados declarativos". *Actas del III Congreso Int. de Fonética Experimental* (Santiago de Compostela, Oct. 2005), Santiago: Xunta de Galicia, 475-488.
- PAMIES A., AMORÓS M. & O'NEILL P. (2007). "Esquemas entonativos en el habla de Almería". In J. Dorta (ed.), *La prosodia en el ámbito lingüístico románico*, Santa Cruz de Tenerife: La página, 299-311.
- ROMANO A. (1999\*2001). "Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento: approche linguistique et instrumentale", *Thèse de Doctorat de l'Université Stendhal de Grenoble* (resp. Michel Contini) (part. publiée en 2001, Lille: Presses Univ. du Septentrion).
- ROMANO A. (2000). "Variabilità degli schemi intonativi dialettali e persistenza di tratti prosodici nell'italiano regionale: considerazioni sulle varietà salentine", In A. Zamboni et al. (éds.), *La dialettologia oggi fra tradizione e nuove tecnologie (Atti del Conv. Int., Pisa, févr. 2000)*, Pisa: ETS, 2001, 73-91.
- ROMANO A. (2005). "Utilisation des données *AMPER* pour une description de la variation linguistique : tests de perception et contrôles statistiques", *Géolinguistique*, 3 h.s., 39-64.
- SOSA J.M. (1999). *La entonación del español. Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra.

Siti web del Progetto *AMPER* :

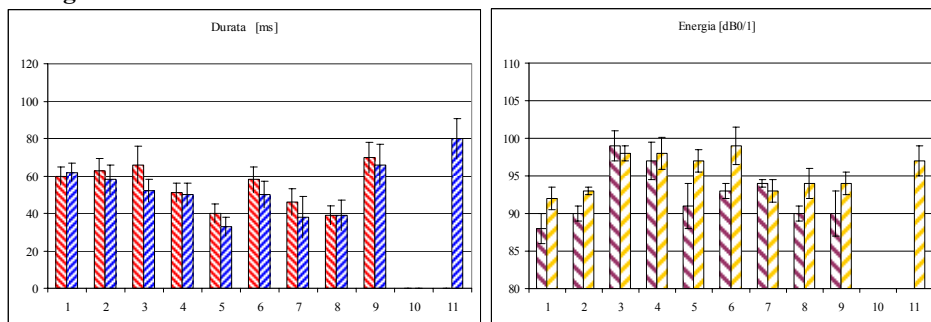
<http://w3.u-grenoble3.fr/dialecto/AMPER/amper.htm>  
<http://w3.u-grenoble3.fr/dialecto/AMPER/BDAMPER/index.php>  
<http://www.personalweb.unito.it/antonio.romano/AMPER.html>  
<http://www2.i.ua.pt/cidlc/gcl/AMPER-POR.htm>  
<http://www.ub.es/labfon/amper/index.html>



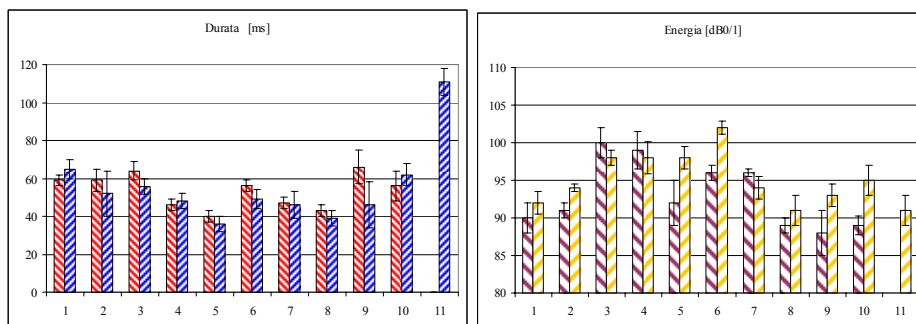
## ARTICOLO I. APPENDICE

Quest'appendice illustra il tipo di grafici su cui è stata condotta l'analisi dei valori degli altri parametri prosodici. In particolare gli istogrammi seguenti esemplificano la rappresentazione grafica usata per l'analisi delle sequenze di valori di durata (a sinistra) ed energia (a destra) per le vocali degli enunciati. Le barre con tratteggio discendente (da sinistra a destra) sono relative alla modalità *D*, quelle col tratteggio ascendente alla modalità *I*.

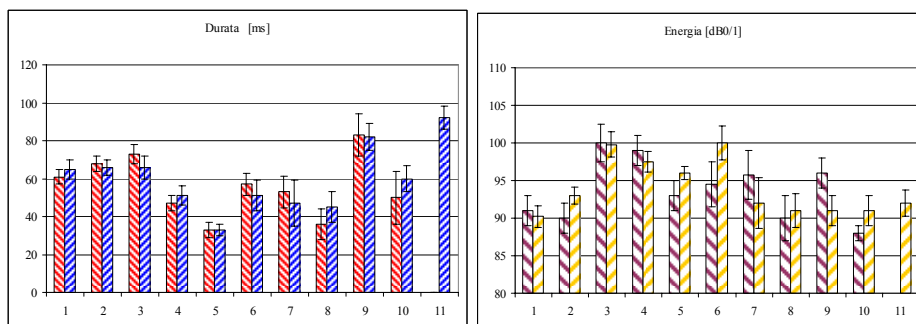
### Malaga



twk

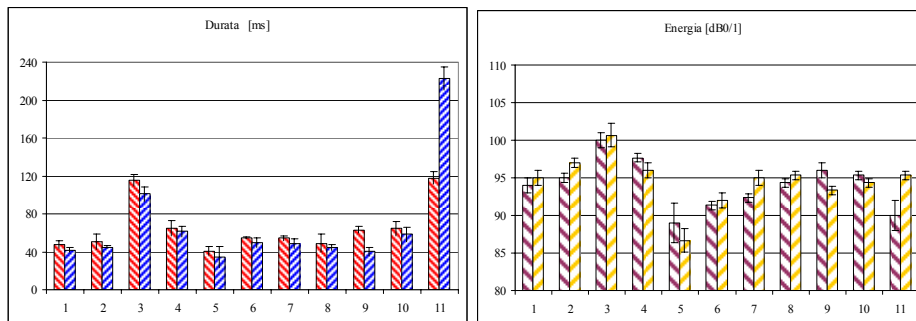


tw

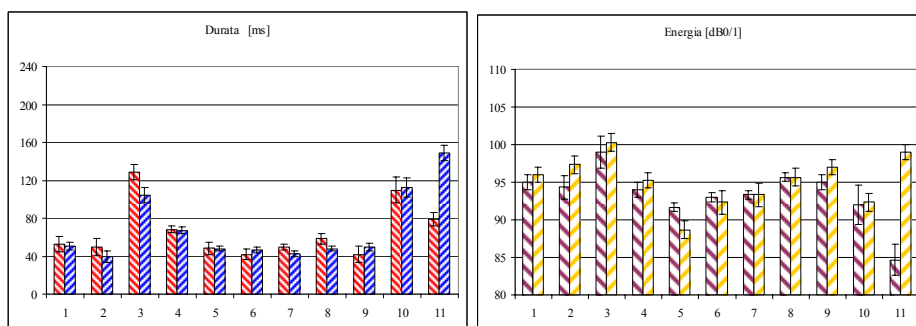


twp

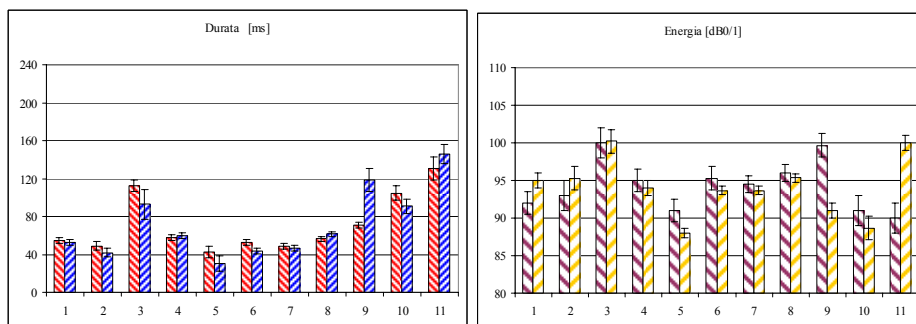
## Motta di Livenza



twk



twt



twp

Un breve commento a questi grafici può limitarsi a segnalare la consistenza dei fenomeni di desonorizzazione (e conseguente esclusione da queste rappresentazioni) delle vocali negli enunciati *D* del locutore andaluso. Come si vede dai primi due istogrammi, questa può interessare talvolta anche vocali per le quali ci si aspettava addirittura un accento di frase. Mentre la prima vocale accentata riceve in tutti i casi un accento di forza, non si può dire che la sua durata relativa sia gestita allo stesso modo nelle due varietà: solo la varietà veneta realizza un significativo allungamento su di essa. Trascurando le influenze dei valori di  $f_0$  tipici delle diverse qualità vocaliche presenti, le durate degli ultimi segmenti sono soggette a un progressivo allungamento 'intonativo' piuttosto indipendente dalla posizione dell'accento (soprattutto alla modalità *I* e nei dati della varietà iberica).

# TEMPI DI ADATTAMENTO IN ELETTROPALATOLOGRAFIA: PRIMI DATI ARTICOLATORI, ACUSTICI E PERCETTIVI

Chiara Bertini, Irene Ricci  
Scuola Normale Superiore  
{c.bertini; i.ricci}@sns.it

## 1. SOMMARIO

L'obiettivo di questo lavoro è quello di descrivere l'impatto che provoca sul parlato la presenza di un palato artificiale. Saranno considerate tre prospettive diverse: articolatoria, acustica e percettiva. Ciò che si vuole individuare è la presenza di un tempo minimo di adattamento al palato, necessario a produrre 'parlato normale' e di conseguenza realizzare un protocollo per la preparazione dei parlanti alle sessioni di registrazioni elettropalatografiche.

Nelle varie ricerche che utilizzano l'elettropalatografo (EPG) come strumento di indagine, il tempo indicato come "adattamento" al palato artificiale varia sensibilmente da ricercatore a ricercatore; si possono leggere indicazioni ad indossare il palato per non più di 20 minuti, fino a vere prescrizioni di preparazioni lunghe anche 3 giorni prima di ogni registrazione.

La nostra analisi si incentra su un unico parlante, che ha effettuato 10 ripetizioni di non-parole del tipo VCV, dove V è sempre la vocale /a/ e C varia fra l'occlusiva alveolare /t/ e la fricativa alveolare /s/. Sono state fatte sessioni di registrazione a 8 intervalli temporali e, nel periodo che precede ogni sessione di raccolta dati, il parlante viene fatto dialogare in modo da consentire un maggiore adattamento al palato stesso.

Da questo studio sembra emergere che l'adattamento al palato è un fenomeno altamente correlato alla specifica consonante e/o alla specifica prospettiva di analisi, ovvero non è possibile determinare un unico tempo di adattamento indipendentemente dalla consonante investigata ed indipendentemente dal punto di vista da cui parte l'analisi.

Da questi primi risultati appare una sostanziale differenza fra le dinamiche relative all'analisi articolatoria e quelle relative all'analisi acustica. In particolare dal punto di vista articolatorio già a tempi brevi dall'inserimento del palato si ha un buon livello di adattamento, mentre per la prospettiva acustica si hanno grandi difficoltà a determinare un protocollo attendibile anche per tempi considerevolmente lunghi.

## 2. INTRODUZIONE

L'elettropalatografia è una tecnica articolatoria di indagine largamente utilizzata per studiare i fenomeni legati alla produzione di parlato; le prime applicazioni di questa tecnica sono in ambito clinico per lo studio delle patologie sul linguaggio. Successivamente l'elettropalatografia viene utilizzata per studi sulle lingue, legati a progetti di ricerca scientifica che mirano ad investigare le problematiche dello 'speech production': coarticolazione, differenze nella produzione fra le lingue,...

Ciò che è stato meno indagato in questi anni è il fenomeno di adattamento al palato artificiale che inevitabilmente si crea nei parlanti: la presenza del palato perturba la produzione spontanea di parlato e ciò crea un processo di adattamento che dovrebbe nel tempo portare il parlante a non percepire più la presenza del palato e a produrre uno *speech* imperturbato.

Nelle varie ricerche che utilizzano l'elettropalatografo (EPG) come strumento di indagine, il tempo indicato come "adattamento" al palato artificiale varia sensibilmente da ricercatore a ricercatore: si possono leggere indicazioni ad indossare il palato per non più di 20 minuti, fino a vere e proprie prescrizioni di preparazioni lunghe anche 3 giorni prima di ogni registrazione.

Fra i lavori che si trovano in letteratura e che si sono occupati di questo fenomeno si distinguono il lavoro di McLeod & Searl (2006) e quello di McAuliffe et al (2007).

Il primo si propone di analizzare il fenomeno di adattamento al palato su sette parlanti Australiani, analizzando gli aspetti acustici, impressionistici e percettivi e registrando ad intervalli di tempo regolari fino ad un massimo di 5 ore, sviluppate su due giorni di registrazione. Gli autori, come nel nostro caso, si sono concentrati sulle due consonanti /t/ e /s/; i risultati registrati sono in parte gli stessi trovati nel nostro caso, con la differenza sostanziale, che impedisce un confronto completo, legata al numero di soggetti registrati.

Il secondo lavoro concentra lo studio su otto parlanti Australiani e analizza il fenomeno dalle due prospettive acustico e percettivo. Le consonanti analizzate sono /t/, /k/, /s/ e /f/ e le analisi effettuate evidenziano che non si hanno variazioni sulla durata temporale della consonante, né sui valori delle formanti delle vocali adiacenti, mentre si notano differenze significative nei risultati dell'analisi spettrale.

Il nostro contributo si inserisce in questo quadro introducendo alcune novità rispetto ai lavori già citati, senza considerare la lingua che rappresenta forse la novità più importante dal momento che nessuna indagine è stata fatta sull'italiano in questo campo.

Ciò che ci proponiamo di indagare maggiormente è l'aspetto articolatorio del fenomeno in esame, ovvero come cambino gli indici articolatori che descrivono l'articolazione della consonante e, nel caso in cui si raggiunga adattamento, quando ciò avviene. In questo modo si cerca di ottenere un quadro più completo del fenomeno, arricchendo i dati provenienti dalle analisi acustica e percettiva.

Presupponendo che il pattern articolatorio della consonante registrato subito dopo aver inserito il palato sia chiaramente distorto rispetto a quello prototipico, si analizzano sia qualitativamente che quantitativamente i miglioramenti dovuti proprio all'adattamento al palato, aspettandoci di raggiungere un livello di bontà che comunque non si può migliorare oltre.

Questo lavoro rappresenta una prima indagine informativa sul fenomeno in questione e come tale ha ancora molti aspetti da approfondire. Sarà sicuramente necessario, per completare il protocollo per la preparazione delle sessioni di registrazione

elettropalatografiche, analizzare il numero minimo di ripetizioni utili ad avere informazioni soddisfacenti sul fenomeno linguistico che si vuole studiare.

Sarà interessante estendere l'indagine ad ulteriori suoni consonantici e a parole più complesse in contesto di frase con addestramenti specifici e sarà necessario, per completare il quadro, analizzare come variano i tempi di adattamento da parlante a parlante.

### 3. METODO

#### 3.1 Partecipanti, materiali e raccolta dati

I dati presentati in questo lavoro si riferiscono all'analisi condotta sulle registrazioni di un unico parlante italiano adulto della variante pisana di sesso femminile. Il locutore ha una dentatura normale, non presenta protesi dentarie, né assenza di dentatura, non ha mai indossato apparecchiature ortodontiche per uso medico e per ricerca e non presenta disturbi neurologici, di eloquio e di ascolto.

Il parlante ha indossato un palato artificiale a 62 elettrodi (modello *Reading EPG*) per la registrazione di 10 ripetizioni di due parole del tipo VCV, dove la vocale è sempre la /a/ e la consonante può essere o la fricativa /s/, o la occlusiva /t/. Ogni serie di registrazione è stata effettuata a tempi diversi a partire dal momento in cui è stato indossato il palato; in particolare sono state scelte 8 condizioni sperimentali corrispondenti a diversi tempi di adattamento (TA), come riportato in Tabella 1.

|    |                                                     |
|----|-----------------------------------------------------|
| T0 | prima dell'inserimento del palato                   |
| T1 | 5 minuti dopo l'inserimento del palato artificiale  |
| T2 | 15 minuti dopo l'inserimento del palato artificiale |
| T3 | 30 minuti dopo l'inserimento del palato artificiale |
| T4 | 45 minuti dopo l'inserimento del palato artificiale |
| T5 | 60 minuti dopo l'inserimento del palato artificiale |
| T6 | 2 ore dopo l'inserimento del palato artificiale     |
| T7 | 3 ore dopo l'inserimento del palato artificiale     |
| T8 | 4 ore dopo l'inserimento del palato artificiale     |

Tabella 1 : Riepilogo dei tempi di adattamento.

Il parlante, durante il tempo di attesa prima di ogni registrazione, è stato sollecitato a parlare in modo da facilitare il processo di adattamento al palato e velocizzare i tempi di riduzione dell'eccessiva salivazione, fisiologicamente inevitabile. È stata fatta anche una registrazione (T0) prima dell'inserimento del palato in modo da aver un riferimento per il confronto nelle analisi.

È stata nostra cura effettuare le 10 ripetizioni nel minor tempo possibile (in media circa 3 minuti), in modo che non ci fossero variazioni di adattamento apprezzabili all'interno di un medesimo TA.

I dati articolatori sono stati registrati con l'elettropalatografo WinEPG della ditta Articulate Instruments con frequenza di campionamento pari a 10 kHz<sup>1</sup>. I dati acustici, sincronizzati con quelli articolatori, sono stati registrati con un microfono unidirezionale

<sup>1</sup> 16 bit, 2 canali, campionamento palati 100 al secondo, input microfono 1 mV su 600Ω, 20-4.5 kHz opto-isolated,  $\pm$  1dB

Shure (del tipo 'head mounted') e digitalizzati con una scheda audio SoundMax a 44 KHz con una precisione di 16 bit.

### 3.2 Analisi articolatoria

Uno degli scopi della nostra ricerca consisteva nell'individuare un tempo di adattamento relativo alla prospettiva articolatoria; a tal fine è stata condotta un'analisi articolatoria dei dati, tramite il software Articulate Assistant (versione 1.12).

Sono stati studiati gli andamenti di vari indici articolatori<sup>2</sup> in funzione dei TA, relativamente all'intera regione palatale e a zone più localizzate, in modo da poter individuare quali parametri evidenziassero meglio variazioni significative e tra questi quali erano i più rappresentativi. Limitatamente ai nostri dati, l'analisi ha portato all'individuazione di tre indici rilevanti: il centro di gravità (COG), il contatto alveolare totale (AlvTot) e la 'closure' (Clos). Essi sono stati calcolati nella zona tipica di tenuta per entrambe le consonanti e per l'occlusiva alveolare anche nella transizione tra la vocale iniziale e la consonante stessa. Il COG, essendo calcolato come una media pesata dei contatti attivati in ogni fila del palato, dà un'indicazione di quanto il contatto lingua-palato è anteriore o posteriore. L'AlvTot, essendo invece una media pesata sull'area dei contatti rappresenta la percentuale dei sensori contattati limitatamente alle prime 2 file di elettrodi (14) del palato artificiale. Infine il calcolo della Clos è il risultato di un algoritmo complesso che misura la connettività tra il lato destro e sinistro del palato, prendendo in considerazione non solo la quantità di contatti, ma anche la loro simultanea attivazione.

Per ognuna di questi tre indici articolatori e per ognuna ripetizione sono stati calcolati i valori medi ad ogni tempo di adattamento.

### 3.3 Analisi acustica

Mediante l'utilizzo del software PRAAT (versione 4.4.24), per ogni ripetizione sono state calcolate la durata totale della parola, la durata della sola parte consonantica e la distribuzione spettrale della consonante stessa.

Per quanto riguarda le durate temporali, a partire dallo spettrogramma a banda larga di ogni registrazione è stata misurata la durata complessiva della bisillaba VCV; la durata della fricativa /s/ è stata determinata prendendo come estremi l'inizio e la fine della zona di fricazione; mentre quella della /t/ è stata calcolata partendo dall'inizio del burst fino all'onset della periodicità della vocale seguente.

Per entrambe le consonanti sono stati calcolati i tre momenti della distribuzione spettrale<sup>3</sup>, SM1 (frequenza del centroide, media), SM3 (skewness) e SM4 (kurtosis); per la fricativa /s/ il calcolo è stato effettuato nel punto centrale della fricazione e per la occlusiva /t/ sull'onset del burst, utilizzando la FFT con una finestra di Hamming di 20 ms.

Tali indici sono stati scelti in quanto è stato dimostrato che variazioni nei momenti spettrali sono legati alla percezione di distorsioni nella produzione di parlato; in particolare SM1 subisce variazioni in presenza di perturbazioni, rappresentando, così, un indice di

---

<sup>2</sup> Il centro di gravità, l'indice di lateralità, l'indice di variabilità, l'indice di contatto alveolare, la 'closure', la 'closure' alveolare, il contatto totale.

<sup>3</sup> Dall'analisi spettrale è stato escluso il secondo momento (varianza SM2) in quanto è stato dimostrato non subire variazioni in questo contesto (Forrest K. et al., 1988).

qualità nella produzione della consonante (Mc Farland D. & S. Baum, 1995, Shadle & Mair, 1997).

Si è deciso di non analizzare i parametri temporali delle consonanti (durata della consonante e della parola), in quanto abbiamo trovato che il fenomeno in questione non ha ripercussioni sul versante temporale come già osservato da Mc Auliffe *et al.* (2007).

### 3.4 Analisi percettiva

È stato realizzato un esperimento di percezione tramite il software *Presentation* su 15 adulti italiani normoudenti (età media 30 anni) non aventi familiarità con studi di fonetica. Il parlante con cui è stato realizzato il corpus non ha partecipato all'esperimento di percezione.

Tra tutti gli stimoli prodotti con il palato artificiale, sono state scelte in modo casuale 3 registrazioni per ogni tempo di adattamento T1, T2, T3, T4, T5 e T7 e una per le registrazioni senza palato, sia per la parola /asa/, sia per /ata/. È stato scelto di non utilizzare tutti gli stimoli registrati sia come numero di ripetizioni, sia come numero di tempi di adattamento in quanto ciò avrebbe reso l'esperimento eccessivamente lungo, con ovvie conseguenze sull'attendibilità dei risultati.

Ai soggetti, collocati in ambiente silenzioso, è stato chiesto di ascoltare in cuffia (Creative HQ-1300) una coppia di suoni separati da un intervallo di silenzio di 1,5 secondi e di esprimere un giudizio di similarità tra i due. Il primo suono si riferiva sempre ad una delle due parole registrate senza palato e il secondo all'analoga parola prodotta con il palato artificiale in uno delle diverse condizioni sperimentali (TA), per un totale di 42 confronti. Gli ascoltatori erano stati informati chiaramente che il primo suono era quello di riferimento rispetto al quale dovevano confrontare il secondo; il giudizio di similarità veniva espresso mediante un voto da 0 (suoni completamente diversi) a 5 (suoni perfettamente uguali). Al termine di ogni coppia di ascolto il soggetto aveva un tempo massimo di 3 minuti per dare una risposta; una volta presa la decisione, dopo 3 secondi di silenzio veniva presentata la coppia di stimoli successiva. È stato scelto di non dare la possibilità del riascolto per rendere più omogenei e paragonabili i giudizi espressi dai partecipanti. Gli stimoli sono stati presentati in ordine casuale per ogni ascoltatore in un'unica sessione della durata media di circa 3 minuti.

I soggetti hanno familiarizzato con la procedura descritta eseguendo preliminarmente un breve e analogo test con 5 coppie di suoni, scelti fra quelli non utilizzati per l'esperimento.

In tutte e tre le prospettive (articolatoria, acustica e percettiva) è stata compiuta, su tutti i parametri e/o indici di ogni registrazione, un'analisi della varianza a misure ripetute ( $F(7,72)$ ,  $p < 0.05$ ) e, successivamente, è stato effettuato il test post-hoc di Bonferroni per valutare i singoli confronti tra i tempi di adattamento al palato.

## 4. RISULTATI

### 4.1 Risultati articolatori

Analizzando qualitativamente l'andamento dei due indici articolatori AlvTot e Clos. medi per la consonante /s/ si nota una netta distinzione tra il tempo di adattamento T1 e tutti gli altri tempi (Figura 1). L'analisi della varianza e il test post-hoc avvalorano la tendenza appena descritta, in quanto i vari TA non sono significativamente distinguibili fra di loro ad eccezione del tempo T1. Questo indica che, limitatamente ai TA analizzati, già a partire da 15 minuti si ha un buon adattamento al palato.

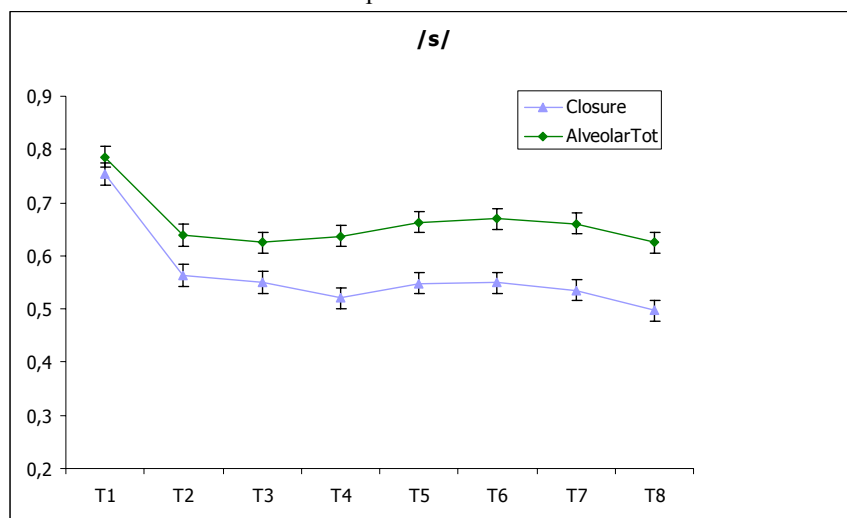


Figura 1: Andamento degli indici AlvTot e Clos. medi in funzione dei tempi di adattamento per la /s/.

Osservando in Figura2 i patterns articolatori registrati a due tempi di adattamento ben lontani (T1 e T8) è evidente la differenza che c'è nell'articolazione della stessa consonante /s/. Si può notare infatti che il numero di contatti che vengono prodotti nelle due condizioni sperimentali estreme è molto diverso e che inoltre a T1 i contatti lingua-palato riguardano tutta la zona alveolare eliminando quasi completamente il canale di flusso dell'aria caratteristico della fricativa, cosa che invece risulta evidente nel caso del pattern registrato a T8.

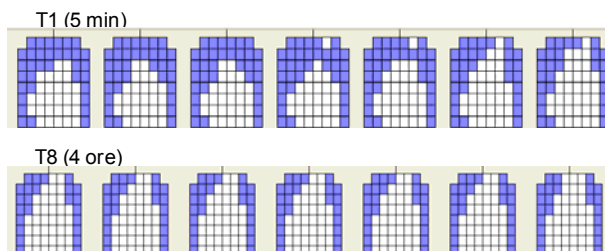


Figura 2 : Confronto fra due patterns articolatori prodotti dal parlante a 5 min (T1) e a 4 ore (T8) per la fricativa in /asa/.



Per entrambi gli indici, anche se nel margine dell'errore della misura, si nota un leggero aumento del loro valore dopo 2-3 ore (T6 e T7) dall'inserimento del palato, cosa che evidenzerebbe un peggioramento nell'andamento. Questo fatto può essere spiegato come un fenomeno di stanchezza e di intolleranza ai ganci metallici che permettono di fissare il palato artificiale alla dentatura, come fatto notare più volte dal parlante stesso.

Il COG medio misurato nella zona di fricazione non ha mostrato variazioni rilevanti, per questo non viene inserito nell'analisi.

Dallo studio degli indici AlvTot e Clos. medi, (Figura 3) relativi alla /t/, nella regione di transizione VC, emerge di nuovo la sostanziale differenza tra i valori di T1 e i restanti TA, ad esclusione di T7 che presenta un andamento singolare. Ciò, come descritto in precedenza, può essere spiegato con un fenomeno di insofferenza al palato artificiale. L'andamento del COG medio per i vari TA nella zona di tenuta della consonante, riportato in Figura 4, indica l'aumento della distribuzione dei contatti anteriori del gesto articolatorio al passare del tempo di adattamento.

L'analisi della varianza e il test post-hoc confermano la significativa differenza tra il primo tempo di adattamento e gli altri, ad eccezione di T7, e la non distinzione tra tutti i tempi di adattamento superiori a T1.

Anche in questo caso come per la fricativa /s/, già dopo pochi minuti dall'inserimento del palato si ottiene un buon adattamento; l'andamento costante degli indici articolatori indica, con molta probabilità, il raggiungimento dell'adattamento.

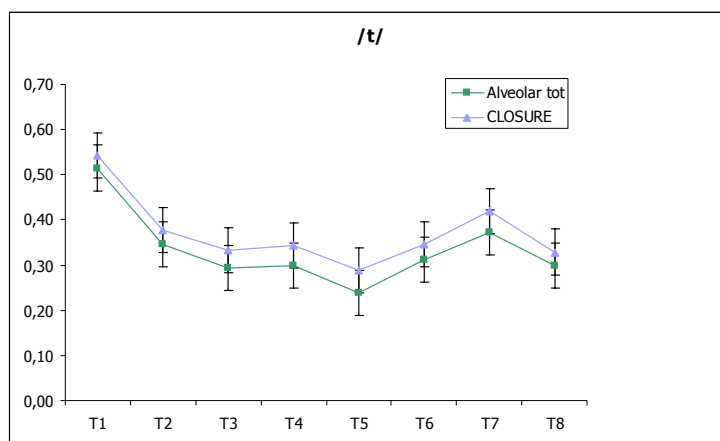


Figura 3 : Andamento degli indici AlvTot e Clos. medi in funzione dei tempi di adattamento per la /t/.

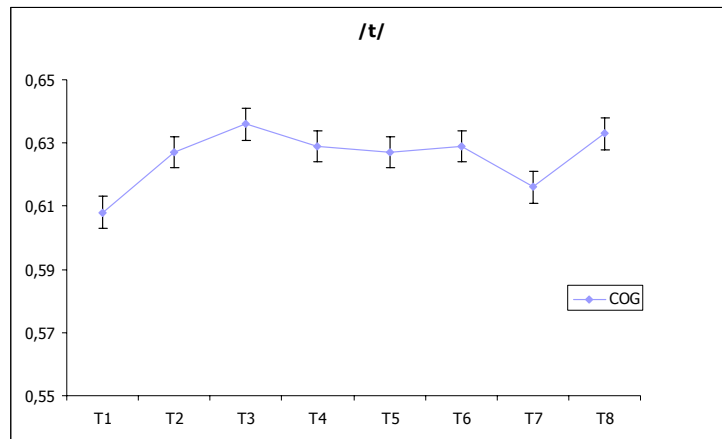


Figura 4: Andamento del COG medio in funzione dei tempi di adattamento per la /t/.

La Figura 5 mostra la differenza fra i patterns articolatori di /t/ ai due tempi di adattamento estremi. È evidente come al passare del tempo la struttura articolatoria dell'occlusiva si sposta maggiormente nella regione alveolare diventando più simmetrica e priva di contatti spuri derivanti principalmente da eccessiva salivazione.

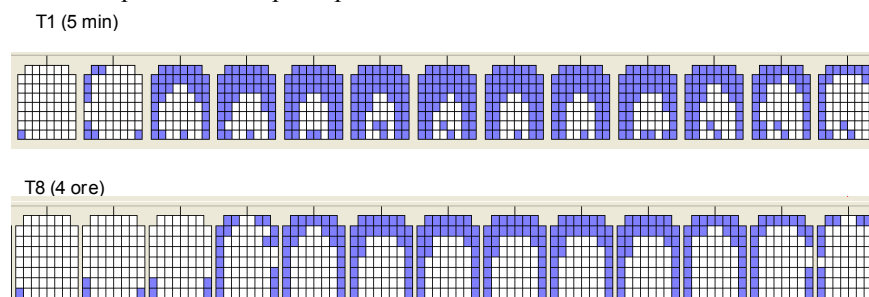


Figura 5: Confronto fra due patterns articolatori a 5 min (T1) e a 4 ore (T8) per l'occlusiva alveolare in /ata/.

#### 4.2 Risultati acustici

Per ognuna delle dieci ripetizioni di /asa/ e /ata/ e per ogni tempo di adattamento sono stati calcolati i tre momenti spettrali (media SM1, skewness SM3 e kurtosis SM4), escludendo, come già detto, il secondo momento (varianza SM2).

Il calcolo è stato eseguito per la occlusiva /t/ nel punto di onset del burst, per la fricativa /s/ invece nel punto di mezzo della fricazione, usando in entrambi i casi una finestra Hamming di 20 ms.

Analizzando il caso della /s/, i valori medi dei tre momenti sono riportati in Figura 6.

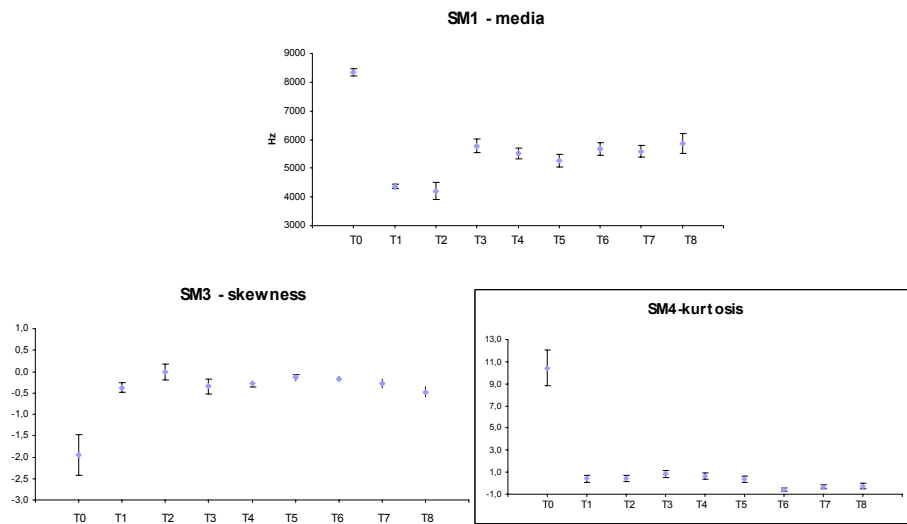


Figura 6: Valori medi dei tre momenti spettrali per la consonante /s/.

Da un'analisi qualitativa dei grafici si nota nell'andamento dei valori dei momenti SM1 SM3 e SM4 una diversità del valore calcolato a T0 rispetto a tutti i valori degli altri TA. Questo dimostrerebbe un'assenza di adattamento al palato, anche se con una lieve tendenza al miglioramento. L'analisi statistica e il test post-hoc mostrano che per il parametro SM1 T0 è significativamente uguale solamente al valore T8, cosa che indica che dopo 4 ore dall'inserimento del palato si può raggiungere un buon livello di adattamento dal punto di visto acustico. Per quanto riguarda SM3 si ha che tutti i valori ai vari TA sono statisticamente uguali.

I valori relativi alla /t/ sono riportati invece in Figura 7.

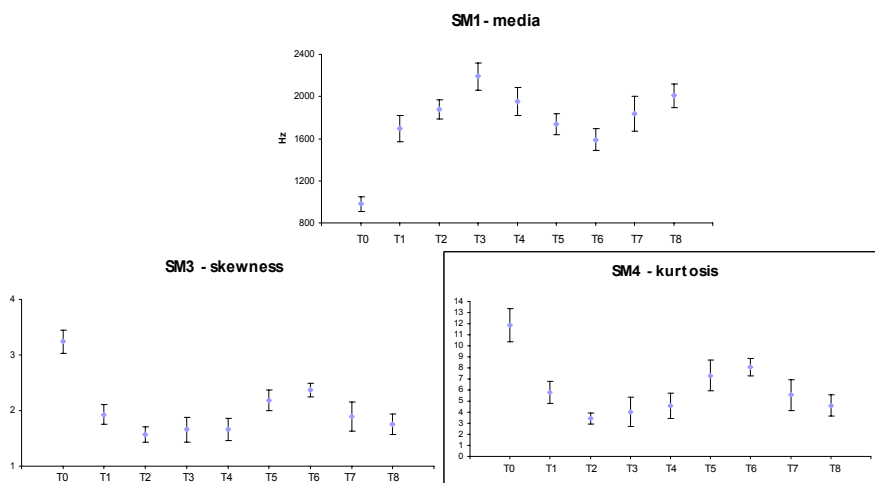


Figura 7: Valori medi dei tre momenti spettrali per la consonante /t/.

Come si può vedere abbastanza chiaramente dai grafici i valori dei tre momenti calcolati in assenza di palato (T0) non sono mai raggiunti dai valori relativi a tutti i tempi di adattamento registrati, nemmeno a quelli più 'alti'; ovvero mai si raggiunge un livello di adattamento al palato tale da poter dire che i dati registrati con e senza palato sono acusticamente indistinguibili.

Questa analisi qualitativa è confermata in parte dai risultati di un'analisi statistica condotta sui dati; per SM1 ciò che si ottiene è che c'è sempre significativa differenza fra i valori calcolati a T0 e tutti quelli calcolati ai vari tempi di adattamento, mentre fra i vari tempi di adattamento non c'è alcuna differenza, ovvero non viene registrato nessun miglioramento. Per SM3 si ha che c'è significativa differenza fra T0 e tutti gli altri TA ad eccezione dei tempi T5, T6 e T7, questo suggerisce che dopo un'ora si raggiunge un buon livello di adattamento. Per quanto riguarda infine SM4 si ha che T0 non è significativamente diverso dai vari TA ad eccezione di T2 e T8, ovvero per quanto riguarda SM4 si raggiunge abbastanza velocemente un valore simile a quello di T0. È opportuno ricordare che le analisi statistiche sono state compiute su tutti i dati calcolati per ogni ripetizione, mentre i grafici riportano solo i valori medi dei parametri analizzati per ogni TA.

Nel trarre le conclusioni dobbiamo tenere presente comunque che la variabile che è maggiormente indicativa del fenomeno in questione è il primo momento spettrale (SM1), come già descritto, ed è da essa che dobbiamo dedurre i risultati; gli altri momenti (SM3 ed SM4) possono servire esclusivamente ad indicare una tendenza di secondo ordine.

Da tutto questo si può concludere che, dal punto di vista acustico, non c'è quasi mai un raggiungimento completo di adattamento al palato e che si possono notare, sia qualitativamente che in parte statisticamente, delle tendenze al miglioramento dei dati registrati.

#### *4.3 Risultati percettivi*

I giudizi di similarità medi sono stati graficati separatamente per le due consonanti ai vari tempi di adattamento (Figura 8). Ciò che si cerca di stabilire è se esista un confine nei tempi di adattamento oltre il quale gli stimoli vengono percepiti come se fossero prodotti senza palato e quindi, se c'è indistinguibilità fra lo stimolo prodotto senza palato e quello prodotto con palato.

Nei grafici riportati nella figura si nota che per la consonante fricativa /s/ solo dopo un tempo di adattamento superiore a 45 minuti si supera la soglia di 3 come giudizio di similarità, ovvero sembrano essere necessari almeno 45 minuti di adattamento per consentire di percepire lo stimolo come 'naturale'. Inoltre si nota che non si arriva mai alla completa indistinguibilità dei due suoni nemmeno a tempi di adattamento pari a 4 ore.

Diversa è la situazione per la consonante occlusiva /t/: il giudizio di similarità medio è sempre piuttosto alto, ovvero gli stimoli risultano percettivamente già di buona qualità anche a tempi di adattamento inferiori a 30 minuti.

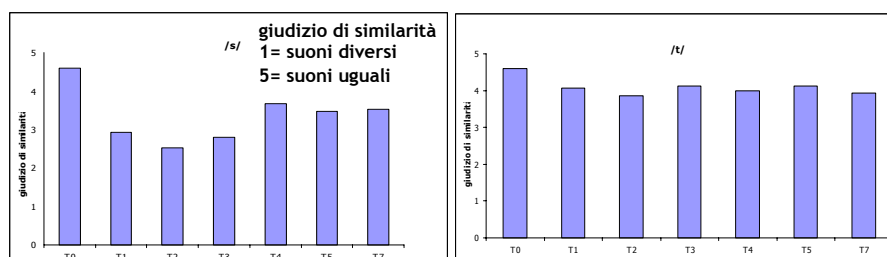


Figura 8: Istogrammi dei valori medi dei giudizi di similarità.

Alle analisi qualitative si è aggiunta un'analisi statistica che conferma i risultati già messi in evidenza precedentemente: per la fricativa /s/ si ha che non c'è significativa differenza fra T0 e T5, mentre c'è per tutti gli altri tempi di adattamento. Quindi si può dire che dopo i 60 minuti si ha indistinguibilità fra i due stimoli. Per la occlusiva /t/ invece si ha che non c'è mai significativa differenza fra i vari tempi di adattamento, ciò a significare che c'è un'assenza di distinzione fra gli stimoli prodotti con e senza palato. Anche in questo caso si può notare la stessa tendenza, trovata nelle analisi precedenti, ad un peggioramento relativo ai tempi di adattamento più alti (T7), spiegabile sempre come un fenomeno di fastidio legato al prolungato periodo di inserimento del palato.

## 5. CONCLUSIONI

Premettendo che il numero di ripetizioni non è particolarmente elevato e che il numero di soggetti è sicuramente poco informativo, si possono avanzare alcune considerazioni in merito all'individuazione di un tempo di adattamento.

Il primo risultato che si ottiene da questo studio è sicuramente il fatto che l'adattamento è un fenomeno dipendente fortemente dal tipo di consonante oggetto di studio e dal tipo di analisi che si vuole intraprendere; è quindi possibile individuare dei tempi minimi di adattamento al palato necessari prima della registrazione degli stimoli strettamente legati al contesto dell'analisi che si vuole intraprendere.

Ciò che colpisce particolarmente è che dal punto di vista articolatorio, l'adattamento avviene in tempi molto più brevi rispetto alle altre due prospettive (acustica e percettiva); questo suggerisce che se i materiali registrati devono essere analizzati esclusivamente dal punto di vista articolatorio possono essere ridotti notevolmente i tempi di adattamento.

I tempi minimi individuati in questo studio per l'adattamento relativo alla consonante occlusiva /t/ sono 15 minuti per l'analisi articolatoria mentre per l'analisi acustica non si individua mai un tempo a cui i parametri analizzati sono mediamente uguali a quelli del tempo di riferimento (anche se si possono notare tendenze al miglioramento); per i risultati percettivi emerge che già le registrazioni ottenute a 5 minuti non mostrano differenze statistiche con quelle registrate senza palato.

Per quanto riguarda la fricativa /s/, si individuano come tempi minimi 15 minuti per l'analisi articolatoria, 4 ore per quella acustica e 60 minuti per quella percettiva.

Un aspetto particolarmente nuovo, e che dovrà essere meglio indagato, è che in tutti i dati registrati emerge la tendenza ad un peggioramento nei livelli di adattamento presenti ai tempi più elevati, ovvero nel caso di tempi pari a 3 e 4 ore (T7 e T8). Questo fatto suggerirebbe un fenomeno di stanchezza e di insofferenza al palato da parte del soggetto. Ciò che è interessante è che esso emerge sempre, lo si ritrova infatti sia nelle analisi

articolatorie che in quelle acustiche e percettive, sia per la occlusiva /t/ che per la fricativa /s/. Sarà necessario indagare se ciò sia un fenomeno peculiare di questo soggetto o se, a tempi diversi, si manifesti in generale.

Ci proponiamo di estendere in futuro l'analisi, confrontando i risultati ottenuti con lo studio di altre consonanti anche in contesto di frase. Inoltre sarà nostra cura incrementare il numero di parlanti in modo da aumentare le dimensioni del campione testato ed avere così una maggiore stabilità statistica.

## 6. BIBLIOGRAFIA

Aasland W., S. Baum, D. McFarland (2006) "Electropalatographic, acoustic, and perceptual data on adaptation to a palatal perturbation" *Journal of the Acoustical Society of America* 119, 2372-81

Baum S. & D. McFarland (1997) "The development of speech adaptation to an artificial palate" *Journal of the Acoustical Society of America* 102, (4) 2353-2359

Baum S. & D. McFarland (2000) "Individual differences in speech adaptation to an artificial palate" *Journal of the Acoustical Society of America* 107, (6) 3572-3575

Calamai S. & P.M. Bertinetto (2006) "Per uno studio articolatorio dei glides palatale, labio-velare e labio-palatale dell'italiano", Atti del 3° Convegno Nazionale AISV Scienze Vocali e del Linguaggio" - Metodologie di Valutazione e Risorse Linguistiche, Povo (Trento) 29 novembre -1 dicembre 2006, CD-rom

Flege, J., Fletcher, S., and Homiedan, A. (1988) "Compensating for a bite block in /s/ and /t/ production: Palatographic, acoustic, and perceptual data," *Journal of the Acoustical Society of America* 83, 212-228.

Forrest K., G.Weismer, P. Milenkovic, R.Dougall (1988) "Statistical analysis of word-initial voiceless obstruents: preliminary data" *Journal of the Acoustical Society of America* 84 (1) 115-123

Hardcastle W., W. Jones , C.Knight (1989) "New developments in electropalatography: A state-of-the-art report" *Clinical Linguistics & Phonetics* 3 (1) 1-38

McAuliffe M., M. Robb & B. Murdoch (2007) "Acoustic and perceptual analysis of speech adaptation to an artificial palate" *Clinical Linguistics & Phonetics* 1- 10

McFarland D., S. Baum, (1995) "Incomplete compensation to articulatory perturbation" *Journal of the Acoustical Society of America* 97, 1865-1873.

McFarland D., S. Baum, C. Chabot (1996) "Speech compensation to structural modifications of the oral cavity" *Journal of the Acoustical Society of America* 100 (2) 1093-1104.

McLeod, S. & J. Searl (2006) "Adaptation to an electropalatograph palate: Acoustic impressionistic, and perceptual data" *American Journal of Speech-Language Pathology*, 15, 192-206.

Shadle, C.H. & Mair, S. (1997), "Quantifying spectral characteristics of fricatives", in *Proceedings ICSLP 1996*, Philadelphia, PA, USA, October 3-6, 1521-1524.

Wrench A. (2003), Articulate Assistant User Guide (vers. 1.12).

# L'ORGANIZZAZIONE TEMPORALE DEI GESTI VOCALICI E CONSONANTICI NELLE CONSONANTI SCEMPIE E GEMINATE DELL'ITALIANO

\*Claudio Zmarich, °Barbara Gili Fivela, ^Pascal Perrier,

^Christophe Savariaux, \*Graziano Tisato

\*ISTC-CNR, sede di Padova, via Martiri della Libertà, 35137, Padova (I),

°Università del Salento, via Taranto 35, 73100 Lecce (I),

^ICP, 46, av. F. Viallet, 38031 Grenoble Cedex 1 (F)

[zmarich@pd.istc.cnr.it](mailto:zmarich@pd.istc.cnr.it), [barbara.gili@ateneo.unile.it](mailto:barbara.gili@ateneo.unile.it), [perrier@icp.inpg.fr](mailto:perrier@icp.inpg.fr), [savariaux@icp.inpg.fr](mailto:savariaux@icp.inpg.fr),  
[tisato@pd.istc.cnr.it](mailto:tisato@pd.istc.cnr.it)

## 1. SOMMARIO

Le differenze nella durata acustica e percettiva dei foni consonantici possono trovare spiegazioni in fatti articolatori di tipo differente, quali le diverse posizioni spaziali dei target e quindi delle ampiezze di movimento, le differenze relative alla velocità dei movimenti e la diversa sincronizzazione di gesti relativi a sequenze di foni.

Un caso particolare e molto interessante è rappresentato dalle consonanti geminate, cioè quei segmenti caratterizzati da durate fonologicamente lunghe. Questi segmenti rivestono particolare interesse perché possono essere confrontati con le loro controparti scempie per controllare varie ipotesi, da quelle che concernono la diversa posizione dei target articolatori a quelle relative a possibili cambiamenti nell'organizzazione temporale dei gesti articolatori in funzione della lunghezza fonologica dei segmenti da allocare nella catena fonica. Quest'ultima ipotesi si connette anche alla questione più generale riguardante il tipo di organizzazione prosodica propria di ciascuna lingua, in particolare per ciò che riguarda le sue caratteristiche ritmiche e la tipologia delle sue unità prosodiche. Per questi motivi, alcuni studiosi si sono interessati recentemente a quelle lingue che evidenziano un contrasto fonologico di lunghezza nel loro sistema consonantico, come il giapponese (Smith, 1995; Löfqvist, 2005; Löfqvist, 2006), lo svedese (Löfqvist, 2006) o l'italiano (Smith, 1995; Zmarich & Gili Fivela 2005, Gili Fivela et alii, 2007, Zmarich et alii, 2007), e che dal punto di vista genealogico e tipologico sono molto diverse tra loro.

Nel sistema fonologico dell'italiano si fa un uso pervasivo del contrasto di lunghezza per le consonanti omorganiche in posizione intervocalica interna alla parola. Studi acustici e percettivi hanno dimostrato da molto tempo che la durata delle geminate è all'incirca doppia della durata delle scempie e che la durata della vocale precedente contribuisce a distinguere scempie e geminate (Bertinetto, 1981). Studi recenti di tipo articolatorio sulla cinematica delle labbra e della lingua nella produzione di consonanti scempie, geminate e in gruppo consonantico raccolti con il sistema EMA presso l'ICP di Grenoble (F) su un'ampia tipologia di gesti labiali e linguali hanno riscontrato [Gili Fivela et alii, 2007] che il gesto consonantico realizzato da due parlanti italiani (Nord-Ovest e Nord-Est) non sembra sovrapposto (*overlapped*) al gesto vocalico contestuale in "mim(m)a", come predetto dal modello coarticolatorio "V-to-V" di Öhman (1967) e riscontrato da Smith (1995). In realtà l'organizzazione temporale reciproca di eventi articolatori quali l'istante di massima costrizione bilabiale rispetto all'inizio e alla fine del movimento della lingua per la vocale e l'intervallo temporale tra i due target vocalici ha richiesto l'ibridazione del modello di Öhman con il modello di organizzazione temporale "V+C" di Browman & Goldstein (1986).

Per quanto riguarda l'ipotesi dell'esistenza di target articolatori più o meno "distanti" per scempie e geminate (Löfqvist & Gracco, 1997; Löfqvist, 2005), i risultati sembrano puntare verso l'esistenza di target articolatori differenziati per posizione in funzione di combinazioni particolari di articolatori e tipi di gesti (di apertura o chiusura). Il dato più forte a favore dell'ipotesi riguarda la velocità significativamente maggiore, con il picco più vicino al rilascio, per il gesto di apertura delle geminate alveolari in entrambi i soggetti.

## 2. INTRODUZIONE

Nel sistema fonologico dell'italiano la maggioranza delle consonanti presenta un contrasto di lunghezza in posizione intervocalica interna alla parola. Le caratteristiche fonetiche delle consonanti geminate dell'italiano sono state presentate in una serie di studi, specialmente per quanto riguarda le caratteristiche acustiche e uditive. Per esempio, molti studi hanno stabilito che la durata delle geminate è all'incirca doppia della durata delle corrispondenti scempie e che la durata della vocale che precede immediatamente la consonante geminata è più breve della vocale che precede la consonante scempia (Bertinetto, 1981; Esposito & Di Benedetto, 1999).

Il primo studio di tipo cinematico sulle consonanti geminate dell'italiano era focalizzato sull'organizzazione temporale dei gesti articolatori consonantici rispetto ai gesti articolatori delle vocali adiacenti (Smith, 1995). L'autrice concludeva la sua analisi affermando che la strategia articolatoria sottostante la produzione delle geminate in italiano era meglio rappresentata dal modello *Vowel-to-Vowel*, proposto da Öhman (1967) invece che dal modello *Vowel & Consonants* di Browman and Goldstein (1986).

Altri studi di tipo cinematico considerano le geminate sotto un altro punto di vista, chiedendosi se siano meglio interpretate da una rappresentazione monosillabica o da una eterosillabica (Gili Fivela & Zmarich, 2005; Zmarich & Gili Fivela, 2005). Dopo aver analizzato le proprietà spaziali e temporali dei movimenti del labbro inferiore nella produzione di consonanti scempie, geminate e in gruppo consonantico (intervocalico), gli autori trovarono che il gesto di apertura (dalla chiusura consonantica alla vocale successiva) aveva durata, ampiezza, profilo di velocità (cfr. *Time-to-Peak*) simili al gesto di apertura del nesso consonantico, ed entrambi erano maggiori rispetto al gesto di apertura della consonante scempia. Inoltre la rigidità del gesto (cfr. *stiffness*) era uguale a quella del nesso consonantico e inferiore a quella per la consonante scempia. Sulla base di tali evidenze, gli autori concludevano per una rappresentazione eterosillabica delle consonanti geminate, grazie alla loro somiglianza articolatoria con i gruppi consonantici (sicuramente eterosillabici).

Questo lavoro prende le mosse dalle analisi comparative menzionate precedentemente tra le consonanti scempie e quelle geminate dell'italiano. Lo scopo del presente studio è quello di discutere i dati tratti da un insieme più ampio di gesti articolatori (delle labbra e della lingua) e testare in questo modo direttamente le ipotesi avanzate finora sull'organizzazione articolatoria delle consonanti geminate. Il primo obiettivo è quello di testare direttamente l'ipotesi di Smith (1995) sull'organizzazione temporale dei gesti nelle coppie di consonanti scempie/geminate dell'italiano, il secondo obiettivo è quello di raccogliere dati che permettano di ricostruire la posizione "virtuale" del target articolatorio per le consonanti scempie e geminate, per scoprire se è vero che i parlanti italiani realizzano le consonanti geminate come se cercassero di raggiungere un target "virtuale" più alto dal punto di vista spaziale di quello delle consonanti scempie (Löfqvist, 2005).



### 2.1 L'organizzazione temporale dei gesti articolatori

Secondo Smith (1995), nelle lingue a isocronia sillabica come l'italiano, la durata dell'intervallo temporale tra i due nuclei di sillabe successive non dipende dal numero delle consonanti intermedie (VCV=VCCV, v. fig.1).

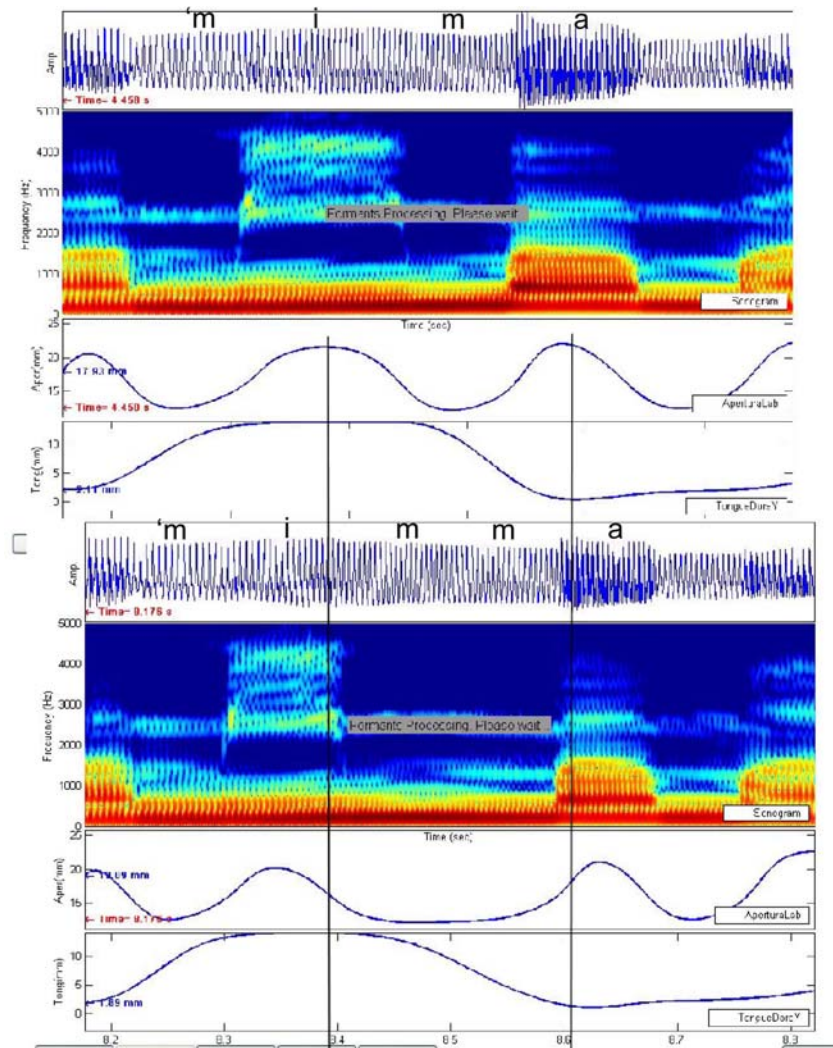


Fig. 1. Nella visualizzazione di *Interface*, sono mostrate due rappresentazioni dei segnali acustici (forma d'onda e sonogramma) e dei segnali cinematici (Apertura Labiale, dorso della lingua) relativi alle frasi “(richiam)a ‘mima/mimma marca(tamente)”. I segnali cinematici relativi alla consonante scempia (sopra) e geminata (sotto) sono stati allineati sul target articolatorio della prima vocale (linea verticale sinistra).

Nel suo esperimento 3 parlanti italiani producevano consonanti scempie e geminate in un contesto V /i/-a/ (per es. “mima” vs. “mimma”). Secondo Smith (1995), il *timing* di due vocali successive è controllato indipendentemente dalla consonante intervocalica (scempia

o geminata), e l'intervallo temporale tra il bersaglio cinematico di V1 e il bersaglio cinematico di C(C) non varia al variare della lunghezza della consonante. Ciò accade perché nella geminata il gesto di chiusura incomincia prima e il gesto di apertura finisce dopo (sovrapponendosi alle vocali contestuali), e il punto mediano del plateau di tenuta della chiusura consonantica è uguale a quello della scempia. Per Smith (1995) il modello che meglio spiega questo tipo di organizzazione articolatoria dell'italiano è il *V-to-V model* (Öhman, 1967), dove il gesto articolatorio per la consonante è considerato come sovrapposto al gesto da vocale a vocale, e quest'ultimo forma la base ritmica. Come esempio di applicazione di un modello alternativo, la Smith (1995) cita il giapponese, che non è una lingua a isocronia sillabica, e i cui dati sperimentali sono meglio spiegati dal modello *Vowel & Consonant* di Browman & Goldstein (1986). In effetti, nel giapponese il gesto vocalico (eseguito dal dorso della lingua) parte alla fine di quello consonantico, e di conseguenza la durata dell'intervallo temporale tra i due nuclei vocalici e la collocazione temporale del punto di massima costrizione consonantica dipendono entrambi dal numero di consonanti intermedie, come se ogni consonante aggiuntiva spostasse sempre un po' più a destra l'ultima vocale (vedi fig. 2; per l'organizzazione temporale del giapponese, cfr. anche Löfqvist, 2006).

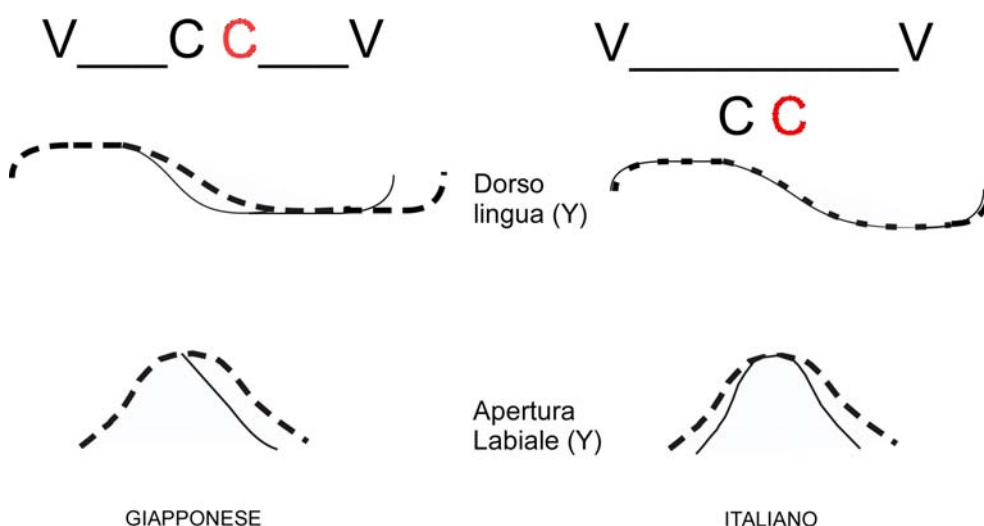


Fig. 2. I due possibili modelli di coordinazione temporale tra i gesti consonantici e vocalici secondo Smith (1995) per C. scempie e geminate: il modello *V & C* esemplificato dal giapponese (a sin.), e il modello *V-to-V* esemplificato dall'italiano (a destra).

L'ipotesi di Smith (1995) fu testata per l'italiano da Zmarich e Gili-Fivela (2005), che analizzarono le caratteristiche cinematiche di consonanti nasali realizzate in contesto vocalico /a/. I risultati furono abbastanza inconcludenti, nel senso che uno dei due soggetti sembrava aderire maggiormente al modello di Öhman (1967) e l'altro al modello di Browman e Goldstein (1986). Ciò potrebbe essere dovuto a strategie articolatorie idiosincratiche o all'influenza di un diverso sostrato linguistico dei due soggetti (Nord-Ovest e Nord-Est). Tuttavia la spiegazione più semplice fa appello a un artefatto metodologico, poiché le misure dell'apertura labiale li utilizzate, pur essendo correlate alla

produzione della vocale in termini maggiore apertura (cfr. Magno-Caldognetto, Vaggies & Zmarich, 1995), non misurano direttamente il movimento del dorso della lingua (che anzi continua a muoversi anche quando la bocca è chiusa, cfr. Löfqvist, 2006), e quindi possono non essere sufficientemente precise per cogliere sottili differenze nell'organizzazione spazio-temporale dei gesti consonantici e vocalici.

## 2.2 La strategia articolatori nel raggiungimento del target consonantico

Per quanto riguarda la differenza nelle strategie articolatorie impiegate per la produzione delle consonanti scempie e geminate, Löfqvist (2005) propose che i parlanti possono controllare la durata del contatto occlusivo tra gli articolatori, e quindi ottenere quelle differenze di durata acustica relative all'intervallo di tenuta dell'occlusione che dal punto di vista percettivo veicolano maggiormente la distinzione di lunghezza fonologica, regolando la posizione spaziale del target articolatorio. Dato che la misurazione diretta del punto spaziale di contatto tra gli articolatori (per es. labbro inferiore e labbro superiore in un'occlusiva bilabiale) non rivela nessuna differenza significativa, cioè le labbra s'incontrano e restano unite più o meno alla stessa altezza per le consonanti scempie e geminate, per verificare l'ipotesi teorica bisogna supporre che i) il target possa essere virtuale, cioè collocato anche al di là della superficie dell'articolatore passivo (volta palatina o labbro superiore) e ii) possa essere misurato. Una di queste misure sfrutta la correlazione positiva relazione tra ampiezza e velocità massima in un movimento: *a parità di durata*, spostamenti maggiori vengono eseguiti a velocità più alte. Dunque secondo Löfqvist (2005), se si trovasse che nelle geminate i picchi di velocità dei gesti di chiusura e di apertura occorrono più vicini ai confini acustici e hanno valori più alti di quelli delle corrispondenti scempie, questo dimostrerebbe indirettamente l'esistenza di un target articolatorio delle geminate posto più in alto di quello delle scempie.

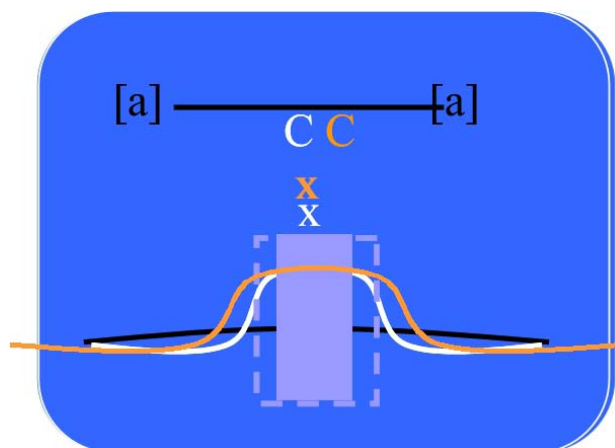


Fig. 3. Il gesto consonantico C(C) si sovrappone al gesto vocalico (a→a). Nella geminata (arancione) il target articolatorio (CC) è più alto che per la scempia (C) (bianco), e il periodo di contatto tra l'articolatore attivo e passivo è maggiore (maggiore durata della chiusura acustica, cfr. rettangolo tratteggiato grigio rispetto al pieno)

E' da notare che una strategia articolatoria che si basa solo sulla diversità dell'ampiezza del movimento per implementare variazioni di durata, presenta, a parità di altri fattori, il

vantaggio dell'economia esplicita (perchè non ricorre alla programmazione temporale esplicita).

I suoi dati sperimentali, tratti dalla cinematica del labbro inferiore di parlanti svedesi e giapponesi, non evidenziavano però differenze significative tra consonanti scempie e geminate nella velocità massima (a parte un soggetto svedese). Quindi Löfqvist (2005) concludeva che i soggetti non cambiano solo la posizione del punto finale del movimento, (il target), ma intervengono anche sull'organizzazione temporale, per es. rallentando il movimento allo scopo di mantenere più a lungo il contatto tra i due articolatori.

### 3. MATERIALI E METODI

#### 3.1 Corpus

Il corpus utilizzato in questa indagine è formato da parole e pseudo-parole CVC(C)V generate con le seguenti regole: a) l'accento cade sulla prima sillaba; b) le due possibili consonanti (occlusive sonore, dentali e velari, e laterali dentali) si oppongono per la lunghezza (per es. /d, l/ vs /dd, ll/); c) il contesto vocalico è sempre la /a/. A queste regole se ne è aggiunta una quarta per la creazione di *cluster* consonantici in cui le consonanti scempie dentali, la liquida laterale e l'occlusiva, comparissero a turno in ciascuna delle due posizioni di un cluster (/dl/, /ld/) (questa parte è già stata oggetto dello studio di Zmarich *et alii* (2007) e non sarà ulteriormente approfondita in questa sede). Si voleva in questo modo permettere il confronto tra le caratteristiche di un segmento scempio, di una geminata e del primo o del secondo segmento di un *cluster* analizzato come eterosillabico, con lo scopo di acquisire dati relativamente allo status mono o bifonematico delle geminate. Per quanto riguarda infine la verifica dell'ipotesi di Smith (1995) sull'organizzazione temporale di consonanti e vocali, sono state impiegate parole con consonanti nasali bilabiali e vocali /i/ ed /a/, come ad esempio nella parola "mim(m)a". Le complessive 12 parole e pseudoparole compaiono in una frase cornice in cui il target è seguita da una parola che comincia con la stessa consonante (per es., 'richiama *lal(l)a* latamente'), e che serve ovviamente a minimizzare possibili influenze coarticolatorie dal contesto non sotto studio. La lista completa del corpus è la seguente: "màma, màmma, làla, làlla, (gàga, gàgga, làlda, dàdla), mima, mimma, dàda, dàdda". In questa sede non verranno analizzate le parole con nessi consonantici o con le occlusive velari. Si può notare come tutte le strutture analizzate costituiscono parole attestate o parole possibili della lingua italiana.

#### 3.2 Soggetti

Le frasi in esame sono state acquisite complessivamente da quattro soggetti. In questo lavoro si presentano i risultati relativi solo a due di loro, entrambe donne e laureate (età: 30-35 anni), una originaria della provincia di Padova (NE) e l'altra di Torino (NW), dichiaratamente esenti da qualsiasi tipo di disturbo o disordine di voce, parola e linguaggio, presente e passato. Le frasi sono state randomizzate e presentate a gruppi di 3 su grandi pannelli posizionati a una distanza di circa due metri dal soggetto. Ai soggetti è stato chiesto di leggere ogni frase prima a velocità normale e subito dopo alla massima velocità sostenibile senza perdere di precisione. Per ogni soggetto sono state ottenute 10 ripetizioni di ciascuna delle 12 frasi a velocità normale e altre 10 a velocità elevata, per un totale di 240 enunciati (120+120) ed un totale complessivo di 960.

#### 3.3 Strumentazione e procedura sperimentale

Il segnale audio è stato acquisito su registratore DAT tramite un microfono professionale, mentre i dati cinematici sono stati ottenuti con il sistema "Electro Magnetic

Articulograph” o EMA AG200 (© Carstens), nella versione 2D in uso presso il laboratorio dell’ICP a Grenoble (F).

In figura 3 si può vedere la configurazione dei sensori utilizzati (e che sono attaccati con colla medica). I sensori alla base del naso e tra gli incisivi superiori servono come riferimento per compensare i movimenti della testa. Quelli posti alla base del confine tra i due incisivi inferiori costituiscono il riferimento per l’articolazione della mandibola. Gli altri sensori erano posizionati nei seguenti punti anatomici: due al centro del labbro inferiore e superiore, e quattro sulla superficie della lingua, a distanza di circa un centimetro l’uno dall’altro, partendo da quello più anteriore, posizionato un centimetro dopo l’apice, e proseguendo verso la radice.

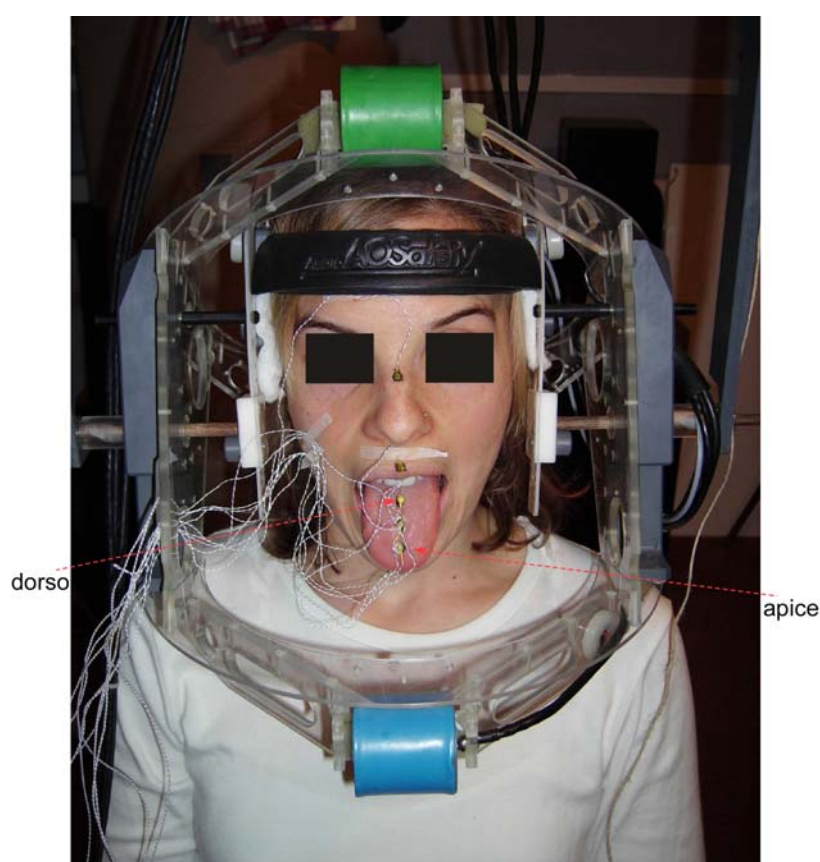


Figura 3: Posizione dei sensori durante la registrazione effettuata nei laboratori dell’ICP di Grenoble, mediante l’*Electro Magnetic Articulograph – 2D*, o EMA (© Carstens).

I dati discussi in questa sede riguardano le misurazioni relative al sensore incollato vicino all’apice della lingua, per le consonanti dentali, mentre per le consonanti labiali si sono utilizzati i sensori posti al centro del labbro superiore e inferiore, derivando la misura della distanza verticale, chiamata Apertura Labiale. Per le vocali, si è utilizzato il sensore incollato al dorso della lingua. Con una buona calibrazione e facendo attenzione che tutti i

sensori siano fra loro allineati lungo il sulcus centrale della lingua, si può ottenere una risoluzione spaziale del sistema con un margine di errore di 0.5 mm.

All'inizio di ogni seduta individuale di registrazione, un piccolo pannello rigido di forma rettangolare su cui erano stati attaccati due sensori è stato appoggiato sulla chiostra dentaria inferiore ed il soggetto è stato invitato a morderlo delicatamente per registrare la posizione della linea di occlusione. Alla fine della registrazione tutti i dati cinematici sperimentali sono stati corretti per il movimento della testa e ruotati fino a far coincidere la linea di occlusione con l'asse delle  $x$ . Questa rotazione è necessaria per ottenere un sistema di coordinate uniforme per tutti i soggetti.

### 3.4 Test uditivo

Mentre nell'italiano standard la durata delle consonanti è distintiva, diversi dialetti parlati nel Nord-Italia non fanno uso di geminate e quindi non sfruttano l'opposizione scempia/geminata. Per questo motivo, in via preliminare abbiamo deciso di verificare con un test di percezione se i nostri due soggetti avevano reso adeguatamente l'opposizione di lunghezza consonantica. Il test ha permesso di individuare le produzioni che risultavano chiaramente attribuibili all'una o all'altra classe di lunghezza. Le analisi acustiche e cinematiche sono state concentrate solo sulle consonanti riconosciute correttamente come scempie o geminate, per mettere in luce quei correlati acustici e articolatori associati al contrasto di lunghezza.

Una selezione dei 480 stimoli prodotti dai due parlanti è stata impiegata per il test di percezione. Per ciascun parlante, sono state scelte due ripetizioni prodotte a velocità normale e tutte le ripetizioni prodotte a velocità rapida (per un totale di 240 stimoli, nessuno dei quali con cluster consonantico). Un *panel* di cinque soggetti provenienti da Torino e cinque da Lecce e Taranto ha preso parte al test di percezione. Con l'ausilio del software *Perceval* (sviluppato presso il laboratorio LPL di Aix-en-Provence, Francia), i soggetti dovevano ascoltare i file audio che contenevano le parole target e giudicare se nelle parole erano presenti consonanti scempie o geminate. I soggetti ascoltavano ogni stimolo per due volte, e, dopo un segnale sonoro, avevano 5 secondi per rispondere, scegliendo tra due opzioni che venivano visualizzate sullo schermo (ad esempio, 1) 'lala' e 2) 'lalla'). Il test prevedeva una fase di addestramento nella quale ai soggetti erano presentati 5 stimoli scelti a caso tra quelli del corpus. Se gli stimoli erano riconosciuti in modo errato da almeno 2 soggetti di entrambe le varietà, o almeno 4 soggetti di una sola varietà, gli stessi stimoli non erano ulteriormente considerati per l'analisi acustica e cinematica.

### 3.5 Analisi acustica e cinematica

Dal punto di vista acustico, sono stati individuati i confini e gli intervalli corrispondenti ai segmenti che componevano le pseudo-parole target del corpus. Le analisi sono state effettuate grazie al software *PRAAT* (P.Boersma e D. Weenink, Institute of Phonetic Sciences of the University of Amsterdam). I risultati sono stati poi importati nell'ambiente di analisi cinematica. Le elaborazioni cinematiche sono state eseguite impiegando il software *Interface*, che è un software interattivo, creato da G. Tisato per l'elaborazione e l'analisi dei dati cinematici (cfr. Tisato, Cosi, Somnavilla, Zmarich, 2007). Delle due dimensioni spaziali in cui è ricostruito il movimento di labbra e lingua, quella verticale e quella longitudinale, è stato scelto come oggetto di analisi la dimensione verticale, più pertinente a rivelare possibili differenze per quanto riguarda le differenze nei movimenti delle labbra, del dorso e dell'apice della lingua, considerati separatamente nelle vocali e nelle consonanti occlusive e laterali che comparivano come scempie o geminate.



In figura 4 si può vedere una schermata tipica di *Interface*, che illustra l'andamento temporale di una serie di parametri acustici e cinematici. Dall'alto in basso si distinguono: il segnale acustico (sez. superiore), relativo a parte dell'enunciato contenente la parola target “(richiam)a 'lalla la(tamente)””; il sonogramma a banda larga; il segnale cinematico relativo all'ampiezza della componente verticale del movimento dell'apice della lingua; il segnale cinematico relativo alla velocità istantanea del movimento dell'apice della lingua.

Sulla curva relativa all'andamento dell'ampiezza verticale dell'apice della lingua abbiamo evidenziato il gesto di chiusura (dalla posizione aperta assunta per la prima /a/ di “lalla” all'occlusione relativa alla consonante intervocalica), e di apertura (dall'occlusione relativa alla consonante intervocalica alla posizione aperta relativa alla seconda /a/ di “lalla”), mentre sulla curva relativa alla velocità istantanea del movimento dell'apice della lingua abbiamo evidenziato (con un circolo pieno) i picchi di velocità del movimento di chiusura e di apertura. In modo analogo a quanto eseguito per es. da Smith (1995), l'inizio e la fine di ogni gesto sono stati collocati nei punti in cui la velocità dell'articolatore scendeva al di sotto del 10% della velocità massima raggiunta dallo stesso gesto (v. Fig.4).

Per ciascuno dei suddetti gesti abbiamo calcolato l'ampiezza (mm), e la durata (ms), che risultano automaticamente dalla differenza nella posizione verticale tra il punto iniziale e quello finale del movimento, ed abbiamo inoltre calcolato per ciascun gesto la velocità massima istantanea (mm/s).

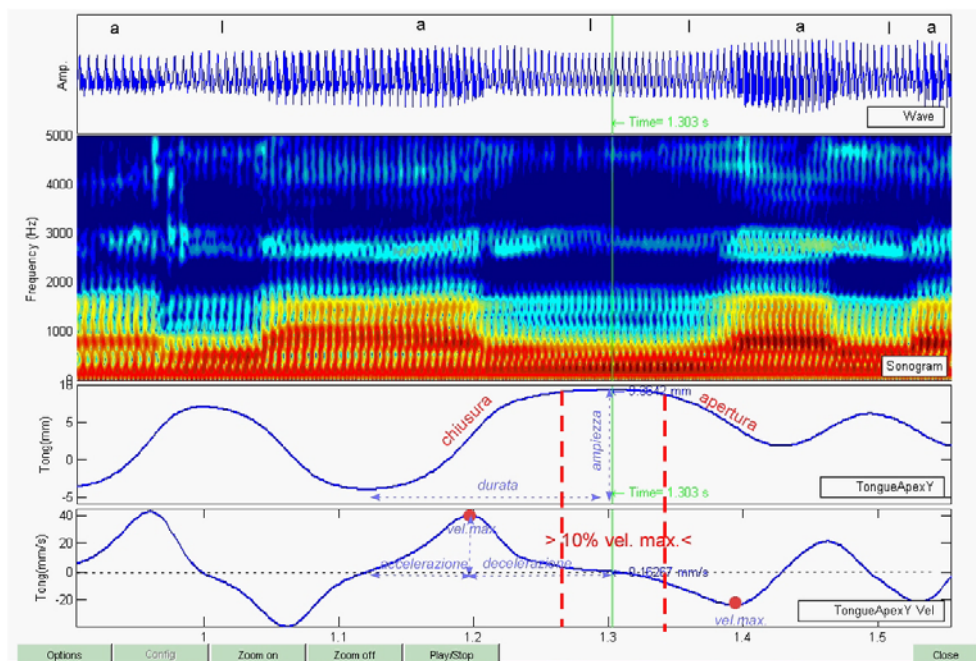


Figura 4. Schermata tipica di *Interface*, che riporta, dall'alto verso il basso, la forma d'onda del segnale acustico relativo alla frase “(richiam)a 'lalla la(tamente)””, il sonogramma a banda larga, il segnale cinematico relativo all'ampiezza della componente verticale del movimento dell'apice della lingua (mm); il segnale cinematico relativo alla velocità istantanea della componente verticale del movimento dell'apice della lingua (mm/s). Sulle curve cinematiche sono visibili i tipi di misurazioni effettuate.

In fig. 5 e fig. 6 vengono illustrati gli intervalli temporali calcolati per poter distinguere tra i due modelli alternativi di organizzazione temporale dei gesti articolatori:

1. tra il target articolatorio della prima vocale /a/ e il centro del ciclo consonantico /m(m)/.  
Laddove per la consonante la fine del gesto di chiusura non coincideva con l'inizio del gesto di apertura, il centro era collocato a metà del plateau di tenuta articolatoria, a bassa o nulla velocità. Se per l'italiano fosse valido il modello di Öhman (1967), questo intervallo nelle strutture VCCV dovrebbe essere uguale a quello per VCV (vedi Fig.5, n. 1).
2. tra il centro del ciclo consonantico /m(m)/ e il target articolatorio della seconda vocale /i/. Se per l'italiano fosse valido il modello di Öhman (1967), nelle strutture VCCV anche questo intervallo, che è speculare al precedente, dovrebbe essere uguale a quello per VCV (vedi Fig.5, n. 2).
3. tra il target articolatorio della prima vocale /i/ (cioè il punto più alto del dorso della lingua) e il target della seconda vocale /a/ (cioè il punto più basso del dorso della lingua): secondo il modello di Öhman (1967), nelle strutture VCCV e VCV dovrebbero essere uguali (vedi Fig.5, n. 3).
4. tra il target della prima vocale /i/ e il momento iniziale della chiusura per la consonante /m(m)/: secondo il modello di Öhman (1967), nelle strutture VCCV dovrebbe essere inferiore a quello per VCV (vedi Fig.6, n.4).
5. tra il momento finale del gesto di apertura della consonante /m(m)/ e la posizione del target per la seconda vocale (a): secondo il modello di Öhman (1967), nelle strutture VCCV dovrebbe essere inferiore a quello per VCV (vedi Fig. 6, n. 5).

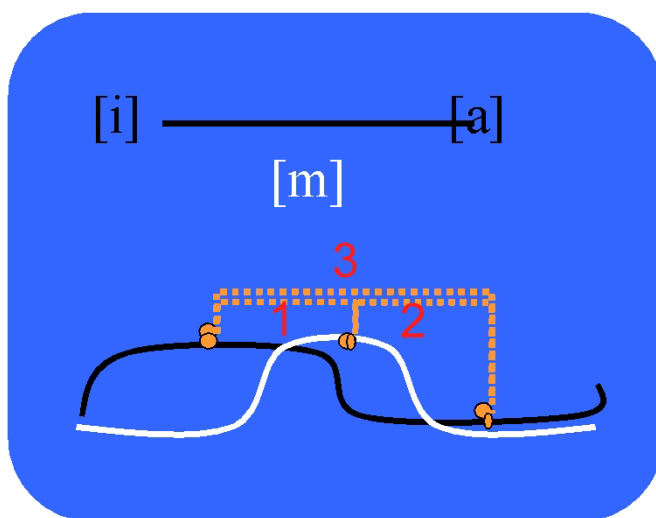


Fig 5. Schema delle misure effettuate per verificare il modello di coordinazione temporale C-V di Öhman (1967) applicato a "mim(m)a". La linea nera si riferisce al movimento del dorso della lingua da [i] ad [a], mentre la linea bianca si riferisce al gesto dell'Apertura Labiale per la realizzazione della consonante [m]/[mm]. Vedi testo per la spiegazione.



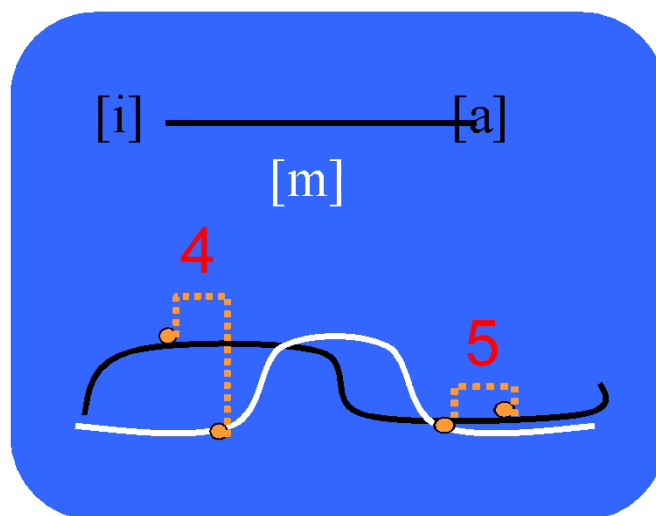


Fig 6. Schema delle misure effettuate per verificare il modello di coordinazione temporale C-V di Öhman (1967) applicato a “mim(m)a”. La linea nera si riferisce al movimento del dorso della lingua da [i] ad [a], mentre la linea bianca si riferisce al gesto dell’Apertura Labiale per la realizzazione della consonante [m]/[mm]. Vedi testo per la spiegazione.

## 4. RISULTATI

### 4.1 L’organizzazione temporale dei gesti fonetici

Per verificare se il fattore consonante singola/geminata fosse significativo per la coordinazione temporale tra consonante e vocale sono state eseguite una serie di analisi ANOVA a un fattore, separatamente per ciascun soggetto e per ciascuna velocità normale e massima.

Per quanto riguarda l’intervallo temporale tra il target articolatorio della prima vocale /a/ e il centro del ciclo consonantico /m(m)/ (vedi Fig. 5, n. 1) dovrebbe essere uguale per scempie e geminate, se per l’italiano fosse valido il modello di Öhman (1967). Infatti, in base ai risultati, non ci sono differenze significative, in entrambi i soggetti, sia a velocità normale, NE:  $[F(1,17)=1.992; p>0.05]$ , NW:  $[F(1,17)=0.345; p>0.05]$  che a velocità massima (NE:  $[F(1, 17)=0.003; p>0.05]$ , NW:  $[F(1, 14)=0.205; p>0.05]$ ).

D’altra parte, l’intervallo temporale tra il centro del ciclo consonantico /m(m)/ e il target articolatorio della seconda vocale /i/ (Fig. 5, n. 2) è maggiore per entrambi i soggetti nella produzione della geminata a velocità normale, sebbene una differenza significativa riguardi solo il soggetto NW ( $[F(1,17)=0.492; p>0.05]$ ,  $[F(1,17)=109.59; p<0.01]$ ). A velocità massima l’intervallo è significativamente maggiore nelle geminate rispetto alla scempie solo per il soggetto NE ( $[F(1,17)=6.867; p<0.05]$ ; NW:  $[F(1,14)=0.002; p>0.05]$ ).

Relativamente all’intervallo temporale tra il target articolatorio per la prima vocale /i/ e il target della seconda vocale /a/ (vedi Fig.5, n. 3), che in base al modello di Öhman (1967), dovrebbe essere uguale per scempie e geminate, è invece significativamente maggiore per entrambi i soggetti nelle geminate prodotte a velocità normale (NE:  $[F(1,17)=4.435;$

$p < 0.05$ ], NW:  $[F(1,17)=49.027; p < 0.05]$ ) e per il soggetto NE anche a velocità massima ( $[F(1,17)=8.859; p < 0.05]$ , NW:  $[F(1,14)=0.000; p > 0.05]$ ) (v. fig. 7).

Sembrerebbe così che l'ipotesi di Smith, che prevede in base al modello di Öhman (1967) al variare della lunghezza fonologica l'invarianza degli intervalli temporali tra il punto centrale di massima costrizione consonantica e i target articolatori delle due vocali contestuali, non sia del tutto confermata. Questi risultati trovano spiegazione nel fatto che, a velocità di produzione normale, sia il gesto di chiusura come quello di apertura durano significativamente più a lungo nelle geminate rispetto alle scempie, (chiusura e apertura per NE:  $[F(1,18)=19.619; p < 0.01]$ ,  $[F(1,18)=32.794; p < 0.01]$ ; chiusura e apertura per NW:  $[F(1,17)=12.405; p < 0.01]$ ,  $[F(1,17)=140.753; p < 0.01]$ ). A velocità massima di produzione, solo il gesto di chiusura risulta significativamente maggiore nelle geminate rispetto alle scempie, e solo per il soggetto NE ( $[F(1,17)=34.297; p < 0.01]$ ).

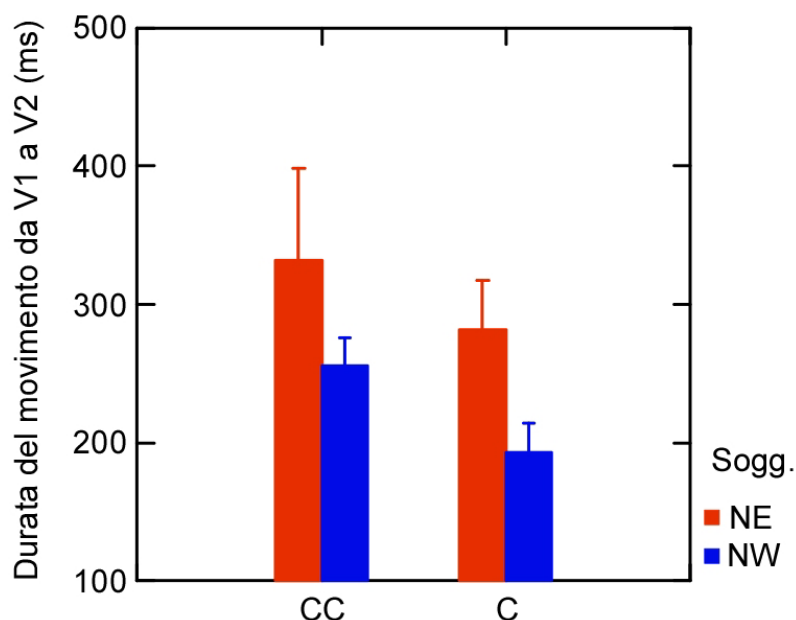


Fig. 7. Intervallo temporale dal target articolatorio della prima vocale /i/ a quello della seconda vocale /a/ per i due soggetti NE e NW e per la scempia (C) e geminata (CC)

La penultima misura, e cioè l'intervallo temporale tra l'istante relativo al target articolatorio del dorso della lingua per [i] e il momento iniziale della chiusura bilabiale per [m] e [mm] (v. Fig. 6, n. 4), di nuovo sembrerebbe soddisfare il modello di Öhman (1967), poiché effettivamente i risultati evidenziano che nelle geminate l'intervallo è significativamente più breve che nelle scempie per ambedue i soggetti a velocità normale (NE:  $[F(1,17)=4.450; p < 0.05]$ , NW:  $[F(1, 17)= 15.994; p < 0.05]$ , v. fig. 8), mentre non lo è a velocità massima (NE:  $[F(1, 17)=3.656; p > 0.05]$ , NW:  $[F(1, 14)=0.058; p > 0.05]$ ).

Invece l'intervallo speculare, che misura la durata tra la fine del gesto di chiusura bilabiale e l'istante relativo alla posizione del dorso della lingua per il target articolatorio [a] (vedi fig. 6, n. 5) non è significativamente diverso tra geminate e scempie, per nessun soggetto a nessuna velocità (normale e massima per NE:  $[F(1,17)=0.998; p > 0.05]$ ,  $[F(1,17)=3.791; p > 0.05]$ ; per NW:  $[F(1,17)=1.151; p > 0.05]$ ,  $[F(1,14)=0.008; p > 0.05]$ ).

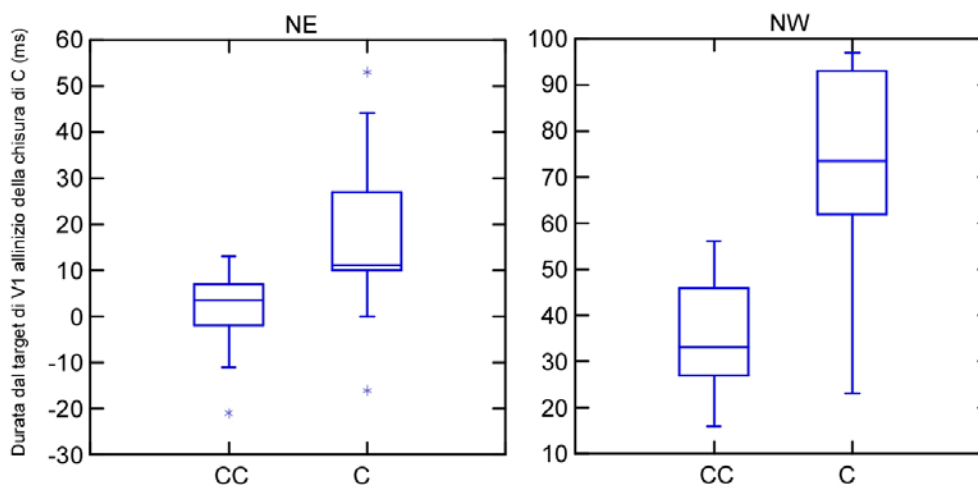


Fig.8. Intervallo temporale tra l'istante relativo al target articolatorio del dorso della lingua per [i] e il momento iniziale della chiusura bilabiale per [m] e [mm], per i due soggetti NE e NW per la scempia (C) e geminata (CC).

Riassumendo il senso di queste ultime due misure, si può dire che dall'analisi di queste due misure risulta che il gesto di chiusura per le geminate è anticipato rispetto a V1 (*ex hypothesis*), ma la fine del gesto di apertura non è ritardato rispetto a V2 (*contra hypothesis*).

Il quadro generale si qui delineato è confermato anche da altri risultati, ottenuti tramite analisi ispirate da un recente lavoro di Löfqvist (2006). Come nel caso di parlanti giapponesi, anche per i nostri soggetti l'inizio e la fine del movimento del dorso della lingua dalla prima alla seconda vocale occorrono rispettivamente prima e dopo il periodo di chiusura articolatoria della consonante, determinata acusticamente, e questo accade sia per le scempie che per le geminate. Solo per i nostri soggetti però, l'inizio della chiusura acustica occorre significativamente prima nelle geminate rispetto alle scempie (velocità normale e massima per NE:  $[F(1,17)=17.583; p<0.05]$ ,  $[F(1,17)=9.804; p<0.05]$  e per NW:  $[F(1,16)=11.069; p<0.05]$ ,  $[F(1,14)=0.605; p>0.05]$ ). Il secondo intervallo speculare, quello che va dalla fine del periodo di chiusura consonantica, determinato acusticamente, e la fine del movimento della lingua per il target della seconda vocale, non rivela invece alcuna differenza.

Per concludere, la predizione di Smith (1995) di un'anticipazione del gesto consonantico nelle geminate rispetto alla prima vocale e di un suo ritardo rispetto alla seconda vocale è confermato solo nel primo caso.

#### 4.2 L'ipotesi di Löfqvist (2006) del target gestuale "virtuale"

Ricordiamo brevemente che Löfqvist (2005) propose che i parlanti possono controllare la durata del periodo di chiusura degli articolatori, e quindi la distinzione acustica e percettiva legata alla differenza di durata tra consonate scempie e geminate, regolando la posizione spaziale del target consonantico.

Per testare l'ipotesi di Löfqvist (2005) abbiamo considerato la curva della velocità istantanea per l'apertura labiale in /mam(m)a/ e per il movimento verticale del dorso della

lingua per /l(l)a/ e /dad(d)a/. Una serie di analisi ANOVA a un fattore sono state eseguite per verificare l'influenza della lunghezza consonantica su:

- il valore della curva di velocità al momento iniziale e al momento finale della chiusura consonantica, individuata dal punto di vista acustico rispettivamente come inizio e fine della fase di segnale a bassa energia per /m(m)/ /l(l)/ /d(d)/ (in pratica quando incomincia e termina il contatto tra l'articolatore attivo e quello passivo);
- la posizione temporale del picco di velocità rispetto a questi confini acustici.

I risultati, sia per i valore spaziali che temporali, sono riassunti in tabella 1 (in alto: velocità normale; in basso: velocità massima), dove S significa che la differenza tra scempie e geminate per quella misura è significativa ( $p < 0.05$ ).

Per il gesto di apertura delle consonanti alveolari, la curva di velocità al rilascio della tenuta articolatoria ha un valore significativamente maggiore nelle geminate (v. fig. 9), e il suo picco di velocità è temporalmente più vicino all'istante di rilascio rispetto alle consonanti scempie (come accade per le bilabiali pronunciate da uno dei soggetti svedesi riportati in Löfqvist (2005)).

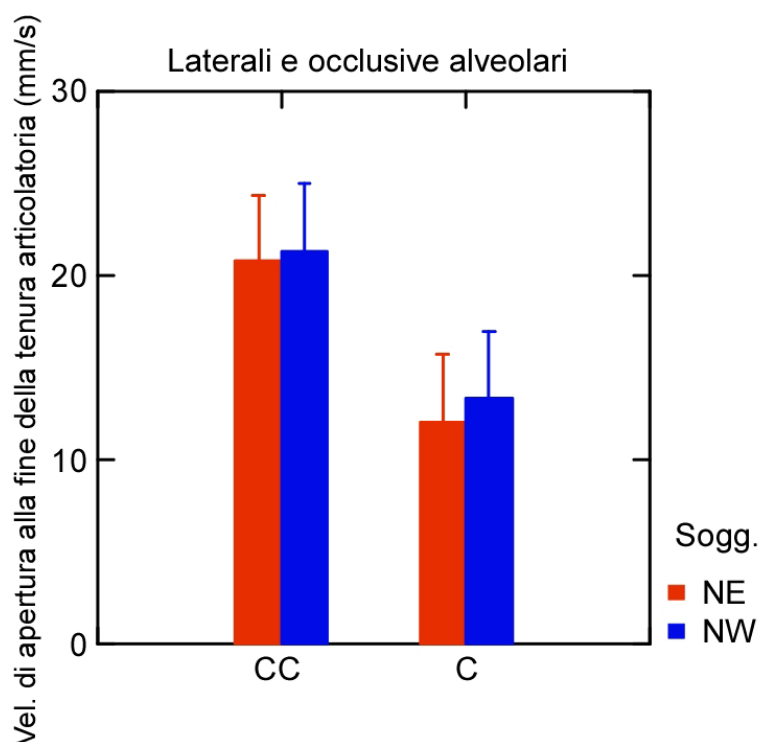


Fig. 9. Valore della velocità di apertura al momento finale della tenuta dell'occlusione alveolare, individuata dal punto di vista acustico, per i soggetti NE e NW, per la scempia (C) e geminata (CC).

A velocità massima, questa differenza tra geminate e scempie continua ad essere verificata solo per il soggetto NE, per tutti i fonemi considerati.

D'altra parte ci sono solo pochi risultati significativi per il gesto di chiusura a velocità normale e non riguardano mai lo stesso soggetto, mentre a velocità massima ci sono risultati significativo solo per il soggetto NE.

|          | Nas. Bilab. |        | Lat e Occl. Alv. |       |
|----------|-------------|--------|------------------|-------|
|          | NE          | NW     | NE               | NW    |
| Chiusura | N : S       | S : S* | N : N            | N : N |
| Apertura | S : N       | N : N  | S : S            | N : S |

|          | Nas. Bilab. |       | Lat e Occl. Alv. |       |
|----------|-------------|-------|------------------|-------|
|          | NE          | NW    | NE               | NW    |
| Chiusura | S : S       | N : S | S : S            | N : N |
| Apertura | S : S       | S : N | S : S            | N : N |

Tab. 1. Risultati delle analisi ANOVA ad un fattore : velocità normale (in alto) e velocità massima (in basso), per entrambi i soggetti NE e NW. "S" significa che la differenza tra scempie e geminate è significativa ( $p < 0.05$ ) per il valore di velocità all'inizio e alla fine del periodo di contatto articolatorio (a sinistra) e per la distanza temporale del suo picco massimo dai suddetti confini (a destra) nelle nasali bilabiali, e laterali e occlusive alveolari.

## 5. DISCUSSIONE E CONCLUSIONE

I risultati e le conclusioni di Smith (1995), che applica il modello di organizzazione temporale di Öhman (1967), non sono pienamente confermate dalle nostre analisi su due soggetti italiani che hanno pronunciato coppie di bisillabi differenziati per la lunghezza della consonante intervocalica. Il dato più forte a favore del modello di Öhman (1967) riguarda l'anticipazione dell'inizio del gesto consonantico per la geminata all'interno di V1. I dati in contrasto con tale modello però sono:

- il mancato ritardo della fine del gesto consonantico per la geminata all'interno di V2;
- l'assenza di invarianza per l'intervallo temporale tra i due target vocalici (V-to-V) . Viceversa esso varia significativamente al variare della lunghezza consonantica (*pro* Browman & Goldstein, *contra* Öhman);
- l'assenza di invarianza per gli intervalli tra il centro temporale dell'occlusione e il target articolatorio di V1 o V2. Essi non sono sempre costanti (come vorrebbe il modello di Öhman), soprattutto nel secondo caso.

Inoltre, a velocità massima, geminate e scempie mostrano poche differenze significative, e queste sono limitate a un singolo soggetto (NE).

Per quanto riguarda l'ipotesi di Löfqvist (2005) di un target articolatorio virtuale posizionato "più distante" per le geminate rispetto alle scempie, questa è parzialmente confermata. A favore ci sono i seguenti dati:

- a velocità normale, il gesto di apertura delle geminate alveolari ha una velocità significativamente maggiore, con il picco più vicino al rilascio.
- a velocità massima ciò è vero per ambedue i gesti di tutti i segmenti, ma solo per il soggetto NE.

A sfavore, la mancanza di differenze sistematiche per il gesto di chiusura, e a velocità massima, l'assenza di differenze anche per il gesto di apertura del soggetto NW.

## BIBLIOGRAFIA

- Bertinetto, P.M. (1981), Strutture prosodiche dell'italiano, *Accademia della Crusca*, Firenze.
- Browman, C.P. & Goldstein, L.M. (1986), Towards an articulatory phonology, *Phonology Yearbook*, 3, 219-252.
- Esposito, A. & Di Benedetto M.G. (1999), Acoustical and perceptual study of gemination in Italian stops, *JASA*, 106, 4, 2051-2062.
- Gili Fivela, B., Zmarich, C. (2005), Italian Geminate under Speech Rate and Focalization Changes: Kinematics, Acoustic, and Perception Data, *InterSpeech 2005*, Lisbon, 2897-2900
- Gili Fivela, B., Zmarich, C., Perrier, P., Savariaux, C., Tisato, G. (2007), Acoustic and kinematic correlates of phonological length contrast in Italian consonants, in *Atti dell'International Conference of Phonetic Sciences (ICPhS'07)* ISBN:978-3-9811535-0-7, Saarbrücken, Germany, 6-10 agosto 2007, 469-472.
- Löfqvist, A. (2005), Lip kinematics in long and short stop and fricative consonants, *JASA*, 117, 2, 858-878.
- Löfqvist, A. (2006), Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese, *JASA*, 120, 5, 2872-2883.
- Löfqvist, A. & Gracco, V. (1997), Lip and jaw kinematics in bilabial stop consonant production, *Journal of Speech, Language and Hearing Research*, 40, 877-893.
- Öhman, S.E.G. (1967), Numerical Model of Coarticulation, *JASA.*, 41, 310-320.
- Smith, C.L. (1995), Prosodic patterns in the coordination of vowel and consonant gestures, *Papers in Laboratory Phonology IV, Phonology and phonetic evidence*, (B. Connell & A. Arvaniti, editors), CUP, 205-222.
- Tisato, G., Cosi, P., Sommovilla, G., Zmarich C. (2007), Recent improvements in *InterFace* software: New tools for Articulatory Signal Processing, *Atti del III° Convegno Nazionale AISV* (V. Giordani, V. Bruseghini, P. Cosi, editors), Povo di Trento, 29-30 novembre-1 dicembre 2006, EDK Editore srl, Torriana (RN), 353-382.
- Zmarich, C., Gili Fivela, B. (2005), Consonanti scempie e geminate in italiano: studio cinematico e percettivo dell'articolazione bilabiale e labiodentale, in, *Misura dei parametri, Atti del I° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV)* (P. Cosi, editor), Padova, 2-4 dicembre, 2004, EDK Editore, Brescia, 429-448.
- Zmarich, C., Gili Fivela, B., Perrier, P., Savariaux, C., Tisato, G. (2007) Consonanti scempie e geminate in italiano: studio acustico e cinematico dell'articolazione linguale e bilabiale, in *Atti del Convegno Nazionale AISV, Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche* (V. Giordani, V. Bruseghini, P. Cosi, editors), Povo di Trento, 29-30 novembre-1 dicembre 2006, EDK Editore srl, Torriana (RN), 151-163.

# LÀ DOVE IL [ʃi] SUONA. UN'ANALISI PARAMETRICA DELLE SIBILANTI NELL'ITALIANO DI CARRARA.

Massimiliano Barbera, Marlen Barth, Diego Frassinelli, Gabriella Lapesa\*  
Università di Pisa

[m.barbera@ling.unipi.it](mailto:m.barbera@ling.unipi.it), [marlen\\_barth@yahoo.com](mailto:marlen_barth@yahoo.com), [d.frassinelli@gmail.com](mailto:d.frassinelli@gmail.com), [g.lapesa@gmail.com](mailto:g.lapesa@gmail.com)

## 1. INTRODUZIONE

L'area Apuana costituisce tradizionalmente un territorio di “frontiera” dal punto di vista sia paesaggistico (demarcato dalla cintura pseudoalpina delle Apuane), sia etnico, sia e soprattutto dialettologico. Estrema propaggine nordoccidentale della Toscana politica, da sempre è stata conteso crocevia di genti e filtro culturale e linguistico tra il Nord gallo-italico e la Toscana. Ancora oggi Massesi, Carraresi e Lunigianesi avvertono la propria identità etnica e dialettologica come la traccia di una specificità che li contraddistingue rispetto alle tre regioni di cui costituiscono il punto di intersezione (Liguria, Emilia-Romagna e appunto Toscana).



Figura 1: La posizione geografica di Carrara

Il presente lavoro si inserisce all'interno di un più esteso progetto di ricerca che intende mettere a fuoco alcune caratteristiche ricorrenti nell'italiano parlato di Carrara e Massa.

\* Massimiliano Barbera è autore dei §§ 5 e 8; Marlen Barth dei §§ 6 e 7; Diego Frassinelli dei §§ 3 e 4; Gabriella Lapesa dei §§ 1 e 2. Il lavoro è stato concepito in maniera unanime e le conclusioni sono state redatte concordemente dai quattro autori.

Innestato su una base dialettale ancora piuttosto vivace anche nelle fasce giovanili, l'italiano di Carrara risente di forti influssi fonetici diatopici, tanto nel vocalismo, patentemente non toscano, quanto nel consonantismo, che presenta interessanti fenomeni imputabili sia alla posizione di isoglossa, sia ad una ipotizzata deriva di tratti articolatori di sostrato variamente identificati, ereditati, forse, dai Liguri Apuani; la presenza di foni retroflessi (Ambrosi, 1956; Luciani, 1999), ad esempio, costituisce un aspetto fonetico idiosincratico rispetto ai parlari dell'Italia centro-settentrionale.

Il *corpus* esaminato è costituito dall'eloquio elicitato (sia in esecuzione spontanea, sia in lettura) di 6 parlanti nativi, 3 maschi e 3 femmine, di età compresa tra i 21 e 69 anni.

In questa fase dell'indagine viene valutato il particolare statuto articolatorio delle sibilanti, che manifestano sensibili peculiarità a livello percettivo e presumibilmente articolatorio. Una prima istanza di ordine sociolinguistico pare emergere dall'analisi empirica dei dati, per cui le specificità acustiche dei fonemi sibilanti denotano forti differenze distribuzionali in funzione del genere. In particolar modo si osserva la tendenza ad un'articolazione peculiare – percettivamente imputabile ad un arretramento articolatorio – delle sibilanti alveolari, sia sorde che sonore, tanto in contesto intervocalico, quanto in posizione pre-/post-consonantica. Il fenomeno potrebbe forse essere interpretato come deriva articolatoria di un generale *setting* che, come si è accennato, favoriva nell'inventario fonologico apuano la presenza di suoni retroflessi (Savoia, 1980; Luciani, 1999). In tal senso l'analisi spettroacustica sembra indicare una particolare conformazione del gesto articolatorio sotteso alla produzione di questi suoni.

La citata differenziazione di genere evidenzia, ad una prima ricognizione uditiva, una sensibile diversificazione nella caratterizzazione dei fenomeni fonetici considerati: i maschi manifestano infatti in maniera marcata la palatalizzazione (o retroflessione?) delle fricative alveolari in qualsiasi contesto sillabico, mentre le femmine accolgono in maniera più massiccia altri localismi fonetici; in particolar modo, sembrano realizzare in misura preponderante rispetto ai maschi la degeminazione consonantica in posizione intervocalica di matrice settentrionale (cfr. Barbera in stampa a e b). Per converso, anche le sibilanti prodotte dalle locutrici denotano elementi spettroacustici particolari, ravvisabili in maniera più rarefatta anche all'analisi percettiva, ma con connotati diversi rispetto a quelle dei soggetti maschili.

L'indagine mira a delineare un quadro parametrico e sperimentale relativo ad un primo modulo (quello delle sibilanti, appunto) all'interno dell'inventario fonetico/fonologico carrarese. Dal punto di vista interpretativo vengono proposte ipotesi classificatorie in relazione ai gesti articolatori sottesi alla produzione dei foni in esame e, in particolar modo, si cerca di chiarire se e in che senso il tratto di marcatezza della realizzazione dei fonemi /s s:/ possa essere identificato come semplice palatalizzazione, o piuttosto come retroflessione.

## 2. MATERIALI E METODI

Come già si è accennato, l'indagine riguarda 6 informatori nativi, 3 maschi e 3 femmine, di età compresa tra i 21 e 69 anni. I soggetti si iscrivono in un segmento di diglossia (lingua nazionale/dialetto), quantomeno per la competenza del dialetto locale, del quale hanno dichiarato una pratica più massiccia nelle interazioni pregresse all'interno delle famiglie di origine. Ad oggi l'uso del dialetto risulta fortemente penalizzato, ovviamente, a vantaggio dell'italiano regionale, ma comunque adoperato in alcuni contesti informali e frequentemente impiegato, magari con funzione enfatica ed espressionistica, in alternanza e/o inserto anche nella produzione in italiano.



I locutori sono stati intervistati seguendo un protocollo di elicitazione del parlato spontaneo, allo scopo di individuare, in un contesto il più possibile naturale, un quantitativo sufficiente di parole contenenti segmenti bersaglio. La lettura di liste di parole e frasi confezionate *ad hoc*, infatti, ha rivelato scarsa attendibilità per quanto riguarda l'esecuzione linguistica di nostro interesse, essendo il livello segmentale notevolmente polarizzato da una pronuncia tendente allo *standard*, laddove la produzione risulti più sorvegliata e iperarticolata. Si è quindi optato per la lettura di ricette della gastronomia locale in margine alle quali si è proceduto a gestire un dialogo, il più possibile informale, sugli stessi temi.

Il materiale audio, prelevato con registratore *DAT* portatile *Sony* e due microfoni professionali, è stato campionato ad una frequenza di 22050 Herz e una quantizzazione di 16 bit con *software Multi-Speech 3700* (versione 2.5). L'analisi è stata condotta applicando una griglia di annotazione realizzata con il *software Praat* (versione 4.6.31), strutturata in 4 livelli dialoganti (ortografico, fonologico, fonetico, sillabico) sul modello del sistema *SAIL2* (*Sistema di Annotazione per l'Italiano L2*) concepito presso il Laboratorio di Fonetica dell'Università di Pisa<sup>1</sup>. I dati quantitativi sono stati rilevati per mezzo di *query* appositamente formulate in interfaccia con i livelli annotati.

I parametri impiegati per l'analisi spettroacustica delle sibilanti vengono indicati di seguito. La categorizzazione allofonica, oltre al modo e al luogo di articolazione, fa riferimento ai parametri di sonorità e di durata, tenendo in conto anche eventuali fenomeni di degeminazione/geminazione secondaria (cfr. Barbera in stampa a e b); comprende pertanto basicamente gli *output* [s] [s:] [ʃ] [ʃ:] [z] [z:] [ʒ] [ʒ:], per i quali si sono rilevati:

- valori frequenziali di F1<sup>2</sup>, F2, F3, F4 rilevati sull'*offset* del segmento vocalico precedente; nello specifico si sono considerati pertinenti gli ultimi 23 ms della vocale (Gordon *et al.*, 2002)
- valore del *centre of gravity* come funzione relativa all'altezza frequenziale media occupata dall'energia nello spettro del segmento fricativo come descritto in Gordon *et al.* (2002), Jongman *et al.* (2000).

In relazione alla qualità timbrica della vocale precedente (sulla quale sono stati estratti i valori formatici in transizione), ovvero al luogo di articolazione della consonante contigua rispetto al segmento *target*, si è operata una distinzione preliminare in tre classi: anteriore, centrale, posteriore, allo scopo di discriminare tra gesti articolatori tendenti verso le tre diverse aree della cavità orale (ad esempio le occlusive velari sono state trattate come foni posteriori al pari delle vocali posteriori).

---

<sup>1</sup> Per i dettagli del sistema di annotazione *SAIL2* si rimanda a Marotta e Soriano (2008, in stampa).

<sup>2</sup> Di fatto il parametro F1 non ha dimostrato alcun rilievo né pertinenza in relazione alla nostra indagine; è stato pertanto escluso dall'analisi e dai computi riportati in tabella nelle pagine seguenti.

### 3. IL SISTEMA DI INTERROGAZIONE

Per le nostre analisi abbiamo sviluppato alcuni *script* usando linguaggi differenti (*Praat* e *PERL*), scelti in relazione alla diversa natura dei parametri da analizzare: spettrografici o quantitativi.

#### 3.1 Gli script Praat per l'estrazione dei parametri spettrografici

L'analisi dei valori formantici e del centro di gravità è stata implementata nel linguaggio di *scripting* incluso nel *software Praat*, che permette di gestire in maniera automatica le analisi acustiche realizzabili manualmente attraverso l'interfaccia utente.

Questi *script* analizzano il solo livello fonetico, estraendo i valori medi di F1, F2, F3, F4 e del centro di gravità degli ultimi 23 ms delle vocali che precedono le sibilanti. All'interno di questa finestra temporale, la media viene calcolata su tre punti equidistanti tra loro e mantenendo invariati i parametri di *default* delle funzioni di *Praat*.

#### 3.2 PERL e l'analisi dell'interfaccia fonetico-fonologica

L'analisi delle caratteristiche quantitative del consonantismo impone che l'analisi automatica prenda in considerazione tutte le possibili realizzazioni allofoniche dei singoli fonemi.

Per raggiungere questo obiettivo è stato realizzato in *PERL (Practical Extraction and Reporting Language)* uno *script* che:

- concatena tutti gli elementi del livello fonologico corrispondenti a segmenti consonantici con gli elementi corrispondenti presenti sul livello fonetico e calcola occorrenze, durata media e deviazione *standard* delle risultanti combinazioni fonema-allofono;
- analizza durata media e deviazione *standard* dei segmenti vocalici (divisi per timbro e sonorità) adiacenti alle consonanti, sempre rappresentate dalla combinazione di fonema e varianti allofoniche.

Abbiamo ritenuto, infatti, che questo tipo di linguaggio fosse più adatto a gestire l'analisi incrociata dei due livelli di annotazione.

Per quanto un'analisi multilivello sia realizzabile anche utilizzando il linguaggio di *scripting* di *Praat*, essa richiederebbe di conoscere in anticipo tutte le possibili realizzazioni allofoniche di un fonema: *Praat*, infatti, non dispone di strutture di dati complesse.

*PERL* consente, invece, di creare delle strutture di dati dinamiche e multidimensionali: questa caratteristica ci ha permesso di realizzare uno script unico, veloce e portatile che considera tutte le combinazioni fonema-allofono presenti nel *corpus*.

Una struttura di dati dinamica si presta ad essere agevolmente estesa ad interrogazioni più complesse senza dover riscrivere l'intero codice e senza aumentare in maniera esponenziale il numero di combinazioni da prendere in considerazione. La stessa struttura di dati è stata già utilizzata, infatti, per alcune indagini più granulari, che coinvolgono tre livelli di annotazione:

- l'analisi delle realizzazioni allofoniche di ogni fonema messe in relazione con la categoria morfologica della parola alla quale il fonema stesso appartiene (questa informazione è presente, nel nostro corpus, sul livello delle parole);
- analisi della durata media e deviazione standard delle vocali adiacenti ad ogni

segmento consonantico (combinazione fonema-allofono), con un collegamento al livello sillabico che permette di individuare sillabe aperte e chiuse.

La prospettiva di sviluppo di un sistema di interrogazione così strutturato è quella di realizzare un'interfaccia tra i due linguaggi di *scripting* in modo che le funzioni di *Praat* siano automaticamente gestite da uno *script PERL*, più adatto ad analisi multilivello complesse e dinamiche.

L'applicazione degli *script* così strutturati ha consentito il rilievo automatico delle misurazioni e la gestione statistica delle stesse. Le interrogazioni hanno riguardato i segmenti fricativi i cui *target* fonologici soggiacenti corrispondono ai fonemi /s, s:/, ʃ:/ dell'italiano. Le parole bersaglio, individuate all'interno del parlato spontaneo, hanno permesso di considerare contesti sia intervocalici, sia preconsonantici; con ciò si è tentato di evitare effetti di standardizzazione della dizione in senso toscano che si erano ravvisati, in una prima fase di registrazione, nella lettura degli stessi soggetti.

#### 4. I DATI QUANTITATIVI: INCIDENZA DEGLI ALLOFONI SIBILANTI

Come si è accennato, già la semplice ricognizione percettiva ha evidenziato all'interno del gruppo di informatori presi in esame una distinzione piuttosto sensibile di genere in relazione alla distribuzione degli allofoni sibilanti.

In un quadro più generale, la valutazione contrastava tra maschi e femmine non manifesta grosse sperequazioni; salvo registrare lievi variazioni per quanto concerne l'incidenza statistica di fenomeni di alterazione della quantità (cfr. Barbera in stampa a e b). L'unico tratto piuttosto divaricato consiste proprio nell'impiego massiccio da parte dei locutori maschili di un allofono della sibilante alveolare caratterizzato indubbiamente da fisionomia articolatoria peculiare. Ad un primo test impressionistico si è ipotizzato che potesse trattarsi di una fricativa alveolare con *locus* articolatorio arretrato (palatalizzazione); di fatto l'analisi comparativa tra [s] [ʃ] ha evidenziato una certa specificità, sia in ordine ai valori parametrici (come discuteremo di seguito), sia in relazione all'*output* fonetico che manifesta una visibile alterità rispetto all'uno e all'altro fono/fonema, anche per quanto concerne il visema corrispondente.

I dati quantitativi mettono in luce due fatti essenziali (si veda fig. 2):

- i fonemi /s/ e /s:/ vengono prodotti con articolazione tendenzialmente *standard* (alveolari) dalle informatrici femmine;
- per i maschi la proporzione si inverte, con una prevalenza assoluta della realizzazione allofonica marcata sia di /s/ che di /s:/.

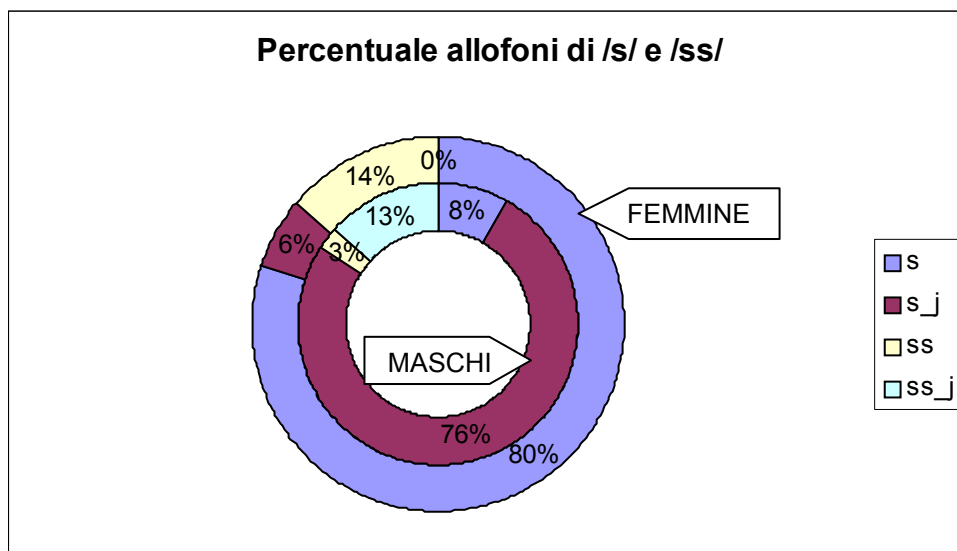


Figura 2: Valori percentuali relativi alla distribuzione degli allofoni<sup>3</sup> di /s/ /s:/ per locutori femminili (anello esterno) e maschili (anello interno).

La realizzazione peculiare delle sibilanti riguarda evidentemente i locutori maschili in misura assolutamente più massiccia, confermando la valutazione impressionistica di una sperequazione connessa al genere.

## 5. SIBILANTI E LATERALI: LO STATUTO PECULIARE DELLE CORONALI

Qualche osservazione merita la valutazione comparativa di due classi di consonanti che manifestano, nel dialetto carrarese, alcune peculiarità acustiche tali da aver sviluppato presumibilmente dei tratti di persistenza articolatoria nell'italiano locale: le laterali da un lato, le sibilanti dall'altro. Il segmento /l/, infatti, è soggetto, nella maggior parte delle varietà dialettali apuane, a processi di retroflessione (cfr. Ambrosi, 1956; Savoia, 1980; Luciani, 1999), con gradi di variabilità allofonica piuttosto sfumati a seconda della distribuzione geografica. In generale, l'area di retroflessione (o cacuminale, secondo la dizione tradizionale) copre una porzione piuttosto estesa del territorio apuano (cfr. fig. 3) dove si registra la presenza a livello dialettale di foni retroflessi nell'esito delateralizzato di \*[l:] originario > [d:] in una zona caratterizzata dall'andamento della cosiddetta isoglossa cacuminale, avente il suo centro geometrico in corrispondenza del monte Pisanino, sulle Apuane (cfr. Savoia, 1980: 260).

<sup>3</sup> La notazione SAMPA *s<sub>j</sub>* è stata impiegata nelle *textgrid* e nei grafici per indicare la pronuncia percepita come palatalizzata.

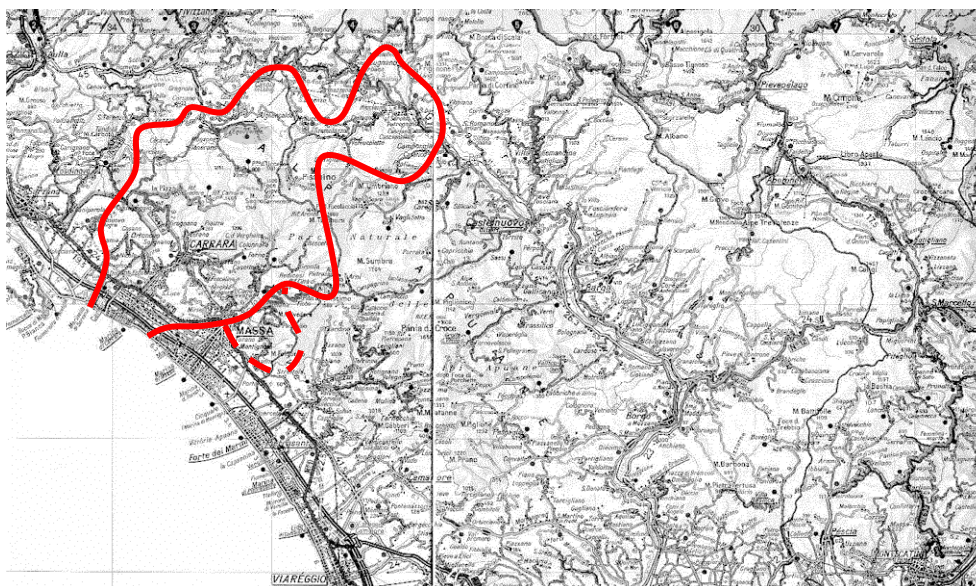
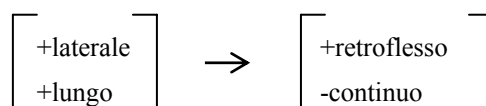


Figura 3: Area di diffusione di [d]; riproduzione sulla base di Ambrosi (1956).

Secondo Savoia (1980: 277) «lo sviluppo della retroflessa o degli esiti ad essa collegati da \*// originaria, sembra essere avvenuto dopo la degeminazione di \*// in certi contesti fonosintattici».

Lo schema seguente rappresenta le regole, sincronicamente alternanti in area carrarina, in base alle quali si registra la conservazione della laterale degeminata in posizione preconsonantica fonosintattica e la retroflessione in contesto prevocalico (Savoia, 1980: 276):

[+laterale] → [+lungo] / \_\_ # C



In altre parole, la pronuncia retroflessa, di matrice secondaria romanza e non sostratica, secondo Savoia, si sarebbe sviluppata come tratto seriore su una tessitura consonantica generale già ampiamente degeminata; una stratificazione di questo tipo giustificerebbe la conservazione della laterale degeminata in certi contesti fonosintattici (diffusamente in confine di parola preconsonantica), per cui il dialetto ammette l'alternanza ['m beq a'neq] e ['m beq a'neɫ'dɔr] (cfr. Savoia, 1980: 276). Nel sistema di distribuzioni allofoniche parrebbe dunque potersi ravvisare un condizionamento negativo esercitato dalla degeminazione delle C: sulla retroflessione, poiché la riduzione della quantità avrebbe agito «privando questo processo del suo input» (Savoia, 1980: 279).

Nello specifico della varietà carrarese di italiano, partiamo dall'osservazione dei dati ricavabili dal nostro campione di inchiesta. Dalla nostra analisi si evincono due dati di interesse:

- a. il fonema /l/ (cfr. fig. 4) risulta colpito da degeminazione in una percentuale assolutamente preponderante di occorrenze (76%);
- b. la proporzione di durata tra /l/ e /l:/ sembra indicare una manifestazione quantomeno peculiare del tratto di lunghezza, dal momento che la laterale geminata raggiunge valori più che raddoppiati rispetto alla scempia e alla degeminata corrispondenti (cfr. tab. 1).

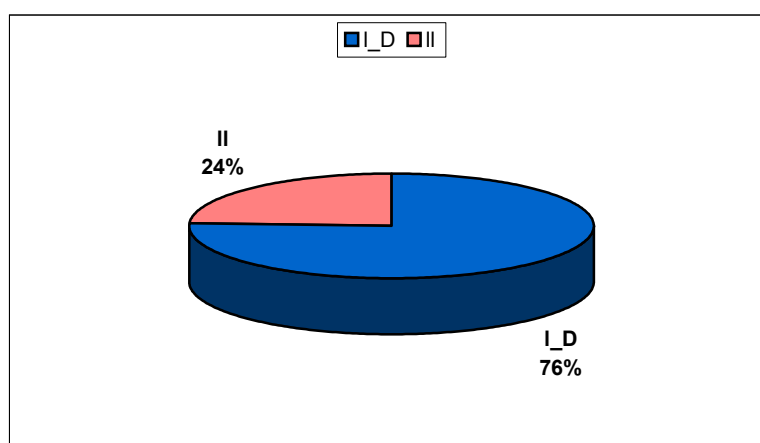


Figura 4: Valori percentuali della distribuzione allofonica di /l/ nelle sue varianti geminata (II) e degeminata (I\_D).

|              | /l/       | /l:/              |            |
|--------------|-----------|-------------------|------------|
|              |           | [l·] <sup>4</sup> | [l:]       |
| <b>media</b> | <b>45</b> | <b>49</b>         | <b>128</b> |
| <b>σ</b>     | <b>5</b>  | <b>7</b>          | <b>37</b>  |

Tabella 1: Valori (in ms) e deviazione *standard* relativi ai fonemi /l/ e /l:/.

Non è forse da escludere che la dotazione articolatoria di /l:/, prima degeminata e successivamente retroflessa nel repertorio fonetico dialettale, possa aver prodotto qualche effetto sulla realizzazione del fonema nell'italiano locale. Condizioni simili potrebbero

<sup>4</sup> La notazione [·] viene impiegata, qui e di seguito, in relazione a consonanti degeminate, non tanto come indicazione di durata foneticamente equivalente a un tempo *half-long*, quanto piuttosto di degeminazione che ha ridotto la quantità del fonema [:] (per i dettagli cfr. Barbera in stampa a).

averlo esposto ad una degeminazione massiva oltre che ad una sporadica attribuzione ipercorrettistica del tratto di lunghezza (conservazione della geminata), deprivandolo parallelamente della pronuncia retroflessa più marcata ed incompatibile con tratto [-lungo] così fortemente caratterizzante.

Per quanto concerne i fonemi /s/ e /s:/, che costituiscono il *focus* della presente indagine, evidenziamo innanzi tutto due aspetti. Il primo – e più macroscopico – è l’elevatissima occorrenza di allofoni che, alla luce della nostra analisi arriveremo a qualificare come retroflessi (cfr. *ultra*); il secondo, in controtendenza rispetto a quanto osservato per /l:/, corrisponde alla bassa incidenza di degeminazione rispettivamente sul fonema /s:/ (cfr. fig. 5), cui fa da contrappunto lo sviluppo massiccio delle varianti allofoniche che ipotizziamo di poter qualificare come retroflesse.

| /s/       |           | /s:/       |           |            |           | parlante     |
|-----------|-----------|------------|-----------|------------|-----------|--------------|
| [s]       | [ʂ]       | [s:]       | [sʰ]      | [ʂ:]       | [ʂʰ]      |              |
| 70        | 93        | 106        | 95        | 157        | 57        | ATm          |
| 93        | 60        | 146        | 102       | 108        | -         | GAf          |
| 89        | 99        | 144        | -         | 113        | -         | GDf          |
| 131       | 100       | 110        | -         | -          | -         | GVm          |
| -         | 105       | -          | -         | -          | -         | RRm          |
| 91        | 136       | 167        | 120       | ..         | ..        | TBf          |
| <b>95</b> | <b>99</b> | <b>135</b> | <b>99</b> | <b>126</b> | <b>57</b> | <b>media</b> |
| 20        | 22        | 23         | 11        | 22         | 0         | $\sigma$     |

Tabella 2: Valori in ms e deviazione *standard* relativi ai fonemi /s/ e /s:/.

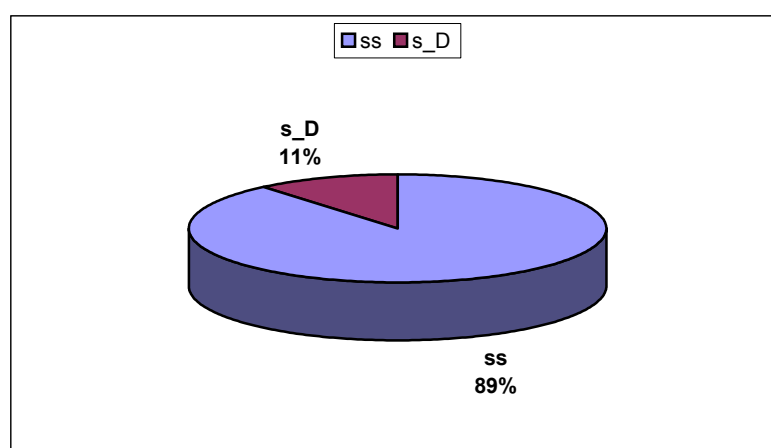


Figura 5: Valori percentuali della distribuzione allofonica di /s/ nelle sue varianti geminata (ss) e degeminata (s\_D).

Questi dati sembrano fotografare una situazione speculare rispetto a quella rilevata per /l:/, poiché, nel caso di /s:/, la geminazione parrebbe bloccata dalla articolazione retroflessa, verisimilmente favorita dalla durata intrinsecamente più lunga di [s] e [s:]. A ciò andrà

aggiunto, con tutta probabilità, l'influsso non trascurabile esercitato delle pronunce palatalizzate di origine settentrionale<sup>5</sup> che potrebbero aver funzionato da rinforzo allo sviluppo di un allofono eteroarticolato del fonema /s/.

In definitiva, sembra potersi rintracciare nella particolare esposizione delle consonanti laterali, sibilanti ed occlusive alveolari (cfr. Barbera in stampa b) a fenomeni di alterazione – in ordine alla quantità consonantica *in primis*, ma anche alle caratteristiche articolatorie – una traccia delle cosiddetta “sindrome delle coronali” (*Coronals Syndrome*: Kenstowicz, 1994: 516 sgg.).

La coronalità si caratterizzerebbe infatti, secondo Kenstowicz, per un particolare *status* di non marcatezza, che potrebbe giustificare diversi fenomeni fonologici cui i segmenti coronali risultano esposti nelle lingue del mondo. Più precisamente:

- le coronali sono presenti a livello cross-linguistico con una elevata frequenza statistica;
- sono l'esito più naturale di processi di neutralizzazione articolatoria;
- sono impiegate diffusamente in funzione epentetica;
- si combinano liberamente in *cluster* consonantici;
- sono ampiamente soggette ad assimilazione del punto di articolazione;
- risultano infine «more likely to be transparent to transconsonantal vowel-echo than labials and velars» (Kenstowicz, 1994: 516-517).

Poste queste premesse e considerando che i fonemi coronali manifestano in alcune varietà linguistiche una tendenza elevata allo sviluppo di allofoni «since phonemes with a coronal place of articulation are the ones most frequently subject to lenition, according to the so-called coronal syndrome» (Marotta e Barth, 2005: 407), non ci sorprende che le coronali (/l/, /s/, /t/) costituiscano, anche nell'italiano di Carrara (e ancor prima nel dialetto), un modulo particolarmente ricettivo a processi di alterazione del modo/luogo di articolazione – ciò che rappresenta l'interesse primario in questa sede – oltre che della quantità.

## 6. VALORI DEL CENTRE OF GRAVITY

L'impiego del Centro di Gravità (COG) come parametro utile a identificare le caratteristiche articolatorie delle fricative funziona sostanzialmente come indicatore della frequenza media alla quale si colloca la massima concentrazione dell'energia spettrale ovvero «a measure for how high the frequencies in a spectrum are on average» (Boersma, 2007, s.v. *Spectrum: get centre of gravity*). Oltre a ciò, precisano Hamann e Avelino (2007: 950):

«The centre of gravity (or: spectral mean) correlates with the size of the front cavity, and is lower for postalveolar articulations and higher for alveolar ones».

I dati in nostro possesso indicano un andamento piuttosto nitido del centro di gravità per gli allofoni impressionisticamente percepiti come palatalizzati. Sia per i maschi che per le

---

<sup>5</sup> In modo particolare l'esposizione ai contatti con l'area emiliana è indubitabile; cfr. Rizzi (1986), Canepari e Vitali (1995), Hajek (1997).



femmine i valori risultano tendenti verso quelli delle prepalatali e si attestano su valori intermedi rispetto a quelli della fricativa prepalatale /ʃ:/ e delle alveolari /s s:/<sup>6</sup>.

|                | maschi |      |      |       |      | femmine |      |      |       |      |
|----------------|--------|------|------|-------|------|---------|------|------|-------|------|
| fono           | ATm    | GVm  | RRm  | media | σ    | GAf     | GDf  | TBf  | media | σ    |
| ʃ              | 3405   | 3113 | 2574 | 3031  | 422  | -       | 2482 | -    | 2482  | -    |
| ʃ:             | 3707   | 2552 | 2976 | 3078  | 584  | 5150    | 4928 | 4660 | 4794  | 245  |
| s              | 4479   | 2557 | -    | 3518  | 1359 | 5404    | 4962 | 5760 | 5361  | 400  |
| s:             | 5116   | 4572 | -    | 4844  | 385  | 5579    | 2535 | 6123 | 4329  | 1934 |
| z              | 1507   | 1240 | -    | 1374  | 189  | 1943    | 1643 | 2218 | 1930  | 288  |
| ʒ              | 4599   | 2707 | 2912 | 3406  | 1038 | 3991    | 1846 | 5667 | 3756  | 1915 |
| ʒ:             | 4681   | 3179 | 2793 | 3551  | 997  | -       | -    | -    | -     | -    |
| ʒ <sub>l</sub> | 1341   | 1349 | 1141 | 1277  | 118  | -       | 1533 | 1593 | 1563  | 42   |

Tabella 3: Valori di COG distinti per singoli locutori; valori medi e deviazione *standard* per gruppo maschile e gruppo femminile.

Significativo per i maschi (cfr. tab. 3) è il *gradatum* [ʃ] < [ʒ] < [s] che indica per [ʒ] un punto di articolazione intermedio (Gordon *et al.*, 2002: 161). Le nostre risultanze paiono dunque convergere con il quadro d'insieme delineato da Sottofattori (2004) in relazione al repertorio fonetico della Val di Magra, zona non a caso limitrofa e strettamente integrata dal punto di vista dialettologico all'Apuania di confine:

«(...) il valore medio riscontrato in questa ricerca è di 3321 Hz. Valori simili si hanno per la retroflessa descritta in una lingua dravidica, la lingua Toda, studiata da Shalev (Shalev *at alii*, 1994), nella quale si nota che il picco di frequenza è di 3000 Hz». (Sottofattori, 2004: 198).

Quanto al *trend* frequenziale complessivo e alla tendenza dei valori femminili di COG verso *target* più elevati, valgono le osservazioni di Cox (2005)<sup>7</sup>:

«/s/ has its lower boundary of noise at about 4kHz peaking at about 5.5kHz with the main antiresonance at about 3kHz. The spectral centre of gravity is higher for females than males. For the palatoalveolars (e.g. /ʃ/) – the main resonance occurs at a lower frequency due to the longer cavity in front of the constriction which is accentuated by the lip protrusion. The main resonance occurs at about 2.5kHz with an antiresonance at about 1kHz. Females have higher frequencies».

<sup>6</sup> Riteniamo che per /s/ il confronto più plausibile sia quello tra l'allofono [ʒ] dei maschi e la realizzazione femminile dell'alveolare [s], visto che le occorrenze sono quantitativamente paragonabili; per /ʃ(:)/ la comparazione può valersi dei i valori sia maschili, sia femminili.

<sup>7</sup> Cfr. <http://www.ling.mq.edu.au/speech/acoustics/consonants/fricweb.html>.

Resta da sottolineare come la letteratura testimoni, pure all'interno di un *pattern* interpretativo del parametro COG piuttosto solido, la persistenza di un margine di variabilità interindividuale non trascurabile, per cui la proporzione paradigmatica tra i valori di COG associati alla produzione dei singoli suoni fricativi può manifestare disomogeneità in relazione al *ranking* da locutore a locutore, anche per una stessa varietà linguistica (cfr. Jones e Nolan, 2007: 875; Hamann e Avelino, 2007: 951). Si veda a proposito la tab. 3 che evidenzia questo aspetto sia per il gruppo di informatori maschili, sia per quelli femminili.

## 7. VALORI FORMANTICI

Allo scopo di analizzare i connotati articolatori degli allofoni sibilanti, abbiamo utilizzato i valori formantici rilevati sulla transizione tra il segmento vocalico precedente e il segmento fricativo di nostro interesse. Nello specifico si è considerato l'*offset* della vocale corrispondente agli ultimi 23 ms della vocale stessa ritenuti pertinenti per valutare il punto di articolazione (Gordon *et al.*, 2002: 143).

Come già si è osservato a proposito del COG, viste le forti sperequazioni statistiche nella distribuzione dell'allofono presumibilmente retroflesso rispetto alla variante fonematica alveolare, abbiamo ritenuto di dover sviluppare il confronto tra [ʃ ʃ:] prodotte dai locutori maschili da un lato e [s s:] prodotte dalle informatrici dall'altro. Questa scelta mira ad attutire effetti di rumore nell'interpretazione dei dati, poiché le poche occorrenze maschili di [s s:] non risultano né statisticamente rilevanti, né, a nostro avviso, del tutto attendibili. L'allofono alveolare sembra infatti detenere per i maschi uno statuto marcato<sup>8</sup>, dunque articolatoriamente meno stabile; viceversa le femmine risultano in larga misura esenti dall'impiego della variante retroflessa.

| MASCHI    | [ʃ]  |      |      | media       | σ   |
|-----------|------|------|------|-------------|-----|
|           | ATm  | GVm  | RRm  |             |     |
| <b>F2</b> | 1836 | 1843 | 1645 | <b>1775</b> | 112 |
| <b>F3</b> | 2762 | 2877 | 2699 | <b>2779</b> | 90  |
| <b>F4</b> | 3811 | 4107 | 3601 | <b>3840</b> | 254 |
|           | [s:] |      |      |             |     |
| <b>F2</b> | 2016 | 1909 | 1851 | <b>1925</b> | 84  |
| <b>F3</b> | 3017 | 2935 | 2784 | <b>2912</b> | 118 |
| <b>F4</b> | 4106 | 4239 | 3729 | <b>4025</b> | 265 |

Tabella 4: Valori di F2, F3, F4 in transizione tra la vocale precedente e i segmenti [ʃ] e [s:].

<sup>8</sup> “Marcato” vs “naturale” vale qui non nell’accezione universalistica di matrice generativista; considerando la realizzazione allofonica [ʃ ʃ:] di /s s:/ come più naturale, in quanto normalmente prodotta dai parlanti carraresi di sesso maschile, si vuole alludere al dato peculiare – verosimilmente di matrice sociolinguistica – per cui i maschi riservano solo sporadicamente l’eventuale pronuncia [s s:] a contesti diafasicamente elevati e non quotidiani.

| FEMMINE   | [s]        |            |            | media       | $\sigma$ |
|-----------|------------|------------|------------|-------------|----------|
|           | <i>GAf</i> | <i>GDf</i> | <i>TBf</i> |             |          |
| <b>F2</b> | 1728       | 1732       | 1543       | <b>1668</b> | 108      |
| <b>F3</b> | 2793       | 2814       | 2785       | <b>2797</b> | 15       |
| <b>F4</b> | 4405       | 4324       | 3917       | <b>4215</b> | 262      |
|           | [s:]       |            |            |             |          |
|           |            |            |            |             |          |
| <b>F2</b> | 1765       | 1574       | 2080       | <b>1806</b> | 256      |
| <b>F3</b> | 2773       | 2890       | 3167       | <b>2943</b> | 202      |
| <b>F4</b> | 4079       | 4029       | 4239       | <b>4116</b> | 110      |

Tabella 5: Valori di F2, F3, F4 in transizione tra la vocale precedente e i segmenti [s] e [s:].

L'esame dei dati proposti nelle tabelle 4 e 5 consente di formulare due osservazioni essenziali:

1. si registra una flessione piuttosto sensibile dei valori di F4 sia per il fonema scempio sia per quello geminato, mentre lo scarto, per quanto concerne i valori di F3, risulta minimo, sia pure indicativo di un tendenziale abbassamento della transizione formantica;
2. per [s s:] F2 mantiene ovvero innalza (di circa 100 Hz) i valori rispetto a [s s:].

L'evidenza segnalata al punto 1 sembra corrispondere ad una *retraction* della radice della lingua senz'altro compatibile con un gesto articolatorio retroflesso (Stevens & Blumstein, 1975; Dart & Nihalani, 1999; Celata, 2006); parallelamente il comportamento di F2 indica un *locus* articolatorio complessivamente anteriore, prossimo a quello della sibilante alveolare, eventualmente accompagnato da retroflessione dell'apice della lingua (cfr. dati di [s] svedese in De Sio & Romano, 2003: 323-324). Del resto è verificato (cfr. Celata, 2006: 12) che i valori di F2 ed F3 detengono un peso relativo nella precisa definizione del punto di articolazione dei foni retroflessi nel caso in cui si tratti di consonanti coronali, «infatti le retroflesse con articolazione non posteriore (es. alveolari) presentano normalmente una minor flessione di F3, rispetto alle retroflesse articolate più indietro». In questa prospettiva, se il fono è anteriore, la posizione del corpo della lingua non manifesterebbe sensibili differenze nel caso di articolazione retroflessa vs non retroflessa, poiché la lingua stessa risulta inevitabilmente sospinta in avanti, mentre il gesto retroflesso si svilupperebbe nella parte apicale della lingua. Si assume dunque che l'indice di maggior interesse sia rappresentato da F4.

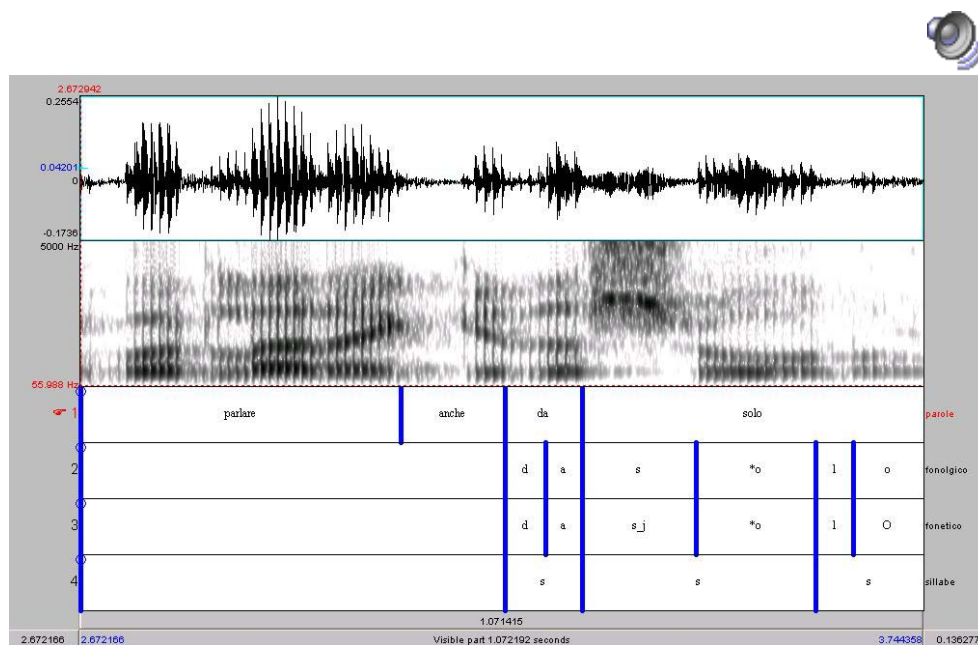


Figura 6: Esempio di sibilante retroflessa prodotta da locutore maschile (RRm): oscillogramma, spettrogramma e annotazione in *textgrid*.

L'assetto articolatorio così definito verrebbe pertanto a giustificare una sorta di effetto "fischio" frequentemente percepibile nell'esecuzione dei maschi (cfr. *file* sonoro associato alla fig. 6).

## 8. UN'IPOTESI ARTICOLATORIA

### 8.1 La fisionomia spettrografica

Partendo da valutazioni puramente acustiche e spettroacustiche, il caso peculiare delle sibilanti prodotte dai parlanti maschili carraresi presenta alcuni aspetti di peculiarità, specie in un'ottica contrastiva rispetto alle realizzazioni allofoniche femminili. Un primo dato è costituito dalla evidente intensità associata ad una durata tendenzialmente lunga anche in produzione corrispondente a segmento fonologicamente scempio (cfr. dati in tab. 2 *supra*). Percettivamente, infatti, ma anche spettrograficamente, [ʃ ʒ:] non sono di fatto mai scempie, né tantomeno degeminate, nonostante il repertorio fonetico carrarese sia diffusamente esposto a fenomeni di alterazione della quantità consonantica, soprattutto in senso degeminante (cfr. Barbera in stampa a e b). Per converso, la sibilante prodotta dalle femmine presenta una morfologia spettrografica diversa da quella dei maschi mediamente corrispondente ad un segmento:

1. tendenzialmente meno lungo;
2. meno intenso;
3. in alcuni casi molto tenue o appena percettibile.

Di seguito si propone un campione spettrografico relativo alla produzione femminile di /s/ (fig. 7).

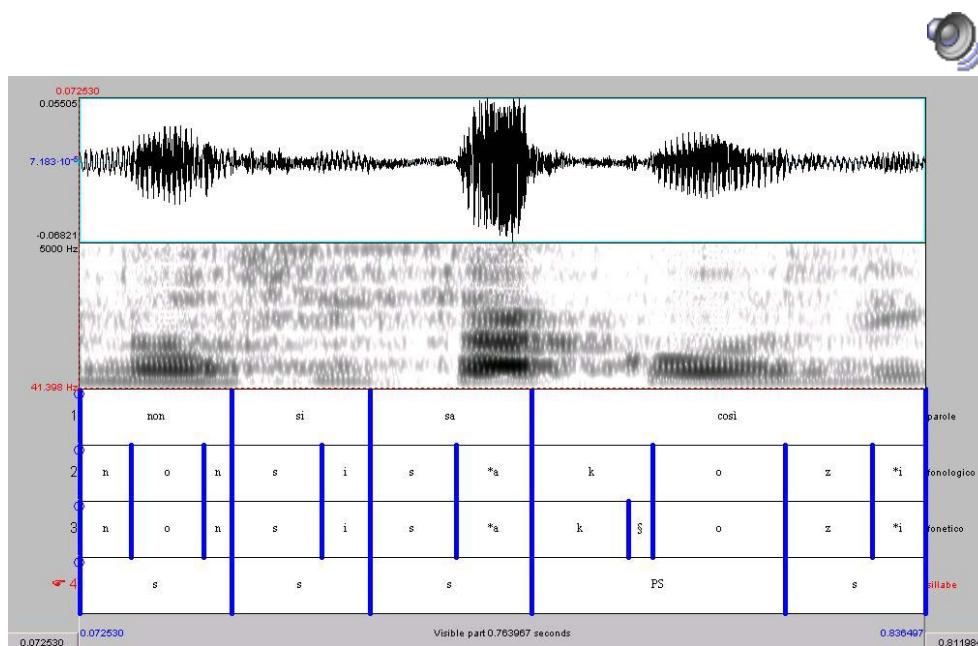


Figura 7: Esempio di sibilanti alveolari prodotte da locutrice femminile (GDf): oscillogramma, spettrogramma e annotazione in *textgrid*.

Le specificità sopra indicate sembrano confermare l'impressione uditiva di una sottile marcatezza rispetto alle sibilanti *standard* dell'italiano, salvo non trovare conferma palese nelle risultanze dell'analisi parametrica che indicano un'articolazione senz'altro alveolare.

Resta eventualmente da considerare il dato secondo cui anche per le femmine – sociolinguisticamente innovatrici nel senso del prestigio? – il gesto articolatorio associato alla produzione di /s/ possa mantenere traccia di uno statuto peculiare delle sibilanti, in quanto classe sovraesposta a varianza allofonica per le ragioni sopra delineate.

## 8.2 I tratti articolatori

Per poter identificare, sia pure indirettamente, i tratti specifici dell'articolazione dei suoni [ʃ ʒ:] presi in esame, le risultanze desunte dall'analisi delle transizioni formatiche e del COG andranno integrati con i dati relativi alla durata. Un primo aspetto interessante, sembra corroborare l'ipotesi di retroflessione: la durata media 99-140 ms rispettivamente per l'allofono scempio e geminato (cfr. *supra* tab. 2) è senz'altro compatibile con un gesto articolatorio complesso, che richiede una maggiore durata del segmento. Come precisa Celata (2006: 78)

«(...) per conservare il carattere apicale e postalveolare che lo caratterizza, il gesto articolatorio di una retroflessa ha bisogno di uno spazio temporale dell'ampiezza di una consonante geminata; nel momento in cui i fenomeni del parlato spontaneo e del *fast speech* spingono verso la riduzione della durata della consonante, questa tende irrimediabilmente alla semplificazione articolatoria e viene pronunciata come semplice alveodentale».

Di fatto all'interno del nostro *corpus* i fenomeni di ipoarticolazione imputabili al *fast speech* tendenzialmente caratterizzante l'eloquio spontaneo non intaccano la connotazione articolatoria retroflessa, né l'impressione di marcatezza percettiva ad essa correlata, proprio perché la lunghezza di [ʃ ʒ:] non risulta sostanzialmente ridotta in alcun contesto.

Sulla base dei dati disponibili, tenendo conto dei valori estratti in relazione ai parametri considerati, sembra potersi delineare un'ipotesi articolatoria funzionale all'*output* fonetico corrispondente all'allofono sibilante basicamente maschile fin qui esaminato. Schematizzando si individuano i seguenti tratti:

- + PROTRUSIONE: aumento del volume d'aria anteriore al punto di costrizione; il dato è confermato dai valori di *gravity centre* (cfr. tab. 3);
- + LIMITATA *RETRACTION*: flessione di F4 accompagnata da una flessione non significativa di F3; il *locus* articolatorio è sostanzialmente anteriore;
- + COSTRIZIONE APICALE: la punta della lingua risulta sollevata e leggermente retroflessa dando luogo ad una costrizione nella regione alveolare o, più frequentemente, post-alveolare (cfr. Ladefoged e Maddieson, 1996: 25). Un simile assetto del gesto linguale comporta peraltro «the creation of an additional volume on the underside of the tongue as a result of the curled and raised tongue blade» (Stevens & Blumstein, 1975: 218)

Rispetto alla morfologia dei gesti articolatori normalmente riconosciuti come tipici della sibilante retroflessa, si propone (cfr. fig. 8), per la [ʃ] del carrarese una rappresentazione

parzialmente differente, come indicato dalle linee in tratteggio che modificano l'andamento del *setting* individuato da Hamann (2002: 107) per la lingua Toda:

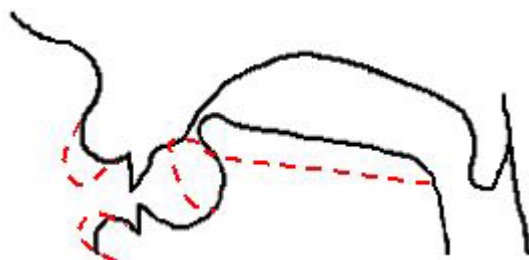


Figura 8: A sinistra: adattato da Hamann (2002: 107); in tratteggio il gesto articolatorio ipotizzato per il fono [s] < /s/ del carrarese.

Per visualizzare in una prospettiva sinottica i diversi punti articolatori discussi per le sibilanti, proponiamo di seguito il diagramma realizzato da Ladefoged e Maddieson (1996: 160) ancora per la lingua Toda; si deve ipotizzare, per la variante carrarese, una posizione ovviamente più avanzata rispetto a quella corrispondente a [s]:

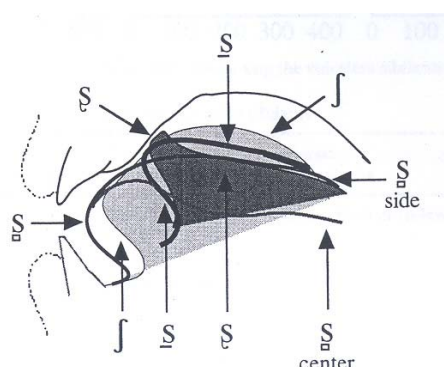


Figura 9: Diagramma composito raffigurante la posizione della lingua nella produzione delle diverse sibilanti in lingua Toda. Da Ladefoged e Maddieson (1996: 160):

Dobbiamo dunque postulare per [ʂ] un *setting* articolatorio peculiare in cui il corpo della lingua assume una conformazione solcata mantenendo una posizione, tipica di /s/, relativamente avanzata (o, viceversa, modestamente ritratta); la lamina risulta protesa verso la zona alveolare e la punta della lingua retroflessa, il che produce una costrizione che si sviluppa su un'area ristretta rispetto all'asse longitudinale mediano dell'apparato vocale (cfr. Nespø, 1993: 59) e definisce un modo articolatorio qualificabile, in termini chomskiani (Chomsky e Halle 1968: 312), con il tratto [-distribuito].

Del resto la compatibilità tra anteriorità/coronalità e retroflessione è stata verificata in diversi repertori fonetici; benché «the placement of the tongue tip and blade to form a coronal constriction is facilitated if the the tongue body moves cooperatively since the

tongue tip and blade ride on the tongue body» (Flemming, 2003: 338), in maniera meno economica è pur sempre possibile dare luogo ad un gesto articolatorio retroflesso – tale sembra il caso delle nostre [ʂ] – che si estrinseca a partire dalla lamina linguale e si realizza sull’apice accompagnandosi con una dinamica procheila che aggiunge volume di frizione frontale.

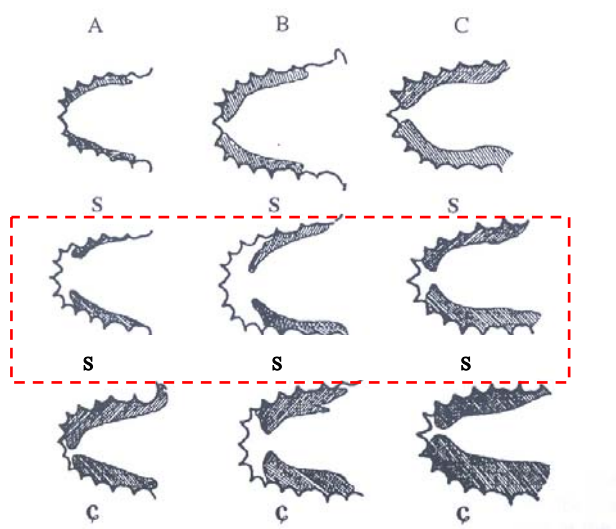


Figura 10: Palatogrammi di sibilanti prodotte da tre locutori di lingua cinese *standard*; da Ladefoged e Maddieson (1996: 152). Al gesto articolatorio carrarese dovrebbe corrispondere sommariamente il palatogramma riquadrato in tratteggio.

Andrà precisato da ultimo che Ladefoged e Maddieson (1996: 25) proponevano di mantenere una notazione a sé stante, [ʂ], per questo tipo di articolazione retroflessa apicale (cfr. palatogrammi in fig. 10) «with a less retroflexed tongue shape», volendone rimarcare la specificità rispetto all’articolazione subapicale (tipica, ad esempio, della lingua Toda) che l’IPA associa al simbolo [ʂ] estendendone di fatto l’impiego anche al caso della variante da noi individuata, corrispondente ad un gesto articolatorio «with a less retroflexed tongue shape» (*ibidem*).



## 9. CONCLUSIONI

La peculiarità percettiva dell'allofono carrarese della sibilante alveolare viene dunque confermata dall'analisi acustica. In prima approssimazione avevamo segnalato che l'*output* percettivo risultava tendente verso una pronuncia di tipo palatalizzato, con tutta probabilità indotta dalla posizione di isoglossa del territorio apuano, in diretta prossimità con l'area gallo-italica e più specificamente emiliana.

La nostra analisi ha evidenziato dal punto di vista acustico alcuni aspetti specifici della realizzazione che abbiamo notato come [ʃ ʂ:], fin dalla semplice osservazione della morfologia spettrografica.

L'indagine parametrica ha poi quantificato e dettagliato l'ipotesi di un preciso *setting* articolatorio qualificabile come retroflesso.

I valori del COG e i valori formatici in transizioni tra i segmenti vocalici precedenti e i segmenti fricativi in esame indicano infatti per [ʃ ʂ:] una distribuzione dell'energia di frizione intermedia tra [s s:] e [ʃ:] e un *locus* articolatorio senz'altro anteriore; per converso l'assenza di palatalizzazione resta testimoniata dai valori medi di F2 che non lasciano intendere un apprezzabile arretramento del punto di articolazione. I valori di durata dei segmenti considerati convergono poi sull'ipotesi di un gesto articolatorio complesso, indirettamente qualificabile come retroflesso. La conformazione orale associata alla produzione di foni fricativi (post-)alveolari retroflessi riesce, a nostro avviso, a soddisfare i requisiti acustici caratterizzanti l'*output* percettivo evidenziato.

Nonostante il nostro *corpus* non abbia un respiro tale da consentire estensioni di ordine sociolinguistico, i dati analizzati seguono comunque una *ratio* distributiva piuttosto omogenea; alla luce della nostra analisi è stato infatti possibile riconoscere una marcata differenziazione della distribuzione allofonica di /s/ che attribuisce ai maschi una assoluta prevalenza nell'utilizzo della variante retroflessa.

Alle conclusioni sopra esposte sospinge anche la suggestione diacronica che riconosce nel territorio apuano, ancora con Ambrosi (1956), un'«area fonetica cacuminale», incline ad implementare nel proprio repertorio fonetico suoni appunto cacuminali, ovvero retroflessi. Il confronto contrastivo dei comportamenti allofonici assunti da /l:/ sembra poi rintracciare nella cosiddetta *sindrome delle coronali* un possibile *background* teorico funzionale.

## BIBLIOGRAFIA

- Ambrosi, A.C. (1956), *Osservazioni sugli attuali limiti dell'area fonetica cacuminale nelle Alpi Apuane*. «Giornale storico della Lunigiana», a. VII, n. 1-2: 5-25.
- Barbera, M. (in stampa a), *Toscani a metà. Degeminazione e geminazione nell'italiano di Carrara*. Atti del IV Dies Romanicus Turicensis, Zürich, 17-18 novembre 2007 (in corso di pubblicazione).
- Barbera, M. (in stampa b), *Terre di mezzo. Osservazioni acustiche e fonologiche sulla durata consonantica nell'italiano di Carrara*. «L'Italia Dialettale», LXVIII.
- Behrens, S. J., Blumstein, S. E. (1988), *Acoustic characteristics of English voiceless fricatives: a descriptive analysis*. «Journal of Phonetics», 16: 295-298.

- Boersma, P. (2007), *Praat (version 5.0.20). Manual*. Pagina web <http://www.fon.hum.uva.nl/praat/>.
- Canepari, L. e Vitali, D. (1995), *Pronuncia e grafia del bolognese*. «Rivista Italiana di Dialettologia», 19: 119-163.
- Celata, C. (2006), *Analisi dei processi di retroflessione delle liquide in area romanza, con dati sperimentali dal corso e dal siciliano*. Tesi di Perfezionamento in Linguistica, SNS di Pisa.
- Chomsky, N. e Halle, M. (1968), *The sound pattern of English*. New York, Harper and Row.
- Dart, S. e Nihalani, P. (1999), *The articulation of Malayalam coronal stops and nasals*. «Journal of the International Phonetic Association», 29:129-142.
- De Sio, L. e Romano, A. (2003), *Caratteristiche acustiche delle fricative dello svedese: uno studio delle transizioni formantiche in contesto VCV*. Atti delle XIII Giornate del GFS, Pisa, 28-30 novembre 2002.
- Flemming, E. (2003), *The relationship between coronal place and vowel backness*. «Phonology», 20: 335-373.
- Giannelli L. e Savoia L.M. (1978), *L'indebolimento consonantico in Toscana I*. «Rivista Italiana di Dialettologia», 2: 23-58.
- Giannelli L. e Savoia L.M. (1979-80), *L'indebolimento consonantico in Toscana II*. «Rivista Italiana di Dialettologia», 4: 38-101.
- Gordon, M., Barthmaier, P., Sands, K. (2002), *A cross-linguistic acoustic study of voiceless fricatives*. «Journal of the International Phonetic Association», 32: 141-174.
- Hajek, R. (1997), *Emilia-Romagna*. In Maiden, M. e Parry, M. (a cura di), *The dialects of Italy*. London, Routledge: 271-278.
- Hamann, S. (2002), *Retroflex fricatives in Slavic languages*. «Journal of the International Phonetic Association», 34: 53-67.
- Hamann, S. e Avelino, H. (2007), *An acoustic study of plain and palatalized sibilants in Ocotepéc Mixe*. «Proceedings of XVI ICPhS», Saarbrücken, 2007: 949-952.
- Jones, M.J. e Nolan, F.J. (2007), *An acoustic study of north welsh voiceless fricatives*. «Proceedings of XVI ICPhS», Saarbrücken, 2007: 873-876.
- Jongman, A., Wayland, R., Wong, S. (2000), *Acoustic characteristics of English fricatives*. «Journal of the Acoustical Society of America», 108: 1252-1263.
- Kenstowicz, M. (1994), *Phonology in Generative Grammar*. Cambridge MA, Blackwell.
- Ladefoged, P. e Maddieson, I. (1996), *The sounds of the world's languages*. Cambridge, Mass. e Oxford, Blackwell.
- Luciani, L. (1999), *Il dialetto carrarese. Suoni, forme, costrutti, parole*. Carrara, Aldus Casa di Edizioni.
- Maffei Bellucci, P. (1977), *Lunigiana*. Pisa, Pacini editore.

- Marotta, G. e Barth, M. (2005), *Acoustic and Sociolinguistic Aspects of Lenition in Liverpool English*. «SLIFO», 3.2: 377-414.
- Marotta, G. e Sorianello, P. (in stampa), *SAIL2: un sistema di annotazione fonetica integrato per l'italiano L2*. In Andorno, C. (a cura di), *Corpora di italiano L2: tecnologie, metodi, spunti teorici*. Pavia, 22-23 novembre 2007, Guerra Editore, Perugia.
- Rizzi, E. (1986), *Variabili consonantiche nell'italiano regionale di Bologna*. «Rivista Italiana di Dialettologia», 10: 111-127
- Shadle, C.H., Mair, S.J. (1996), *Quantifying spectral characteristics of fricatives*. «Proceedings of the Fourth International Conference on Spoken Language» (ICSLP).
- Nespor, M. (1993), *Fonologia*. Bologna, Il Mulino.
- Rohlf, G. (1966), *Grammatica storica della lingua italiana e dei suoi dialetti. Fonetica*. Torino, Einaudi.
- Savoia, L.M. (1980), *Fonologia delle varietà apuane e garfagnine: consonantismo*. «Studi Urbinati di Storia, Filosofia».
- Shadle, C.H., Mair, S.J. (1996), *Quantifying spectral characteristics of fricatives*. «Proceedings of the Fourth International Conference on Spoken Language» (ICSLP).
- Sorianello, P. (2002), *I suoni fricativi dell'italiano fiorentino*. «Quaderni del Laboratorio di Linguistica», 3: 26-39.
- Sottofattori, S. (2004), *Le sibilanti scibilanti della Val di Magra*. Atti delle XIV Giornate del GFS, Viterbo, 4-6 dicembre 2003.
- Stevens, K. e Blumstein, S. (1975), *Quantal aspects of consonant production and perception: a study of retroflex stop consonants*. «Journal of Phonetics», 3: 215-233.
- Stevens, P. (1960), *Spectra of fricative noise in human speech*. «Language and Speech», 3: 32-49.

# TRASCRIZIONE DI SEGNALI RUMOROSI

Andrea Paoloni\*, Marco Medda°, Adriano Soccorsi°

\*Fondazione Ugo Bordon - °Università "Sapienza" Roma

pao@fub.it

## SOMMARIO

Qual è la soglia di intelligibilità sotto la quale un'intercettazione audio dovrebbe essere dichiarata dal perito trascrittore "non trascrivibile"? Ma soprattutto, esiste una soglia di questo tipo? E nel caso in cui non esista, non sarebbe opportuno delinearne una?

### 1. PREMESSA

In numerose applicazioni viene richiesto di trasporre il segnale vocale registrato nel corrispondente testo scritto. Questo cambio di codice presenta alcune difficoltà di natura pratica e di natura teorica. Queste difficoltà sono anche dipendenti dalla natura del segnale da trascrivere e dall'ambito applicativo.

Due sono le aree in cui viene prevalentemente richiesta la trascrizione: la resocontazione di assemblee e dibattiti e la trascrizione di comunicazioni intercettate.

Nel primo caso il problema principale è quello di rendere in forma scritta il parlato con i suoi anacoluti, le sue ripetizioni, le sue incertezze lessicali e grammaticali rispettando nel contempo il modo di esprimersi del parlatore, mentre nel secondo caso il problema è comprendere tutte le parole e trascriverle senza alcuna alterazione. Gli ostacoli a una completa comprensione o meglio ad una completa intelligibilità, sono la lingua o dialetto dei parlanti, l'assenza dei riferimenti visivi e situazionali e la qualità insufficiente del segnale in termini di rapporto segnale/rumore (d'ora in poi S/N), banda passante e distorsione.

Obiettivo del presente lavoro è contribuire a valutare la probabilità di trascrivere correttamente segnali aventi caratteristiche analoghe a quelli riscontrabili in casi reali di intercettazione con particolare riferimento alle intercettazioni ambientali.

### 2. CONSIDERAZIONI SULLA PERCEZIONE DEL PARLATO

La corretta comprensione ed interpretazione di un qualunque tipo di enunciato linguistico prevede una modalità di ricostruzione del significato fortemente vincolata dall'integrazione di dati, memorie, o conoscenze pregresse e/o presupposte, che non sono necessariamente riconducibili al contenuto proposizionale dell'enunciato in questione.

#### 2.1 La comprensione come processo lineare

Viene qui illustrato un modello di tipo lineare a *tappe* (ispirato a De Mauro, 1998) particolarmente adatto alla discussione e all'interpretazione dei risultati ottenuti tramite l'esperimento riportato nelle seguenti pagine. Il modello in questione è stato rivisitato ed ampliato, anche attraverso una sintesi del lavoro di diversi linguisti.

Le *tappe* della comprensione da parte dell'ascoltatore, in una situazione idealizzata, sono quelle che seguono:

1. *individuazione sonora*: fase della percezione fisiologica del suono;
2. *individuazione semiotica*: capire che si tratta di segnale semiotico piuttosto che di rumore di fondo;

3. *individuazione “linguaggistica”*: comprendere che si tratta del linguaggio umano e non di altro codice semiotico;
4. *individuazione linguistica*: comprendere a quale lingua o dialetto appartiene l'individuato;
5. *individuazione fonetico/fonologica*: effettuare una mappatura del parlato sul sistema fonologico della lingua, individuando quindi i fonemi, le varianti obbligatorie e quelle libere, i confini di monemi, i fenomeni sovra segmentali etc.;
6. *individuazione lessicale*: capire i significati dei monemi e le loro eventuali valenze semantiche;
7. *individuazione della struttura frasale*: interpretare in modo corretto la combinazione dei monemi lessicali nella frase tramite regole morfologiche e/o sintattiche;
8. *individuazione del significato astratto della frase completa*, cioè del suo valore proposizionale;
9. *individuazione dei referenti concreti*: capire a quale realtà si riferisce il significato astratto della frase;
10. *individuazione pragmatica*: valore illocutivo, atto linguistico;
11. *individuazione delle “implicature” del discorso*: cosa bisogna sapere già per capire a cosa allude il senso della frase;
12. *individuazione del valore informativo*: riconoscere cosa aggiunge la frase a quello che sappiamo già (tramite categorie comunicative come tema/rema e dato/nuovo).

L'evidente problema del trascrivere una registrazione di pessima qualità è che non ci si trova chiaramente, per una serie di motivi che prenderemo in esame, in un contesto comunicativo ideale, per cui non tutte le tappe della comprensione normale del discorso vengono completate. Vediamo pertanto quali sono i fattori che giocano un ruolo fondamentale nei processi di mancata o errata comprensione.

## 2.2 *Quando la comprensione non avviene*

Nel caso in cui l'ascoltatore fallisca nell'identificare una delle *tappe*, la comprensione e la comunicazione saranno inevitabilmente compromesse. In questo caso si procederà effettuando delle ipotesi sulle *tappe* che risultano ignote, operando “a cascata” e ripartendo, quindi, da quelle a monte.

Tali ipotesi sono altresì indotte da ulteriori fattori, quali i seguenti:

- extralinguistici (mimica, prossemica etc.): data la struttura dell'esperimento, nel caso in esame non sussistono;
- frames* comportamentali (cosa ci si aspetta che si dica nella situazione corrente<sup>1</sup>): sussistono solo qualora la traccia in esame generi un proprio contesto linguistico intrinseco. Nell'es.: “Prima giri a (destra ?) poi a destra e poi di nuovo a (sinistra ?)” (cfr. Allegato 1 etc); l'ascoltatore ha ipotizzato, dopo un primo ascolto nel quale ha percepito chiaramente il senso generale dell'enunciato, la presenza di ulteriori riferimenti ad indicazioni stradali, evidente tema della traccia.
- bagaglio nozionale (cosa si sa sull'argomento tanto da poter intuire ciò che il nostro interlocutore può dirne): data la struttura dell'esperimento, nel caso in esame non sussiste;

---

<sup>1</sup> Sulla teoria dei *frames* vedi C. Fillmore, 1976.

- *massime conversazionali* (si presuppone che l'interlocutore sia sano di mente, che parli la nostra lingua, dica cose che sono effettivamente utili da dire, ossia che comunichino qualcosa, ecc.): giocano un ruolo di grande importanza, applicate o meno.

Poiché anche le ipotesi interpretative, a loro volta, possono fallire, si può arrivare, nei punti finali della sequenza lineare, a delle conclusioni assurde; di conseguenza l'ipotesi è da considerarsi fallita. Se ciò avviene si ritorna ai livelli più bassi ma, ad ogni modo, i fattori a monte avranno lasciato una forte traccia su quelli sottostanti. Ad esempio, possono cambiare le massime discorsive: si può supporre, infatti, che non si tratti della nostra lingua, oppure che l'interlocutore non sia sano di mente. In questo caso si riparte quasi da zero, riesaminando tutto dal punto di vista fonologico e semantico.

### 2.2.1 La teoria delle massime conversazionali di Grice

P. Grice identifica come funzionanti all'interno della lingua quelle che egli chiama *massime conversazionali*, cioè delle norme razionali che regolerebbero la comunicazione così da renderla il più efficiente possibile. Da tener presente che, in accordo con Grice, il parlante non aderisce sempre e totalmente a tali norme, che restano funzionanti sempre e comunque ad un certo livello. Egli identifica le seguenti quattro *massime conversazionali*:

- “Il principio di Cooperazione  
fornite il vostro contributo così come è richiesto, al momento opportuno, dagli scopi o dall'orientamento del discorso in cui siete impegnati.
- La massima di Qualità  
cercate di fornire un contributo vero; in particolare:  
non dite cose che credete false;  
non dite cose per le quali non avete prove adeguate.
- La massima di Quantità  
fornite un contributo che soddisfi la richiesta di informazioni in modo adeguato agli scopi del discorso;  
non fornite un contributo più informativo del necessario.
- La massima di Relazione  
fornite contributi pertinenti.
- La massima di Modo  
siate perspicui ed in particolare:  
evitate oscurità di espressione;  
evitate le ambiguità;  
siate brevi;  
procedete in modo ordinato.” (Levinson, 1993, pp. 113 e sg.)

### 2.3 Analisi del modello lineare

Un'importante premessa generale riguarda il fatto che le *tappe* sopra elencate, sono spesso tra loro interdipendenti; ad esempio non è possibile procedere ad un'individuazione linguistica senza avvalersi degli elementi prosodici facenti parte del processo di individuazione fonetico/fonologico. Un altro dato importante è che, sempre attenendosi alla tesi di Grice relativa alle massime conversazionali, le prime quattro *tappe* del processo vengono spesso sottintese dall'ascoltatore.

In un processo che appare, quindi, notevolmente meno lineare di quanto prospettato, le *tappe* rappresentano, come indicato, una situazione comunicativa ideale, ma in questa sede è opportuno riferirsi a parametri strettamente funzionali all'esperimento in questione.

Già a partire dall'individuazione semiotica è stato infatti dimostrato come a questo punto del processo si possa cadere in errore, finendo col trascrivere in lingua italiana ciò che originariamente era solo un ronzio interrotto a cadenza di sillaba mascherato da rumore rosa (Paoloni, 2006).

Trovandosi in una condizione particolarmente sfavorevole dal punto di vista della pulizia del canale, assume importanza cruciale l'analisi del livello fonologico, la cui errata percezione porta come naturale conseguenza all'interpretazione forzata, nonché incorretta dei punti che lo seguono.

### **3. DESCRIZIONE DELL'ESPERIMENTO**

In seguito ai sorprendenti risultati ottenuti da un esperimento molto simile a quello che presentiamo in questa sede, nel quale campioni di tre frasi (voce italiana, voce araba e voce artificiale) venivano sottoposti ad un pubblico esperto, col compito di operare trascrizioni il più fedeli possibili all'originale (Paoloni, 2006), si è voluto procedere ad un secondo più attento ed ampio esperimento, con 12 frasi diverse per struttura e livelli di rapporto segnale/rumore (S/N).

In particolare sono state acquisite varie frasi in differenti lingue: giapponese, romeno, polacco e italiano. Tali registrazioni sono state tutte effettuate con le seguenti specifiche audio: 16000 Hz ; 16 bit; mono.

I parlanti, tutti madrelingua, sono stati intervistati con l'ausilio di diversi supporti di registrazione e solo nel caso del giapponese, le tracce sono state estratte da cd, sempre e comunque in condizioni di pulizia del segnale il più possibile ottimali.

I campioni sono stati poi equalizzati con un Vu-meter ad una intensità media di 78 dB. Successivamente, a suddette tracce è stato aggiunto un rumore rosa anch'esso con frequenza di campionamento pari a 16000 Hz, con 16 dB di quantizzazione e con intensità variabile in modo tale da ottenere due distinti valori di rapporto S/N: -1.3 dB e -3.6 dB. Tutte le tracce così ottenute sono state successivamente filtrate in banda telefonica (300/3400 Hz).

I segnali rumorosi sono stati sottoposti ad un gruppo eterogeneo di 23 persone e divisi in tre distinti blocchi che riportavano le tracce in diverso ordine. Ciascun ascoltatore ha ascoltato i file utilizzando una cuffia di qualità professionale o comunque buona.

Ai partecipanti non era stato detto nulla circa la lingua o il dialetto utilizzato nelle registrazioni e sono stati invitati a scrivere su un apposito foglio quanto riuscivano a sentire.

In sede d'esperimento, agli ascoltatori è stato richiesto un ulteriore dato: il riconoscimento del sesso del parlante.

### **4. ANALISI DEI RISULTATI**

I risultati dell'esperimento descritto nel precedente paragrafo sono stati organizzati in modo tale da ottenere una visione organica dell'intero corpus in esame. In questo modo si sono potuti analizzare i dati sulla base di diversi criteri: comprensione individuale dei singoli soggetti e comprensione globale del singolo campione; il tutto ricordando la distinzione tra frasi in lingua italiana ed in lingue straniere e quella tra ascoltatori esperti e non.

Infatti, dei 23 ascoltatori, 16 sono stati considerati *esperti*, in quanto svolgenti professionalmente lavori di trascrizione; contrariamente i restanti 7, non possedendo tale caratteristica, sono stati considerati *non esperti*.

Tenendo presente le difficoltà sottese al problema della comprensione di un qualunque continuum acustico, relativamente a segmentazioni, valutazioni ed attribuzioni di valori numerici alle trascrizioni, si è optato per un conteggio convenzionale effettuato sulla base dell'individuazione di *parole* intese nella loro accezione intuitiva di corrispondenti fonici di sequenze di grafemi, che per lo meno nella scrittura dell'italiano, sarebbero separate da uno spazio. Nel processo di valutazione, si è deciso di trattare in modo differente i dati ottenuti dai campioni stranieri e italiani. Per quanto riguarda quelli italiani, si è deciso di contare le *parole* corrette, quelle errate e quelle omesse. In modo differente si è operato per i campioni stranieri: infatti, dal momento che lo scopo dell'esperimento è quello di trovare una soglia minima sotto la quale è impossibile trascrivere una registrazione in modo affidabile, dal momento che in ambito giudiziario è di estrema importanza non cadere in errore e dal momento che non ci sono stati risultati in cui parole straniere sono state trascritte in modo corretto, si è optato per considerare come corrette tutte le *parole* che sono state tralasciate, mentre errate tutte quelle che sono state trascritte.

Le seguenti tabelle hanno lo scopo di mostrare, nel modo più organico possibile, i risultati ottenuti nella trascrizione dei due gruppi di campioni (italiani e stranieri) dai due gruppi di ascoltatori (*esperti* e *non esperti*). Vengono, perciò, proposte quattro tabelle che mostrano nel dettaglio le varie possibilità di combinazione. In ognuna di esse viene considerata la media di ciò che è stato compreso, omesso o sbagliato, per ogni singolo ascoltatore. L'ultima riga di ogni tabella mostra la quantificazione di ciò che, in media, è stato compreso da ogni gruppo di ascoltatori per ogni gruppo di campioni.

| Ascoltatore  | Corrette (n°) | Corrette (%) | Errate (n°) | Errate (%) | Omesse (n°) | Omesse (%) |
|--------------|---------------|--------------|-------------|------------|-------------|------------|
| 1            | 3,3           | 27,1         | 5,8         | 51,2       | 2,3         | 21,7       |
| 3            | 3,5           | 30,9         | 3,5         | 31,9       | 4,5         | 37,3       |
| 5            | 2             | 18,8         | 8           | 66,8       | 1,5         | 14,4       |
| 7            | 1,8           | 17           | 6           | 52,6       | 3,7         | 30,4       |
| 8            | 4,3           | 38,7         | 2,7         | 24,1       | 4,5         | 37,2       |
| 9            | 3             | 29,3         | 2,3         | 19,4       | 6,2         | 51,4       |
| 10           | 2,8           | 23,3         | 4,7         | 40,5       | 4           | 36,2       |
| 12           | 3,2           | 28,4         | 5,3         | 48,7       | 3           | 22,9       |
| 13           | 3,5           | 26,3         | 3,7         | 35,5       | 4,3         | 38,3       |
| 14           | 4,3           | 37,9         | 2,7         | 22,7       | 4,5         | 39,4       |
| 16           | 3             | 28,6         | 3,3         | 28,5       | 5,2         | 43         |
| 17           | 3,5           | 31,3         | 4,8         | 39,8       | 3,2         | 28,9       |
| 18           | 4             | 36,2         | 4,2         | 34,6       | 3,3         | 29,2       |
| 20           | 3,5           | 31,8         | 4,2         | 36,2       | 3,8         | 32         |
| 21           | 4,2           | 35,7         | 6           | 52,1       | 1,3         | 12,3       |
| 22           | 2,8           | 25,3         | 2,3         | 20,5       | 6,3         | 54,2       |
| <b>Media</b> | 3,3           | 29,4         | 4,3         | 37,5       | 3,9         | 33,1       |

Tab. 1: medie matematiche dei risultati individuali degli *esperti*, per l'italiano.



| Ascoltatore  | Corrette (n°) | Corrette (%) | Errate (n°) | Errate (%) | Omesse (n°) | Omesse (%) |
|--------------|---------------|--------------|-------------|------------|-------------|------------|
| 2            | 1,8           | 15,6         | 3           | 25,5       | 6,7         | 58,9       |
| 4            | 1,8           | 15,6         | 4,3         | 34,4       | 5,3         | 50         |
| 6            | 3,3           | 28,7         | 4,7         | 42         | 3,5         | 29,4       |
| 11           | 2             | 17           | 5,2         | 44,7       | 4,3         | 38,4       |
| 15           | 2,3           | 19,1         | 1           | 8,9        | 8,2         | 72         |
| 19           | 2,8           | 24,5         | 6,5         | 58,7       | 2,2         | 16,8       |
| 23           | 1,2           | 9            | 4,2         | 36,5       | 6,2         | 54,6       |
| <b>Media</b> | 2,2           | 18,5         | 4,1         | 35,8       | 5,2         | 45,7       |

Tab. 2: medie matematiche dei risultati individuali dei *non esperti*, per l'italiano.

| Ascoltatore  | Errate (trascritte) | Corrette (omesse) |
|--------------|---------------------|-------------------|
| 1            | 100,0               | 0,0               |
| 3            | 83,3                | 16,7              |
| 5            | 100,0               | 0,0               |
| 7            | 100,0               | 0,0               |
| 8            | 0,0                 | 100,0             |
| 9            | 16,7                | 83,3              |
| 10           | 66,7                | 33,3              |
| 12           | 100,0               | 0,0               |
| 13           | 66,7                | 33,3              |
| 14           | 66,7                | 33,3              |
| 16           | 33,3                | 66,7              |
| 17           | 100,0               | 0,0               |
| 18           | 66,7                | 33,3              |
| 20           | 83,3                | 16,7              |
| 21           | 83,3                | 16,7              |
| 22           | 83,3                | 16,7              |
| <b>Media</b> | <b>71,9</b>         | <b>28,1</b>       |

Tab. 3: medie matematiche dei risultati individuali degli *esperti*, per le lingue straniere.

| Ascoltatore  | Errate (trascritte) | Corrette (omesse) |
|--------------|---------------------|-------------------|
| 2            | 33,3                | 66,7              |
| 4            | 33,3                | 66,7              |
| 6            | 100,0               | 0,0               |
| 11           | 100,0               | 0,0               |
| 15           | 0,0                 | 100,0             |
| 19           | 50,0                | 50,0              |
| 23           | 66,7                | 33,3              |
| <b>Media</b> | <b>54,8</b>         | <b>45,2</b>       |

Tab 4: medie matematiche dei risultati individuali dei *non esperti*, per le lingue straniere.

Ad esempio, la prima riga della tab. 1 mostra che l'ascoltatore numero 1, per quanto riguarda i campioni italiani, ha compreso in media 3,3 *parole* (il 27,1%), ne ha sbagliate 5,8 (il 51,2%) e omesse 2,3 (il 21,7%).

Il numero identificativo associato ad ogni singolo ascoltatore è stato del tutto casuale ed è stato a lui assegnato nel momento in cui si è proceduto all'elaborazione dei dati ottenuti dopo l'ascolto e la trascrizione.

I risultati relativi all'attribuzione del sesso dei parlanti sono stati quasi del tutto corretti: l'esperimento mostra, quindi, che questo aspetto della comunicazione, anche qualora questa avvenga in circostanze proibitive, risulta essere sempre sufficientemente intellegibile. La tabella seguente mostra il dettaglio dei risultati ottenuti.

| Ascoltatore | S/N = -1,3 dB |     |     | S/N = -3,6 dB |     |     | S/N = -1,3 dB |     |     | S/N = -3,6 dB |     |     |
|-------------|---------------|-----|-----|---------------|-----|-----|---------------|-----|-----|---------------|-----|-----|
|             | i 1           | i 2 | i 3 | i 4           | i 5 | i 6 | G 1           | P 1 | R 1 | G 2           | P 2 | R 2 |
| 1           |               |     |     |               |     |     |               |     |     |               |     |     |
| 2           |               |     |     |               |     |     |               |     |     |               |     |     |
| 3           |               |     |     |               |     |     |               |     |     |               |     |     |
| 4           |               |     |     | M             |     |     |               |     |     |               |     |     |
| 5           |               |     |     |               |     |     |               |     |     | F/M           |     |     |
| 6           |               |     |     |               |     |     |               | M/F |     | M             |     |     |
| 7           |               |     |     | M             |     |     |               |     |     |               |     |     |
| 8           |               |     |     |               |     |     |               |     |     |               |     |     |
| 9           |               |     |     |               |     |     |               |     |     |               |     |     |
| 10          |               |     |     | M             |     |     |               |     |     |               |     |     |
| 11          |               |     |     |               |     |     |               | M   |     | F/M           |     |     |
| 12          |               |     |     |               | 2M  |     |               |     |     |               | M   |     |
| 13          |               |     |     |               |     |     |               |     |     | F/M           |     |     |
| 14          |               |     |     |               |     |     |               |     |     |               |     |     |
| 15          |               |     |     |               |     |     |               |     |     |               |     |     |
| 16          |               |     |     |               |     |     |               |     |     |               |     |     |
| 17          |               |     |     |               |     |     |               |     |     |               |     |     |
| 18          |               |     |     |               |     |     |               |     |     |               |     |     |
| 19          |               |     |     |               |     |     |               | M   |     |               | M   |     |
| 20          |               |     |     |               |     |     |               |     |     |               |     |     |
| 21          |               |     |     |               |     |     |               |     |     |               |     |     |
| 22          |               |     |     |               |     |     |               |     |     |               |     |     |
| 23          |               |     |     |               |     |     |               |     |     |               |     |     |

Tab 5: panoramica della percezione relativa al sesso dei parlanti.

| <b>Legenda</b>                                                                              |                                                                                            |
|---------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
|  Omissione |  Corretto |
| M: parlante donna scambiata per uomo                                                        | I : italiano                                                                               |
| 2M: percepite due distinte voci maschili invece di una                                      | G: giapponese                                                                              |
| M/F: percepita una voce maschile seguita da una femminile                                   | P: polacco                                                                                 |
| F/M: percepita una voce femminile seguita da una maschile                                   | R: romeno                                                                                  |

## 5. CONCLUSIONI

Da questo database non molto ampio ma comunque significativo, a seguito dell'esperimento si evince che i comportamenti degli ascoltatori di fronte alle tracce con S/N – 1,3 dB e con S/N – 3,6 dB sono molto simili: le differenze, infatti, sono nell'ordine del 5 % circa.

L'esperimento ha dimostrato inoltre come le massime conversazionali siano effettivamente operative anche in casi di scarsa intelligibilità e come, così come affermava Grice, non siano né strettamente necessarie, né superficiali, ma che agiscano ad ogni modo a un certo livello (per l'appunto anche in casi estremi come quelli presentati). E' chiaro infatti come la presupposizione relativa a una comunicazione possibile e non ostacolata da altri fattori oltre la bassa qualità del segnale, abbia giocato un ruolo fondamentale, in particolare nella percezione (chiaramente errata) di frasi in lingua straniera, ricostruite come frasi in lingua italiana.

Il senso compiuto di molte di queste frasi dimostra altresì l'assunto che l'ascoltatore ritenga spontaneamente che il suo interlocutore sia sano di mente e che stia comunicando un'informazione utile e pertinente.

Al contrario, frasi trascritte come “Quella cammina a settanta [inc.] perché 31 e” (si vedano le appendici per ulteriori informazioni) dimostrano quanto affermato da Grice relativamente alla non sistematicità dell'influenza delle massime conversazionali.

E' possibile quindi spiegare la trascrizione in italiano di tracce in lingua straniera anche sulla base di *principi di cooperazione*, talvolta trascurati dall'orecchio più attento e in grado di seguire con precisione ritmo, intonazione e *pattern* delle vocali toniche della propria L1, ma spesso inevitabilmente coinvolti nei processi percettivi.

H. Fraser (2003) mostra una cosa apparentemente banale, ma di grande importanza: l'ascoltatore, nonostante presupposizioni di vario tipo, conoscenze pregresse o massime conversazionali, non deve comunque essere libero di sentire ciò che vuole, soltanto perché è ciò che si aspetta<sup>2</sup>.

Un'importante constatazione relativa ai risultati dell'esperimento, riguarda ancora una volta le tesi sostenute da Grice relativamente alle *massime conversazionali*. Infatti, le considerazioni degli ascoltatori relativamente ai campioni in esame e le conseguenti presupposizioni ed essi legate non sono state mutate nemmeno in virtù del fatto che l'ascolto poteva essere liberamente ripetuto fino a completa soddisfazione. L'idea quindi che le informazioni contenute nelle diverse tracce dell'esperimento siano state

<sup>2</sup> Cfr. a tal proposito l'esperimento condotto da Bruce (1958).

effettivamente “utili ai fini della cooperazione comunicativa” non è stata cambiata neanche dopo ripetuti ascolti.

E’ necessario tenere presente come i casi presentati in questo articolo siano in realtà dei casi limite: situazioni estreme, ma ancora possibili. E’ stato ampiamente dimostrato infatti come il rumore infici l’intelligibilità di un segnale molto di più qualora si tratti di un parlato “imprevedibile” (Miller, 1954).

Pertanto non resta che ricordare quanto sia importante che i periti trascrittori possiedano reali competenze fonetiche e psicolinguistiche in modo da poter correttamente valutare il rischio che la trascrizione effettuata possa differire anche sostanzialmente dalle parole pronunciate. Con questo non stiamo vogliamo asserire che le probabilità di fornire una trascrizione errata verrebbero ridotte a zero (e questo esperimento ne è stato una dimostrazione palese), ma che la consapevolezza dell’ errore condurrebbe ad un’analisi ancor più accurata o, in alcuni casi, a dichiarare un segmento audio non trascrivibile.

Per concludere, riassumiamo in un’unica tabella i risultati complessivi dell’esperimento.

Nel caso delle frasi in lingua straniera (*Non it.*), le voci relative alle percentuali di correttezza ed omissione, figurano in un’unica casella, proprio per la decisione, nel caso corrente, di considerare una risposta lasciata in bianco, come una risposta esatta.

| Risultati per campioni con S/N pari a -1,3 dB |          |                     |                   |                   |
|-----------------------------------------------|----------|---------------------|-------------------|-------------------|
| Ascoltatore                                   | Stimolo  | Parole corrette (%) | Parole omesse (%) | Parole errate (%) |
| Esperto                                       | Italiano | 30,5                | 28,5              | 41                |
| Non esperto                                   | Italiano | 23                  | 46,7              | 30,3              |
| Esperto                                       | Non it.  | 29,2                |                   | 70,8              |
| Non esperto                                   | Non it.  | 42,9                |                   | 57,1              |
| Risultati per campioni con S/N pari a -3,6 dB |          |                     |                   |                   |
| Ascoltatore                                   | Stimolo  | Parole corrette (%) | Parole omesse (%) | Parole errate (%) |
| Esperto                                       | Italiano | 28,3                | 37,8              | 33,9              |
| Non esperto                                   | Italiano | 14                  | 44,7              | 41,3              |
| Esperto                                       | Non it.  | 27,1                |                   | 72,9              |
| Non esperto                                   | Non it.  | 47,6                |                   | 52,4              |
| Risultati globali                             |          |                     |                   |                   |
| Ascoltatore                                   | Stimolo  | Parole corrette (%) | Parole omesse (%) | Parole errate (%) |
| Esperto                                       | Italiano | 29,4                | 33,1              | 37,5              |
| Non esperto                                   | Italiano | 18,5                | 45,7              | 35,8              |
| Esperto                                       | Non it.  | 28,1                |                   | 71,9              |
| Non esperto                                   | Non it.  | 45,2                |                   | 54,8              |

Tab. 6: sommario dei dati conclusivi

## RINGRAZIAMENTI

Un sentito ringraziamento va al Dottor Artemij Keidan e alla Professoressa Claudia Angela Ciancaglini, docenti presso l'Università "Sapienza" di Roma, senza i quali questo lavoro non sarebbe stato lo stesso.

Un ulteriore ringraziamento va alla Professoressa Mariko Saito, lettrice di giapponese della Facoltà di Studi Orientali dell'Università "Sapienza" di Roma, dal libro della quale provengono le registrazioni in lingua giapponese, utilizzate durante l'esperimento.

## BIBLIOGRAFIA

Bruce, D.J. (1958) 'The effect of listeners' anticipations on the intelligibility of heard speech', *Language and Speech*, n°1, pp. 79-97

De Mauro, Tullio (1998), *Capire le parole* (seconda edizione), Laterza.

Fraser, Helen (2003), *Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases*, in *Forensic Linguistics* 10 (2), University of Birmingham Press.

Levison, Stephen C. (1993), *La Pragmatica*, Il Mulino, Bologna.

Miller, G.A. (1954) *Language and Communication*, New York: McGraw-Hill

Paoloni A. (2006), Limiti alla trascrizione giudiziaria, Atti 3°, Convegno Nazionale AISV, Trento 2006, pp. 215-221

Paoloni A., Zavattaro D. (2007), *Intercettazioni telefoniche e ambientali*, Centro Scientifico Editore.

Paoloni A. (2007), *Con la multimodalità si moltiplicano le informazioni*, Media Duemila, n°244 marzo 2007, pp. 53-59.

## 6. APPENDICE 1: TRASCRIZIONI DELLE TRACCE IN LINGUA ITALIANA

### 6.1. Italiano, prima frase

Frase pronunciata: “Prima giri a sinistra, poi a destra e poi di nuovo a sinistra”.

Parlante: Maschio.

Trascrizione IPA ['pri:mãd'dzi:rɛssi'nĩ:stra'pɔɛæd'de:straep'pɔ:edi'nwɔ:oãssi'nĩ:stra].

Prima giri a sinistra poi a destra e poi di nuovo dritto

1. Io ho le indicazioni poi a destra e poi di nuovo [inc.]
2. Prima gira qui, poi a destra e poi di nuovo [inc.]
3. Io ho le indicazioni [inc.] e poi a destra e poi di nuovo [inc.]
4. Io ho un'indicazione! E quale sarebbe? Transelvético.
5. Prima gira qui, poi a destra e poi di nuovo [inc.]
6. Io non ci tengo [inc.] a fare sempre [inc.] e poi dico [inc.]
7. Prima gira a sinistra poi a destra e poi di nuovo a sinistra
8. Io non ti danneggio [inc.]
9. Primi giri a sinistra, poi a destra e poi di nuovo a sinistra
10. Io mi giro a sinistra, poi a destra, e poi [inc.]
11. Gira a sinistra, poi a destra e poi di nuovo [inc.]
12. [inc.] poi a destra poi di nuovo a sinistra
13. Prima giri a sinistra poi a destra e poi di nuovo a sinistra [inc.]
14. Prima giri a (destra ?) poi a destra e poi di nuovo a (sinistra ?)
15. Prima in direzione, poi a destra, poi rivolti [inc.]
16. Io me [inc.] a sinistra e poi a destra e poi dico [inc.]
17. Prima abitazione poi a destra e poi di nuovo dritto
18. Prima giri a [inc.] poi a destra poi di nuovo a sinistra
19. Io non ti chiamo più, poi a destra e poi di nuovo giù
20. Prima di fronte a te poi a destra e poi di nuovo [inc.]
21. [inc.] tu hai ragione però poi dico [inc.]
22. Io [vendisette] poi a destra (!)<sup>3</sup> e poi di nuovo [inc.]

| Italiano, prima frase, 13 parole |          |       |        |       |        |      |
|----------------------------------|----------|-------|--------|-------|--------|------|
| S/N = -1,3 dB                    | Comprese |       | Errate |       | Omesse |      |
| Ascoltatore                      | n°       | %     | n°     | %     | n°     | %    |
| 1                                | 11       | 84,6  | 1      | 7,7   | 1      | 7,7  |
| 2                                | 7        | 53,8  | 4      | 30,8  | 2      | 15,4 |
| 3                                | 8        | 61,5  | 3      | 23,1  | 2      | 15,4 |
| 4                                | 7        | 53,8  | 4      | 30,8  | 2      | 15,4 |
| 5                                | 0        | 0,0   | 13     | 100,0 | 0      | 0,0  |
| 6                                | 8        | 61,5  | 3      | 23,1  | 2      | 15,4 |
| 7                                | 2        | 15,4  | 9      | 69,2  | 2      | 15,4 |
| 8                                | 12       | 92,3  | 1      | 7,7   | 0      | 0,0  |
| 9                                | 0        | 0,0   | 4      | 30,8  | 9      | 69,2 |
| 10                               | 13       | 100,0 | 0      | 0,0   | 0      | 0,0  |
| 11                               | 7        | 53,8  | 2      | 15,4  | 4      | 30,8 |

<sup>3</sup> Alcune annotazioni degli ascoltatori sono state riportate fedelmente dai manoscritti.

|              |            |             |            |             |            |             |
|--------------|------------|-------------|------------|-------------|------------|-------------|
| 12           | 9          | 69,2        | 1          | 7,7         | 3          | 23,1        |
| 13           | 8          | 61,5        | 0          | 0,0         | 5          | 38,5        |
| 14           | 12         | 92,3        | 1          | 7,7         | 0          | 0,0         |
| 15           | 11         | 84,6        | 2          | 15,4        | 0          | 0,0         |
| 16           | 5          | 38,5        | 5          | 38,5        | 3          | 23,1        |
| 17           | 7          | 53,8        | 3          | 23,1        | 3          | 23,1        |
| 18           | 8          | 61,5        | 5          | 38,5        | 0          | 0,0         |
| 19           | 11         | 84,6        | 0          | 0,0         | 2          | 15,4        |
| 20           | 7          | 53,8        | 6          | 46,2        | 0          | 0,0         |
| 21           | 9          | 69,2        | 2          | 15,4        | 2          | 15,4        |
| 22           | 1          | 7,7         | 5          | 38,5        | 7          | 53,8        |
| 23           | 7          | 53,8        | 4          | 30,8        | 2          | 15,4        |
| <b>Media</b> | <b>7,4</b> | <b>56,9</b> | <b>3,4</b> | <b>26,2</b> | <b>2,2</b> | <b>16,9</b> |

Tab. 7: sinossi

## 6.2 Italiano, seconda frase

Frase pronunciata: “Non so, giorni, settimane, mesi... penso che siano passati dei mesi”.

Parlante: Maschio.

Trascrizione IPA [nõn'tso'dzo:rni'settĩ'mã:nə'mě:si'pentsoyessjãno'ðas'sa:ðeðe'me:si].

1. [inc.] anni è il [inc.] portato [inc.]
2. [inc.]
3. Non so venti anni [inc.]
4. [inc.]
5. Ci sto per magni per [inc.] tocca a te [inc.]
6. [inc.] anni [inc.] Beppe Grillo [inc.]
7. Non so [inc.] peccato che viene
8. [inc.]
9. Non so, giorni, settimane non lo so / può darsi che
10. Un po' [inc.] dodici anni per trentuno con tasso dieci
11. Non posso tutti i giorni [inc.], metto vicino qualcosa che [inc.]
12. Urtò [inc.] volevo dire [inc.] ma il bastoncino con il coso di pietra
13. Giusto [inc.] cartoncino per cartoncino [inc.]
14. Non so [inc.] perché nessuno [inc.]
15. A Quinto ↑ [inc.] B [inc.] [□]
16. [inc.] quattro e dieci [inc.]
17. Non so, devo ved [inc.] maggio? Pe'er trentuno qualcosa pe' ti esce
18. Ci stò [inc.] per anni [inc.] faccio un giro per fatti miei [inc.]
19. Giusto? [inc.] sette e ventuno per sette e venti [inc.]
20. Giusto? Sì, domani, cartoncino per [inc.]
21. Non so se [inc.] penso che prima è abbastanza difficile
22. [inc.] sette e ventuno per sette e dieci
23. Nun [inc.] (?) [bənə] [inc.]

| Italiano, seconda frase, 11 parole |            |            |            |             |            |             |
|------------------------------------|------------|------------|------------|-------------|------------|-------------|
| S/N = -1,3 dB                      | Comprese   |            | Errate     |             | Omesse     |             |
| Ascoltatore n°                     | n°         | %          | n°         | %           | n°         | %           |
| 1                                  | 0          | 0,0        | 4          | 36,4        | 7          | 63,6        |
| 2                                  | 0          | 0,0        | 0          | 0,0         | 11         | 100,0       |
| 3                                  | 2          | 18,2       | 1          | 9,1         | 8          | 72,7        |
| 4                                  | 0          | 0,0        | 0          | 0,0         | 11         | 100,0       |
| 5                                  | 0          | 0,0        | 8          | 72,7        | 3          | 27,3        |
| 6                                  | 0          | 0,0        | 3          | 27,3        | 8          | 72,7        |
| 7                                  | 2          | 18,2       | 3          | 27,3        | 6          | 54,5        |
| 8                                  | 0          | 0,0        | 0          | 0,0         | 11         | 100,0       |
| 9                                  | 4          | 36,4       | 2          | 18,2        | 5          | 45,5        |
| 10                                 | 0          | 0,0        | 10         | 90,9        | 1          | 9,1         |
| 11                                 | 2          | 18,2       | 7          | 63,6        | 2          | 18,2        |
| 12                                 | 0          | 0,0        | 10         | 90,9        | 1          | 9,1         |
| 13                                 | 0          | 0,0        | 8          | 72,7        | 3          | 27,3        |
| 14                                 | 2          | 18,2       | 3          | 27,3        | 6          | 54,5        |
| 15                                 | 0          | 0,0        | 2          | 18,2        | 9          | 81,8        |
| 16                                 | 0          | 0,0        | 3          | 27,3        | 8          | 72,7        |
| 17                                 | 2          | 18,2       | 9          | 81,8        | 0          | 0,0         |
| 18                                 | 0          | 0,0        | 8          | 72,7        | 3          | 27,3        |
| 19                                 | 0          | 0,0        | 7          | 63,6        | 4          | 36,4        |
| 20                                 | 0          | 0,0        | 6          | 54,5        | 5          | 45,5        |
| 21                                 | 4          | 36,4       | 5          | 45,5        | 2          | 18,2        |
| 22                                 | 0          | 0,0        | 6          | 54,5        | 5          | 45,5        |
| 23                                 | 0          | 0,0        | 2          | 18,2        | 9          | 81,8        |
| <b>Media</b>                       | <b>0,8</b> | <b>7,1</b> | <b>4,7</b> | <b>42,3</b> | <b>5,6</b> | <b>50,6</b> |

Tab. 8: sinossi

### 6.3 Italiano, terza frase

Frase pronunciata: “Mi è sembrato che la persona che ho visto volesse fuggire”.

Parlante: Femmina.

Trascrizione IPA [mjɛssem'bra:doɣelaper'sɜ:nɛɣovissoʋo'lessefud'dʒi:re].

1. Recuperate le adozioni a registro come [inc.]
2. [inc.]
3. [inc.] registratore consuma [inc.]
4. [inc.]
5. Mi dispiace per le persone che ho visto perché [inc.]
6. Il rifugiato della persona che ha scritto [inc.]
7. Se si tratta della persona che ho visto [inc.]
8. Il magistrato: che la persona che ho visto l'ho visto in ufficio
9. Mi ha assicurato che era/la [inc.]
10. Se si tratta [inc.] le do un numero fisso o [inc.]
11. Quel contratto è a doppione al registro per così dire



12. Mi dispiace aveva persone non candidacy<sup>4</sup>
13. Considerato che la questione con [inc.]
14. Mi è sembrato che il [inc.] che ho visto
15. [inc.]
16. Il più grande [inc.] che ho visto
17. E assicuratevi che la persona che ho visto [inc.]
18. Considerato che la persona che ho visto \*\*\* vestiti blu
19. Ho recuperato il dottor Magrebito e [inc.]
20. Considerato che l'adozione [inc.] è un [inc.]
21. Se si tratta della persona che ho visto l'ho visto così
22. [inc.] persona che ho visto [inc.]
23. Nessun [grazie] [inc.] tume [inc.] ditto [inc.]

| Italiano, terza frase, 11 parole |            |             |            |           |            |             |
|----------------------------------|------------|-------------|------------|-----------|------------|-------------|
| S/N = -1,3 dB                    | Comprese   |             | Errate     |           | Omesse     |             |
| Ascoltatore n°                   | n°         | %           | n°         | %         | n°         | %           |
| 1                                | 0          | 0,0         | 10         | 90,9      | 1          | 9,1         |
| 2                                | 0          | 0,0         | 0          | 0,0       | 11         | 100,0       |
| 3                                | 0          | 0,0         | 6          | 54,5      | 5          | 45,5        |
| 4                                | 0          | 0,0         | 0          | 0,0       | 11         | 100,0       |
| 5                                | 4          | 36,4        | 6          | 54,5      | 1          | 9,1         |
| 6                                | 2          | 18,2        | 7          | 63,6      | 2          | 18,2        |
| 7                                | 4          | 36,4        | 5          | 45,5      | 2          | 18,2        |
| 8                                | 6          | 54,5        | 5          | 45,5      | 0          | 0,0         |
| 9                                | 2          | 18,2        | 3          | 27,3      | 6          | 54,5        |
| 10                               | 0          | 0,0         | 6          | 54,5      | 5          | 45,5        |
| 11                               | 0          | 0,0         | 11         | 100,0     | 0          | 0,0         |
| 12                               | 1          | 9,1         | 8          | 72,7      | 2          | 18,2        |
| 13                               | 2          | 18,2        | 6          | 54,5      | 3          | 27,3        |
| 14                               | 7          | 63,6        | 1          | 9,1       | 3          | 27,3        |
| 15                               | 0          | 0,0         | 0          | 0,0       | 11         | 100,0       |
| 16                               | 3          | 27,3        | 4          | 36,4      | 4          | 36,4        |
| 17                               | 6          | 54,5        | 3          | 27,3      | 2          | 18,2        |
| 18                               | 6          | 54,5        | 5          | 45,5      | 0          | 0,0         |
| 19                               | 0          | 0,0         | 10         | 90,9      | 1          | 9,1         |
| 20                               | 1          | 9,1         | 6          | 54,5      | 4          | 36,4        |
| 21                               | 4          | 36,4        | 7          | 63,6      | 0          | 0,0         |
| 22                               | 4          | 36,4        | 0          | 0,0       | 7          | 63,6        |
| 23                               | 0          | 0,0         | 5          | 45,5      | 6          | 54,5        |
| <b>Media</b>                     | <b>2,3</b> | <b>20,6</b> | <b>5,0</b> | <b>45</b> | <b>3,8</b> | <b>34,4</b> |

Tab. 9: sinossi

<sup>4</sup> L'interpretazione di questa parola è dubbia, causa grafia ambigua.

#### 6.4 Italiano, quarta frase

Frase pronunciata: “Meglio far passare qualche giorno prima di metterla alle strette”.

Parlante: Femmina.

Trascrizione IPA ['mɛλλofarpas'sa:re'warke'dʒi:rnɔ'bri:mãdi,mĩθθer'la:lle'strette].

1. Ennio Baldassare è più di un amico [inc.]
2. Meglio far saltare qualche [inc.] prima che
3. Meglio far passare qualche giorno prima di [inc.]
4. Meglio far saltare qualche [inc.] prima che [inc.]
5. Meglio far saltare qualche punto [inc.]
6. Meglio far passare qualche giorno e finita qui
7. E noi per saltare qualche turno dovremmo [inc.]
8. Meglio far passare una persona [inc.]
9. Meglio far passare una persona [inc.]
10. Elio va a staccare ora [inc.]
11. Meglio far saltare qualche [inc.] a te
12. Meglio far passare qualche giorno prima di richiamarlo
13. [inc.] e io per bloccare [inc.]
14. [inc.] deve aspettare qualche giorno [inc.]
15. Meglio frequentare qualche giorno (xxx)
16. Meglio far passare qualche giorno che [inc.]
17. E' anche per tracciare qualche firma su [inc.]
18. Meglio far passare qualche giorno [inc.]
19. Ennio Baldassarò è più di un buon amico [inc.]
20. Vengo per passare qualche giorno [inc.]
21. E io per passare qualche giorno con [inc.]
22. Meglio far passare qualche giorno [inc.]
23. E' suto [attaccare/a staccare] una [inc.]

| Italiano, quarta frase, 10 parole |          |      |        |      |        |      |
|-----------------------------------|----------|------|--------|------|--------|------|
| S/N = -3,6 dB<br>Ascoltatore n°   | Comprese |      | Errate |      | Omesse |      |
|                                   | n°       | %    | n°     | %    | n°     | %    |
| 1                                 | 0        | 0,0  | 5      | 50,0 | 5      | 50,0 |
| 2                                 | 4        | 40,0 | 2      | 20,0 | 4      | 40,0 |
| 3                                 | 7        | 70,0 | 0      | 0,0  | 3      | 30,0 |
| 4                                 | 4        | 40,0 | 2      | 20,0 | 4      | 40,0 |
| 5                                 | 3        | 30,0 | 2      | 20,0 | 5      | 50,0 |
| 6                                 | 5        | 50,0 | 2      | 20,0 | 3      | 30,0 |
| 7                                 | 1        | 10,0 | 6      | 60,0 | 3      | 30,0 |
| 8                                 | 3        | 30,0 | 2      | 20,0 | 5      | 50,0 |
| 9                                 | 3        | 30,0 | 2      | 20,0 | 5      | 50,0 |
| 10                                | 0        | 0,0  | 4      | 40,0 | 6      | 60,0 |
| 11                                | 3        | 30,0 | 2      | 20,0 | 5      | 50,0 |
| 12                                | 7        | 70,0 | 1      | 10,0 | 2      | 20,0 |
| 13                                | 0        | 0,0  | 3      | 30,0 | 7      | 70,0 |
| 14                                | 2        | 20,0 | 3      | 30,0 | 5      | 50,0 |
| 15                                | 3        | 30,0 | 2      | 20,0 | 5      | 50,0 |

|              |            |             |            |             |            |             |
|--------------|------------|-------------|------------|-------------|------------|-------------|
| 16           | 5          | 50,0        | 2          | 20,0        | 3          | 30,0        |
| 17           | 1          | 10,0        | 4          | 40,0        | 5          | 50,0        |
| 18           | 5          | 50,0        | 0          | 0,0         | 5          | 50,0        |
| 19           | 0          | 0,0         | 10         | 100,0       | 0          | 0,0         |
| 20           | 3          | 30,0        | 2          | 20,0        | 5          | 50,0        |
| 21           | 3          | 30,0        | 3          | 30,0        | 4          | 40,0        |
| 22           | 5          | 50,0        | 0          | 0,0         | 5          | 50,0        |
| 23           | 0          | 0,0         | 4          | 40,0        | 6          | 60,0        |
| <b>Media</b> | <b>2,9</b> | <b>29,1</b> | <b>2,7</b> | <b>27,4</b> | <b>4,3</b> | <b>43,5</b> |

Tab. 10: sinossi

### 6.5 Italiano, quinta frase

Frase pronunciata: “Il tuo problema è proprio questo. A te non interessa quello che dicono gli altri”.

Parlante: Maschio.

Trascrizione IPA [ətʷoproˈbl̩ːmæˈfrɔːboˈgwɛssottenɔ̃nteˈrɛssəˈɡl̩ɡediˈgoːnoʎˈlaltʁi].

1. Il suono strano è proprio questo cioè non c'è la fuga di ridondanza
2. Ti ho telefonato apposta per dirti che comunque [inc.]
3. C'hai ragione è proprio che [inc.]
4. Ti ho telefonato per dirti che comunque ho perso la credibilità
5. La situazione è proprio questa. Già è scaduto tutto il bando
6. Se c'è qualcuno è proprio questo che [inc.] di ricordarlo
7. Cinque per cento interprete glieli devi da [inc.]
8. Questa persona è [inc.] difficoltà
9. La situazione è proprio questa [inc.]
10. La situazione è un po' così [inc.] piena di [inc.]
11. C'era prima quello dietro, che non c'era [inc.] ace
12. M1: Tutto sommato [inc.] M2: sono scarpe ingombranti
13. Il tuo problema è proprio questo a te non interessa [inc.]
14. Cioè puoi fare quello che [inc.] indipendente
15. [inc.]
16. In quella zona [inc.] circondati
17. Perché qualcuno ha proprio chiesto: “perché? Non c'è la fila per [inc.]?”
18. Anche per è la cosa giusta [inc.] ricordare
19. Si ragiona con [inc.] poi non ci interessa la difficoltà
20. Il suo problema è quello di [inc.]
21. La questione è proprio questa perciò non penso di avere più problemi
22. Il problema è proprio questo! [inc.]
23. Te l'ho detto/chiesto pech'[#] <<vvs>> [inc.] devi [inc.] ricordare (!)

| Italiano, quinta frase, 15 parole |          |      |        |      |        |      |
|-----------------------------------|----------|------|--------|------|--------|------|
| S/N = -3,6 dB<br>Ascoltatore n°   | Comprese |      | Errate |      | Omesse |      |
|                                   | n°       | %    | n°     | %    | n°     | %    |
| 1                                 | 5        | 33,3 | 10     | 66,7 | 0      | 0,0  |
| 2                                 | 0        | 0,0  | 7      | 46,7 | 8      | 53,3 |

|              |            |             |            |             |            |             |
|--------------|------------|-------------|------------|-------------|------------|-------------|
| 3            | 2          | 13,3        | 4          | 26,7        | 9          | 60,0        |
| 4            | 0          | 0,0         | 15         | 100,0       | 0          | 0,0         |
| 5            | 2          | 13,3        | 13         | 86,7        | 0          | 0,0         |
| 6            | 3          | 20,0        | 6          | 40,0        | 6          | 40,0        |
| 7            | 0          | 0,0         | 7          | 46,7        | 8          | 53,3        |
| 8            | 0          | 0,0         | 4          | 26,7        | 11         | 73,3        |
| 9            | 2          | 13,3        | 3          | 20,0        | 10         | 66,7        |
| 10           | 1          | 6,7         | 7          | 46,7        | 7          | 46,7        |
| 11           | 0          | 0,0         | 7          | 46,7        | 8          | 53,3        |
| 12           | 0          | 0,0         | 5          | 33,3        | 10         | 66,7        |
| 13           | 10         | 66,7        | 0          | 0,0         | 5          | 33,3        |
| 14           | 0          | 0,0         | 6          | 40,0        | 9          | 60,0        |
| 15           | 0          | 0,0         | 0          | 0,0         | 15         | 100,0       |
| 16           | 0          | 0,0         | 4          | 26,7        | 11         | 73,3        |
| 17           | 1          | 6,7         | 10         | 66,7        | 4          | 26,7        |
| 18           | 1          | 6,7         | 6          | 40,0        | 8          | 53,3        |
| 19           | 1          | 6,7         | 8          | 53,3        | 6          | 40,0        |
| 20           | 3          | 20,0        | 3          | 20,0        | 9          | 60,0        |
| 21           | 3          | 20,0        | 12         | 80,0        | 0          | 0,0         |
| 22           | 5          | 33,3        | 0          | 0,0         | 10         | 66,7        |
| 23           | 0          | 0,0         | 6          | 40,0        | 9          | 60,0        |
| <b>Media</b> | <b>1,7</b> | <b>11,3</b> | <b>6,2</b> | <b>41,4</b> | <b>7,1</b> | <b>47,3</b> |

Tab. 11: sinossi

#### 6.6 Italiano, sesta frase

Frase pronunciata: “Il direttore sa che lei ha mandato un civile?”

Parlante: Maschio.

Trascrizione IPA [ldirɛt'to:rəsakellejamman'de:tuntʃi'vi:re ↗ ].

- Il direttore sa chi deve mandare in civile
- Professore in cattedra [inc.] del lavoro
- Il direttore del carcere non lo hanno fatto uscire
- Professore in cattedra nella disciplina
- Il direttore sa che hanno la pizzeria?
- Il direttore del carcere glielo manda a dire
- Il ministero sa che durante la verifica [inc.] di lavoro
- Il direttore sa che mi hanno mandato in licenziamento
- Il direttore sa che lei ha mandato [inc.] ?
- Il direttore sa chi [inc.]
- [inc.] la disciplina
- Il direttore sa chi è della Sicilia
- Tu lo sai che io nella [inc.]
- Il direttore sa [inc.] deve dire
- [inc.]

- Il direttore sa chi l'ha mandato [inc.]
- Il direttore sa che [inc.]
- Il direttore sa che [inc.] pizzeria
- Il dottore sa che lei non la deve seguire
- Il direttore sa che lei ha mandato un'ispettiva?
- Il difensore sa però e lo capisce pure
- Il direttore sa che [inc.]
- Gli dicevo grazie [inc.] voi (!)

| S/N = -3,6 dB<br>Ascoltatore n° | Italiano, sesta frase, 9 parole |             |            |             |            |             |
|---------------------------------|---------------------------------|-------------|------------|-------------|------------|-------------|
|                                 | Comprese                        |             | Errate     |             | Omesse     |             |
|                                 | n°                              | %           | n°         | %           | n°         | %           |
| 1                               | 4                               | 44,4        | 5          | 55,6        | 0          | 0,0         |
| 2                               | 0                               | 0,0         | 5          | 55,6        | 4          | 44,4        |
| 3                               | 2                               | 22,2        | 7          | 77,8        | 0          | 0,0         |
| 4                               | 0                               | 0,0         | 5          | 55,6        | 4          | 44,4        |
| 5                               | 3                               | 33,3        | 6          | 66,7        | 0          | 0,0         |
| 6                               | 2                               | 22,2        | 7          | 77,8        | 0          | 0,0         |
| 7                               | 2                               | 22,2        | 6          | 66,7        | 1          | 11,1        |
| 8                               | 5                               | 55,6        | 4          | 44,4        | 0          | 0,0         |
| 9                               | 7                               | 77,8        | 0          | 0,0         | 2          | 22,2        |
| 10                              | 3                               | 33,3        | 1          | 11,1        | 5          | 55,6        |
| 11                              | 0                               | 0,0         | 2          | 22,2        | 7          | 77,8        |
| 12                              | 2                               | 22,2        | 7          | 77,8        | 0          | 0,0         |
| 13                              | 1                               | 11,1        | 5          | 55,6        | 3          | 33,3        |
| 14                              | 3                               | 33,3        | 2          | 22,2        | 4          | 44,4        |
| 15                              | 0                               | 0,0         | 0          | 0,0         | 9          | 100,0       |
| 16                              | 5                               | 55,6        | 2          | 22,2        | 2          | 22,2        |
| 17                              | 4                               | 44,4        | 0          | 0,0         | 5          | 55,6        |
| 18                              | 4                               | 44,4        | 1          | 11,1        | 4          | 44,4        |
| 19                              | 5                               | 55,6        | 4          | 44,4        | 0          | 0,0         |
| 20                              | 7                               | 77,8        | 2          | 22,2        | 0          | 0,0         |
| 21                              | 2                               | 22,2        | 7          | 77,8        | 0          | 0,0         |
| 22                              | 4                               | 44,4        | 0          | 0,0         | 5          | 55,6        |
| 23                              | 0                               | 0,0         | 4          | 44,4        | 5          | 55,6        |
| <b>Media</b>                    | <b>2,8</b>                      | <b>31,4</b> | <b>3,6</b> | <b>39,6</b> | <b>2,6</b> | <b>29,0</b> |

Tab. 12: sinossi

## 7. APPENDICE 2: TRASCRIZIONE DELLE TRACCE IN LINGUA STRANIERA

Nella valutazione delle trascrizioni relative alle frasi in lingua straniera, si è operato in senso dicotomico, considerando corrette le risposte lasciate in bianco; al contrario, poiché quando vi sono state trascrizioni, esse sono state fatte sempre in italiano, queste sono state considerate come completamente errate.

Alcuni ascoltatori, come si nota dalla Tab. 13, avevano ipotizzato che le tracce in questione non appartenessero alla lingua italiana. Un'analisi più approfondita delle conoscenze linguistiche di ciascun partecipante (lingua madre, eventuale bilinguismo,

dialetto d'origine, etc.) avrebbe probabilmente prodotto risultati più significativi in questo senso, indicando un'eventuale relazione tra competenze e corretta comprensione dei campioni in esame.

### 7.1 Giapponese, prima frase

Frase pronunciata 看護婦さんに注射させました。

In trascrizione *rōmaji*: “Kangohu san ni chūshasasemashita” (traduzione: ha fatto fare l'iniezione all'infermiera).

Parlante: Femmina.

Trascrizione IPA [kẽŋŋó:ɸũ̃ːsẽnnitẽy:ẽẽsẽsẽmãẽːtã].

1. A Lucca registrano prima di me [inc.]
2. [inc.]
3. A nu [inc.] spareggio [inc.]
4. [inc.]
5. Hanno un parente [inc.]
6. Al mio [inc.] intendono rilasciare [inc.]
7. A lui nun fa niente rispetto [inc.]
8. [inc.]
9. [inc.]
10. [inc.]
11. A n [inc.]
12. A nun fa niente non serve a tuo marito
13. “Annusali”/”sta lì” che mi sembra che manchi [inc.]
14. [inc.]
15. [inc.]
16. A nu fa nien ti serve e [inc.]
17. Annuk<sup>5</sup> m’ha detto: m’è piaciuto ma [inc.]
18. Hanno smarrito \*\*\* la strada
19. [inc.]
20. [inc.]
21. a me sai [inc.]
22. all’ultima riga [inc.]
23. [inc.]

### 7.2 Giapponese, seconda frase

Frase pronunciata: 家の子はすぐ外の子と同じ玩具をほしがります。

In trascrizione *rōmaji*: “Uchi no ko wa sugu hoka no ko to onaji omocha wo hoshigarimasu” (traduzione: Il nostro bambino vuole subito i giocattoli uguali a quelli degli altri bambini).

Parlante: Femmina.

Trascrizione IPA

[ũttẽnõːkò:æssú:hõːhõ:kàndõːkòtõ:nẽːɖi?ðmõːtẽãð:hõẽĩgãĩmãːss].

1. [inc.] non tanto bene [inc.]
2. [inc.]
3. [inc.]
4. [inc.]

<sup>5</sup> <k> nel manoscritto in questione può essere interpretato anche come <n>.

5. Nel novantuno si sa che [inc.]
6. [inc.] novantuno [inc.]
7. Ogni tanto va [inc.]
8. [inc.]
9. [inc.]
10. [inc.] un po' assurda [inc.]
11. (voce femminile) [inc.] chiami [inc.] per quella mezz'ora mi chiami. (voce maschile) [inc.] quelli
12. Lo assumi tramite call opitmal<sup>6</sup>
13. (voce femminile) dopo a tu non chiami [inc.] (voce maschile) ti chiamo pure io
14. Un po' assurda [inc.] come motivazione
15. [inc.]
16. [inc.]
17. [inc.] qualcuno mi chiami per dirmi "sta divorziando Marini"
18. qualcuno mi chiami [inc.]
19. [inc.]
20. Se sente qualcuno mi chiami sul telefonino di [inc.]
21. Andiamo va tu non chiami [inc.]
22. Oh! Attenzione alle compagne [inc.] ?
23. [inc.]

### 7.3 Romeno, prima frase

Frase pronunciata: "Azi am fost sã cumpãr pâine" (traduzione: oggi ho comprato il pane).

Parlante: Femmina.

Trascrizione IPA ['a:zjaŋ'fostə: 'səḵũm'pəɾə'pəĩ.nẽ].

1. Agli appunti come ti scrivi [inc.]
2. Hai fatto i fori [inc.] ?
3. Alza un po' il volume [inc.]
4. Hai fatto i fori [inc.] ?
5. Asha perché [inc.]
6. Hai fatto [inc.] insieme a Tarquini?
7. Guarda che [inc.]quello lì
8. [inc.]
9. [inc.]
10. Alzati un poco [inc.] vieni qui
11. Alza un po' di più, voglio sentire
12. Alta pressione FPT
13. Ah se consegna i [inc.]
14. [inc.]
15. [inc.]
16. [inc.]
17. Alza un po' di più il bordo dei fiumi
18. [inc.]
19. Hai fatto una spesa e preso un impegno

<sup>6</sup> Altre possibile interpretazioni della trascrizione di questa parola sono "opttmaly" o "otitmaly".

20. Asia [inc.]
21. Alza un po' [inc.]
22. [inc.]
23. Hai sapu' [#] se [inc.] (?)

#### 7.4 Romeno, seconda frase

Frase pronunciata: “Băiatul meu acest an a fost promovat” (traduzione: Mio figlio quest'anno è stato promosso).

Parlante: Femmina.

Trascrizione IPA [bwe'ja:tulmewatʃe'sta:ne::af'fostə::promo'va:].

1. E gli altri non li ho accettati [inc.] eh [inc.]
2. Se tanto non mi vuoi accettare [inc.]
3. E' alto fino a [inc.] perché è troppo basso
4. Se tanto non mi vuoi accetta [inc.] chi troppo vuol
5. Quella cammina a settanta [inc.] perché 31 e.
6. Se l'altro non mi fa aspettare [inc.]
7. Ogni tanto veniva a [inc.] verso
8. [inc.]
9. [inc.]<sup>7</sup>
10. [inc.] a cercarmi [inc.]
11. [inc.] per cui [inc.]
12. Tra gli altri meno accettante<sup>8</sup>
13. [inc.] mia su tanti di questi [inc.]
14. In altra maniera c'è qualche [inc.]
15. Ragazzo veniva accercarti [e::] doppio le trenta volte
16. [inc.]
17. Di notte non mi menire<sup>9</sup> a (cercare/cercarmi ???) perché è un casino, eh!
18. [inc.]<sup>10</sup>
19. [inc.]
20. Un altro mi [inc.]
21. [inc.]
22. Hai anche [inc.]
23. E gli altri me l'[aggettavi] [inc.]

#### 7.5 Polacco, prima frase

Frase pronunciata: “Dziwię się że jeszcze nie zszedł. Zawołali go do telefonu” (traduzione: mi meraviglio che non sia ancora sceso. L'hanno chiamato al telefono).

Parlante: Femmina.

Trascrizione IPA ['dʒivʲɛdʒe'jɛtɕɛɲe s'tɕɛ: dʒɔ'vɔ'wa: rɪɡodɔtɛlɔ'fɔ:nu ].

1. Di di codesti amici [inc.]
2. [inc.]
3. Digli se deve [inc.]
4. [inc.]

<sup>7</sup> L'ascoltatore qui annota {non sembra italiano}.

<sup>8</sup> Interpretazione dubbia.

<sup>9</sup> *Idem*.

<sup>10</sup> L'ascoltatore qui annota {francese??}.



5. Mi dice che non sa niente, mi dice che sta da solo
6. (voce maschile) Lui dice che non sai mentire (voce femminile) [inc.] la do.
7. (vedi se hanno detto) niente [inc.] all'alba dei trentun anni
8. [inc.]
9. Il giudice [inc.]
10. [inc.]
11. Mi disse che non sa nient [inc.], dovrebbe metterla fuori
12. Digli che venisse dietro [inc.] come preferisco fare
13. [inc.]
14. Gli dice che non c'è niente [inc.]
15. [inc.]
16. Si dice che [inc.] e anche ch [inc.]
17. [inc.] avere esperienza [inc.]
18. Mi dice che sa niente [inc.] dovrebbe essere ancora qui
19. Digli se abbiamo finito che lancio il detonatore
20. Lui dice che non sa niente [inc.]
21. Gli dice che [inc.] rompe questo rapporto
22. Lui dice che [inc.]
23. Gli disse di non [s'e n'jì scegli] [inc.] hanno detto che sei sola (!)

#### 7.6 Polacco, seconda frase

Frase pronunciata: “Dobrze się czujesz? Masz to co potrzebujesz?” (traduzione: stai bene? Hai quel che ti serve?).

Parlante: Femmina.

Trascrizione IPA ['dɔ:bʒɛʃɛ'tɕu:jɛʃ] / 'mäʃtɔtsopotɕə'bu:jɛɕ].

- La diessoppi? Mai sentito dirmi
- [inc.]
- A Giuseppina [inc.]
- [inc.]
- Orientativi [inc.]
- A Giuseppina [inc.] altro che pettegulesso!
- L'ambiente qui [inc.] essere curiosi
- [inc.]
- [inc.]<sup>11</sup>
- Per mia fortuna [inc.]
- Voglio capire, anche se non ci riescono, [inc.]
- Aggiò capio<sup>12</sup>
- [inc.]
- Voglio capire [inc.]
- [inc.]
- [inc.]
- Vorrei capire [inc., forse “che mi ha detto”] su di me
- Ho detto qui [inc.]
- La DFT in questo procedimento

<sup>11</sup> L'ascoltatore qui annota {non sembra italiano}.

<sup>12</sup> L'ascoltatore qui annota {siciliano?}.

- Con questo qui? [inc.]
- E' questo qui? Hai detto questo qui
- [inc.] che comincia [inc.] ?
- e [gli insettini] [inc.]

La seguente tabella mostra, in sintesi, i risultati ottenuti dall'ascolto dei campioni in lingua straniera, seguendo i principi enunciati ad inizio paragrafo.

| Ascoltatore              | S/N = -1,3 dB  |               |               | S/N = -3,6 dB |               |               |
|--------------------------|----------------|---------------|---------------|---------------|---------------|---------------|
|                          | G 1            | P 1           | R 1           | G 2           | P 2           | R 2           |
| 1                        |                |               |               |               |               |               |
| 2                        | X              | X             |               | X             | X             |               |
| 3                        |                |               |               | X             |               |               |
| 4                        | X              | X             |               | X             | X             |               |
| 5                        |                |               |               |               |               |               |
| 6                        |                |               |               |               |               |               |
| 7                        |                |               |               |               |               |               |
| 8                        | X              | X             | X             | X             | X             | X             |
| 9                        | X              |               | X             | X             | O             | O             |
| 10                       | X              | X             |               |               |               |               |
| 11                       |                |               |               |               |               |               |
| 12                       |                |               |               |               |               |               |
| 13                       |                | X             |               |               | X             |               |
| 14                       | X              |               | X             |               |               |               |
| 15                       | X              | X             | X             | X             | X             | X             |
| 16                       |                |               | X             | X             | X             | X             |
| 17                       |                |               |               |               |               |               |
| 18                       |                |               | X             |               |               | O             |
| 19                       | X              |               |               | X             |               | X             |
| 20                       | X              |               |               |               |               |               |
| 21                       |                |               |               |               |               | X             |
| 22                       |                |               | X             |               |               |               |
| 23                       | X              |               |               | X             |               |               |
| <b>Corrette - errate</b> | <b>10 - 13</b> | <b>6 - 17</b> | <b>7 - 16</b> | <b>9 - 14</b> | <b>7 - 16</b> | <b>7 - 16</b> |

Tab. 13: sinossi

|                                |                                                                                 |
|--------------------------------|---------------------------------------------------------------------------------|
| <b>Legenda:</b>                |                                                                                 |
| G 1: giapponese, prima frase   | P 1: polacco prima, frase                                                       |
| G 2: giapponese, seconda frase | P 2: polacco seconda frase                                                      |
| R 1: romeno, prima frase       | X: non trascritto (corretto)                                                    |
| R 2: romeno, seconda frase     | O: non trascritto (l'ascoltatore pensa che non si tratti della lingua italiana) |
|                                | <input type="checkbox"/> : trascritto (errato)                                  |

# VOCALI ACCENTATE BREVI PRODOTTE CON INCREMENTO DI F0: COME E PERCHÉ?

Arianna Uguzzoni

Alma Mater Studiorum – Università di Bologna

*arianna.uguzzoni@unibo.it*

## 1. SOMMARIO

Sono discusse alcune questioni cruciali che riguardano, da un lato, il fenomeno universale della F0 intrinseca, dall'altro, l'incremento di F0 riscontrato in vocali rilassate-brevi (vs. vocali tese-lunghe) del tedesco e in vocali brevi (vs. lunghe) del danese. Sulla base di una ricca bibliografia sono esposte in modo sintetico le principali ipotesi che sono state avanzate sulle ragioni e sui meccanismi dei due fenomeni, di cui uno è "language-independent", l'altro è "language-dependent". Sul primo, che è stato oggetto da tempo di numerose indagini, non è stata raggiunta una soluzione ampiamente condivisa. Il secondo, su cui è stata attirata l'attenzione da Eli Fischer-Jørgensen nel 1990, richiede ulteriori ricerche e approfondimenti.

Ritengo importante mettere in rilievo la differenza che sussiste tra le problematiche inerenti alla F0 intrinseca, che è universalmente in correlazione con la "vowel height", e le problematiche inerenti all'incremento di F0, che è in correlazione con la brevità della vocale in lingue che fanno un uso distintivo della durata.

Per quale ragione la F0 dovrebbe essere più elevata nelle vocali brevi? Per cercare una plausibile risposta a tale interrogativo è necessario, a mio parere, che il fenomeno non sia studiato in isolamento, ma venga inserito nella cornice generale della opposizione di quantità vocalica e della sua implementazione. Le singole "quantity languages" ingaggiano per così dire una serie, più o meno ampia, di dimensioni che simultaneamente, in misura più o meno grande, collaborano allo stesso fine: il rafforzamento della distinzione breve/lungo.

E' sembrata opportuna una breve trattazione di alcune proprietà fonetiche concomitanti che cooperano con la durata nella realizzazione dell'opposizione di quantità vocalica in tedesco, in danese, in frignanese. La differenza nei valori di F0 che è stata trovata in queste lingue si può considerare una di tali proprietà compresenti e coagenti.

Perché il frignanese? Una parte dell'articolo è dedicata alla esposizione dei risultati di una ricerca pilota su una varietà italo-romanza parlata nel Frignano (Appennino modenese). Sono stati misurati vari parametri, compresa la F0, delle vocali accentate brevi /ø/, /e/, /ɔ/, /a/ e delle loro controparti lunghe /ø:/, /e:/, /ɔ:/, /a:/, limitatamente a parole e pseudoparole /'pVpa/ vs. /'pV:pa/ prodotte in frase cornice da un soggetto femminile. Dalla comparazione dei valori delle singole vocali è emersa una regolarità sistematica: in tutte le coppie la F0 è più elevata nell'elemento breve che nell'elemento lungo, con un incremento che, nell'ordine, è pari a 7.71%, 18.40%, 15.70%, 18.73%.

## 2. LA FREQUENZA DELLA FONDAMENTALE

### 2.1 Il fenomeno della F0 intrinseca

Prendo l'avvio da una citazione che pone in rilievo i due aspetti compresenti nella struttura delle singole vocali: "Vowels exhibit dual acoustic patterns in speech: formant patterns that derive from articulatory settings, and intrinsic prosodic features that arise from the influence of articulation" (Honda et al., 2007: 348). Per quel che riguarda il secondo aspetto, è noto che quando i parametri durata, intensità, F0 vengono considerati in relazione alla qualità di una vocale si parla, rispettivamente, di durata intrinseca, intensità intrinseca, F0 intrinseca.

Dall'esame di numerose lingue sono emerse correlazioni regolari tra il valore di ciascun parametro e la collocazione della vocale lungo l'asse alto-basso ("tongue height"). A parità di condizioni, le vocali alte presentano valori di durata e valori di intensità che sono minori di quelli delle vocali basse; a parità di condizioni, nelle vocali alte i valori di F0 sono maggiori rispetto a quelli delle vocali basse (Lehiste, 1970). Facendo riferimento alle frequenze formantiche le tre generalizzazioni sono espresse di solito nel seguente modo: la durata è in correlazione positiva con F1, l'intensità è in correlazione positiva con F1, la F0 è in correlazione negativa con F1.

Questi cenni intendono soltanto rammentare in quali direzioni, articolatorie e acustiche, si realizza l'influenza esercitata dal tipo di articolazione di una determinata vocale sulle variazioni di durata, di intensità, di F0 della vocale stessa. Non entro invece nella tematica dei vari fattori che possono essere responsabili di tali manifestazioni fonetiche, argomento sul quale è disponibile una bibliografia specifica. Per esempio nel recente articolo di Honda et al. (2007), sono descritti e discussi in modo sintetico i processi biomeccanici e aerodinamici coinvolti nella produzione di ciascuno dei suddetti fenomeni intrinseci.

Sulla universalità della F0 intrinseca delle vocali non ci sono dubbi (Whalen & Levitt, 1995). Ben condivisa è l'opinione che essa contribuisca alla identificazione della qualità delle vocali tramite la tendenza generale delle vocali alte ad un innalzamento della F0. Un grande disaccordo si trova invece se si esaminano le ipotesi formulate in merito ai meccanismi che sono sottesi alla realizzazione di tale fenomeno. Perché in lingue numerose e tipologicamente diverse le vocali alte hanno un valore della frequenza fondamentale che è superiore a quella delle vocali basse? Per rispondere a questo interrogativo vari studiosi hanno svolto molteplici ricerche, ma non sono ancora state raggiunte soluzioni del tutto soddisfacenti e ampiamente condivise.

È stato opportunamente rilevato che "the phenomenon of intrinsic vowel F0 appears simple and universal, although its implementation mechanisms may not be" (Honda, 1995: 228). Se vogliamo dare un rapido sguardo ad alcune spiegazioni della F0 intrinseca può essere utile tracciare una linea divisoria tra quelle che la vedono come un effetto meccanico della produzione e quelle che la considerano un fenomeno regolato da un controllo attivo.

Comune alle proposte del primo tipo (per es., Honda, 1983; Ohala & Eukel, 1987; Honda, 1995; Honda et al., 2007) è la convinzione che il meccanismo soggiacente al fenomeno della IF0 consiste nell'associazione biomeccanica tra la laringe e gli articolatori sovralaringali. La maggiore F0 delle vocali alte (vs. le vocali basse) è vista come una conseguenza automatica dell'interazione tra la muscolatura della laringe e la muscolatura degli articolatori orali. Responsabili del fenomeno sono ritenuti i movimenti sovralaringali che agiscono meccanicamente sulla configurazione laringale ("tongue pull theory"/"tongue root advancement theory"). Semplificando al massimo si può dire che i gesti orali implicati

nella articolazione di vocali relativamente più alte provocano un aumento della tensione (verticale/orizzontale) delle corde vocali che quindi vibrano in modo più veloce.

Le proposte del secondo tipo sono animate dall'idea che la maggiore F0 delle vocali alte (vs. le vocali basse) sia il risultato di una pianificazione deliberata (per es., Diehl & Kluender, 1989; Kingston, 1992; Kingston & Diehl, 1994; Fahey & Diehl, 1996). Si muove dalla premessa che la distanza tra F1 e F0 è rilevante per la percezione delle differenze tra vocali nella dimensione alto-basso (Traunmüller, 1981). Si constata che il valore più elevato della F0 in una vocale alta, la cui F1 è piccola, rende meno larga la distanza tra F1 e F0; d'altra parte, il valore meno elevato della F0 in una vocale bassa, la cui F1 è grande, rende più larga la distanza tra F1 e F0. In tale modo la distinzione percettiva tra vocali alte e vocali basse viene irrobustita e questo effetto sarebbe la vera ragione d'essere della covariazione tra F1 e F0.

Con l'espressione "speech enhancement" si intende che il parlante producendo la F0 intrinseca contribuisce a rafforzare la separazione delle qualità vocaliche da parte dell'ascoltatore. Viene sottolineato il fatto che il beneficio percettivo non è "the fortuitous effect of a physiological linkage between tongue height and rate of vocal fold vibration", ma è invece "produced deliberately, by coordinating independent articulations" (Kingston, 1992:100-101).

I sostenitori di questa posizione ipotizzano che la pianificazione, volta a regolare il valore della F0 di una vocale in funzione del suo grado di altezza, si espliciti attraverso il controllo diretto del muscolo cricotiroideo (CT). Tale ipotesi si appoggia su due generi di connessioni che si sono trovate tra variazioni nel muscolo CT e variazioni nelle proprietà vocaliche: da un lato il fatto che il CT risulta attivo quando la F0 viene innalzata, dall'altra il fatto che l'attività del CT risulta maggiore in caso di vocali alte.

Una discussione intorno a "automaticità" vs. "intenzionalità" viene condotta in maniera efficace da Whalen et al. (1998). La ricerca specifica consiste nell'esame EMG dell'attività del muscolo cricotiroideo in vocali alte e in vocali basse, pronunciate da quattro soggetti inglesi. I risultati ottenuti forniscono l'appoggio empirico per respingere la prospettiva della intenzionalità e sostenere invece quella della automaticità. Con le loro stesse parole, la F0 intrinseca non è dunque "a deliberate enhancement of the speech signal", bensì "a consequence of successful vowel articulation" (Whalen et al., 1998: 141). ).

Gli autori, fra l'altro, richiamano l'attenzione sulla presenza della F0 intrinseca nel babbling di bambini di 6-12 mesi : "a result that would be difficult to account for if IF0 where anything other than automatic" (Whalen et al., 1998: 126). A ciò si collega l'ipotesi che la "intrinsic vowel F0 found in infants and deaf children implies some innate mechanism of neural connections which causes simultaneous muscle activities in the tongue and the larynx" (Honda, 1995: 228).

## 2.2 Il problema della F0 in vocali rilassate-brevi vs. vocali tese-lunghe

Secondo un uso corrente con i termini rilassato e teso si fa riferimento a differenze qualitative (F1 e F2) che si accompagnano a differenze di durata nelle vocali di alcune lingue germaniche (in particolare, tedesco, inglese, olandese). A scopo di maggiore chiarezza, in questa sede si abbineranno, da una parte, gli aggettivi rilassato e breve, dall'altra, gli aggettivi teso e lungo per richiamare entrambe le manifestazioni, quella spettrale e quella temporale (prescindendo dalla nota controversia sulla priorità linguistica dell'uno o dell'altro aspetto).

Il problema della F0 in vocali rilassate e brevi vs. vocali tese e lunghe è stato affrontato per la prima volta dalla Fischer-Jørgensen (1990) in un saggio che ritengo magistrale da

vari punti di vista. La ricerca sperimentale verte sulle vocali tedesche /ɪ, ε, a/ e /i:, e:, a:/ nel contesto /d-də/, prodotte due volte da sei soggetti (cinque maschili, uno femminile).

Per quel che concerne la durata, tutte le differenze tra brevi e lunghe sono altamente significative e il rapporto medio V/V: risulta pari a 61.00. A fini illustrativi mi è sembrato utile estrarre dalle numerose tabelle i valori medi riguardanti la durata, la F1, la F0 del soggetto maschile V.P. e collocarli in un unico quadro sinottico: le vocali sono messe nell'ordine usato dalla Fischer-Jørgensen (1990) per la frequenza fondamentale.

|        | i:     | ɪ      | e:     | ε      | a      | a:     |
|--------|--------|--------|--------|--------|--------|--------|
| durata | 111.00 | 69.00  | 133.00 | 84.00  | 98.00  | 197.00 |
| F1     | 259.00 | 344.00 | 312.00 | 436.00 | 645.00 | 746.00 |
| F0     | 126.50 | 128.00 | 117.80 | 120.00 | 116.90 | 113.60 |

Tabella 1: Durata (ms), F1 (Hz), F0 (Hz) di un soggetto maschile tedesco.

Se si tengono separate l'una dall'altra le due serie delle vocali analizzate, il crescere o il calare dei valori di F1 e di F0 sono riassumibili nei seguenti schemi: (A) /ɪ < ε < a /; /i: < e: < a: / per F1; (B) /ɪ > ε > a /; /i: > e: > a: / per F0. Leggendo con questo criterio i dati della tabella 1, si constata dunque che all'interno di ciascuna serie la frequenza fondamentale è in correlazione negativa con la frequenza della prima formante; e ciò è conforme al modello descrittivo della F0 intrinseca.

Una chiara difformità da tale modello si riscontra invece se le vocali rilassate-brevi vengono comparate con le vocali tese-lunghe. La coppia /a/ vs. /a:/ va lasciata da parte perché in tedesco la prima unità presenta una F1 inferiore e una F0 superiore a confronto con la seconda unità. Esamineremo il comportamento di F1, prima, e di F0, poi, nelle coppie /ɪ/ vs. /i:/, /ε/ vs. /e:/, come è rappresentato nei dati della tabella 1.

E' evidente che nelle rilassate-brevi i valori di F1 sono molto maggiori che nelle loro controparti tese-lunghe: /ɪ/ > /i:/ (344.00 Hz vs. 259.00 Hz); /ε/ > /e:/ (436.00 Hz vs. 312.00 Hz). Si nota inoltre che la F1 di /ɪ/ supera non solo quella di /i:/ ma anche quella di /e:/.

In base al modello classico della F0 intrinseca ci aspetteremmo che i valori di F0 mostrassero una direzione inversa rispetto a quella osservata per F1. Invece non è così: l'entità delle differenze nella F0 è piccola, ma nel soggetto V.P. si può dire che esse sono sempre in correlazione positiva con le differenze nella F1. Infatti nelle rilassate-brevi i valori di F0 sono tutti un po' più elevati che nelle corrispondenti tese-lunghe: /ɪ/ > /i:/ (128.00 Hz vs. 126.50 Hz); /ε/ > /e:/ (120.00 Hz vs. 117.80 Hz.). Si vede inoltre che la F0 di /ɪ/ supera non solo quella di /i:/ ma anche quella di /e:/.

La differenza tra la F0 di /ɪ/ e la F0 di /e:/ merita di essere sottolineata perché si manifesta in tutti i soggetti ed è statisticamente significativa. Per agevolare le comparazioni calcolo la percentuale sulla base dei dati di cui dispongo. Se consideriamo V.P. come singolo (tabella 1), il valore di F0 in /ɪ/ (128.00 Hz) presenta un incremento dell'8.65% rispetto al valore di F0 in /e:/ (117.80 Hz). Se guardiamo alla media dei valori di F0 dei cinque soggetti maschili, compreso V.P. (Fischer-Jørgensen, 1990: 123), l'incremento della F0 di /ɪ/ (124.40 Hz) rispetto alla F0 di /e:/ (117.20 Hz) risulta del 6.14%.

Questa presentazione, sia pure selettiva e sintetica, dei risultati dell'indagine sul tedesco mostra un indubbio "breakdown" del modello classico secondo il quale tra F0 e F1 c'è correlazione negativa. Il caso delle vocali rilassate-brevi confrontate con le vocali tese-lunghe è in conflitto diretto con il modello descrittivo della F0 intrinseca e con quelle proposte esplicative che fanno dipendere gli incrementi di F0 dall'azione meccanica esercitata dai movimenti della lingua sulle strutture laringali (cfr. paragrafo 2.1).

Credo che siano emersi elementi convincenti per portarci a condividere l'affermazione che "there are other factors involved in intrinsic F0 other than tongue height" (Fischer-Jørgensen, 1990: 135). Non va dimenticato che pagine importanti sono dedicate a discutere le relazioni tra il valore relativamente maggiore di F0 osservato in vocali rilassate-brevi e i seguenti fattori: la posizione della mandibola, il flusso dell'aria, la durata.

Bisogna ammettere che intorno ai meccanismi e alle ragioni di una situazione così complessa non possediamo ancora conoscenze adeguate e proposte risolutive. Dodici anni dopo in un bilancio sul tema sono delineati due scenari possibili che reputo degni di citazione e di attenzione: "the IF0 patterns in German remain intriguing because *either* they indicate the presence of mechanical effects on F0 that have hitherto escaped our understanding *or* they reveal – contrary to the view of Wahlen et al. (1998) – that there can be an *active* laryngeal component in IF0 control" (Hoole & Mooshammer, 2002: 140).

### 2.3 Il problema della F0 in vocali brevi vs. vocali lunghe

Ritengo interessante e utile esaminare ora qualche aspetto della relazione tra frequenza fondamentale e durata.

Per esempio è stato osservato che un valore più elevato della F0 di /ɪ/ rispetto a quello di /i:/ si trova più frequentemente in tedesco che in inglese. Ciò può essere attribuito alla misura della brevità delle vocali rilassate-brevi tedesche. Nel caso specifico di questa coppia tedesca, facendo la media delle realizzazioni dei sei soggetti (elencate vocale per vocale in Fischer-Jørgensen, 1990: 115), risulta che la durata della rilassata-breve /ɪ/ è di 67 ms mentre la durata della tesa-lunga /i:/ è di 110 ms (con una V/V: ratio pari a 60.90).

Per cercare di comprendere natura e importanza della correlazione tra differenze in F0 e differenze sistematiche di durata è illuminante fare una comparazione fra tedesco e danese.

Come si sa da precedenti studi della Fischer-Jørgensen (1972), in danese opera in sede accentata una opposizione generale tra vocali fonologicamente brevi e vocali fonologicamente lunghe (con un rapporto medio V/V: pari a 50.50). D'altra parte in danese le differenze temporali non sono accompagnate in modo coerente e consistente da differenze qualitative: quando le brevi e le lunghe differiscono anche nelle frequenze formantiche, l'entità degli scarti è inferiore a quella del tedesco, e anche a quelle dello svedese e del norvegese.

Dal saggio del 1990 traggio e commento alcuni dati emersi dall'analisi delle vocali /i, e, ε, a/ e /i:, e:, ε:, æ:/ nel contesto /d-də/, pronunciate otto volte da tre soggetti.

I valori di F1 e F2 confermano che le differenze spettrali tra le brevi e le lunghe di questo settore del vocalismo danese sono molto piccole. Per esempio, comparando le brevi /i/, /e/ con le lunghe /i:/, /e:/, risulta che le prime hanno in media 10 Hz in più nella prima formante (lieve abbassamento) e 16 Hz in meno nella seconda formante (lieve centralizzazione). Dal punto di vista della qualità nelle coppie /i/ vs. /i:/ e /e/ vs. /e:/ le unità brevi e le unità lunghe sono dunque molto simili e talora identiche.

Le misurazioni dei valori di F0 non forniscono esiti coerenti riguardo alle vocali /ɛ/, /a/ e /e:/, /æ:/, mentre indicano con chiarezza che nelle brevi /i/, /e/ essi sono di regola più elevati che nelle lunghe /i:/, /e:/. L'entità della differenza varia nei tre soggetti: si osservano, rispettivamente, scarti di 6.0 Hz, 7.8 Hz, 5.7 Hz tra i membri della coppia /i/ vs. /i:/ e scarti di 7.2 Hz, 6.0 Hz, 1.1 Hz tra i membri della coppia /e/ vs. /e:/.

Tali risultati concernenti la F0, a mio parere, sono molto utili sia sul piano generale sia sul piano particolare del confronto tra tedesco e danese. La differenza nei valori di F0 tra le brevi e le lunghe in danese (/i/ vs. /i:/ e /e/ vs. /e:/) è, in termini assoluti, abbastanza consistente; e, in termini comparativi, è superiore a quella trovata in tedesco tra le rilassate-brevi e le corrispondenti tese-lunghe (/ɪ/ vs. /i:/ e /ɛ/ vs. /e:/) (cfr. tabella 1).

Il notevole interesse della digressione sulle vocali danesi deriva soprattutto da due fatti. Essa, da un lato, mostra la presenza in danese di una significativa correlazione tra differenze in F0 e differenze di durata; dall'altro, porta ad escludere che in danese F1 possa avere un ruolo nelle variazioni della frequenza fondamentale.

Le problematiche inerenti al caso danese e al caso tedesco compaiono insieme nel breve paragrafo finale, di cui riporto testualmente il passo conclusivo: "the F0 relations between short and long vowels cannot be explained by the current theories of tongue pull or tongue root advancement, and...they are in direct conflict with these theories when they are also distinguished by degree of tenseness" (Fischer-Jørgensen, 1990, 135).

#### 2.4 Interrogativi nuovi, ipotesi nuove

Difficilmente si potrà contestare la validità della conclusione che si è tratta dall'analisi delle coppie danesi /i/ vs. /i:/ e /e/ vs. /e:/. Con termini più specifici e chiari, parlerei allora di interrelazione tra incremento di F0 e brevità, e ciò nel quadro di un vocalismo tonico caratterizzato da una opposizione distintiva tra vocali brevi e vocali lunghe.

A questo punto conviene fare qualche osservazione di carattere generale. Quando tra due fatti viene individuata una connessione, questa, per quanto sia fondata e sistematica, non deve essere intesa come una relazione causale in cui uno sarebbe l'elemento che domina e l'altro sarebbe l'elemento che è dominato. Anche quando un fattore risulta rilevante, esso non è da ritenere l'unica fonte di spiegazione. Dal punto di vista metodologico con l'uso del vocabolo "interrelazione", si vuole, da una parte, rispecchiare una impostazione che prende le distanze da un rigido determinismo, dall'altra, indicare la scelta della strada della multidimensionalità in cui più fattori possono coagire e interagire. Per esempio la connessione tra F0 e durata è palese, essa però non implica che il valore più elevato di F0 dipenda necessariamente dalla durata breve delle vocali e solo da questa.

Se si è d'accordo su tali linee, il compito di trovare le possibili ragioni e i possibili meccanismi del comportamento della F0 nelle vocali si ripresenta in tutta la sua complessità. La interrelazione tra incremento di F0 e brevità pone domande cruciali che si possono sintetizzare in questo modo: "why short vowels should be produced on a higher F0, and how it is brought about" (Fischer-Jørgensen, 1990: 133).

Il compito sarà differente, dato che, a ben vedere, le problematiche sono differenti da quelle che concernono il fenomeno universale della F0 intrinseca. Gli studiosi, che per molti anni si sono impegnati a cercare di spiegare perché una vocale alta, a parità di condizioni, presenta una F0 più elevata di quella di una vocale bassa, condividono un tratto comune. Intendo dire che il centro delle loro indagini è costituito essenzialmente (se non proprio esclusivamente) dal ruolo della "vowel height", identificata con la "tongue height", correlata a sua volta alla frequenza della prima formante. (cfr. paragrafo 2.1).



La Fischer-Jørgensen (1990) ha analizzato e discusso a fondo una situazione specifica, non studiata prima, che riguarda due lingue germaniche: da un lato la F0 in vocali rilassate-brevi, che si oppongono a vocali tese-lunghe, in tedesco; dall'altro la F0 in vocali brevi, che si oppongono a vocali lunghe, in danese. Questo saggio fa sorgere problemi interpretativi nuovi: perché vocali rilassate-brevi tedesche e vocali brevi danesi sono prodotte con un incremento della frequenza fondamentale? (cfr. paragrafi 2.2 e 2.3).

E' emerso con chiarezza che in vocali rilassate-brevi tedesche e in vocali brevi danesi la F0 presenta, anche se in misura diversa nelle due lingue, un relativo innalzamento del suo valore senza che ciò sia in corrispondenza con una diminuzione del valore di F1. In questa specifica situazione non è quindi adeguato il modello creato per descrivere la IF0 quando di IF0 si tratta: quando cioè, a parità di condizioni, la F0 è più elevata in vocali (le alte) che hanno una F1 più piccola e la F0 è meno elevata in vocali (le basse) che hanno una F1 più grande.

Penso che con il termine "condizione" si debba abbracciare anche lo status linguistico delle vocali sotto esame. Come si è visto, se si osservano separatamente i valori della frequenza fondamentale nella serie delle rilassate-brevi e nella serie delle tese-lunghe del tedesco, all'interno di ciascuna serie si riscontra la attesa correlazione negativa tra la F0 e la F1: "a parità di condizioni" appunto. Il "breakdown" del modello descrittivo della IF0 si ha quando si mettono a confronto vocali appartenenti a due classi fonologiche diverse (cfr. paragrafo 2.2).

Questi fatti e queste considerazioni rendono necessaria una differente prospettiva di lavoro. Bisogna scovare il maggior numero possibile di fattori plausibili e pertinenti che consentano di proporre, anche se in linea provvisoria, una interpretazione consona al nuovo explanandum. Credo che sia opportuno sottolineare che l'oggetto principale di indagine non è più un fenomeno "language-independent" come generalmente si riconosce che sia la F0 intrinseca, ma una fenomenologia che è "language-dependent".

### 3. UNA RICERCA PILOTA SULLE VOCALI BREVI /ø/, /e/, /ɔ/, /a/ VS. LE VOCALI LUNGHE /ø:/, /e:/, /ɔ:/, /a:/ IN FRIGNANESE

#### 3.1 Premessa

Nell'ambito di studi recenti e in corso intorno ad una varietà italo-romanza parlata nel Frignano (Appennino modenese) è emerso un dato acustico che ritengo interessante e nel contempo problematico per vari aspetti. La ricerca pilota svolta finora mostra che i valori della frequenza fondamentale delle vocali brevi sono più elevati rispetto a quelli delle loro controparti lunghe. Nel tentativo di cercare di comprendere questo fenomeno è opportuno non considerarlo in isolamento, ma collegarlo ad altre proprietà che caratterizzano la distinzione tra vocali accentate brevi e vocali accentate lunghe in frignanese.

#### 3.2 Vocali accentate brevi vs. vocali accentate lunghe nel contesto /'p-pa/

In sede accentata il vocalismo del dialetto di Crocette (Pavullo nel Frignano) è costituito di tredici unità distintive: nove lunghe /i:, y:, u:, e:, ø:, o:, ɛ:, ɔ:, a:/, e quattro brevi /e, ø, ɔ, a/. L'opposizione tra le vocali /e, ø, ɔ, a/ e le vocali /e:, ø:, ɔ:, a:/ è operante in tre strutture di parola: /'CV(:)CV/, /CV(:)C/, /CV(:)/.

In queste pagine saranno presentati soltanto i risultati dell'analisi relativa alle suddette vocali in parole e pseudoparole con struttura /'CV(:)CV/, dove la consonante, sia protonica sia postonica, è /p/ e la vocale finale è /a/: quindi /'pVpa/ vs. /'pV:pa/. Ogni parola target,

inserita nella frase cornice “declama .... vivamente”, è stata prodotta cinque volte da un soggetto femminile. Le analisi acustiche sono state eseguite con il sistema MultiSpeech.

Raccolgo nella tabella 2 i valori medi delle misurazioni di sette parametri: la durata, la frequenza della prima formante, la frequenza della seconda formante, la frequenza della prima armonica (F0), l'ampiezza della prima armonica (H1), l'ampiezza della seconda formante (A2), l'ampiezza della terza formante (A3).

|        |       | ø (·)   | e(·)    | ɔ(·)   | a(·)    |
|--------|-------|---------|---------|--------|---------|
| durata | breve | 70.72   | 61.59   | 83.03  | 74.09   |
|        | lunga | 167.91  | 147.96  | 178.33 | 181.37  |
| F1     | breve | 493.33  | 532.00  | 591.33 | 1083.40 |
|        | lunga | 552.00  | 479.00  | 565.50 | 1149.20 |
| F2     | breve | 1460.00 | 2014.00 | 922.00 | 1683.00 |
|        | lunga | 1498.75 | 2597.75 | 889.16 | 1619.20 |
| F0     | breve | 318.46  | 289.92  | 279.57 | 258.32  |
|        | lunga | 295.65  | 244.86  | 241.62 | 217.56  |
| H1     | breve | 58.33   | 55.66   | 55.00  | 45.80   |
|        | lunga | 53.00   | 52.00   | 44.66  | 40.60   |
| A2     | breve | 43.66   | 44.66   | 51.66  | 42.40   |
|        | lunga | 35.00   | 41.66   | 42.66  | 39.20   |
| A3     | breve | 37.00   | 41.66   | 25.33  | 28.40   |
|        | lunga | 25.25   | 35.75   | 18.33  | 14.80   |

Tabella 2: Valori medi di durata (in ms); di F1, F2, F0 (in Hz); di H1, A2, A3 (in dB)(soggetto femminile frignanese).

In generale la tabella 2 fa vedere che ciascuna delle unità classificate come brevi (/ø, e, ɔ, a/) presenta valori differenti rispetto a ciascuna delle corrispondenti unità classificate come lunghe (/ø:, e:, ɔ:, a:/).

### 3.3 Differenze nella durata e nella frequenza delle prime due formanti

Se si osservano i dati della tabella 2 concernenti la durata e la frequenza delle prime due formanti, il quadro che ne emerge non si discosta sostanzialmente da quello delineato in base a precedenti indagini sui dialetti frignanesi di Benedello e di Crocette che sono state pubblicate in varie sedi. L'unica cosa da segnalare è il valore della F1 di /a/ (1083.40 Hz) e di /a:/ (1149.20 Hz) che è superiore, come era da aspettarsi, a quello dei soggetti maschili analizzati in Uguzzoni & Busà (1995): per esempio i valori medi del soggetto SG sono 854.00 Hz in /a/ e 893.00 Hz in /a:/

Sul piano temporale le differenze sono notevoli: la durata delle singole vocali brevi è sempre meno della metà della durata delle loro controparti lunghe. Il rapporto medio V/V: è pari a 42.78, mentre in precedenti ricerche il calcolo della V/V: ratio in parole /'CV(:)CV/ dava risultati varianti nei tre soggetti: 41.84 in RI, 45.61 in GB, 51.49 in SG. L'esame delle durate assolute mostrava escursioni maggiori nelle vocali lunghe rispetto a quelle delle vocali brevi: una conferma della intuizione di Trubetzkoy che considerava “dehnungsfähig” la lunga e “undeihnbar” la breve? (Uguzzoni & Busà, 1995: 23-28).

L'entità delle differenze qualitative rappresentate dai valori di F1 e F2 è più modesta, ma è indubbio che tali differenze sono sistematiche dal punto di vista della loro direzione. Confrontando i dati concernenti le coppie /e/ vs. /e:/, /ɔ/ vs. /ɔ:/, /a/ vs. /a:/, si trova che l'elemento breve differisce sempre dal corrispondente elemento lungo secondo queste linee: in /e/ F1 è maggiore e F2 è minore che in /e:/; in /ɔ/ F1 è maggiore e F2 è maggiore che in /ɔ:/; in /a/ F1 è minore e F2 è maggiore che in /a:/. Pertanto lo spazio acustico delle suddette vocali brevi si configura come meno periferico rispetto a quello delle lunghe corrispondenti. Modalità analoghe si riscontrano in altre lingue che a differenze temporali associano differenze qualitative (Uguzzoni & Busà, 1995: 29-36).

### 3.4 Differenze nell'ampiezza delle prime due formanti

Una trattazione particolareggiata degli altri parametri inseriti nella tabella 2 è stata fatta in articoli recenti a cui rinvio (Uguzzoni, 2006a: 579-588; Uguzzoni, 2006b: 124-134). Ai fini del presente articolo basterà un breve sunto.

Per ogni replica delle vocali in esame sono stati creati uno spettro FFT e uno spettro LPC sovrapponibili l'uno all'altro. Sullo spettro FFT ho misurato: la frequenza della prima armonica (F0); l'ampiezza della prima armonica (H1); l'ampiezza dell'armonica più forte (A2 e A3) che sulla scala frequenze si trova ad essere più vicina al secondo picco formantico (F2) e al terzo picco formantico (F3) del corrispondente spettro LPC. Si considera che A2 e A3 esprimano i valori di ampiezza, rispettivamente, della seconda e della terza formante.

Nella tabella 2 si vede che in tutte le coppie vocaliche i valori di A2 e A3 differiscono a seconda che la vocale appartenga alla classe delle brevi o alla classe delle lunghe. Rispetto all'unità lunga corrispondente, l'unità breve mostra in modo sistematico un'ampiezza maggiore: lo scarto è, nell'ordine, di 8.66 dB, 3.00 dB, 9.00 dB, 3.20 dB per A2 e di 11.75 dB, 5.81 dB, 7.00 dB, 13.60 dB per A3. Sia pure con la cautela richiesta dai limiti di una ricerca pilota, ritengo degna di nota questa differenziazione di intensità associata alla opposizione tra /ø, e, ɔ, a/ e /ø:, e:, ɔ:, a:/ in parole di tipo /'CV(:)CV/.

Che cosa significa precisamente dire che in frignanese le vocali brevi sono più intense delle lunghe dal punto di vista acustico? Occorre sottolineare che qui non si sta parlando dell'intensità quale era concepita e misurata fino alla metà degli anni '90. Le differenze che sono emerse da questa analisi vanno inserite nella cornice di un modo nuovo di intendere e quantificare l'intensità, che mi piace denominare "intensità rivisitata". Come ho illustrato dettagliatamente in altre occasioni (Uguzzoni, 2003; 2006a; 2006b), molti studiosi, in vari settori di indagine, sono passati dalla misurazione dell'energia globale ("overall signal amplitude") alla misurazione dell'energia presente in differenti zone della scala delle frequenze ("frequency-sensitive amplitude measurement").

Tra i numerosi lavori sperimentali condotti con questa impostazione è opportuno ora fare un rapido riferimento ad un notevole studio di Jessen (2002), a cui sono debitrice di importanti stimoli. La ricerca ha come oggetto le vocali tedesche /ɪ, ɛ, ʊ, ɔ, a/ e /i:, e:, u:, o:, a:/, che, riassumendo, si contrappongono in quanto le prime sono rilassate, brevi, "abruptly cut", mentre le seconde sono tese, lunghe, "smoothly cut". Il risultato su cui voglio richiamare l'attenzione è questo: in sede accentata tra le une e le altre intercorrono differenze di intensità che sono significative. Tali differenze non riguardano l'intensità globale dell'intero spettro; si manifestano invece nelle parti dello spettro che corrispondono alla collocazione della seconda e della terza formante. Nel tedesco settentrionale dunque le vocali rilassate-brevi-con taglio brusco, rispetto alle vocali tese-lunghe-con taglio piano,

sono caratterizzate anche da un nuovo attributo: il grado più elevato dell'energia presente nella zona delle medie e alte frequenze. L'espressione più felice per designare questa proprietà acustica è, a mio parere, "mid-to-high frequency emphasis", che in bibliografia si trova in alternativa con altre.

Per tornare alla domanda posta sopra per il caso frignanese, si può precisare che le vocali brevi sono acusticamente più intense delle loro controparti lunghe nel senso che mostrano un grado maggiore di energia nelle parti dello spettro corrispondenti alla seconda e alla terza formante.

### 3.5 Differenze nella frequenza fondamentale

I valori di F0 sono stati ricavati dalla misurazione della frequenza della prima armonica su spettri istantanei FFT. Ma per ciascuna replica delle vocali studiate sono stati eseguiti due tipi di controllo in corrispondenza del punto del segnale di cui si è creato lo spettro FFT: da un lato l'esame del valore che compare nel tracciato del "pitch", dall'altro il calcolo dell'inverso della durata del periodo che compare nella forma d'onda.

Se si confrontano i valori riportati nella tabella 2 per le singole vocali brevi /ø/, /e/, /ɔ/, /a/ con quelli delle vocali lunghe /ø:/, /e:/, /ɔ:/, /a:/ si coglie facilmente una regolarità sistematica: in tutte le coppie la F0 è più elevata nell'elemento breve che nell'elemento lungo. A scopo di comodità costruisco la tabella 3 in cui vengono ritrascritti i dati e per i singoli tipi vocalici è espresso in percentuale l'incremento del valore medio di F0 nella breve rispetto al valore medio di F0 nella lunga.

|    |       | ø(:)   | e(:)   | ɔ(:)   | a(:)   |
|----|-------|--------|--------|--------|--------|
| F0 | breve | 318.46 | 289.92 | 279.57 | 258.32 |
|    | lunga | 295.65 | 244.86 | 241.62 | 217.56 |
|    |       | 7.71%  | 18.40% | 15.70% | 18.73% |

Tabella 3: Percentuali di incremento nei valori medi di F0 (in Hz) delle brevi confrontati con quelli delle lunghe (soggetto femminile frignanese).

Il calcolo della percentuale è consigliato dal fatto che si tratta di produzioni di un soggetto femminile, che, dal punto di vista della F0, sarebbero meno agevolmente comparabili con quelle di soggetti maschili. Secondo la generalizzazione più diffusa, la frequenza fondamentale nelle femmine adulte è in media superiore di un fattore 1.7 a quella dei maschi adulti (Klatt & Klatt, 1990). Se prendiamo questo rapporto come valido punto di riferimento, possiamo aspettarci che, in condizioni uguali, il valore medio di F0 di un soggetto frignanese maschile sia 170.54 Hz nella breve /e/ e 144.03 Hz nella lunga /e:/. Quando per questa coppia, addotta come esempio, si calcola la percentuale di incremento di F0 del membro breve rispetto al membro lungo, l'incremento risulta del 18.40%, che è ovviamente uguale a quello riportato nella tabella 3.

Inserisco ora una figura in cui due spettri istantanei FFT (512 punti), presi come esempi, mostrano che la frequenza della prima armonica è maggiore nella vocale breve /e/ (289.64 Hz) che nella vocale lunga /e:/ (252.27 Hz). Come è riportato nella tabella 2, il valore medio delle cinque repliche di tali vocali è, rispettivamente, 289.92 Hz e 244.86 Hz e l'incremento medio della F0 nel membro breve della coppia è del 18.40%.

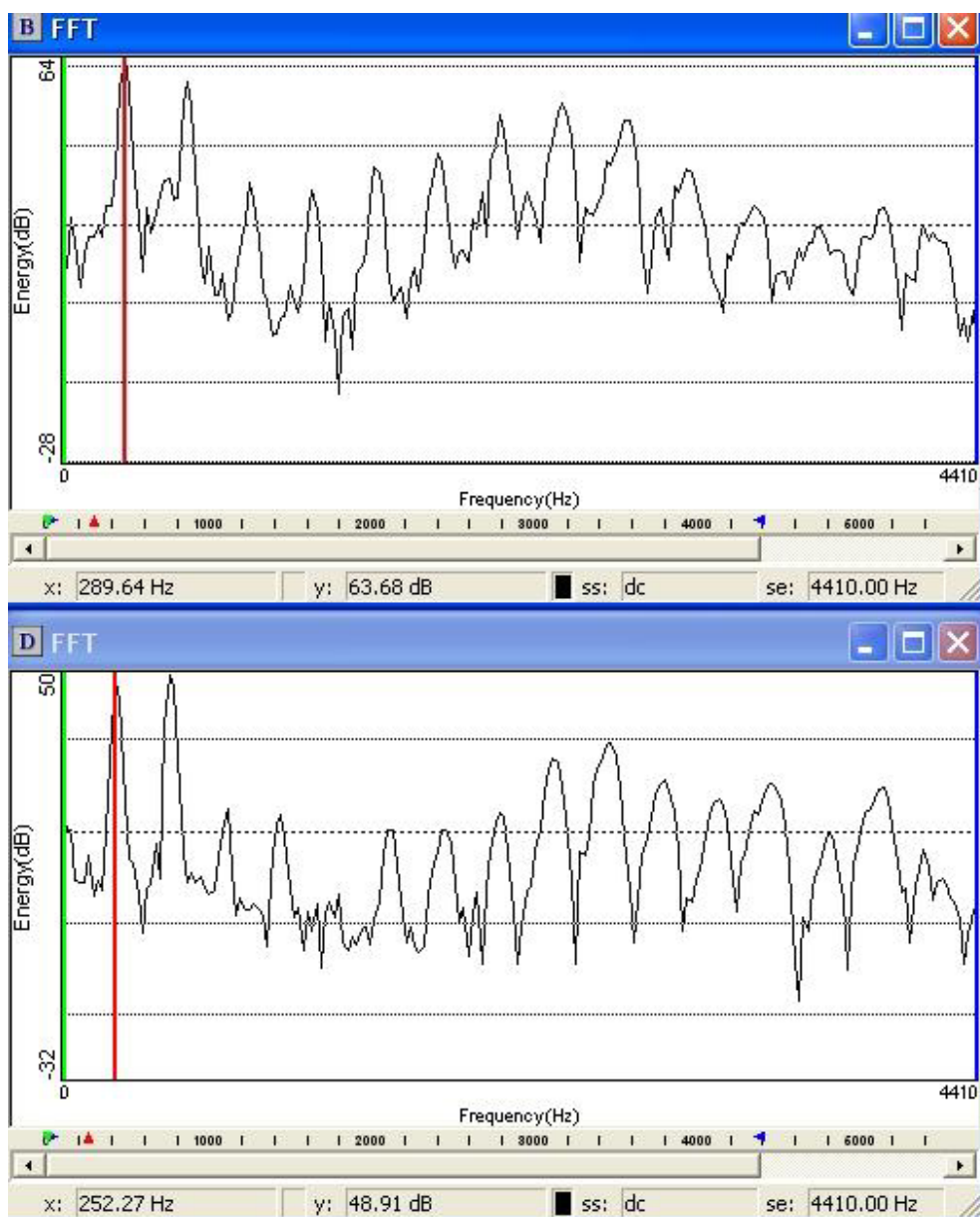


Figura 1: Spettri istantanei FFT della vocale /e/ (sopra) e della vocale /e:/ (sotto)

#### 4. ALCUNE CONSIDERAZIONI

##### 4.1 *Excursus sulla opposizione di quantità vocalica e sulla sua implementazione*

Le divergenze nei valori della frequenza fondamentale presentate in questo articolo (cfr. paragrafi 2.1, 2.2, 3.5) sono associate ad una caratteristica strutturale di tedesco, danese, frignanese che in sede accentata distinguono fonologicamente una classe di vocali brevi da una classe di vocali lunghe. Ragioni espositive mi inducono a parlare per le tre lingue di opposizione di quantità vocalica nel senso tradizionale, pur conoscendo le motivazioni che, per il tedesco settentrionale, hanno portato a postulare invece una opposizione di taglio sillabico : “abruptly cut vowels” vs. “smoothly cut vowels” (Uguzzoni, 2002; 2006a; 2006b; Uguzzoni et al., 2003).

L’opposizione di quantità vocalica è un fenomeno multifacciale e multiparametrico, nel quale sono coinvolte molteplici proprietà che cooperano con le differenze di durata e che si combinano tra di loro in maniera diversa nelle singole lingue. Le condizioni e le modalità in cui e con cui viene realizzata l’opposizione tra vocali brevi e vocali lunghe possono differire nelle varie lingue, come mostrano anche le lingue considerate in questa sede che fanno tutte un uso distintivo della durata.

Una condizione che è comune al danese, al tedesco e al frignanese è il vincolo dell’accento lessicale, per cui l’opposizione breve/lungo si trova solo in sillabe accentate.

Tra le modalità che risultano diverse passando da una lingua all’altra si deve ricordare in primo luogo l’entità delle differenze qualitative che si accompagnano alle differenze di durata. In merito a questo aspetto dai dati riportati si evince la seguente graduatoria: tedesco > frignanese > danese. Il maggiore peso delle differenze di qualità in tedesco è rispecchiato dall’uso degli aggettivi rilassato e teso abbinati talora agli aggettivi breve e lungo (cfr. paragrafi 2.2, 2.3, 3.3).

Un altro concomitante fonetico è costituito dalla possibilità che la consonante successiva alla vocale accentata sia più lunga quando è preceduta da vocale breve rispetto a quando è preceduta da vocale lunga. Da una ricerca sul tedesco settentrionale risulta che sia in monosillabi sia in bisillabi tutti i soggetti mostrano “longer consonants after short vowels than after long vowels, the general average being a prolongation of 27%” (Fischer-Jørgensen & Jørgensen, 1969: 65). Invece in danese “consonants are not usually longer after short vowels than after long vowels” (Fischer-Jørgensen & Jørgensen, 1969: 69). In frignanese un allungamento della consonante successiva a vocale breve si constata in parole monosillabiche (/CVC/), ma non in parole bisillabiche (/’CVCV/) (Uguzzoni & Busà, 1995; Uguzzoni, 2006a; Uguzzoni, 2006b).

In connessione con la distinzione fonologica tra brevi e lunghe si osservano valori differenti nell’intensità acustica delle vocali. Per esempio in parole bisillabiche ungheresi l’intensità delle brevi (e rilassate) è maggiore di quella delle loro controparti lunghe (e tese): si va da un minimo di 1.20 dB (nella coppia /ø/ vs. /ø:/) a un massimo di 2.60 dB (nella coppia /y/ vs. /y:/) (Fonagy, 1966). In danese valori maggiori di intensità compaiono nell’elemento breve delle coppie /i/ vs. /i:/ e /e/ vs. /e:/ (Fischer-Jørgensen, 1990: 134).

Nell’ambito di un nuovo modo di concepire e misurare l’intensità del segnale è risultato che nel tedesco settentrionale e nel frignanese le vocali brevi si differenziano dalle corrispondenti lunghe per il fatto che le prime sono caratterizzate dalla presenza di un grado superiore di energia nelle zone alte dello spettro: “mid-to-high frequency emphasis” (cfr. paragrafo 3.4).

Una caratteristica fisiologica che si collega alla opposizione breve/lungo è il flusso dell’aria. In parole /CV(:)C/ e /’CV(:)CV/ del tedesco e del danese si è riscontrato che al

momento della transizione tra la vocale e la consonante successiva esso è più forte dopo vocale breve. La maggiore forza del flusso dell'aria non è limitata alla implosione, si trova infatti in tutta la vocale accentata breve e anche nella consonante iniziale della parola. Riguardo al flusso dell'aria, l'entità della differenza tra parole con vocale breve e parole con vocale lunga in danese è più piccola che in tedesco : "in German the difference may be enhanced by the laxness of the short vowels" (Fischer-Jørgensen & Jørgensen, 1969: 69). Mi pare che ciò costituisca una chiara illustrazione del concetto di collaborazione e di interazione tra parametri, visto come un fatto "language-dependent".

#### *4.2 Implicazioni per il problema della F0 in vocali accentate brevi*

La interrelazione tra l'incremento della frequenza fondamentale e la brevità della vocale accentata può essere collocata a buon diritto nella cornice generale ora delineata. Come si è visto, in tedesco, in danese, in frignanese le vocali brevi sono prodotte con valori di F0 che sono più elevati rispetto a quelli delle corrispondenti lunghe. Per quel che concerne l'entità della differenza i dati riportati sembrano indicare una graduatoria di questo genere: frignanese > danese > tedesco (cfr. paragrafi 2.1, 2.2, 3.5).

Chiedersi per quale ragione la F0 dovrebbe essere più elevata nella breve o per quale motivo tale innalzamento sarebbe importante per la breve è legittimo (cfr. paragrafo 2.4); ma penso che questi interrogativi vengano ridimensionati se il fenomeno in questione non è visto in modo isolato, ma è inserito in quella sorta di comprensorio plurimo che è la implementazione di una vocale fonologicamente breve.

Le suddette domande potrebbero essere fatte anche riguardo agli altri concomitanti fonetici menzionati nell'exkurs: la tendenza ad una configurazione formantica di tipo non periferico, un certo allungamento della consonante successiva, una intensità maggiore, un flusso d'aria più forte.

Queste considerazioni non implicano che debba essere trascurata la diversa misura in cui si manifesta, quando si manifesta, la presenza addizionale di questa o di quella proprietà. Sono convinta che nel caso della opposizione di quantità vocalica, come accade per altri fatti linguistici, ogni lingua ingaggi per così dire una serie, più o meno ampia, di dimensioni che simultaneamente, in misura più o meno grande, collaborano allo stesso fine: il rafforzamento della distinzione breve/lungo.

D'altra parte andrebbero fatti adeguati approfondimenti su tale combinazione di proprietà: per esempio sui modi spesso complessi in cui un fattore interagisce con l'altro, sui meccanismi per cui un fattore può compensare l'altro, sulle forme di influsso unilaterale o bilaterale.

Una volta che siano stati chiariti i vari aspetti a cui si è accennato, la interpretazione della frequenza fondamentale, considerata in relazione con lo status breve della vocale, potrebbe risultare differente a seconda delle lingue. Penso alle connessioni possibili tra F0 e intensità, tra F0 e flusso dell'aria, tra F0 e pressione sottoglottidale.

Richiamando le osservazioni fatte sopra (cfr. paragrafo 4.1), nelle vocali brevi delle lingue considerate l'incremento della frequenza fondamentale coesiste, da un lato, con un grado maggiore di intensità (tedesco, danese, frignanese), dall'altro con una forza maggiore del flusso dell'aria (tedesco, danese). E' stata riferita con cautela l'ipotesi che "the intensity of short vowels is increased by heightened subglottal pressure in order to compensate for their short duration, and a higher subglottal pressure might also increase F0" (Fischer-Jørgensen, 1990: 134). Si sa da tempo che una maggiore pressione sottoglottidale può influire sia sul valore dell'intensità sia sul valore della F0, ma si hanno poche conoscenze

riguardo a differenze di pressione sottoglottidale associate alla distinzione vocale breve/vocale lunga.

## RINGRAZIAMENTI

Sono molto grata a Stephan Schmid, Nadia Nocchi, Enzo Lumiella, Luciano Romito, Gabriele Azzaro.

## BIBLIOGRAFIA

- Bucella, F.; Hassid, S.; Beeckmans, R.; Soquet, A.; Demolin, D., 2000. Pression sous-glottique et débit d'air buccal des voyelles en français. In *Actes des XXIIIèmes Journées d'Etudes sur la Parole*, Aussois, 449-452.
- Diehl, R.L.; Kluender, K.R., 1989. On the objects of speech perception. *Ecological Psychology*, 1, 121-144.
- Dyhr, N., 1990. The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. *Phonetica*, 47, 141-154.
- Fahey, R.P.; Diehl, R.L., 1996. The missing fundamental in vowel height perception. *Perception and Psychophysics*, 58, 725-733.
- Fant, G., 1960. *Acoustic theory of speech production*. The Hague: Mouton.
- Ferrero, F. E., 1997. Semiologia avanzata della funzione vocale (laringe e condotto vocale). *Quaderni del Centro di Studio per le Ricerche di Fonetica*, 16, 192-210.
- Fischer-Jørgensen, E., 1972. Formant frequencies of long and short Danish vowels, In E. Fierchow et al. (a c. d), *Studies for Einar Haugen*, The Hague: Mouton, 189-213.
- Fischer-Jørgensen, E., 1990. Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica*, 47, 99-140.
- Fischer-Jørgensen, E.; Jørgensen, H. P., 1969. Close and loose contact (Anschluss) with special reference to North German. *Annual Report of the Institute of Phonetics of the University of Copenhagen*, 4, 43-80.
- Fónagy, I., 1966. Electro-physiological and acoustic correlates of stress and stress perception. *Journal of Speech and Hearing Research*, 9, 231-244.
- Gandour, J.; Weinberg, B., 1980. On the relationship between vowel height and fundamental frequency: evidence from esophageal speech. *Phonetica* 37, 344-354.
- Hanson, H. M., 1997. Glottal characteristics of female speakers: acoustic correlates. *Journal of the Acoustical Society of America*, 101, 466-481.
- Hirose, H., 1997. Investigating the physiology of laryngeal structures. In Hardcastle, W.J. e J. Laver (a c.d.), *The Handbook of Phonetic Sciences*, Oxford: Blackwell, 116-136.
- Holmberg, E. B.; Hillman, R. E.; Perkell, J. S.; Guiod, P.; Goldman, S. L., 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research*, 38, 1212-1223.



- Honda, K., 1983. Relationship between pitch control and vowel articulation. In Bless, D. e J. Abbs (a c. d.), *Vocal fold physiology: Contemporary research and clinical issues*, San Diego: College-Hill Press, 286-297.
- Honda, K., 1995. *Laryngeal and extra-laryngeal mechanisms of F0 control*. In Bell-Berti, F. e L.J. Raphael (a c. d.), *Producing speech: Current issues*, New York: American Institute of Physics, 215-232.
- Honda, K.; Maeda, S.; Sugito, M., 2007. Articulation changes in different voicing patterns. In *Proceedings of the XVI International Congress of Phonetic Sciences*, Saarbrücken, 237-240 (CD-rom).
- Hoole, P.; Mooshammer, C., 2002. Articulatory analysis of the German vowel system. In P. Auer et al. (a c.d.), *Silbenschnitt und Tonakzente*, Tübingen: Niemeyer, 129-152.
- Jessen, M., 2002. Spectral balance and its relevance for syllable cut theory. In P. Auer et al. (a c. d.), *Silbenschnitt und Tonakzente*, Tübingen: Niemeyer, 153-179.
- Kingston, J., 1992. The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech*, 35, 99-113.
- Kingston, J.; Diehl, R.L., 1994. Phonetic knowledge. *Language*, 70, 419-454.
- Klatt, D.; Klatt, L., 1990. Analysis, synthesis, and perception of voice quality variation among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.
- Ladefoged, P., 1967. *Three areas of experimental phonetics*. London: Oxford University Press.
- Ladefoged, P., 2003. *Phonetic data analysis. An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.
- Ladefoged, P., 2005. Speculations on the control of speech. In Hardcastle, W.J. e J. Beck (a c. d.), *A Figure of Speech: A Festschrift for John Laver*, London: Lawrence Erlbaum Associates, 3-21.
- Lehiste, I., 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
- Ohala, J.J.; Eukel, B. 1987. Explaining the intrinsic pitch of vowels. In Channon, R. e L. Shockey (a c. d.), *In Honor of Ilse Lehiste*, Dordrecht: Foris, 207-215.
- Pettorino, M., 1987. Intrinsic pitch in vowels: an experimental study on Italian. In *Proceedings of the XI International Congress of Phonetic Sciences*, Tallinn, 1, 138-141.
- Repetti, L., 1992. Vowel length in Northern Italian dialects. *Probus*, 4, 155-182.
- Sawashima, M.; Hirose, H., 1983. Laryngeal gestures in speech production. In MacNeilage, P.F., (a c.d.), *The production of speech*, New York: Springer-Verlag, 11-38.
- Shadle, C. H., 1997. The aerodynamics of speech. In Hardcastle, W.J. e J. Laver (a c.d.), *The Handbook of Phonetic Sciences*, Oxford: Blackwell, 33-64.
- Sluijter, A. M. C.; van Heuven, V. J., 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100, 2471-2485.
- Spiekermann, H. 2000. *Silbenschnitt in Deutschen Dialekten*. Tübingen: Niemeyer.

- Stevens, K. N., 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Traunmüller, H. 1981. Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69, 1465-1475.
- Traunmüller, H.; Eriksson, A., 2000. Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America*, 107, 3438-3451.
- Trubetzkoy, N. S., 1939. Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague*, 7, 1-268.
- Uguzzoni, A., 2002. Fester vs. loser Anschluss. Appunti per una storia di un concetto secolare. *Lingue e Linguaggio*, 1, 327-340.
- Uguzzoni, A., 2003. In margine ad una rivisitazione della intensità. In P. Cosi et al. (a c. d.), *Voce, canto, parlato. Studi in onore di Franco Ferrero*, Padova: Unipress, 299-302.
- Uguzzoni, A., 2006a. I valori di H1-A2 e H1-A3 come correlati della intensità “rivisitata”. Aspetti e problemi. In *Atti del 2° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce*, Fisciano (Salerno), 30 novembre-2 dicembre, 2005, 566-592 (CD-rom).
- Uguzzoni, A., 2006b. Produzione, acustica, percezione della “intensità rivisitata”. Ricerche in area germanica e in area italo-romanza. *Rivista Italiana di Dialettologia*, 30, 103-138.
- Uguzzoni, A.; Azzaro, G.; Schmid, S., 2003. Short vs long and/or abruptly vs smoothly cut vowels. New perspectives on a debated question. In *Proceedings of the XV International Congress of Phonetic Sciences*, Barcelona, 3, 2717-2220.
- Uguzzoni, A.; Busà, M. G., 1995. Correlati acustici della opposizione di quantità vocalica in area emiliana. *Rivista Italiana di Dialettologia*, 19, 7-39.
- Whalen, D. H.; Levitt, A. G., 1995. The universality of intrinsic  $F_0$  of vowels. *Journal of Phonetics*, 23, 349-366.
- Whalen, D.; Gick, B.; Kumada, M.; Honda, K., 1998. Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic  $F_0$ . *Journal of Phonetics*, 27, 125-142.
- Zmarich, C.; Uguzzoni, A.; Ferrari, V., 2003. Controllo articolatorio della opposizione di quantità vocalica in area emiliana: analisi cinematica dei gesti labiali. In *Atti delle XIII Giornate di Studio del Gruppo di Fonetica Sperimentale*, Pisa: Edizioni ETS, 295-306.

# L'ETÀ DELLA VOCE

Antonella Giannini - Massimo Pettorino  
Università degli Studi di Napoli L'Orientale  
[agiannini@unior.it](mailto:agiannini@unior.it) [mpettorino@unior.it](mailto:mpettorino@unior.it)

## 1. SOMMARIO

Numerosi lavori sperimentali hanno dimostrato che, al variare dell'età, la voce subisce profondi cambiamenti sia sul piano segmentale che soprasegmentale. Tuttavia uno dei punti più controversi di questo tipo di ricerche è nella scelta del corpus, in quanto non è stata mai condotta un'analisi su uno stesso parlante che, in età diverse, abbia prodotto il medesimo testo. Per superare questa difficoltà, seguendo una metodologia già sperimentata in altri nostri lavori, abbiamo proceduto ad un esperimento in cui un noto giornalista televisivo, Piero Angela, ha letto, nella stessa situazione contestuale e a distanza di 40 anni, lo stesso testo.

I dati raccolti evidenziano un decremento della velocità di articolazione, della velocità di eloquio e della fluenza. Il numero e la durata media delle pause di respirazione presentano sostanziali variazioni. Le durate delle sillabe, sia toniche che atone, subiscono un incremento anche se in misura diversa a seconda della tipologia sillabica. Il tono è mediamente più alto e più variato e i pattern formantici delle vocali mostrano, per le vocali posteriori, un ampliamento ed un arretramento dell'area vocalica.

## 2. INTRODUZIONE

Il complesso rapporto tra caratteristiche della voce ed età del parlante è stato studiato in numerosi lavori sperimentali. Varie indagini, condotte sia a livello acustico che percettivo, hanno dimostrato che, al variare dell'età, la voce subisce profondi cambiamenti sia sul piano segmentale che soprasegmentale. Le principali cause di tali cambiamenti sono da attribuirsi ad alterazioni di ordine fisiologico e anatomico che coinvolgono non solo i polmoni, ma anche la laringe ed il canale epilaringeo. Con l'avanzare degli anni si determina, infatti, un decremento delle funzioni respiratorie causato da vari fattori, tra i quali un rilassamento dei tessuti polmonari, una ridotta capacità vitale, un irrigidimento delle pliche vocali che diventano più sottili, più rigide e meno elastiche, la calcificazione di alcune cartilagini della laringe, nonché un irrigidimento complessivo dell'apparato fonatorio che coinvolge, in particolar modo, le labbra, la lingua e la mandibola (Linville 2004; Awan 2006, Strom et alii 2006). Tali alterazioni si riflettono, in diversa misura, sui parametri acustici che caratterizzano il parlato.

Ad esempio Hollien & Ship (1972) trovano, in uno studio condotto su 175 parlanti americani, che la  $f_0$  di una voce maschile diminuisce mediamente, tra i 20 e i 69 anni, da 120 a 112 Hz (anche se in modo non progressivo) per poi aumentare, fino a raggiungere i 146 Hz, dagli 80 agli 89 anni. Pegoraro-Krook (1988), in uno studio condotto su 198 parlanti svedesi, rileva un decremento di  $f_0$  tra i 20 e i 59 anni (da 112 a 109 Hz) con successivo incremento che raggiunge il valore massimo (124 Hz) tra i 70 e i 79 anni. Per la voce femminile, al contrario, il cambiamento della  $f_0$  sembra essere strettamente connesso al periodo della menopausa. In tale periodo, infatti, la  $f_0$  inizia a decrescere fino a raggiungere, per Pegoraro-Krook (1988), il massimo decremento, tra i 60 e i 69 anni, di circa 15 Hz, per Stoiceff (1981) di circa 19 Hz e, per Russell et alii, (1995) di circa 48 Hz. Le variazioni, comunque, sembrano collegate, più che a fattori cronologici, a motivi di ordine fisiologico.

Risultati contrastanti sono stati ottenuti anche a livello percettivo. Ad esempio alcuni autori hanno dimostrato che per la voce maschile “the listeners tended to associate higher mean  $f_0$  with more advanced age and vice versa” e “the mean  $f_0$  is one of the cues that influence the perception of aging in the voice” (Hori & Ryan 1981, p. 231). Altri, come Hartman & Danhauer (1976), hanno trovato che è il *low pitch* che caratterizza la voce senile, altri ancora affermano che “spectral cues are probably more important than the prosodic cues of  $F_0$  and duration” (Schötz 2003, p. 4).

L’invecchiamento della voce dipende, quindi, da numerosi fattori e varia notevolmente da persona a persona: alcuni individui possono mantenere un suono di voce stabile fino ad 80 anni, mentre altri iniziano a cambiarla verso i 50 anni. Il confronto fra i dati, inoltre, come è stato sottolineato da Russell et alii (1995) e da Price (2006), è reso ancor più complicato a causa delle diverse metodologie adoperate, dai tipi di corpora utilizzati, dalle lingue considerate, nonché dai cambiamenti diacronici del parlato. Di conseguenza è alquanto difficile stabilire quali siano i parametri maggiormente coinvolti e ricavare conclusioni univoche.

Nonostante ciò sembra esserci una certa concordanza di opinioni sul fatto che, in generale, una voce, col trascorrere del tempo, sia soggetta ad un:

- rallentamento articolatorio
- progressivo abbassamento tonale
- rallentamento della velocità di eloquio
- incremento di jitter e shimmer e della percentuale di rumore
- abbassamento delle frequenze formantiche
- allungamento delle vocali e della tenuta della occlusive
- aumento della deviazione standard di  $f_0$

Uno dei punti deboli di queste ricerche è che non vi è mai la coincidenza a livello di enunciato prodotto, nel senso che non è stata mai condotta un’analisi su uno stesso parlante che, in età diverse, abbia prodotto il medesimo testo<sup>1</sup>. I materiali più largamente usati per le indagini sono essenzialmente di due tipi: o uno stesso testo (brani, frasi, parole o vocali isolate) letto da diversi parlanti appartenenti a fasce d’età diverse, o un corpus di parlato spontaneo (racconti in risposta ad una domanda), sempre prodotto da parlanti di diverse fasce d’età. Tuttavia a nostro avviso tali metodologie non assicurano che i risultati delle analisi siano legati esclusivamente alla diversa età del/dei parlante/i. Va infatti considerato che il tipo di parlato che utilizziamo dipende da numerose variabili, quali, ad esempio, la situazione contestuale, il rapporto con l’interlocutore, lo stile formale/informale, l’argomento, il parlato letto o spontaneo. Per poter confrontare parlati diversi e poter attribuire le cause delle eventuali differenze ad uno specifico parametro, è necessario che tutte le altre variabili siano state precedentemente azzerate: solo in questo modo, infatti, è lecito concludere che le variazioni riscontrate sono state determinate proprio da quell’unico parametro lasciato in libera variazione. Ma è possibile condurre un esperimento del genere? Occorrerebbe che uno stesso parlante si ritrovasse, a distanza di molti anni, a pronunciare le stesse frasi, nella medesima situazione comunicativa, ad uno stesso interlocutore con il

---

<sup>1</sup> Unica eccezione Russell et alii (1995) i quali mettono a confronto due frasi registrate nel 1945 e nel 1993 dalle stesse donne australiane. L’analisi si limita allo studio dei cambiamenti della frequenza fondamentale.

quale abbia lo stesso grado di familiarità. Molto improbabile ma, forse, non impossibile. È quanto abbiamo provato a fare in questo esperimento.

Seguendo una metodologia già sperimentata in nostri precedenti lavori, abbiamo proceduto ad un esperimento in cui un noto giornalista televisivo, Piero Angela, ha letto, nella stessa situazione contestuale e a distanza di 40 anni, lo stesso testo da lui letto nel corso di un telegiornale del 1968. Grazie alla collaborazione delle TECHE RAI, ci siamo procurati un telegiornale “antico”, degli anni ‘60, nel quale le notizie erano lette in studio da più speakers. Abbiamo quindi estratto le notizie presentate da Piero Angela, a quel tempo quarantenne, e le abbiamo trascritte ortograficamente dotandole della punteggiatura che, sintatticamente e percettivamente, ci sembrava più adeguata. Il testo così trascritto è stato letto, presso uno studio di registrazione della sede RAI di Saxa Rubra, dallo stesso parlante che, con grande spirito di collaborazione, ha acconsentito a simulare un nuovo telegiornale. Va detto che tutto si è svolto cercando di rispettare il modello originale. Ad esempio quando nel testo era previsto un servizio esterno, il parlato veniva interrotto per dare spazio, allora come ora, ad un corrispondente esterno, anche se in maniera puramente virtuale.

Questo artificio ci ha permesso di azzerare le numerose variabili dovute sia al testo che al parlante e di ottenere due registrazioni, in un certo senso sovrapponibili, che consentono un confronto puntuale, anche se limitato ad un solo parlante.

### 3. MATERIALE E METODO

Dal corpus così ottenuto abbiamo selezionato brani di parlato letto da Piero Angela a distanza di quarant’anni ([un esempio](#)). L’analisi acustica è stata effettuata mediante il software *Wavesurfer* che consente la visualizzazione dello spettrogramma a banda larga e dell’andamento intonativo. Complessivamente sono stati analizzati 350 secondi di parlato, sono state computate 1724 sillabe distribuite in 128 catene foniche. Per ciascuna porzione di parlato sono stati misurati i seguenti parametri:

- durata della catena fonica
- durata delle pause silenti
- durata sillabica
- durata della vocale
- durata consonantica
- $f_0$  min/max per ciascuna catena fonica
- andamento intonativo della frase ( $f_0$  sillaba per sillaba)
- F1 e F2 delle vocali toniche e atone.

In questa ricerca abbiamo fissato la soglia minima delle pause silenti a 80 ms. Tale soglia è stata portata a 150 ms quando le pause erano seguite da occlusive sorde.

Dai dati ottenuti sono stati calcolati i seguenti indici:

- velocità di articolazione, ottenuta dividendo il numero di sillabe per la durata delle catene foniche;
- velocità di eloquio, calcolata dividendo il numero di sillabe per la durata totale dell’enunciato;
- fluenza, data dal rapporto tra numero di sillabe e numero di catene foniche
- composizione dell’enunciato, calcolata in valori percentuali tra porzione di parlato e porzione di silenzio.

## 4. RISULTATI E DISCUSSIONE

### 4.1 Indici prosodici e composizione dell'enunciato

In Tav. I sono riportati i dati relativi alla velocità di articolazione (VDA), alla velocità di eloquio (VDE) e alla fluenza. Come si vede, i dati del 2007 mostrano, rispetto a quelli del 1968, un decremento di circa una sillaba al secondo sia per la VDA che per la VDE e di circa tre sillabe per quanto riguarda la fluenza. Con il trascorrere degli anni il parlato di Piero Angela presenta, quindi, una decelerazione, strettamente connessa all'aumento della durata sillabica, un incremento della durata e del numero delle pause non silenti e un accorciamento delle catene foniche.

|                | 1968 | 2007 |
|----------------|------|------|
| VDA            | 6,3  | 5,5  |
| VDE            | 5,3  | 4,5  |
| <u>fluenza</u> | 15   | 12   |

Tabella I: Dati relativi alla velocità di articolazione e di eloquio e alla fluenza.

Nonostante le differenze rilevate, se andiamo a considerare la composizione complessiva dei due enunciati, prodotti a distanza di circa 40 anni, possiamo notare come i due elinqui appaiono produttivi in egual misura in quanto vi è una differenza minima, pari al 2%, tra la percentuale di parlato e di silenzio (figura 1). Tale dato indica che con l'età, l'aumento in durata delle sillabe e quella dei silenzi avviene in maniera proporzionale e bilanciata. Questo dato ci sembra particolarmente interessante, in quanto è il primo, tra quelli che abbiamo riscontrato nel corso dell'esperimento, a testimoniare non tanto le "differenze" dovute all'età, quanto piuttosto le "somiglianze" che persistono nonostante l'età. Le sillabe si allungano, la velocità diminuisce, i silenzi, come vedremo, diventano più frequenti e più lunghi, ma alla fine la percentuale di silenzio rispetto alla porzione di parlato resta immutata. Quasi come se vi fosse un modello di parlato che il parlante adotta in una certa fase della sua vita e che, anche se a livello inconsapevole, continua a rispettare negli anni. Ma su questo punto torneremo più avanti.

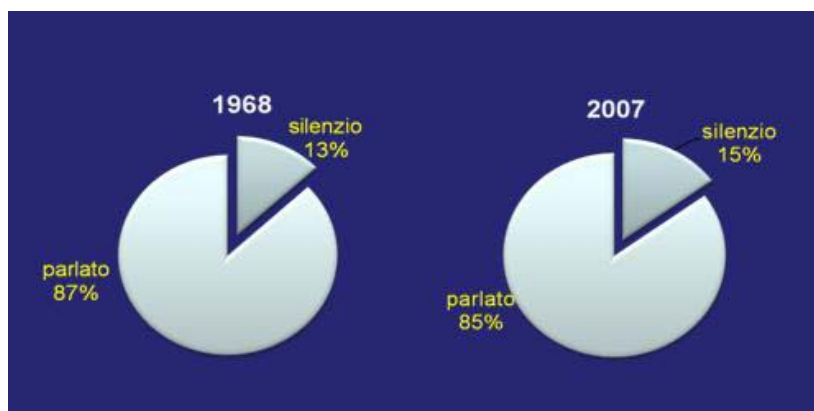


Figura 1: Composizione dell'enunciato

A questo punto è lecito porsi una domanda: le differenze riscontrate possono essere attribuite esclusivamente ai cambiamenti dovuti all'età o, trattandosi nel nostro caso di uno speaker televisivo, possono dipendere da un cambiamento più generale, occorso nello stile del parlato trasmesso? Se andiamo a confrontare i dati relativi agli indici prosodici e alla composizione dell'enunciato di Angela nel '68 con quelli da noi riscontrati in altri lavori condotti sul parlato dell'informazione televisiva degli anni '60, vediamo che essi concordano pienamente, mentre i dati relativi al 2007 si discostano totalmente da quelli rilevati negli odierni telegiornali caratterizzati da un incremento della VDA, della VDE e della fluenza (Giannini 2004, 2006; Pettorino & Giannini 2005).

Tali indici prosodici, d'altronde, trovano conferma con quanto riportato in letteratura dove spesso si fa riferimento alla velocità di eloquio e di articolazione come parametri discriminanti l'età, seppur in maniera impressionistica e non quantitativa. A volte, infatti, si parla genericamente di *rapid rate* o *slow rate* come parametro percettivamente distintivo tra parlanti giovani e anziani (Hartman & Danhauer 1976, p. 714; Ptacek & Sander 1966), altre volte si fa riferimento alla velocità calcolata in base al numero di parole prodotte in un certo intervallo di tempo (Mysak 1959)<sup>2</sup> e, altre volte ancora, ci si riferisce esclusivamente alla durata delle vocali: "...the older subjects produced vowels that were longer than those produced by the younger speakers... Consequently, the lengthening of vowels by older speakers may contributed to the slow rate of speech that is generally perceived in the speech of the elderly" (Benjamin 1982, pp. 162-163). Certamente un allungamento vocalico corrisponde ad una durata maggiore della sillaba con conseguente rallentamento della VDA, ma nessun dato quantitativo accompagna questi risultati. Gli indici prosodici da noi calcolati possono, quindi, essere confrontati solo in maniera generale con quelli ottenuti negli altri lavori.

#### 4.2 Le sillabe

Nei due enunciati prodotti da Piero Angela quarantenne e ottantenne, la durata media della sillaba è, rispettivamente, di circa 158 ms e a 188 ms. Tale aumento di durata sillabica riscontrato nel 2007, coinvolge in egual misura tutte le sillabe o varia a seconda della loro

<sup>2</sup> Mysak trova che con l'avanzare dell'età, il numero di parole lette in un minuto scende da circa 172 a 123, mentre nel parlato spontaneo il numero di parole sale da circa 122 a 128.

tipologia? E ancora, gli elementi consonantici e vocalici contribuiscono in egual misura all'incremento sillabico o, come è stato dimostrato per lingue diverse dall'italiano, è la durata delle vocali (soprattutto quelle toniche), a discriminare maggiormente l'età del parlante?

Per rispondere alla prima domanda abbiamo calcolato, in ms, le durate medie delle sillabe toniche e atone, aperte e chiuse e il loro incremento percentuale (figg. 2-3).

Il confronto evidenzia come le sillabe subiscano tutte un incremento in durata anche se in misura diversa a seconda della loro tipologia: il minor incremento è presente nelle sillabe toniche chiuse (11%), mentre il maggiore incremento nelle sillabe atone aperte (24%). Tale dato ci permette di affermare che l'eloquio di Piero Angela ottantenne tende ad essere più isocrono rispetto a quello prodotto da Angela quarantenne, in quanto vengono ridotte le differenze in durata delle sillabe.

Per rispondere alla seconda domanda riportiamo, in figura 4, i valori percentuali della durata consonantica e vocalica, suddivisi per posizione e complessità sillabica.

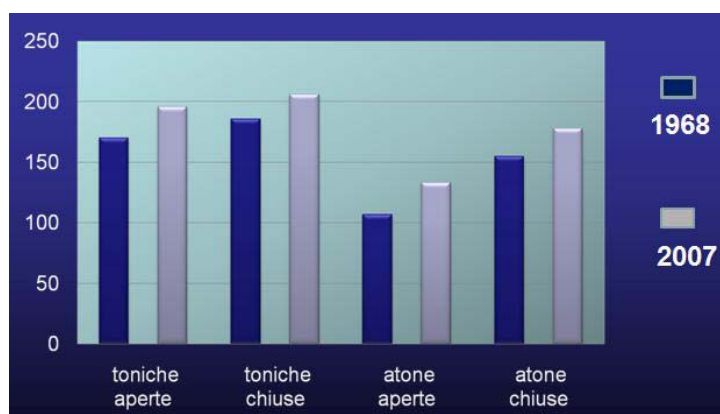


Figura 2: Durata media delle sillabe in ms

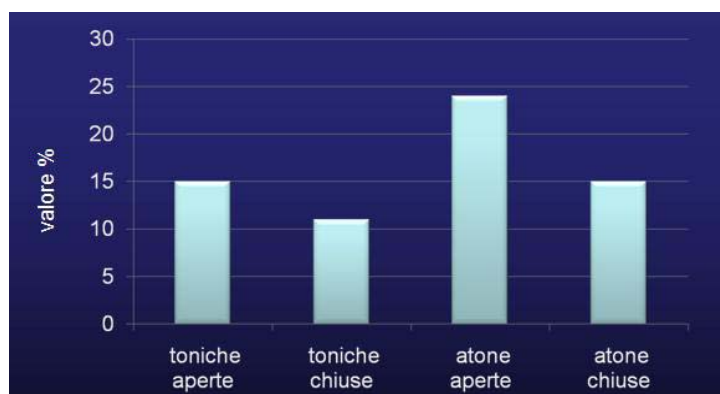


Figura 3: Incremento percentuale di durata



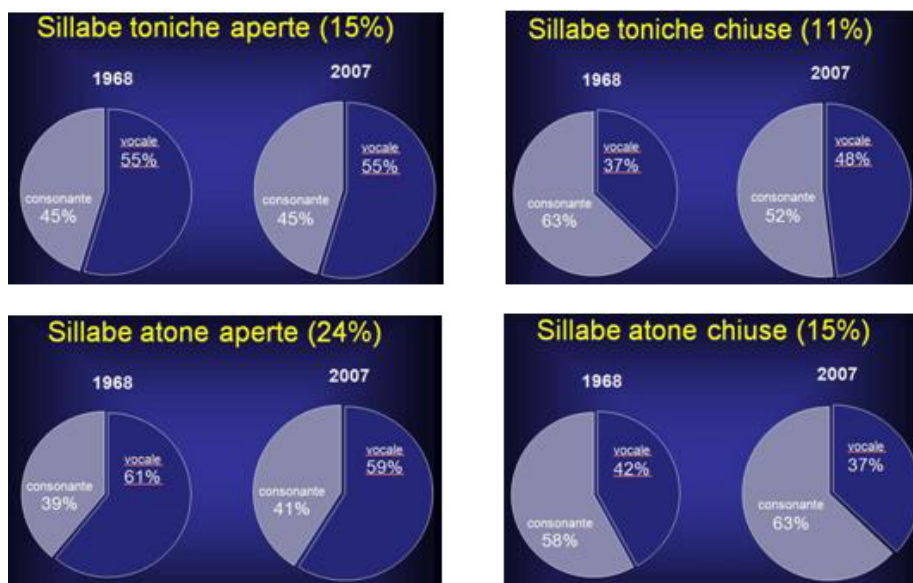


Figura 4: Distribuzione percentuale degli elementi vocalici e consonantici in base alla tipologia e posizione sillabica

Come si nota, la variazione temporale a favore dell'elemento vocalico interessa, in maniera significativa, soltanto le sillabe toniche chiuse che, come abbiamo visto, presentano, complessivamente, il minore incremento percentuale (fig. 3) mentre, nelle altre tipologie, l'incremento è equamente distribuito tra porzione vocalica e consonantica (uguale incremento nelle toniche aperte e variazioni non significative nelle sillabe atone).

#### 4.3 Le pause silenti

La VDE e la fluenza sono strettamente correlate alla durata delle pause e al loro numero. In tab. II-III sono riportati i dati relativi al numero delle pause e alla loro durata media, divisi per tipologia.

|                                         | 1968      | 2007      | %          |
|-----------------------------------------|-----------|-----------|------------|
| <b>numero pause di respirazione</b>     | <b>38</b> | <b>52</b> | <b>+36</b> |
| <b>numero pause non di respirazione</b> | <b>20</b> | <b>18</b> | <b>-10</b> |
| <b>Totale silenzi</b>                   | <b>58</b> | <b>70</b> | <b>+20</b> |

Tabella II: Numero di pause silenti e relativa variazione percentuale.

Le pause di respirazione aumentano del 36%, mentre diminuiscono le pause sintattiche. Va detto che le pause sintattiche vengono spesso sostituite e quindi sfruttate come pause di respirazione, il che sta a significare che con l'età vi è bisogno di ispirare più di frequente,

bisogno strettamente connesso, come abbiamo già detto, alla minore funzionalità del mantice polmonare.

Un'ulteriore differenza si riscontra nella durata delle pause. Infatti, sebbene entrambe le tipologie subiscano nel 2007 un incremento, ancora una volta l'incremento maggiore si ha nelle pause di respirazione (30% vs 18%).

|                                               | 1968       | 2007       | %          |
|-----------------------------------------------|------------|------------|------------|
| <b>durata media pause di respirazione</b>     | <b>357</b> | <b>466</b> | <b>+30</b> |
| <b>durata media pause non di respirazione</b> | <b>166</b> | <b>196</b> | <b>+18</b> |

Tabella III: Durata media delle pause in ms

Le pause silenti sono quindi uno degli elementi discriminanti dell'età. Hartman e Danhauer (1976), in una ricerca condotta sul parlato spontaneo prodotto da soggetti maschili caucasici, affermano che la maggiore durata delle pause è, dal punto di vista percettivo, uno dei parametri più importanti per il riconoscimento dell'età del parlante, dato confermato da Oyer e Deal (1985) per il parlato letto. Questi autori sottolineano come non solo la durata media, ma anche il tempo totale delle pause e la loro maggiore frequenza caratterizzino l'eloquio di parlanti americani di età superiore ai 60 anni. Un'ulteriore differenza è da Oyer e Deal (1985) riscontrata tra voce maschile e femminile: mentre la durata media delle pause, tra il brano letto da voci maschili e femminili, non presenta sostanziali differenze, il tempo totale e la frequenza delle pause mostrano valori maggiori per la voce maschile: "...the males showed greater total pause time than did females... The frequency of pauses for males was significantly greater than for females" (p. 111- 112).

Abbiamo inoltre preso in considerazione la distribuzione delle pause: soltanto il 45% dei silenzi ricorre nella stessa posizione. Per fare un esempio riportiamo la stessa porzione di testo prodotto da Angela quarantenne (1) e Angela ottantenne (2):

- (1) In Francia invece al largo di Tolone le ricerche si sono concentrate in un punto ben preciso // dove il sonar il sonar è una specie // di radar subacqueo // ha registrato la presenza di una grossa massa metallica. // Questa massa metallica potrebbe essere il sottomarino francese Minerve // [in francia\\_68.wav](#)
- (2) In Francia invece // al largo di Tolone le ricerche si sono concentrate // in un punto ben preciso dove il sonar // il sonar è // una specie di radar subacqueo // ha registrato la presenza di una // grossa massa metallica. // Questa massa metallica potrebbe // essere il sottomarino francese Minerve // [in francia\\_07.wav](#)

Come si vede, soltanto in due casi la pausa silente occorre in uno stesso punto del testo. In entrambi i casi si tratta di pause sintattiche, una di confine debole di frase ("subacqueo // ha registrato...") e una di confine forte ("massa metallica. //"). In tutti gli altri casi i silenzi si spostano lungo il testo in maniera indipendente da esso, evidentemente in funzione della diversa meccanica respiratoria.

Sebbene nelle ricerche effettuate sui silenzi la distinzione tipologica tra pause di respirazione e pause sintattiche e la loro diversa distribuzione all'interno di uno stesso testo siano state raramente e molto marginalmente osservate,<sup>3</sup> ci sembra di poter comunque dire che, complessivamente, i nostri dati concordano con quelli riscontrati negli altri lavori.

#### 4.4 $f_0$ e andamento intonativo

In Fig. 5 sono riportati i valori medi di  $f_0$ , minimi e massimi, misurati per catena fonica, e la loro deviazione standard. Mentre la  $f_0$  minima presenta, nel 2007, un incremento del 12%, la  $f_0$  massima aumenta del 26%. Il valore medio della frequenza fondamentale, ottenuto misurando la  $f_0$  relativa a ciascuna sillaba è, di circa 109 Hz nel 1968 (deviazione standard 19 Hz) e di circa 141 Hz nel 2007 (deviazione standard 31 Hz). Come si vede, Angela ottantenne parla con un tono più alto e più variato rispetto ad Angela quarantenne. Il modesto incremento del valore minimo di frequenza glottidale crediamo vada attribuito alla minore elasticità delle pliche vocali, che non sono più in grado di vibrare se non ad un livello di tensione più elevato. Esse iniziano così a vibrare periodicamente intorno ai 90c/s invece degli 80 c/s. Si tratta quindi di un aumento non volontario da parte del parlante, in quanto si verifica per ragioni di ordine anatomo-fisiologico. L'incremento significativo del valore massimo di frequenza glottidale è invece determinato da fattori di tipo volontario: il soggetto ottantenne parla deliberatamente con un tono più alto (sia di picco tonale che di valore medio) per mantenere inalterata la sensazione propriocettiva della propria voce in relazione alla minore funzionalità dell'apparato uditivo che, anch'esso, con gli anni subisce naturali trasformazioni.

|                                    | 1968       | 2007       |
|------------------------------------|------------|------------|
| <b><math>f_0</math> minima</b>     | <b>84</b>  | <b>94</b>  |
| <b><math>f_0</math> massima</b>    | <b>150</b> | <b>190</b> |
| <b>SD <math>f_0</math> minima</b>  | <b>10</b>  | <b>15</b>  |
| <b>SD <math>f_0</math> massima</b> | <b>27</b>  | <b>34</b>  |

Figura 5: Valori medi di  $f_0$  per catena fonica e deviazione standard misurati in Hz.

Altro dato che emerge dal confronto è che le vocali atone prodotte nel 2007 presentano, rispetto a quelle del 1968, un aumento di tono di circa il 33% mentre quelle toniche del 26%. Come spiegare questo dato? Forse si può trattare di una caratteristica idiosincratICA del parlante, se si osserva che anche l'eloquio di Piero Angela quarantenne, seppure in minor misura, presenta un aumento di tono su alcune sillabe atone. Tale

<sup>3</sup> Oyer e Deal (1985) scrivono: "Gordon [1] examined frequency and location of pauses and found that the two groups did not differ in terms of pause location" (p.109).

caratteristica segna dunque lo stile di eloquio del parlante (Angela utilizza il picco tonale come marca di enfasi) e si conserva, accentuandosi, con l'età. Un esempio viene proposto in figura 5.<sup>4</sup>

- (1) "...l'offensiva **sim**ultanea contro **nu**merosi centri attaccando le guarnigioni americane" [l'offensiva 201968.wav](#)
- (2) "...l'offensiva // **sim**ultanea contro **nu**merosi centri attaccando // le guarnigioni americane" [l'offensiva 202007.wav](#)

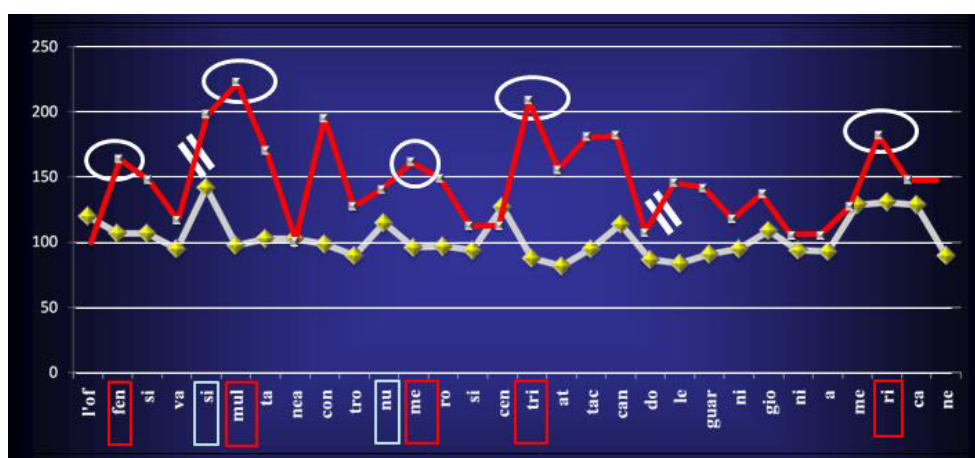


Figura 6: Andamento intonativo: in bianco 1968, in rosso 2007.

Al di là di queste differenze di  $f_0$  che si riflettono, ovviamente, sull'andamento intonativo che appare molto più variato, si notano a volte sorprendenti somiglianze, tali da rendere quasi sovrapponibili i due elinqui. Un esempio di questo tipo è la frase riportata in figura 7 ("Il caso della Pueblo, la nave americana bloccata nella Corea del Nord"). Dobbiamo confessare che quando abbiamo sovrapposto i due grafici siamo rimasti davvero sorpresi. Certo, il testo è lo stesso, è stato letto dalla stessa persona nella stessa situazione contestuale, eppure la quasi perfetta sovrapponibilità dei tracciati lascia perplessi e pone molti interrogativi. Una sera del 1968 Piero Angela, che legge in studio il telegiornale, si trova a dover pronunciare una determinata frase: in quanti modi avrebbe potuto leggerla? Sicuramente tanti, modificando l'andamento dell'intensità, del tono, inserendo uno o più silenzi, enfatizzando una sillaba piuttosto che un'altra. A distanza di quarant'anni la stessa persona (ma "stessa" fino a un certo punto, in quanto il tempo trascorso avrà pure lasciato il segno nel suo modo di pensare e di agire) si ritrova davanti a un foglio dove è scritta una certa frase e in quel momento, tra i tanti modi in cui può pronunciarla, decide di leggerla nello stesso identico modo in cui l'aveva letta nel 1968. A noi tutto questo sembra molto poco casuale e pensiamo che vi debbano essere delle ragioni profonde in grado di spiegare tali somiglianze. Sembra quasi che nel rapporto tra voce ed età le differenze e le somiglianze siano dovute a fenomeni distinti che si riferiscono a livelli diversi: un livello

<sup>4</sup> Le sillabe atone, caratterizzate da un innalzamento di tono, sono segnalate con un rettangolo di colore diverso (nel testo sono in grassetto).

per così dire “periferico”, al quale appartengono le differenze, e un livello, che potremmo definire “centrale”, al quale sono ancorate le somiglianze. Il primo riguarda i cambiamenti sopraggiunti negli organi fonatori con il trascorrere degli anni (elasticità dei tessuti, tensione muscolare, motilità articolatoria etc) e le conseguenti modifiche occorse nella produzione fonica (rallentamento dell’eloquio, variazione tonale, diverso uso delle pause etc). Il secondo, che non cambia con l’età, riguarda invece il modello di parlato che ogni soggetto possiede e che tende, più o meno inconsapevolmente, a riprodurre.

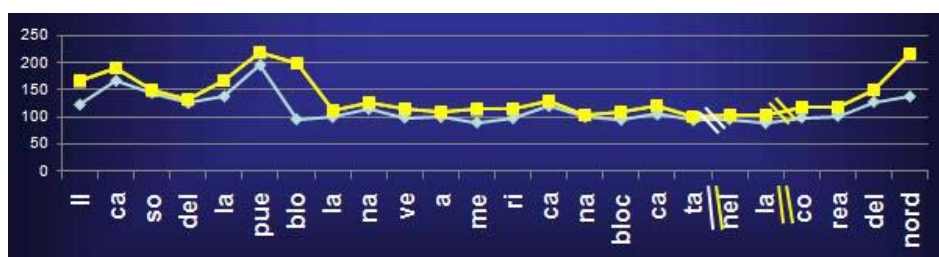


Figura 7. Andamento intonativo: in bianco 1968, in giallo 2007

#### 4.5 Pattern formantici

Come si vede in figura 8, l’analisi relativa ai pattern formantici evidenzia che, con l’età, l’area vocalica di Piero Angela ha subito un ampliamento dovuto alle vocali posteriori e alla vocale [a] (fig. 8). Tale ampliamento, che coinvolge in egual misura sia le vocali toniche che atone, è causato da un aumento in frequenza della F1 e da un abbassamento di F2.

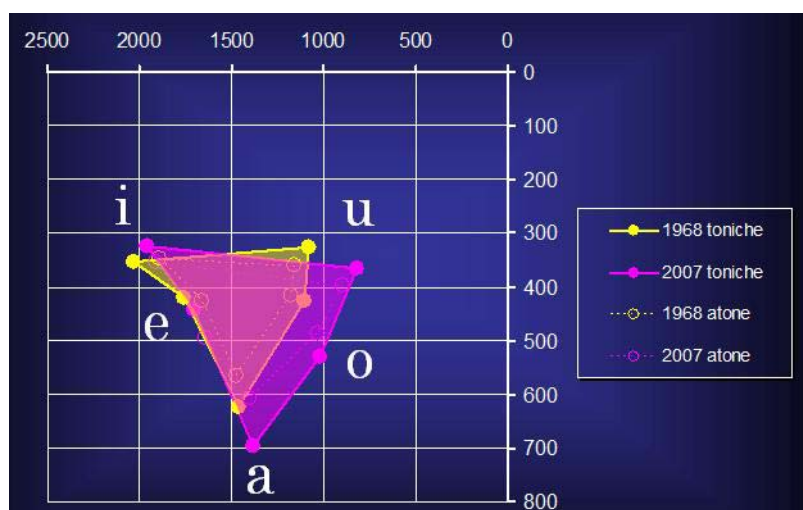


Figura 8: Area delle vocali toniche e atone

L’innalzamento di F1 è da mettere in relazione con un minore sollevamento del dorso della lingua e, di conseguenza, con una maggiore stretta a livello faringale e l’abbassamento di F2 con l’arretramento della lingua e con un ampliamento di volume della cavità orale. Presi di per sé, tali valori testimoniano complessivamente una postura della lingua più

arretrata. Tuttavia la stabilità delle vocali anteriori ci induce a pensare che la causa delle variazioni riscontrate sia da addebitare ad uno specifico muscolo, lo stiloglosso, che con la sua contrazione determina il sollevamento e arretramento della parte posteriore del dorso della lingua verso il velo palatino. Una lieve ipotonia dello stiloglosso può avere come conseguenza proprio le variazioni fonetiche da noi riscontrate nel corpus esaminato.

Per quanto riguarda la F1 i nostri dati non concordano con quelli riscontrati in altri lavori, in cui si è evidenziato che con l'età tutte le formanti tendono ad abbassarsi. Tale andamento è stato messo in relazione o con un allungamento e allargamento del canale fonatorio o con un mancato raggiungimento del target articolatorio che provoca una centralizzazione dell'area vocalica (Endres et alii 1971; Linville 1987, 2004; Linville & Rens 2001)<sup>5</sup>, centralizzazione che si riflette, dal punto di vista percettivo, in una "imprecise articulation" (Hartman & Danhauer 1976). Al contrario, Benjamin (1981) asserisce che "the vowel formants did not show a general centralization or reduction of vowels towards the neutral phoneme, the schwa" (p.164) ma sottolinea che, con l'età, vi è una notevole dispersione delle altezze formantiche delle vocali.

Altre ricerche trovano che, con l'età, nelle voci maschili, vi è un incremento di F1 per le vocali /i/ e /u/ e un incremento di F2 per /ae/ e /a/ (Rastatter & Jacques 1990; e Rastatter et alii 1997) e, altre ancora, considerando le stesse vocali, trovano che sia la F1 di /i/ e /u/ che la F2 di /ae/ e /a/ non mostrano significative differenze (Xue & Hoo 2003). Inoltre, Xue & Hoo (2003) trovano che, l'intero tratto vocale non subisce, con l'età nessun allungamento, ma solo un aumento di volume come conseguenza dell'aumento del volume nella cavità orale e faringale e che soltanto "oral cavity tend to lengthen significantly as speakers age, whereas such lengthening changes were largely absent from speakers' pharyngeal lumen" (p. 698). Tali cambiamenti anatomico-fisiologici provocano un significativo abbassamento formantico soprattutto nella F1. I nostri dati, come abbiamo visto, vanno in un'altra direzione in quanto attestano un aumento di F1, almeno per le vocali posteriori. Va detto, comunque che, sistemi di lingua diversi, corpora diversi, situazioni contestuali diverse, speaker diversi (Piero Angela è uno speaker professionista) possono influire notevolmente sui risultati. Infine, poiché nessun lavoro di questo tipo è stato condotto sulla lingua italiana, resta aperta la questione relativa alla natura linguospecifica del fenomeno.

## 5. CONCLUSIONI

Grazie alla collaborazione offertaci da Piero Angela, abbiamo potuto condurre per la prima volta un esperimento su un corpus di parlato prodotto, a distanza di 40 anni, dallo stesso parlante nella medesima situazione contestuale. In questo modo abbiamo potuto azzerare, nel confronto tra i due enunciati, una variabile molto importante, quella relativa al testo. Poiché tutte le altre variabili erano controllate e mantenute costanti (parlato letto, situazione formale, notiziario trasmesso in uno studio RAI), le differenze riscontrate mediante l'analisi acustica possono essere attribuite all'unico elemento che differenzia i due enunciati, vale a dire la diversa età del parlante. Gli esami condotti hanno messo in risalto varie differenze sia a livello segmentale che a livello prosodico, differenze che possono essere così riassunte:

---

<sup>5</sup> Alcune ricerche evidenziano che mentre la F1 si abbassa con valori significativi in entrambi i sessi, la F2 subisce un maggiore decremento nella voce femminile e una tendenza ad abbassarsi nella voce maschile

- un rallentamento, di circa una sillaba, della velocità di articolazione e della velocità di eloquio;
- una fluenza più bassa (tra un silenzio e il successivo vengono prodotte mediamente tre sillabe in meno);
- una stessa composizione dell'enunciato (rapporto percentuale tra parlato e silenzio);
- un aumento della durata sillabica;
- un aumento della porzione vocalica nelle sillabe toniche chiuse;
- un parlato più isocrono (la durata aumenta in misura maggiore nelle sillabe atone aperte e in misura minore nelle sillabe toniche chiuse);
- un incremento della durata e del numero delle pause sia silenti, in particolar modo delle pause di respirazione;
- un accorciamento delle catene foniche;
- una  $f_0$  più alta;
- un andamento intonativo più variato;
- un ampliamento dell'area vocalica causato da un arretramento e abbassamento delle vocali posteriori.

Sono proprio queste variazioni a determinare, sul piano uditivo, la percezione dell'età del parlante da parte dell'ascoltatore. Una possibile conferma a questa ipotesi può essere ricercata mediante un esperimento di manipolazione sintetizzata dei suddetti parametri e da relativo test percettivo, esperimento che ci proponiamo di effettuare per verificare se è possibile “invecchiare” o “ringiovanire” una voce.

Infine dobbiamo sottolineare che, accanto alle differenze, abbiamo riscontrato anche sorprendenti similarità: alcune frasi sono state pronunciate, a distanza di 40 anni, con lo stesso andamento tonale, la stessa velocità, le stesse pause, le stesse esitazioni. Se le differenze trasmettono informazione sull'età del parlante, le similarità ci dicono evidentemente che quel parlante è Piero Angela in entrambi i casi. Se le prime sono dovute alle trasformazioni fisiologiche dovute al passare degli anni, le seconde sono dovute a caratteristiche che si mantengono costanti nel tempo e che sono frutto delle scelte fonatorie che ciascun parlante opera molto presto, presumibilmente già durante la fase dell'apprendimento del linguaggio, e che restano poi, in buona misura, immutate nel tempo.

## BIBLIOGRAFIA

- Awan, S. N. (2006), The aging female voice: acoustic and respiratory data, *Clinical Linguistics & Phonetics*, vol 20, 2 & 3, April 2006, 171-180
- Benjamin, B. J. (1982), Phonological performance in gerontological speech, *Journal of Psycholinguistic Research* 11 (2), 159-167
- Endres, W., Bambach, W., Floessner, G. (1971), Voice spectrograms as a function of age, Voice Disguise and Voice Imitation, *Journal of the Acoustical Society of America*, 49, 6 (2), pp. 1842-1848
- Giannini, A. (2004), Analisi acustica del parlato televisivo, In “Misura dei Parametri” aspetti tecnologici ed implicazioni nei modelli linguistici, CD-ROM Proceedings del I convegno Nazionale AISV, a cura di Piero Cosi, pp. 49-61, Padova

- Giannini, A. (2006), I silenzi del telegiornale. In *Congresso internazionale "La Comunicazione Parlata"*, Liguori, Napoli
- Hartman, D. E. & Danhauer, J.L. (1976), Perceptual features of speech for males in four perceived age decades", *Journal of the Acoustical Society of America*, 59 (3), 713-715
- Hollien, H. & Shipp, T. (1972), Speaking fundamental frequency and chronological age in males, *Journal of Speech and Hearing Research*, 15:155-159
- Horri, Y. & Ryan, W. J. (1981), Fundamental frequency characteristics and perceived age of adult male speakers, *Folia Phoniatrica*, 33: 227-233.
- Linville, S. E. (1987), Acoustic-perceptual studies of aging voice in women, *Journal of voice*, 1 (1), pp. 44-48
- Linville, S.E. ( 2004), The Aging Voice, *American Speech-Language Hearing Association*, pp. 12-21
- Mysak, E. (1959, Pitch and duration characteristics of older males, *Journal of Speech and Hearing Research*, 2:46-54
- Oyer, E. & Deal, L. (1985), Temporal aspects of speech and the aging process, *Folia Phoniatrica (Basel)*, 37:109-112
- Pegoraro-Krook, M. I. (1988), Speaking fundamental frequency characteristics of normal Swedish Subjects Obtained by Glottal Frequency Analysis, *Folia Phoniatrica*, 40: 82-90
- Pettorino, M., Giannini, A. (2005), Italian TV speech: a diachronic analysis, *Proceedings of IX International Symposium on Social Communication*, Actas-II, pp. 849-853, Santiago de Cuba
- Ptacek, P. H., Sander, E. K. (1966), Age recognition from voice, *Journal of Speech and Hearing Research*, 9:273-277
- Rastatter, M. P. & Jacques, R. D. (1990), Formant frequency characteristics of elderly speakers in contextual speech, *Folia Phoniatrica*, 42: 312-319
- Rastatter, M. P., McGuire, R. A., Kalinowski, J., Stuart, A. (1997), Formant frequency structure of the aging male and female vocal tract, *Folia Phoniatrica*, 49: 1-8
- Russell, A., Penny, L., Pemberton, C. (1995), Speaking Fundamental Frequency Changes Over Time in Women. A Longitudinal Study, *Journal of Speech and Hearing Research*, 38: 101-109
- Shipp, T. & Hollien, H. (1969), Perception of the aging male voice", *Journal of Speech and Hearing Research*, 12:703-710
- Stoicheff, M. L., (1981), Speaking fundamental frequency characteristics of nonsmoking female adults, *Journal of Speech of Hearing Research*, 24:437-441
- Xue, S. A. & Hoo, G. J. (2003), Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study, *Journal of Speech, language and Hearing Research*, 46: 689-701



# LA PREASPIRAZIONE NELLE LINGUE DEL NORD EUROPA E IN VARIETÀ DI ITALIANO: DATI DI BASE PER UN CONFRONTO TRA ISLANDESE E SENESE<sup>1</sup>

P. Mairano<sup>1</sup>, A. Romano<sup>2</sup>, M. Stevens<sup>3</sup> e J. Hajek<sup>4</sup>

<sup>1,2</sup> Università di Torino, <sup>3,4</sup> Università di Melbourne

<sup>1</sup>paolomairano@gmail.com, <sup>2</sup>antonio.romano@unito.it, <sup>3</sup>mes@unimelb.edu.au,

<sup>4</sup>j.hajek@unimelb.edu.au

## 1. SOMMARIO

Recentemente è stata riscontrata la presenza di preaspirazione in senese da parte di Stevens & Hajek (2004b, 2007) e Hajek & Stevens (2005), che ne hanno descritto le caratteristiche e la distribuzione. Questo ci ha spinti a intraprendere un'analisi contrastiva di questo fenomeno, confrontando come esso si presenta in senese e in islandese, la lingua più tipicamente associata alla preaspirazione. Abbiamo quindi raccolto dei dati di islandese e abbiamo analizzato le variabili prese in considerazione da Hajek e Stevens nei loro precedenti lavori. In particolare, sono state calcolate la percentuale di realizzazioni preaspirate rispetto al totale dei casi potenziali e la lunghezza dei segmenti facenti parte del nesso: il tutto in funzione dei singoli parlanti, del sesso e della presenza/assenza di accento lessicale. Per quanto riguarda l'islandese, sono state prese in considerazione la qualità della vocale precedente l'occlusiva preaspirata e il luogo d'articolazione di quest'ultima. Inoltre, si è pensato di passare in rassegna i materiali d'italiano parlato raccolti nell'ambito del progetto *CLIPS* e *AVIP*. Deboli forme di preaspirazione compaiono qua e là soprattutto nella resa di occlusive sorde geminate ma, presentandosi in maniera incostante e disuniforme, sembrano prevalentemente legate a caratteristiche idiosincratiche dei parlanti. D'altra parte, un fenomeno spesso legato alla presenza di preaspirazione, la desonorizzazione di consonanti sonoranti, sembra riguardare parzialmente i cluster /rt/, /lt/, /nt/ e si presenta solo come una fricativizzazione transitoria di /r/, /l/, /n/ che interessa meno del 20% della durata dell'intera consonante e con un'energia talmente contenuta da renderla poco rilevante nella descrizione di questo tipo di parlato.

Per quanto riguarda l'islandese, sono stati raccolti dei campioni di parlato letto da 10 parlanti madrelingua registrati da uno degli autori in Islanda. Ai primi 5 parlanti sono stati somministrati due testi in cui erano presenti complessivamente 34 casi in cui la norma dell'islandese standard prescrive preaspirazione, mentre agli altri 5 parlanti sono stati somministrati due testi in cui erano presenti 49 casi. Le percentuali di realizzazioni preaspirate sono sensibilmente più alte rispetto a quelle riscontrate in senese – bisogna però tenere conto del fatto che Hajek e Stevens hanno usato campioni di parlato spontaneo – e, in generale, essa sembra pesantemente correlata con la velocità di eloquio. Come già riscontrato in letteratura a proposito di altre lingue (in particolare alcuni dialetti svedesi – Helgason 2003 – e nella varietà inglese del Tyneside – Docherty & Foulkes 1999 e Docherty 2003), le due locutrici hanno presentato una percentuale di realizzazioni

---

<sup>1</sup> Il presente lavoro è stato concepito congiuntamente dai quattro autori ma i dati islandesi e le sezioni relative alla loro analisi e discussione sono da attribuire all'autore PM, responsabile della redazione finale dei §§ 2.1 e 2.2 e dei §§ 3 e 4. Il § 2.3 va attribuito agli autori MS e JH, mentre il § 2.4 va attribuito all'autore AR. I restanti paragrafi sono invece il risultato di una redazione congiunta.

preaspirate leggermente superiore rispetto a quelle degli otto locutori maschi. Le misurazioni dei segmenti hanno permesso di mettere in luce notevoli differenze tra i parlanti e hanno dimostrato il fatto, già notato da Thráinsson (1978), che la preaspirazione ha una durata paragonabile a quella di un intero segmento. Inoltre, i nostri dati suggeriscono che la preaspirazione tenda ad avere una durata maggiore, oltre a essere percettivamente più saliente, in presenza di vocali anteriori (soprattutto se non labializzate) e, in misura minore, di occlusive alveolari. Possiamo provvisoriamente concludere che la preaspirazione debba essere considerata come microfenomeno in alcune varietà di italiano e come macrofenomeno in islandese: questo dato è confermato anche dal fatto che, mentre i parlanti islandesi sono a conoscenza della presenza di preaspirazione (infatti, nonostante le occlusive preaspirate non sembrano avere statuto di fonema in islandese – problema che ha dato adito a numerose discussioni – sono comunque presenti coppie minime di superficie), i parlanti italiani ne sono del tutto ignari. Invece, in base ai dati discussi in più occasioni, la situazione del senese va forse paragonata a quella di alcuni dialetti svedesi e norvegesi, per i quali Helgason (2002) usa il termine di preaspirazione non normativa.

## 2. INTRODUZIONE

### 2.1 *La preaspirazione in Europa e nel mondo*

La preaspirazione viene definita da Helgason come segue:

“Preaspiration is a period of (usually glottal) friction that occurs between a vocalic and a consonantal interval. Phonotactically within a language, this type of friction noise occurs only before a limited subset of consonant types, typically voiceless stops” (Helgason, 2002: 11).

Dunque, la preaspirazione può essere vista come un periodo di tempo in cui le pliche vocali, entrate in vibrazione durante la produzione della vocale, smettono di vibrare prima che nella cavità orale abbia avuto tempo di formarsi un’occlusione: questo provoca un periodo di frizione, che può manifestarsi come fricativa o approssimante laringale [h] o assimilarsi al luogo di articolazione della vocale o dell’occlusiva, soprattutto in contesto palatale [ç] o velare [x]. Come suggerito da Helgason (2002 e seguenti) sono anche possibili realizzazioni con laringalizzazioni, *creaky voice* o *breathy voice*.

La preaspirazione è presente in lingue e dialetti del nord Europa come l’islandese, il feringio, numerosi dialetti svedesi e norvegesi, l’irlandese, il gaelico scozzese, in alcuni dialetti Saami e nell’inglese del Tyneside (per quest’ultimo si veda Docherty & Foulkes 1999). Alcuni autori ne hanno riscontrata la presenza anche in lingue non europee, come hopi, uiguro, nenets, zapoteco, ceceno e altre (v. Helgason, 2002, per un excursus con maggiori dettagli sulla presenza di preaspirazione nelle lingue del mondo).

### 2.2 *La preaspirazione in islandese*

Secondo quanto riportato da Helgason (2002), la preaspirazione in islandese era stata notata già da Henry Sweet nel *Handbook of Phonetics* (1877). In questa lingua, spesso considerata prototipica a questo riguardo, la preaspirazione è percettivamente molto evidente, raggiungendo in alcuni casi, nei nostri materiali, durate dell’ordine di oltre 100 ms. La varietà di possibili realizzazioni fonetiche da noi riscontrata include fricative o approssimanti laringali, fricative velari e palatali nonché periodi di laringalizzazione della vocale in contesti di minor prominenza.

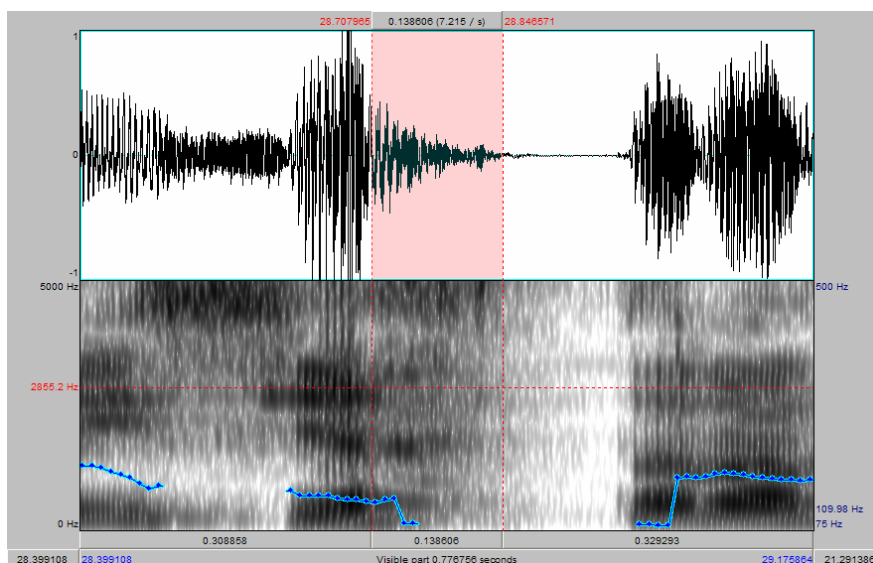


Figura 1: Oscillogramma e spettrogramma della parola *þéttara* ['θjehtərə] (comparativo di *þéttur* - denso).

Nella figura 1 illustriamo uno dei casi più evidenti fra quelli presenti nei nostri dati: è realizzato con una fricativa laringale della durata di 139 ms chiaramente visibile nello spettrogramma (inclusa una breve fase di transizione iniziale con sonorità irregolare). Si tratta della parola *þéttara* ['θjehtərə] (comparativo di *þéttur* - denso), (teniamo a far notare che non si tratta di una parola pronunciata in isolamento, ma di parlato letto all'interno della storia *Il vento di tramontana e il sole*).

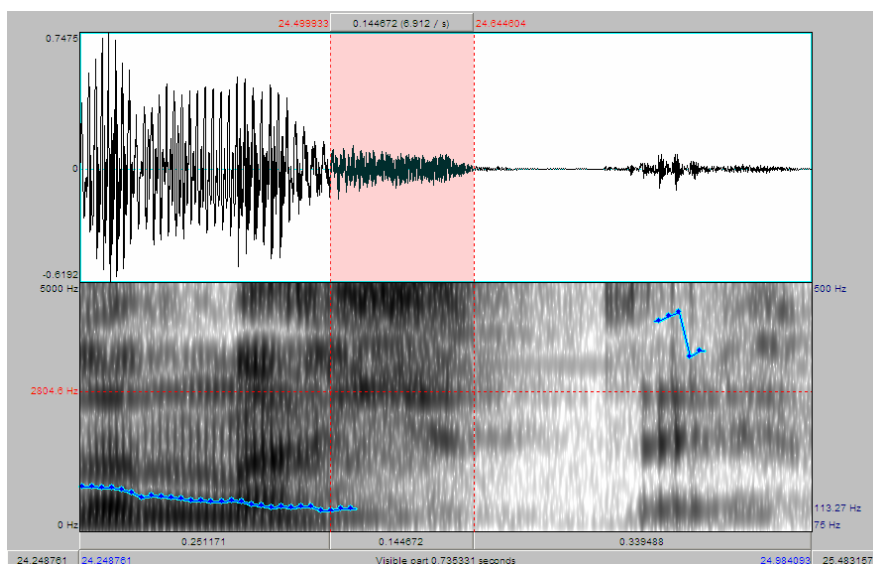


Figura 2: Oscillogramma e spettrogramma della parola *mætti* ['maɪçtɪ] (dalla locuzione *af öllum mætti* – con tutte le proprie forze).

La figura 2 mostra un altro esempio di preaspirazione in islandese, questa volta realizzata come fricativa palatale sorda a causa della coarticolazione con la vocale che precede. La parola interessata è *mætti* ['maɪçtɪ] (dalla locuzione *af öllum mætti* – con tutte le proprie forze) e in questo caso la durata della preaspirazione è di 145 ms. Da questi esempi si nota chiaramente che, come suggerito in vari studi (cfr. per esempio Thráinsson, 1978), la preaspirazione in islandese ha durata equivalente a quella di un intero segmento, fatto ulteriormente dimostrato dai dati relativi alla durata dei segmenti adiacenti alla preaspirazione presentati più avanti.

Per quanto riguarda la fonologia, in islandese esistono opposizioni fonologiche di superficie fra occlusive sorde non aspirate scempie, occlusive sorde post-aspirate scempie, occlusive sorde non aspirate geminate e occlusive sorde preaspirate. Si è spesso sostenuto che a questa opposizione di superficie corrisponda una opposizione profonda più simmetrica che veda la presenza di occlusive sorde scempie non aspirate, occlusive sorde scempie post-aspirate, occlusive sorde geminate non aspirate e occlusive sorde geminate post-aspirate. Queste ultime, che non sono mai presenti in superficie, sarebbero quindi realizzate come preaspirate solo al livello fonetico (superficiale) in cui il tratto di aspirazione viene realizzato come pre-aspirazione piuttosto che come post-aspirazione. Quest'ipotesi, che si riflette nell'ortografia dell'islandese moderno e, forse, nella competenza dei parlanti, può essere così schematizzata:

|                                           |                                             |
|-------------------------------------------|---------------------------------------------|
| <b> /p/ [p]                               | <bb> /pp/ [p:]                              |
| <d> /t/ [t]                               | <dd> /tt/ [t:]                              |
| <g> /k/ [k]                               | <gg> /kk/ [k:]                              |
| <gj> /kj/ [c]                             | <ggj> /kkj/ [c:]                            |
| <p> /p <sup>h</sup> / [p <sup>h</sup> ]   | <pp> /pp <sup>h</sup> / [p <sup>h</sup> ]   |
| <t> /t <sup>h</sup> / [t <sup>h</sup> ]   | <tt> /tt <sup>h</sup> / [t <sup>h</sup> ]   |
| <k> /k <sup>h</sup> / [k <sup>h</sup> ]   | <kk> /kk <sup>h</sup> / [k <sup>h</sup> ]   |
| <kj> /k <sup>h</sup> j/ [c <sup>h</sup> ] | <kkj> /kk <sup>h</sup> j/ [c <sup>h</sup> ] |

Tabella 1: Quadro della preaspirazione in islandese.

Tuttavia, il quadro è complicato dalla presenza (comunque marginale) di preaspirazione in altri contesti, ovvero nei casi in cui un'occlusiva sorda precede una laterale o una nasale, es. *lækni* ['læ<sup>h</sup>knɪ] opp. ['læ<sup>h</sup>knɪ] (medico) e *ætla* ['æ<sup>h</sup>tlə] opp. ['æ<sup>h</sup>tlə] (avere in mente di) (cfr. Helgason, 2002). In alcuni casi, l'abbiamo riscontrata anche quando un'occlusiva sorda precede una vibrante, es. *betra* ['pɛ<sup>h</sup>trə] (comparativo di *góður* - buono).

### 2.3 La preaspirazione in senese

Per quanto riguarda la preaspirazione in senese, essa è stata notata per la prima volta da Stevens & Hajek (2004a). I due autori hanno esaminato un corpus di sei parlanti senesi, riscontrando la presenza di preaspirazione in tutti i soggetti (in percentuali diverse, 46% in media) nella realizzazione di occlusive sorde geminate.

La figura 3 riporta un esempio di preaspirazione nella frase “mio marito ha fatto il militare lassù” pronunciata da un locutore senese (nei dati di Hajek & Stevens, 2005).

Contrariamente a quanto avviene in islandese, in senese si tratta di un fenomeno non obbligatorio non registrato nella letteratura tradizionale. Per quanto concerne la distribuzione della preaspirazione in senese, sono presenti alcune differenze rispetto all'islandese: in particolare, essa può presentarsi tanto all'interno quanto a confine di

parola, in posizione tonica e atona, la frequenza del fenomeno non sembra infatti essere condizionata dalla presenza o assenza dell'accento di parola (rispettivamente 47% e 46%).

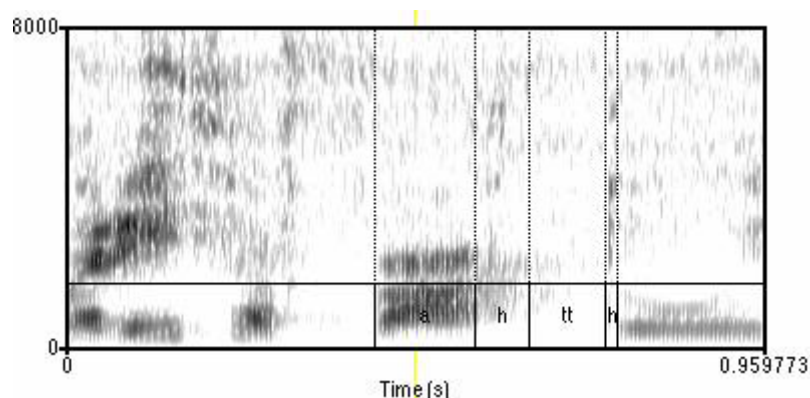


Figura 3: Spettrogramma che illustra un caso di preaspirazione nella parola “fa<sup>[h]</sup>tto”.

Inoltre, Stevens & Hajek (2004b) hanno riscontrato la presenza di sonoranti desonorizzate in senese, un fenomeno presente anche in islandese e spesso associato alla preaspirazione (cfr., per esempio, Helgason 2002). La media globale contrasta con i risultati sulla preaspirazione in svedese (Helgason 1999) e in inglese (Jones & Llamas 2003) e con quelli sull'islandese analizzati in questo nuovo studio e presentati in questa sede. Una dettagliata analisi acustica non dimostra un effetto sistematico del luogo di articolazione dell'occlusiva seguente, ma rivela invece un'interazione fra vocale e l'elemento preaspirato dell'occlusiva seguente. Innanzitutto la preaspirazione è più frequente e ha una durata più consistente dopo la vocale centrale /a/. In secondo luogo, la precisa natura acustico-articolatoria dell'elemento preaspirato può variare a seconda della vocale precedente: è più frequente la voce mormorata dopo /a/ ma la fricativizzazione palatale dopo /i/.

Qui sotto riportiamo una tabella riassuntiva che presenta le caratteristiche della preaspirazione in islandese e senese a confronto.

|                                                                    | Islandese                         | Senese                          |
|--------------------------------------------------------------------|-----------------------------------|---------------------------------|
| Ha status fonologico                                               | No                                | No                              |
| È obbligatoria                                                     | Sì                                | No                              |
| Realizza occlusive sorde geminate fonologiche in sillaba accentata | Sì                                | Sì                              |
| Realizza occlusive sorde geminate fonologiche in sillaba atona     | No                                | Sì                              |
| Realizza altri nessi                                               | Sì (occl. sorda + l, n e forse r) | No                              |
| Realizza consonanti non occlusive                                  | No                                | No                              |
| Presenza a confine di parola                                       | No                                | Sì                              |
| Presenza di sonoranti desonorizzate                                | Sì                                | Sì (cfr. Stevens & Hajek 2004b) |

Tabella 2: Confronto della preaspirazione in islandese e senese.

#### 2.4 La preaspirazione in altre varietà di italiano

A seguito delle pubblicazioni di Hajek e Stevens riguardo alla presenza di preaspirazione in senese, ci è sembrato utile indagare l'eventuale presenza di questo fenomeno in altre varietà di italiano. Con questo proposito è stato condotto uno spoglio di parte dei materiali d'italiano parlato raccolti nell'ambito dei progetti *CLIPS* e *AVIP*. In alcune varietà si è effettivamente riscontrata la saltuaria presenza di brevi periodi di frizione glottidale in corrispondenza di occlusive sorde geminate, tuttavia la portata di questo fenomeno risulta inferiore rispetto ai dati del senese o, in alcuni casi, decisamente irrilevante rispetto ai dati dell'islandese. Ad esempio, nell'ambito dei *map-task* delle 15 località di *CLIPS*, alcune fasi iniziali di frizione (più o meno debole, con forme di sonorizzazione irregolari) compaiono soprattutto nella resa di occlusive sorde geminate o di nessi intervocalici di sonorante più occlusiva sorda. Questi indici possono forse essere classificati come manifestazioni di forme di preaspirazione, tuttavia – presentandosi in maniera incostante e disuniforme – sembrano prevalentemente legate a caratteristiche idiosincratiche dei parlanti. In un campione di 120 turni dialogici, è stato possibile registrare ad esempio una leggera forma di preaspirazione nel caso di una parlante di Lecce (nel 25% dei nessi da lei realizzati); più che riguardare le occlusive sorde geminate (dove è presente solo per /kk/), il fenomeno sembra manifestarsi solo come una fricativizzazione transitoria di /r/, /l/, /n/ nei cluster /rt/, /lt/, /nt/ che interessa meno del 20% della durata dell'intera consonante. Molto più consistente in termini percentuali e di durata relativa si è rivelata invece la frequenza di fenomeni di questo tipo nel caso di locutori baresi, ad es. nella locutrice *g* del dialogo *b02* di *AVIP* (v. figure 4 e 5). Anche la locutrice *p1* del dialogo *DGmtB02B* (v. es. in fig. 6) presenta un'incidenza relativamente alta di realizzazioni preaspirate. I rilevamenti che attribuiscono a questa fase misure di durata maggiore ai 30 ms raggiungono il 25% delle occorrenze (mentre, tenendo conto anche dei casi in cui i valori sono superiori a 20 ms, la percentuale lambisce il 50%). Nel caso del locutore dello stesso dialogo invece le stesse percentuali si attestano tra il 6 e il 15%.

Il 50% dei contesti disponibili per la locutrice è tuttavia relativo a casi in cui la vocale precedente è una /a/ accentata, in un solo caso si tratta di /o/ mentre nel restante 41,7% si tratta di /e/<sup>2</sup>.

Quanto al tipo consonantico interessato, nel 92% dei casi osservati si tratta di /tt/, mentre nel restante 8% di /ttʃ/.

In questa casistica, la presunta preaspirazione interessa indistintamente /a/ e /e/ (50%) con durate, vocaliche, dell'ordine di 130±20 ms e, consonantiche (inclusa l'eventuale frizione), dell'ordine di 190±20 ms. Nei casi in cui è presente la fase fricativa (>20 ms), la sua durata è mediamente 40±20 ms, con durata vocalica attestata su valori leggermente più bassi (120±20 ms) ma durata consonantica (esclusa la frizione) decisamente ridimensionata 155±20 ms: questo fa propendere ad attribuire maggiormente l'elemento di rumore, nei casi in cui è stato rilevato, a caratteristiche di realizzazione della consonante<sup>3</sup>.

<sup>2</sup> *Freccia*, *barchetta*, *oggetto* hanno infatti /e/ nello standard. È forse rilevante in questo caso far notare però che, in questi contesti (sillaba chiusa), i locutori realizzano entrambi [ɛ]. Lo stesso vale per /o/ (le uniche parole presenti sono occorrenze di *sotto* e *rotto*).

<sup>3</sup> Pur essendo decisamente meno rilevanti, nel caso del locutore (che pure presenta una casistica di contesti più variegata, con 8,8% di /i/ e 26,5% di /o/, oltre a un 47,1% di /e/ e a un restante 17,6% di /a/), nell'80% dei casi in cui il fenomeno è degno di rilievo si verifica dopo /e/ (e solo nel 20% nel caso di /a/): anche per questo locutore il fenomeno riguarda in 3 casi su 5 /tt/ (vs. /ttʃ/).

Tuttavia, in tutti i casi osservati, pur essendo queste fasi ben visibili all'analisi spettrografica, il riscontro uditivo è abbastanza contenuto.

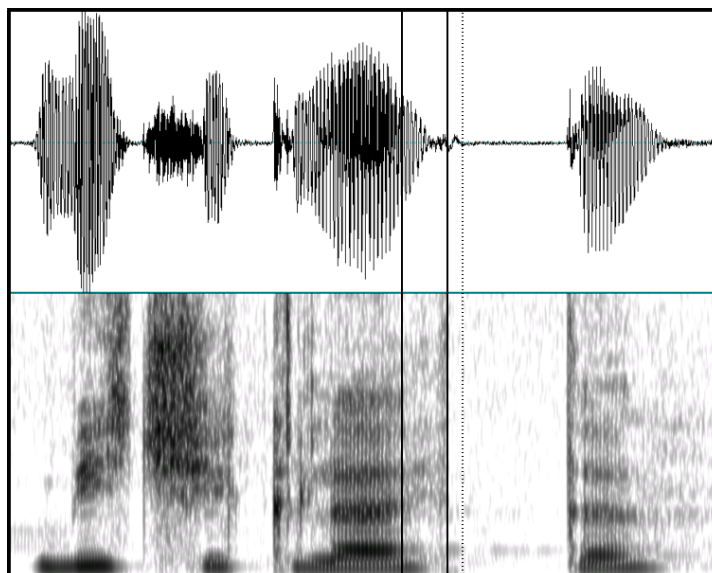


Figura 4: Oscillogramma e spettrogramma di *biciclette* [bitʃi'kle<sup>h</sup>te] pronunciato da una locutrice di Bari (toponomi *b02gtopb*, dati *AVIP*).

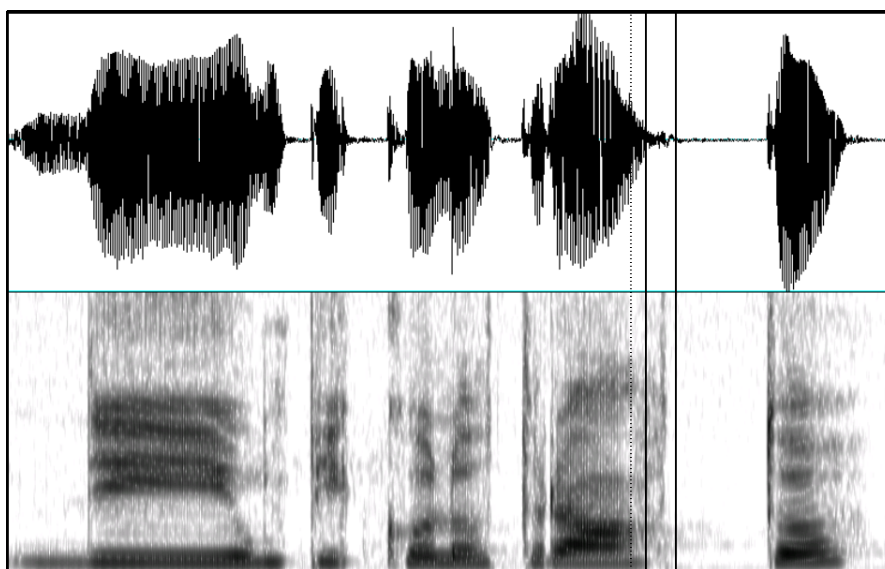


Figura 5: Oscillogramma e spettrogramma di “di un piccolo tra<sup>h</sup>” pronunciato da una locutrice di Bari (turno *b02g135b*, dati *AVIP*).

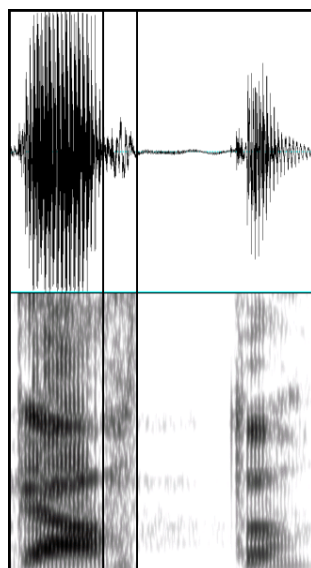


Figura 6: Oscillogramma e spettrogramma di “(g)a<sup>[h]</sup>tto” pronunciato da una locutrice di Bari (turno *DGmtB02B\_p1G9*, dati *CLIPS*).

### 2.5 Obiettivi

Lo scopo dell’esperimento è quello di mettere a confronto i dati relativi alla preaspirazione in varietà d’italiano con quelli di una lingua in cui la presenza di preaspirazione è considerata un dato acquisito e indiscutibile, come l’islandese. A questo proposito, per i dati di islandese si sono calcolate la frequenza di realizzazioni preaspirate nei casi potenziali (in funzione dei singoli parlanti, del sesso dei parlanti e della presenza/assenza di accento lessicale) e le durate dei segmenti adiacenti alla preaspirazione, analogamente a quanto già fatto da Hajek e Stevens per il senese nei lavori precedenti. Inoltre, per l’islandese, sono state analizzate le durate dei segmenti in funzione del luogo di articolazione della vocale e dell’occlusiva<sup>4</sup>.

## 3. L’ESPERIMENTO

### 3.1 I dati

Per quanto riguarda l’islandese, sono stati registrati 10 parlanti madrelingua, nove nati e cresciuti a Reykjavík, uno ad Akureyri (Islanda del nord), tutti con un alto livello di istruzione (cinque sono studenti fra i 20 e i 25 anni, gli altri cinque professori fra i 40 e i 65 anni del Háskóli Íslands – Università d’Islanda), otto maschi e due femmine. Otto parlanti sono stati registrati a Reykjavík in condizioni discrete (nello studio di un professore universitario) da uno degli autori, gli altri due sono stati registrati nella cabina silente del Laboratorio di Fonetica Sperimentale “Arturo Genre” dell’Università di Torino.

<sup>4</sup> Le operazioni di confronto in base ai dati presentati nelle sezioni precedenti, che pur attestano fenomeni analoghi alla preaspirazione in altre varietà di italiano, sono rese difficili dalla diversa tipologia dei dati raccolti e comportano un lavoro molto più dispendioso che è tuttora in fase di completamento ed è, quindi, rinviato a ulteriori contributi.



È importante puntualizzare che si tratta di parlato letto. A cinque parlanti è stato dato un testo in cui erano presenti 30 casi in cui la norma dell'islandese standard prevede preaspirazione, agli altri cinque parlanti è stato somministrato un altro testo in cui erano presenti 44 casi di possibile preaspirazione. Inoltre, tutti i parlanti hanno letto la storia de *Il vento di tramontana e il sole*, in cui sono presenti 5 casi di possibile preaspirazione.

In ultimo, va ricordato che i parlanti non conoscevano l'argomento e lo scopo della ricerca, anche se erano al corrente che si trattava di un'indagine di tipo fonetico e hanno dato il loro consenso scritto all'uso delle registrazioni a fini di ricerca.

Per quanto concerne i dati utilizzati per lo studio della preaspirazione in senese, essi sono già stati presentati nelle precedenti pubblicazioni di Hajek e Stevens, alle quali si rimanda per maggiori dettagli. Si tratta comunque di registrazioni della durata di 5-10 minuti di sei parlanti senesi (tre uomini e tre donne) fatte a Siena nel 1997 da M. Stevens. In questo caso, sono registrazioni di parlato spontaneo, il che pone sicuramente problemi per il confronto con i dati islandesi di parlato letto. Sarà quindi necessaria molta cautela nel trarre conclusioni a questo proposito.

### 3.2 Metodi

Nel presente studio sono state prese in considerazione esclusivamente le occlusive sorde geminate, sia per l'islandese, sia per il senese (e anche per lo spoglio di *AVIP* e *CLIPS* per quanto riguarda le altre varietà di italiano). Dapprima si sono identificate tutte le occorrenze di occlusive sorde geminate (accentate e non) e, successivamente, si è proceduto all'analisi acustica con *Praat*. Analogamente a quanto fatto da Hajek e Stevens (2005), per ogni occorrenza sono stati misurati:

- la durata della vocale;
- la durata di eventuale frizione glottidale o orale (nei casi in cui si è effettivamente manifestata preaspirazione);
- la durata della tenuta dell'occlusiva;
- la durata del rilascio di quest'ultima.

Non è stata presa in considerazione nessuna misura che riguardi il cambio di voce modale, contrariamente a quanto fatto da Helgason (2002 e 2003)<sup>5</sup>: l'inizio della preaspirazione è stato fatto coincidere con la fine della seconda formante vocalica e l'inizio di una fase di frizione; nei casi di sovrapposizione dei due fenomeni, la fine di  $F_2$  è stata considerata indice prioritario dell'inizio di preaspirazione (quindi i periodi di voce modale sono stati conteggiati come appartenenti alla vocale piuttosto che alla preaspirazione).

## 4. RISULTATI

### 4.1 Frequenza della preaspirazione

Come già detto, un primo obiettivo dell'esperimento consisteva nel confrontare la frequenza della preaspirazione nei nostri dati di islandese e senese. Ciò è stato fatto calcolando la percentuale di realizzazioni preaspirate rispetto al totale dei casi potenziali, cioè i casi in cui si trovavano delle occlusive sorde geminate e, in islandese, i nessi

---

<sup>5</sup> Nei suoi lavori, Helgason sostiene che, in molti casi, l'inizio della preaspirazione sia da attribuire non tanto all'istante in cui le corde vocali cessano di vibrare lasciando solo rumore di frizione glottidale o orale, ma piuttosto al cambio di voce modale a causa della sua salienza percettiva. Questa differenza metodologica è probabilmente alla base dei valori di durata tendenzialmente più corti ottenuti nel nostro studio rispetto a quelli di Helgason per altre lingue nordiche.

costituiti da occlusiva sorda seguita da nasale o liquida. I risultati sono illustrati nelle tabelle 3 (per l'islandese) e 4 (per il senese).

| M1  | F1  | F2   | M2  | M3  | M4  | M5  | M6  | M7  | M8  | incidenza |
|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----------|
| 71% | 96% | 100% | 96% | 91% | 91% | 89% | 90% | 94% | 61% | 88%       |

Tabella 3: Percentuale di realizzazioni preaspirate per ogni parlante islandese.

| F1  | M1  | F2  | M2  | F3  | M3  | incidenza |
|-----|-----|-----|-----|-----|-----|-----------|
| 47% | 82% | 48% | 82% | 33% | 38% | 46%       |

Tabella 4: Percentuale di realizzazioni preaspirate per ogni parlante senese.

Come si può notare, la percentuale di realizzazioni preaspirate in islandese è quasi doppia rispetto a quella del senese. Tuttavia, bisogna tenere conto di alcuni fattori: in primo luogo, ricordiamo ancora una volta che i dati di islandese sono di parlato letto, mentre quelli di senese sono di parlato spontaneo. Inoltre, va notato che le percentuali relative all'islandese rendono conto esclusivamente della presenza di preaspirazione in sillabe toniche (infatti, come è stato chiarito nell'introduzione, i nessi potenzialmente preaspirati in sillaba tonica vengono realizzati senza preaspirazione in sillaba atona). Al contrario, le percentuali relative al senese includono sia sillabe toniche, sia atone:

“La presenza dell'accento non ha alcun riflesso né sulla preaspirazione ( $p=0,613$ ) né sul rilascio consonantico ( $p=0,688$ ); entrambi sembrano mostrare un'alta stabilità a prescindere dall'accento. La presenza di questo ha comunque effetto su altre misurazioni di durata” (Hajek & Stevens, 2005:4).

Un'altra considerazione che scaturisce dall'osservazione delle tabelle riguarda la frequenza di realizzazioni preaspirate in funzione del sesso dei parlanti. Per quanto riguarda i dati di islandese, essi sembrano confermare le ipotesi sociofonetiche presenti in letteratura, secondo cui in alcune lingue e dialetti scandinavi (ma anche in Tyneside English – v. Docherty, 2003) la preaspirazione sarebbe più frequente e più prominente nei soggetti femminili: in effetti, le due parlanti islandesi femmine mostrano una percentuale di realizzazioni preaspirate molto alta (rispettivamente 100% e 96%), ma si tratta comunque soltanto di due parlanti ed è quindi molto rischioso generalizzare. Al contrario, in senese la preaspirazione sembrerebbe più frequente nei soggetti maschi (54%) che in quelli femminili (42%), ma le differenze non sono significative ( $p=0,076$ ) (v. Hajek & Stevens, 2005, per maggiori dettagli).

#### 4.2 Durata della preaspirazione

Un secondo obiettivo dell'esperimento era quello di confrontare la durata della preaspirazione in islandese e senese. A questo scopo, per ogni occorrenza di preaspirazione sono state calcolate le durate della vocale, di frizione glottidale o orale (preaspirazione), della tenuta dell'occlusiva e del rilascio di quest'ultima. I risultati sono presentati nella tabella 5 e in figura 6 (ricordiamo che le misure dell'islandese sono tutte relative a posizioni toniche, poiché in questa lingua la preaspirazione è prevista solo sotto accento).

|                  | Voc. (A) | Preasp. (B) | Ten. (C) | Ril. (D) | C+D | B+C+D | Tutto |
|------------------|----------|-------------|----------|----------|-----|-------|-------|
| Islandese        | 82       | 79          | 55       | 28       | 83  | 162   | 244   |
| Senese (toniche) | 92       | 54          | 49       | 32       | 81  | 136   | 228   |
| Senese (atone)   | 54       | 51          | 64       | 28       | 92  | 142   | 196   |

Tabella 5: Durata in ms dei segmenti adiacenti alla preaspirazione in islandese e senese (sul modello di Hajek & Stevens, 2005).

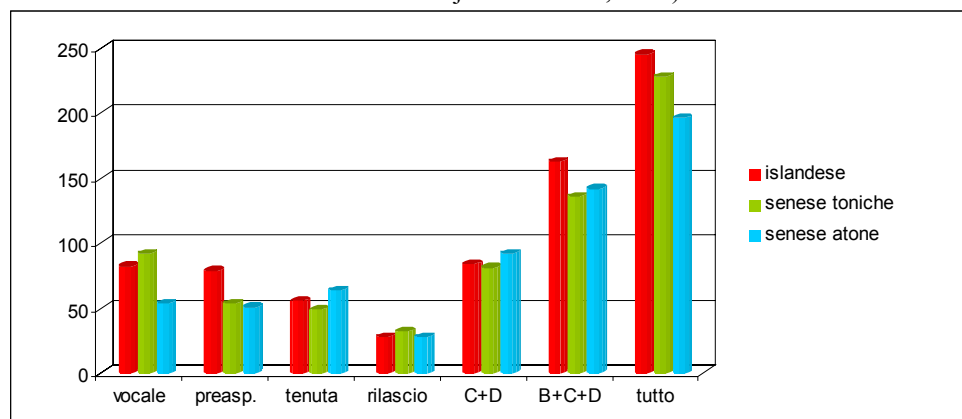


Figura 7: Durata in ms dei segmenti adiacenti alla preaspirazione in islandese e senese relativa ai dati esposti nella tabella sopra.

I dati relativi al senese stupiscono a) per il fatto che l'assenza di accento non abbia rilevanza neanche per la durata della preaspirazione (oltre che per la sua percentuale di realizzazione, come accennato nel paragrafo precedente) e b) per il fatto che una maggior durata della tenuta delle occlusive in posizione tonica sembri favorita dall'assenza dell'accento lessicale. D'altra parte, come ci si aspettava, la durata della vocale è correlata positivamente alla presenza di accento.

Procedendo a un confronto tra i dati del senese con quelli dell'islandese, si nota come la preaspirazione in islandese tenda ad avere durata maggiore (un media di 79 ms contro i 54 ms delle toniche senesi). I circa 25 ms di differenza rispetto alla durata della preaspirazione islandese vengono controbilanciati in senese da una vocale tendenzialmente più lunga (92 ms contro gli 82 ms dell'islandese).

Va comunque tenuto presente che la durata della preaspirazione è un parametro che può variare notevolmente rispetto ai parlanti. Al fine di illustrare questa variabilità, mostriamo in figura 8 le durate dei segmenti analizzati sopra, questa volta calcolando i valori medi di durata separatamente per ogni parlante. Va anche notato che questi dati confermano quanto già affermato in vari studi sulla preaspirazione in islandese (cfr. per esempio Thráinsson, 1978) ovvero che la preaspirazione ha durata maggiore della postaspirazione ed è paragonabile a quella di un intero segmento. Nel caso di alcuni parlanti (Asgeir, Jón e Siggí), essa supera persino la durata della vocale diventando il segmento più lungo del nesso, soprattutto in contesto palatale.

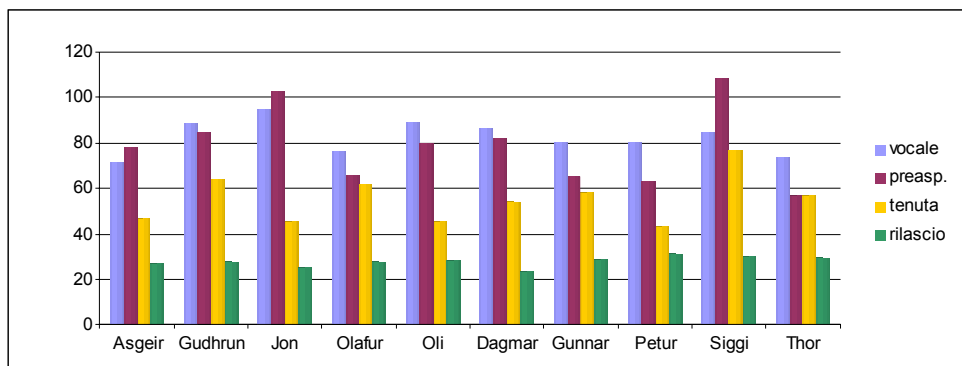


Figura 8: Grafico che illustra la variabilità della durata della preaspirazione nei dati di islandese (valori in ms).

Si può facilmente intuire che questi risultati rischiano di essere influenzati dalla velocità d'eloquio di ogni parlante: per questo motivo abbiamo sentito l'esigenza di normalizzarli calcolando la percentuale di tempo occupata da ognuno dei quattro segmenti all'interno del nesso. I risultati così ottenuti sono mostrati nella figura 9, in cui si nota che il segmento la cui durata oscilla maggiormente fra i 10 parlanti è proprio quello della preaspirazione.

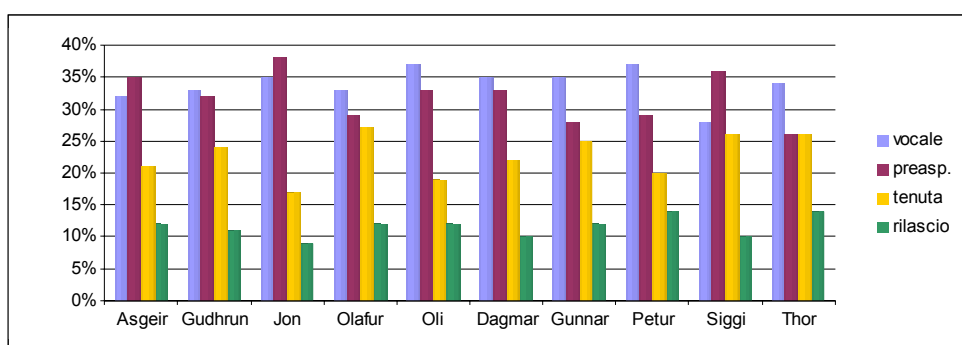


Figura 9: Grafico che illustra la variabilità della durata della preaspirazione nei dati di islandese (valori in percentuale).

#### 4.3 La preaspirazione in islandese nel contesto fonetico

In questo paragrafo si vuole dare un contributo allo studio della preaspirazione in islandese, adottando un punto di vista che, fino ad ora, sembra non essere stato preso in considerazione nella letteratura tradizionale. Si vuole verificare in quale modo il contesto fonetico interferisca con la durata della preaspirazione. Abbiamo quindi usato i nostri dati per confrontare la durata del segmento di preaspirazione in relazione al luogo di articolazione dell'occlusiva (figura 10) e della vocale (figura 11).

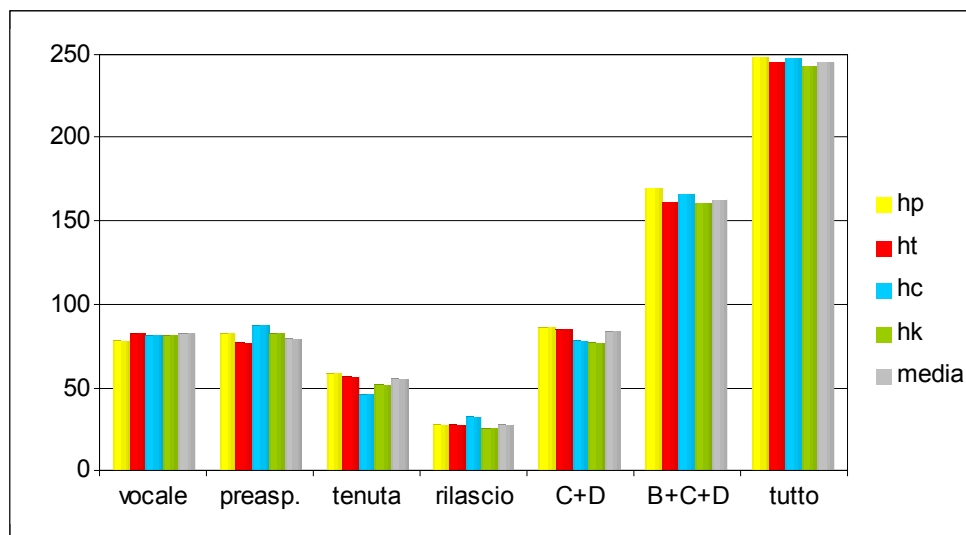


Figura 10: Durata in ms dei segmenti adiacenti alla preaspirazione in islandese a seconda del luogo di articolazione dell'occlusiva

Alcune considerazioni scaturiscono dall'osservazione del grafico in figura 10. In primo luogo, si può notare che la durata della vocale non sembra affatto influenzata dal luogo di articolazione dell'occlusiva. Per quanto riguarda gli altri segmenti del nesso, tuttavia, vi sono alcune piccole differenze, di cui la più visibile riguarda il nesso con l'occlusiva palatale. È risaputo che il rilascio dell'occlusiva palatale ha una durata tendenzialmente più lunga rispetto alle altre occlusive, fatto che è verificato (sebbene in misura ridotta) dai nostri dati di islandese; tuttavia, lo stesso dovrebbe valere per l'occlusiva velare, il cui rilascio si presenta, al contrario, con una durata di poco inferiore rispetto alle occlusive bilabiali e alveolari. È anche interessante notare che la durata della preaspirazione risulta più lunga nel caso delle occlusive palatali (oltre che molto più prominente data la sovente realizzazione come fricativa palatale per assimilazione all'occlusiva), mentre la tenuta è leggermente più corta: si può ipotizzare che la preaspirazione, in questo contesto, "rubì" qualche millisecondo alla tenuta dell'occlusiva.

La figura 11 mostra le durate dei segmenti adiacenti alla preaspirazione a seconda del timbro della vocale o del dittongo precedente. Le vocali sono state separate in: anteriori alte, posteriori alte non labializzate, posteriori alte labializzate, basse. I dittonghi sono stati separati in: anteriori labializzati, anteriori non labializzati, posteriori.

In primo luogo è possibile confrontare la media della lunghezza totale del nesso a seconda che in esso sia presente una vocale o un dittongo: si nota che, come è ragionevole attendersi, la durata totale del nesso aumenta in presenza di un dittongo, anche se in misura limitata. In effetti, osservando la barra relativa alla durata della preaspirazione + la tenuta + il rilascio dell'occlusiva (ovvero la barra B+C+D del grafico), non vi è quasi differenza: questo dimostra che la maggior durata del nesso in presenza di un dittongo è dovuta solo alla maggior durata di quest'ultimo rispetto a una vocale, mentre gli altri segmenti non presentano differenze rilevanti.

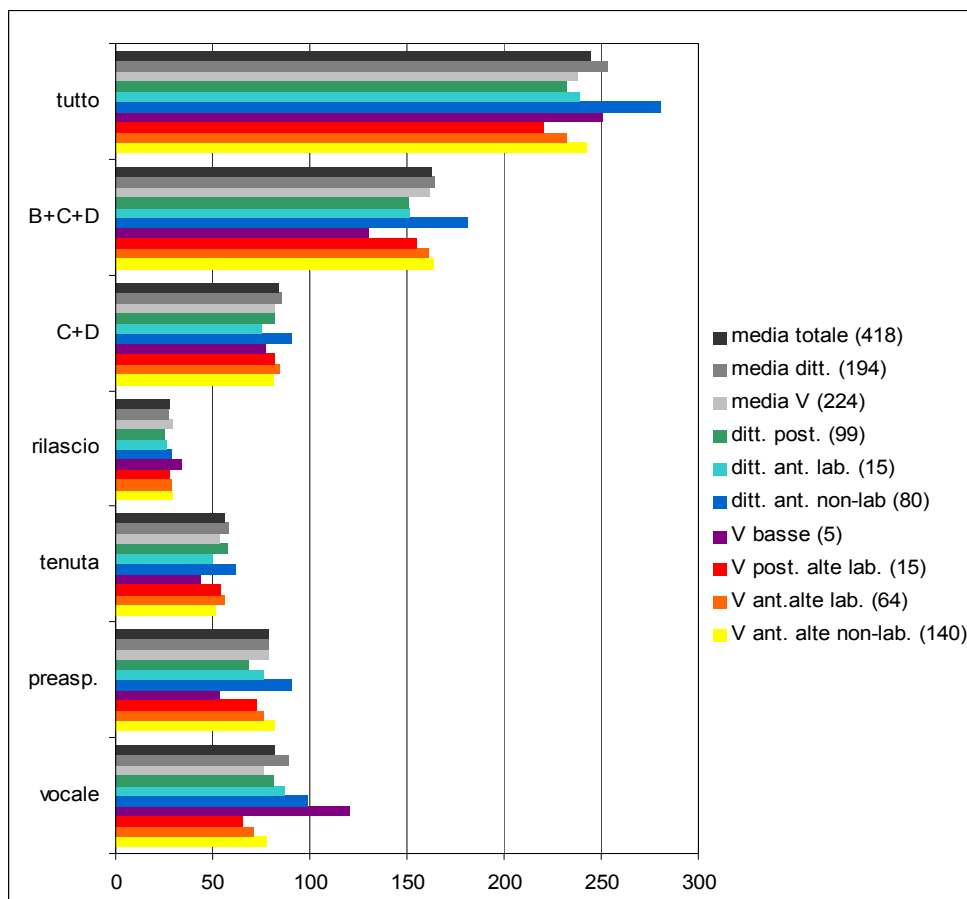


Figura 11: Durata in ms dei segmenti adiacenti alla preaspirazione in islandese a seconda del timbro della vocale che precede. Il numero fra parentesi nella legenda indica il numero di misure su cui sono state calcolate le medie.

Inoltre, salta immediatamente all'occhio la maggior durata delle vocali basse, a cui corrisponde una minor durata della preaspirazione e della tenuta dell'occlusiva. Tuttavia, questo valore, seppur molto evidente, è stato calcolato solo su cinque misure (le uniche 5 vocali basse presenti nei nostri dati) e sono quindi prive di valore statistico.

Tuttavia, quello che più ci preme far notare in questi grafici, riguarda la maggior durata della preaspirazione in contesto anteriore, soprattutto se non labializzato. Come si può osservare nel grafico, ciò si verifica sia per le vocali sia per i dittonghi lungo la stessa scala: dittonghi anteriori non labializzati > dittonghi anteriori labializzati > dittonghi posteriori > vocali anteriori alte non labializzate > vocali anteriori alte labializzate > vocali posteriori alte. Inoltre, teniamo a far notare che questi dati confermano le impressioni percettive ricavate dall'ascolto dei dati e sostengono quanto riscontrato a proposito del luogo di articolazione dell'occlusiva (v. figura 10), secondo cui la preaspirazione tenderebbe ad avere maggior durata in presenza di un'occlusiva palatale. Ci portano infatti a ipotizzare che, in generale, un contesto palatale (vocalico e/o consonantico) sia correlato con una maggior durata della preaspirazione.

## 5. CONCLUSIONI

Prima di trarre conclusioni sulla base dei materiali presentati, teniamo a sottolineare ancora una volta la diversa natura dei dati di islandese (parlato letto) rispetto a quelli di senese (parlato spontaneo) e di altre varietà di italiano (*map-task*).

Per quanto riguarda la frequenza della preaspirazione nei materiali, riportiamo ancora i risultati, che mostrano una percentuale di realizzazioni preaspirate molto alta in islandese (88% dei casi potenziali), media in senese (46%) e molto bassa in altre varietà di italiano (27,5%). Inoltre, per quanto riguarda la durata della preaspirazione, essa ha una media di 79 ms in islandese, 54 ms (posizione tonica) e 51 ms (posizione atona) in senese; in altre varietà di italiano, essa ha durata piuttosto variabile, ma tendenzialmente abbastanza limitata (in generale < 40ms). D'altra parte, abbiamo visto che i contesti in cui si presenta preaspirazione sono simili, ma non identici: la differenza più evidente tra le due lingue riguarda l'accento, poiché l'islandese ammette (o, meglio, richiede) preaspirazione solo in posizione tonica, mentre in senese essa è permessa sia in posizione atona sia tonica, senza differenze rilevanti: questo suggerisce forse che la preaspirazione in islandese, in quanto fenomeno obbligatorio, sia presente in contesti molto ben stabiliti, mentre in senese appaia più irregolarmente (seppur in tutti e sei i parlanti registrati). Nelle altre varietà di italiano in cui abbiamo riscontrato casi di preaspirazione, l'irregolarità con cui essa compare è ancora più evidente, essendo presente con frequenza e durata molto ridotte e solo in alcuni parlanti.

Tutto questo ci porta a ipotizzare un diverso statuto della preaspirazione in queste lingue. In islandese, data la sua obbligatorietà e la notevole frequenza e durata con cui essa compare, unitamente al fatto che i parlanti sembrano esserne coscienti, possiamo affermare che si tratti di un *macrofenomeno*, o, per usare la terminologia di Helgason (1999 e seguenti), di *preaspirazione normativa*. Al contrario, pensiamo che la preaspirazione in varietà di italiano come il barese possa essere descritta come *microfenomeno* in virtù della sua irregolarità, minore salienza e non obbligatorietà. Infine, per quanto riguarda la preaspirazione in senese, ci sembra forse riduttivo definirla come microfenomeno e crediamo di poterla avvicinare ai casi descritti da Helgason (1999 e seguenti) come *preaspirazione non normativa*, per esempio per i dialetti svedesi centrali. Questa etichetta si riferisce soprattutto alla non obbligatorietà e alla conseguente minore sistematicità della presenza di preaspirazione in questi dialetti, ma Helgason (2002 e 2003) nota che esistono anche differenze nella sua realizzazione: per esempio, la preaspirazione non normativa sembra più restia alla coarticolazione, essendo per lo più realizzata come fricativa glottidale, molto meno frequentemente come fricativa orale (come è invece tipico di lingue come l'islandese e il feringio – per quest'ultima cfr. Helgason 2003).

Rimangono tuttavia da chiarire due aspetti. In primo luogo, è necessario verificare in che modo e in che misura la modalità di elicitazione dei dati influisca sulla frequenza e sulla durata della preaspirazione. È probabile che, come hanno suggerito Hajek & Stevens (2005), in senese la preaspirazione sia più frequente nel parlato spontaneo. Al contrario, ipotizziamo che in islandese, essendo normativa, la preaspirazione tenda a occorrere con maggior frequenza e (forse) maggior durata nel parlato controllato. In secondo luogo, rimane da verificare l'eventuale presenza di preaspirazione (presumibilmente a livello di microfenomeno) in altre varietà di italiano rispetto a quelle già analizzate.

Ricordando, infine, i risultati a cui siamo giunti studiando la realizzazione della preaspirazione in islandese a seconda del contesto fonetico in cui essa si presenta, ci limitiamo a osservare che i dati suggeriscono una parziale correlazione tra la durata della preaspirazione e il luogo di articolazione della vocale e, sebbene in misura minore, della consonante.

## RINGRAZIAMENTI

Desideriamo ringraziare Sigurður Pétursson e Íris Gunnarsdóttir per la loro gentilezza e disponibilità in Islanda e a Torino. Ringraziamo anche tutti i locutori che hanno prestato la loro voce per quest'esperimento.

## BIBLIOGRAFIA

Árnason, K. (1980), The Segmental and Suprasegmental Status of Preaspiration in Modern Icelandic, *Nordic Journal of Linguistics*, 9, 1-23.

Bombien, L. (2006), Voicing Alternations in Icelandic – a photoglottographic and acoustic investigation. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, 37, 63-82.

CLIPS - Corpora e Lessici dell'Italiano Parlato e Scritto: <http://www.clips.unina.it/>

Docherty, P. (2003), Speaker, Community, Identity: Empirical and Theoretical Perspectives on Sociophonetic Variation, in *Proceedings of the International Congress of the Phonetic Sciences*, Barcellona, Spagna, 3-9 Agosto 2003, 11-16.

Docherty, P. & Foulkes, G.J. (1999), Instrumental phonetics and phonological variation: case studies from Newcastle upon Tyne and Derby, in *Urban voices: Accent studies in the British Isles* (P. Foulke, & G.J. Docherty, editors), Londra: Arnold, 47-71.

Hajek, J. & Stevens, M. (2005), La preaspirazione delle occlusive sorde geminate nel senese. In *Misura dei Parametri - aspetti tecnologici ed implicazioni nei modelli linguistici. Atti del 1° Convegno Nazionale AISV 2004* (P. Cosi editor), Università di Padova, 2-4 Dicembre 2004, Padova: EDK Press, 89-100.

Helgason, P. (1999), Phonetic preconditions for the development of normative preaspiration, in *Proceedings of the International Congress of Phonetic Sciences*, San Francisco, USA, 1-7 Agosto 1999, 1851-1854.

Helgason, P. (2002), Preaspiration in the Nordic Languages: Synchronic and Diachronic Aspects, *Doctoral Thesis*, Università di Stoccolma ([http://www.lingfil.uu.se/personal/petur/pdf-filer/helgason\\_thesis.pdf](http://www.lingfil.uu.se/personal/petur/pdf-filer/helgason_thesis.pdf)).

Helgason, P. (2003), Faroese preaspiration, in *Proceedings of the International Congress of the Phonetic Sciences*, Barcellona, Spagna, 3-9 Agosto 2003, 2517-2520.

Helgason, P., Stölten, K. & Engstrand, O. (2003), Dialectal and sociophonetic aspects of preaspiration, in *Proceedings of the International Congress of Phonetic Sciences*, Barcellona, Spagna, 3-9 Agosto 2003, 17-20.

Indriðason, Þ., Eyþórsson, V. G., Halldórsson, Þ., Jónsson, J. G. & Bjarnadóttir, K. (1991), Mál er að mæla: um samhljóðalengd í íslensku. *Íslenskt mál og almenn málfræði*, 12-13, 143-190.

Jessen, M. & Pétursson, M. (1998), Voiceless nasal phonemes in Icelandic, *Journal of the International Phonetic Association*, 28, 43-53.

Jones, M. & Llamas, C. (2003), Fricated preaspirated /t/ in Middlesborough English: an acoustic study, in *Proceedings of the International Congress of Phonetic Sciences*, Barcellona, Spagna, 3-9 Agosto 2003, 655-658.



- Ladefoged, P. & Maddieson, I. (1996), *Sounds of the World's Languages*, Oxford: Blackwell.
- Schreier, D. (2005), On the Loss of Preaspiration in Early Middle English, *Transactions of the Philological Society*, 103/1, 99-112.
- Stevens, M. & Hajek, J. (2004a), Raddoppiamento sintattico and word-medial gemination in Italian: are they the same or are they different? The evidence from spontaneous speech, in *Theoretical and Experimental Approaches to Romance Linguistics* (R. Gess & E. J. Rubin editors), Amsterdam: John Benjamins, 257-271.
- Stevens, M. & Hajek, J. (2004b), How pervasive is preaspiration? Investigating sonorant devoicing in Sienese Italian, in *Proceedings of the 10th Australian International Conference on Speech Science and Technology* (S. Cassidy, F. Cox, R. Mannell & S. Palethorpe, editors), Sydney: ASSTA, 334-339.
- Stevens, M. & Hajek, J. (2007), Towards a phonetic conspectus of preaspiration: acoustic evidence from Sienese Italian. In *Proceedings of the 16th International Congress of Phonetics Sciences* (J. Trouvain & W. J. Barry, editors), Saarbrücken, Germania, 6-10 Agosto 2007, 429-432.
- Thráinsson, H. (1978), On the Phonology of Icelandic Preaspiration, *Nordic Journal of Linguistics*, 1, 3-54.

# **L'INSIEME È LA SOMMA DELLE PARTI CONSIDERAZIONI PRELIMINARI SU DATI ARTICOLATORI DELLA LINGUA OTTENUTI TRAMITE ULTRASUONI ED ARTICULOGRAFO ELETTROMAGNETICO 3D**

Mirko Grimaldi

Centro di Ricerca Interdisciplinare sul Linguaggio ([CRIL](#)) – Università del Salento (Lecce)

[mirko.grimaldi@ateneo.unile.it](mailto:mirko.grimaldi@ateneo.unile.it); [grimaldi.mirko@alice.it](mailto:grimaldi.mirko@alice.it)

## **1. SOMMARIO**

Lo studio articolatorio del linguaggio è fortemente condizionato dalle caratteristiche fisiologiche degli organi dell'apparato fonatorio da un lato, e dalle caratteristiche tecniche della strumentazione attualmente disponibile dall'altro. Tuttavia negli ultimi anni i dati ottenuti grazie ad alcune di queste tecnologie hanno consentito di rivoluzionare il modo di vedere il tratto vocale. Informazioni che prima si potevano solo inferire dal segnale acustico sono ora accessibili direttamente.

Più in particolare le tecniche di immagine in uso per monitorare i movimenti della lingua hanno portato ad ipotizzare una deformazione più complessa di questo organo. Ricorrendo anche a ricostruzioni tridimensionali della superficie della lingua, sono state avanzate alcune ipotesi che vedono la lingua come un insieme di segmenti coronali controllati in modo indipendente: i movimenti della lingua sarebbero cioè controllati dalla coordinazione sinergica di *segmenti muscolari funzionali*. Tali segmenti sembrerebbero essere organizzati ortogonalmente lungo l'asse longitudinale del tratto vocale e sarebbero composti da un sistema muscolare multiplo.

In questo lavoro, partendo da alcuni cenni di anatomia e fisiologia della lingua, discuteremo di due delle tecnologie più promettenti per lo studio articolatorio del linguaggio: gli Ultrasuoni (US) e l'Articulografo Elettromagnetico (EMA). Queste due tecniche, utilizzabili presso il CRIL, sono state sincronizzate, insieme ad un Elettroglossografo, per le esigenze sperimentali del progetto europeo [CONTACT](#) permettendoci di ottenere dati articolatori della lingua diversificati. Benché gli obiettivi del progetto CONTACT non fossero direttamente indirizzati allo studio articolatorio di determinate caratteristiche fonetico-fonologiche, i dati acquisiti ci offrono ugualmente la possibilità di fare riflessioni preliminari su possibili protocolli specificamente dedicati allo studio di segmenti o gruppi di segmenti funzionali all'interno delle lingue naturali. Oltre alle immagini ottenute con gli US, e alle posizioni dei sensori EMA sulla lingua, faremo ricorso alla ricostruzione tridimensionale della sua superficie, cercando di correlare le diverse informazioni ottenute.

Benché le potenzialità di queste due tecnologie possono essere sfruttate singolarmente, il perfezionamento del processo di acquisizione sincronizzata di dati EMA ed US offrirà opportunità di analisi articolatorie più potenti indirizzate, per esempio, a tutta la vasta gamma dei processi coarticolatori che caratterizza le varietà romanze di area italiana, già rivelatisi di enorme interesse per le teorie fonologiche contemporanee, e che ora – insieme ad altri livelli del linguaggio ancora poco esplorati – potrebbero essere guardati sotto una nuova luce e meglio interpretati anche in una prospettiva multidisciplinare.

## 2. LO STUDIO ARTICOLATORIO DEL LINGUAGGIO: PROBLEMI E PROSPETTIVE MULTIDISCIPLINARI

Mentre il segnale acustico prodotto dai movimenti dell'apparato fonatorio assume la forma di un'onda sinusoidale continua, permettendo così un'agevole analisi delle caratteristiche acustiche del linguaggio, lo studio degli articolatori che tale segnale producono si presenta irto di innumerevoli difficoltà. Gli articolatori che compongono il tratto vocale, infatti, non si muovono tutti alla stessa velocità, sono strutturati in modo differenziato, e presentano dimensioni, forma e capacità di movimento diverse a seconda se siano formati da tessuti molli (lingua, labbra, velo) o da tessuti duri (mandibola, palato). Le peculiarità fisiologiche degli articolatori, peraltro, incidono sulla velocità con cui essi si muovono, per cui uno strumento adeguato a catturare le risposte motorie della mandibola non si rivela utile per analizzare i movimenti della lingua. Allo stesso modo articolatori che sono visibili ad una analisi superficiale, come le labbra, sono molto più facili da monitorare rispetto a strutture collocate in profondità nella cavità orale, come per esempio il velo.

Ma più di ogni altro fattore, è l'interazione fra gli articolatori a generare serie complicazioni per lo studio articolatorio del linguaggio. Alcuni gesti articolatori sembrano molto più correlati di altri, per cui diventa complicato distinguere il contributo di un articolatore rispetto agli altri. Come osserva Stone (1997), se per esempio il movimento della mandibola è correlato con i movimenti della parte anteriore della lingua (in particolare della punta), tale correlazione viene progressivamente a mancare con i movimenti che avvengono posteriormente.

Un altro caso esemplificativo è quello della /u/ dell'inglese americano studiato da Perkell *et alii*, 1993 con l'Articulografia Elettromagnetica: è stato infatti notato che nella produzione ripetuta di questo fonema – caratterizzato da una leggera restrizione prodotta dall'innalzamento della lingua nella regione velo-palatale e dall'arrotondando delle labbra – non persiste una correlazione sistematica fra i movimenti della lingua e quelli delle labbra. Gli articolatori in questione, quindi, sembrerebbero muoversi attraverso uno scambio motorio reciproco. Gli autori ipotizzano che, data la reciproca variazione di due articolatori controllabili indipendentemente, alla base della realizzazione della /u/ vi sia un goal acustico/uditivo piuttosto che un goal gestuale (motorio).<sup>1</sup>

Ci sono poi anche problematiche legate agli strumenti attualmente a disposizione, per cui se l'elettropalatografo permette di determinare in modo eccellente i contatti della lingua con il palato, allo stato attuale non abbiamo nessuno strumento che, per quanto flessibile, individui i punti della lingua dove questi contatti avvengono (anche perché uno strumento del genere limiterebbe l'espansione, la contrazione, la flessione e la torsione della superficie della lingua). Infine – benché negli ultimi anni siano stati fatti alcuni passi avanti – risulta ancora molto problematico individuare la correlazione fra i complessi movimenti dei muscoli orofacciali e la produzione dei suoni della lingua naturali (cf. Hardcastle, 1999; Jou *et alii*, 2006).

Per questi motivi gli strumenti per studiare il livello articolatorio del linguaggio devono essere non invasivi, in grado di adattarsi a una determinata superficie (come l'elettropalatografo), avere dimensioni ridottissime per essere posizionati su superfici molli (come i sensori per l'articulografo elettromagnetico e i sensori per l'elettromiografia), o capaci di monitorare i movimenti della lingua dall'esterno (come le diverse tecniche di immagine). A causa di tutte queste difficoltà per la maggior parte degli studiosi risulta più

---

<sup>1</sup> Dati equivalenti sono stati ottenuti per la /r/ da Guenther *et alii*, 1999.

agevole analizzare il livello acustico del linguaggio e inferire dai dati ottenuti gli eventi fisiologici ad essi correlati.

Tuttavia negli ultimi anni gli studi articolatori sono stati protagonisti di uno sviluppo importante, producendo ricadute significative anche a livello dei modelli teorici. Alcune tecniche di immagine a disposizione hanno rivoluzionato il modo di vedere il tratto vocale e hanno permesso di avere informazioni anche su strutture profonde come quelle della laringe. Nello stesso tempo le tecniche articulografiche si sono sempre più evolute e hanno trasformato le idee circa un fenomeno universale come quello della coarticolazione, mostrando relazioni fra gli articolatori a cui prima era riservata solo una speculazione teorica (cf. Hoole *et alii* 1993; Hardcastle *et alii*, 1996; Hardcastle & Nigel, 1999; Hoole & Nguyen, in press; Cho, 2002; Recanses 2002).

Fra le tecniche d'immagine più diffuse bisogna menzionare i *Raggi-X*, la *Xeroradiografia*, la *Tomografia*, la *Risonanza Magnetica*, e gli *Ultrasuoni* (US), mentre fra le tecniche articulografiche, vanno citati l'*Articulografia Elettromagnetica* (EMA) e i vari sistemi *Optotrak* (per approfondimenti si veda Stone, 1997). In questa costellazione di tecniche strumentali vanno ovviamente inseriti le già menzionate *Elettropalatografia* ed *Elettromiografia*, mentre un discorso a parte meriterebbero le tecniche di indagine dell'attività laringea, come *Elettroglottografia* e l'*Analisi Stroboscopica* (cf. Hirose, 1997).

Grazie a queste nuove possibilità tecnologiche anno visto la luce una gran quantità di ricerche sul controllo motorio del parlato e sui meccanismi alla base della produzione dei suoni delle lingue naturali: scopo principale è l'identificazione delle invarianti motorie, ovvero di quegli elementi che stanno alla base della programmazione motoria del parlato. Per fare ciò è diventato essenziale rispondere alla seguente domanda, così formulata da Perkell (2007: 6): "What is the *task space*, or the domain of the fundamental control parameters?"

Uno degli approcci più incisivi a riguardo si fonda sulla constatazione, ben nota, che ogni parlante produce stringhe sonore sotto forma di segnale acustico e che tali stringhe, attraverso il sistema uditivo, possono essere trasformate in strutture intelligibili da parte di un ascoltatore: "These acoustic cues consist mainly of time-varying patterns of formant frequencies for vowels and glides, and noise bursts, silent intervals, aspiration and frication noises, and rapid formant transitions for consonants" (Perkell 2007: 6). Le proprietà di tale segnale sono determinate da parametri che possono essere osservati in diversi domini, come per esempio:

- nei livelli della tensione muscolare
- nei movimenti degli articolatori
- nei cambiamenti funzionali che avvengono all'interno del tratto vocale
- negli eventi aerodinamici.

Ogni possibile combinazione di questi parametri può ipoteticamente rappresentare una variabile importante del controllo motorio del linguaggio.

Tale approccio combinato con metodologie di ricerca neurocognitive ha prodotto numerose evidenze a supporto dell'idea che nella programmazione articolatoria fonemica siano pienamente coinvolti i domini uditivi e somatosensoriali. Proprio da questo filone di ricerca è nato uno dei più promettenti modelli neurocognitivi della pianificazione motoria denominato DIVA. Si tratta di un sistema basato sulle reti neurali che impara a controllare i movimenti di un apparato vocale virtuale per produrre e percepire i suoni delle lingue naturali. In questo modello i goal fonemici sono codificati sotto forma di proiezioni (mappature) neurali dalla corteccia premotoria alla corteccia sensoriale: tali mappature

delineano regioni corticali e sottocorticali (come il cervelletto) all'interno di aree uditive-temporali e somatosensoriali-temporali. Il modello utilizza la codifica degli errori del feedback uditivo per apprendere e sviluppare corrette proiezioni motorie. Una volta che il sistema linguistico acquisito diventa fluente, suoni, sillabe e parole possono essere decodificate in modo autonomo senza ricorrere al feedback uditivo (cf. Guenther *et alii* 2006).

Un tentativo ancora più pionieristico di combinare dati articolatori con approcci neurocognitivi è quello condotto all'interno del progetto europeo [CONTACT](#). Obiettivo del progetto CONTACT è l'identificazione delle invarianti motorie e neurali che sono contemporaneamente coinvolte durante lo sviluppo della percezione e produzione del linguaggio e che, secondo la teoria dei neuroni specchio, sarebbero le stesse interessate nei processi di manipolazione (cf. Rizzolatti & Craighero 2004). L'esistenza di tali invarianti neurali verrà testata attraverso un sistema di apprendimento artificiale (basato sul modello delle reti neurali), ovvero un baby-robot, che cercherà, partendo dall'elaborazione di dati articolatori grezzi, di dati motori e dati audio-visivi, di sviluppare autonomamente capacità sensorimotorie diversificate.

Per l'acquisizione dei dati articolatori il progetto CONTACT ha fatto ricorso al sistema US ed EMA 3D, insieme all'Elettroglottografo presenti al CRIL. Questi strumenti (vedi 5) sono stati sincronizzati, consentendo l'acquisizione simultanea di informazioni relative al contorno sagittale della lingua e al tracciato prodotto da un set di sensori collocati sulla superficie della stessa.<sup>2</sup>

In questa sede ci concentreremo esclusivamente su due di queste tecniche strumentali: gli US e l'EMA 3D. Trattandosi di protocolli sperimentali pionieristici e non propriamente finalizzati allo studio diretto delle caratteristiche articolatorie del linguaggio, i dati che discuteremo oltre hanno solo la funzione di illustrare le potenzialità di questi strumenti per lo studio articolatorio del linguaggio. Tale limitazione è anche dovuta alla scelta, ipoteticamente plausibile per il progetto CONTACT, di acquisire i dati US senza stabilizzare la testa dei parlanti. Dal momento che, come vedremo più avanti in 5.1, gli US non riproducono le immagini delle strutture ossee o di altre strutture fisse, è assolutamente necessaria la stabilizzazione della testa del parlante e della sonda perché si abbiano dei parametri di riferimento fissi del tracciato sagittale della lingua da analizzare in studi di tipo articolatorio (cf. Stone 2005)<sup>3</sup>.

### 3. LA LINGUA E IL SUO CONTROLLO MUSCOLARE: IPOTESI RECENTI

L'osservazione della complessa deformazione della lingua durante la produzione del parlato, dovuta (fra le altre cose) al già menzionato fenomeno della coarticolazione, solleva un problema rilevante: quali sono i meccanismi che controllano un tessuto muscolare altamente deformabile per produrre gesti motori così fini, essenziali per la realizzazione dei suoni delle lingue naturali?

I classici lavori sulla coarticolazione (Ohman, 1967, Mermelstein, 1973, Hardcastle, 1976, Hardcastle, 1985) hanno portato ad ipotizzare una suddivisione fra la punta e il corpo della lingua che produrrebbero dei movimenti quasi indipendenti in posizione avanzata o

---

<sup>2</sup> Per maggiori dettagli e informazioni inerenti anche i dati elettroglottografici si veda Grimaldi *et alii*, 2007, Grimaldi *et alii*, 2008.

<sup>3</sup> Al CRIL è stato di recente realizzato un sistema in legno, che non provocando perturbazioni del campo magnetico del cubo EMA (cf. 5.2), immobilizza la testa dei soggetti durante l'acquisizione sincronizzata di dati US ed EMA.

arretrata, e verso l'alto o verso il basso. Questa suddivisione è stata anche assunta sia dai modelli geometrici della lingua (Coker & Fujimura, 1965; Mermelstein, 1973), che da quelli della sintesi articolatoria (Rubin, 1981). I modelli basati sull'analisi fattoriale hanno invece individuato due o tre fattori basilari che, combinati fra loro, permettono di catturare la forma completa del contorno della lingua, in special modo per quanto concerne le vocali (cf. Harshman, Ladefoged & Goldstein, 1977; Maeda, 1989; Stone, Goldstein & Zhang, 1997; Slud, Stone, Smith & Goldstein, 2002).

Tuttavia, grazie alle recenti tecniche di immagine – che hanno permesso di visualizzare piani multipli della lingua – è possibile ipotizzare che la deformazione di questo organo è sicuramente più complessa. Questi dati sono poi stati già usati per ricostruzioni tridimensionali della superficie della lingua, che mostrano deformazioni molto più complesse di quelle presupposte da un modello basato esclusivamente sulle funzionalità del corpo e della punta della lingua (cf. Wilhelms-Tricarico, 1995; Stone & Lundberg, 1996; Yang & Stone, 2002).

Ipotesi ancora più recenti tendono a considerare la lingua come un insieme di segmenti coronali controllati in modo indipendente (Stone, Epstein & Iskarous, 2004): i movimenti della lingua sarebbero cioè controllati dalla coordinazione sinergica di *segmenti muscolari funzionali*. Tali segmenti sono organizzati ortogonalmente lungo l'asse longitudinale del tratto vocale e sono composti da un sistema muscolare multiplo.

A supporto di questa ipotesi gli stessi autori evidenziano che alcune ricerche nel campo della neuroanatomia hanno dimostrato che la lingua è caratterizzata da un sistema di innervazione molto complesso. Alcuni studi hanno messo in evidenza che il nucleo ipoglosso (il complesso di nervi che irradia la lingua) è composto di circa 13.000 motoneuroni (Wozniak & Young, 1969; Atsumi & Miyatake, 1987; O'Kusky & Norman, 1995). Per esempio ogni fibra del muscolo genioglosso contiene un proprio motoneurone (Mu & Sanders, 1999; Takemoto, 2001): dal momento che ogni motoneurone è potenzialmente un'unità di controllo indipendente, il genioglosso potrebbe avere segmenti di controllo multipli (cf. Miyawaki, 1975). Per queste ragioni la neuroanatomia della lingua fornisce evidenze per ipotizzare un controllo selettivo della lingua, opposto a una visione della lingua come un blocco unico.

Partendo da questi presupposti alcuni gruppi di ricerca stanno attualmente verificando la possibilità che la lingua sia suddivisa in segmenti funzionali correlati alla sua conformazione fisiologica (e in parte controllati anche in modo indipendente), piuttosto che vedere l'organo suddiviso in blocchi muscolari che si muovono nel tratto vocale lungo gli assi verticale ed orizzontale. L'idea di base è che questi segmenti sarebbero controllati o in modo indipendente o si aggregerebbero in unità più ampie per creare determinate strutture articolatorie funzionali all'interno di una determinata lingua naturale (cf. Stone, Epstein, Iskarous, 2004).

#### 4. CENNI DI ANATOMIA E FISIOLOGIA DELLA LINGUA

Caratteristica principale della lingua è quella di essere un organo muscolare che non ha struttura cartilaginea od ossea, e per questo motivo riesce a mantenere un volume costante quando i suoi muscoli si contraggono. Attraverso una selettiva contrazione di determinati muscoli la lingua può creare un supporto rigido che permette la contrazione di certe fibre muscolari, generando così movimenti selettivi mirati (Takemoto, 2001).

I muscoli della lingua sono suddivisi in *intrinseci* ed *estrinseci* (vd. Figure 1-3). I muscoli intrinseci hanno origine e sono inseriti all'interno della lingua stessa, formando una matrice tridimensionale di fibre motorie così suddivisa (cf. Kent 1997: 171-203):

- *muscolo superiore longitudinale* (SL): si colloca immediatamente sotto la superficie della lingua, e consente il suo accorciamento e la curvatura verso l'alto della lamina.
- *muscolo inferiore longitudinale* (IL): si trova al di sotto del muscolo superiore longitudinale e consente di abbassare la punta della lingua.
- *muscolo trasverso e verticale* (TV): il primo va da lato a lato della lingua, mentre il secondo si estende inferolateralmente dal dorso della lingua: entrambe consentono l'abbassamento e l'appiattimento della superficie linguale.

In sintesi, questo complesso muscolare è responsabile dei movimenti fini della lingua ed è coinvolto nei più delicati adattamenti necessari alla lingua per assumere forme diverse.

I muscoli estrinseci connettono la lingua alle strutture vicine, come la mandibola, l'osso ioide, il palato, l'osso temporale e la faringe e si possono schematicamente descrivere come segue:

- *muscolo genioglosso* (GG): ha origine nella spina mentale della mandibola e s'inserisce nell'osso ioide. Si tratta di un complesso differenziato di fibre che comprende la maggior parte della massa della lingua e, a seconda di quali fibre si contraggono, produce la protrusione della punta della lingua, oppure movimenti in avanti e verso l'alto.
- *muscolo stiloglosso* (SG): ha forma allungata, origina dall'apice del processo stiloideo (osso temporale) e dalla parte alta del legamento stilomandibolare e da qui si porta in basso, in avanti e medialmente per inserirsi nell'apice e nel setto linguale. Contraendosi muovono la lingua superiormente e dorsalmente.
- *muscolo ioglosso* (IG): origina dalla parte laterale della faccia superiore dell'osso ioide e dal grande corno, da qui si porta in alto e medialmente per inserirsi sulle facce laterali del setto linguale fino all'apice. Contraendosi ritrae e abbassa il corpo della lingua.
- *muscolo geniioioide e mioioide* (GI) e (MI): originano dall'osso ioide e si estendono verso la parte anteriore della mandibola. Si trovano alla base della bocca e permettono l'elevamento dell'osso ioide e la lingua.

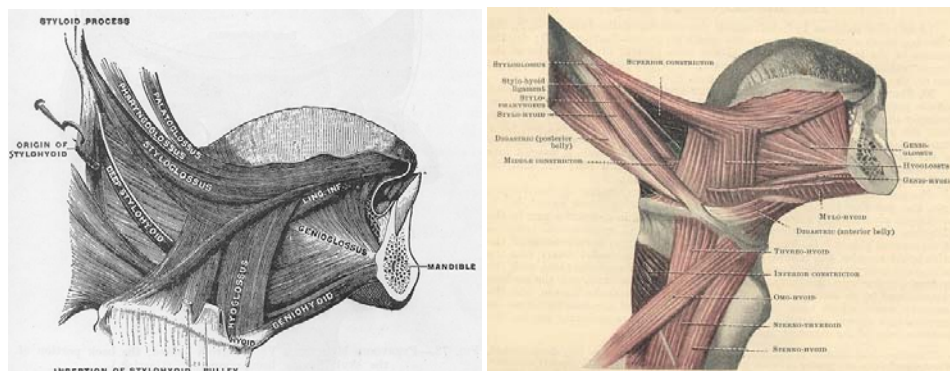


Figure 1-2: muscoli intrinseci ed estrinseci della lingua. Immagini adattate da Gray (1918).



Figura 3: immagine sagittale della lingua (adattata da Stone, Epstein, Iskarous 2004).

## 5. IL SISTEMA AD ULTRASUONI E L'ARTICULOGRAFO ELETTROMAGNETICO (EMA)

### 5.1 Il sistema ad ultrasuoni

Gli ultrasuoni per monitorare e studiare i movimenti della lingua stanno diventando un sistema sempre più accessibile e diffuso (cf. Stone 2005). Benché l'utilizzo dei classici ecografi in un laboratorio di linguistica sia ancora limitato dal costo eccessivo, sono da poco stati immessi sul mercato sistemi portatili che consentono ugualmente buone analisi ecografiche. Aspetti della produzione del linguaggio, precedentemente dedotti da dati parziali o indiretti, possono ora essere osservati direttamente. Dal momento che gli ultrasuoni non sono invasivi, forniscono un segnale abbastanza agevolmente interpretabile, e sono per giunta diventati portatili, si prestano molto bene sia per l'uso prettamente clinico che per vari aspetti della ricerca sperimentale. Negli ultimi anni si è perciò iniziato a riflettere su come impiegarli anche nelle indagini geolinguistiche sul campo, o in settori teorico-applicativi come lo studio articolatorio dei processi di acquisizione della seconda lingua (cf. Gick *et alii*, 2002, Gick, 2008).

Un sistema ad ultrasuoni emette onde sonore ad alta frequenza (2-14 MHz) attraverso un trasduttore, o sonda (in genere convessa), che contiene cristalli piezoelettrici stimolati da tensione elettrica. Quando la sonda è collocata contro la pelle del mento, gli ultrasuoni attraverso i tessuti molli (come quelli della lingua), e vengono riflessi quando raggiungono un alto gradiente di impedenza acustica (tipicamente quella generata dall'interfaccia lingua-aria o lingua-palato). I cristalli piezoelettrici emettono in modo sequenziale nel tempo le onde sonore e ricevono gli echi riflessi: in questo modo è possibile catturare sezioni di tessuto piuttosto che un singolo punto. Gli echi sono quindi processati da un computer per essere trasmessi in forma di immagini continue sul sistema video dell'ecografo, oppure possono essere registrati per ulteriori analisi (Hedrick *et alii* 1995).

Siccome, come abbiamo detto, gli ultrasuoni sono riflessi dall'interfaccia lingua-aria o lingua-palato è possibile visualizzare solo la lingua e non, per esempio, il palato, la mandibola, o le pareti posteriori della faringe. A seconda se la sonda venga collocata in modo perpendicolare od orizzontale sotto il mento si può, rispettivamente, ottenere l'immagine sagittale e coronale della lingua (vd. Figure 4 e 5). L'area più luminosa rappresenta la lingua del soggetto: più luminoso è il dettaglio, maggiore è il gradiente di densità del tessuto e, di conseguenza, l'eco acquisita. La lamina della lingua è chiaramente



visibile in quanto presenta un alto gradiente di densità. Tuttavia la mandibola e l'osso ioide creano un'"ombra acustica" in entrambe i margini dell'immagine, perché rifrangono l'ultrasuono prima che raggiunga la superficie della lingua: in questo modo possono essere oscurate parti della punta (soprattutto quando si trova in posizione elevata) o della radice della lingua.

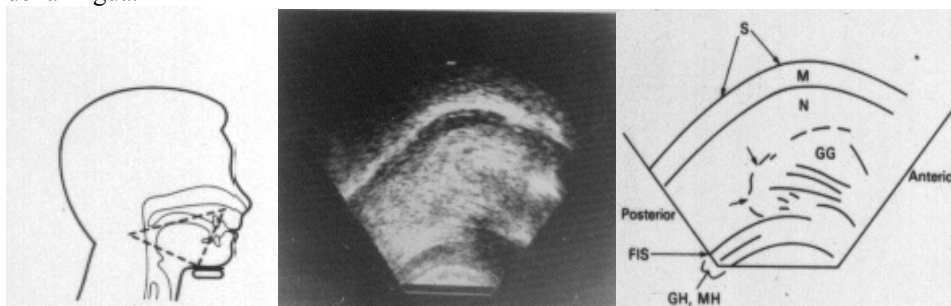


Figura 4: collocazione della sonda a sinistra, immagine sagittale della lingua al centro e sua rappresentazione schematica a destra adattata da Stone (2005): S = Tongue Surface; M = mucosa; N = Superior, Inferior longitudinalis, and Vertical, Horizontal network of fibres; GG = Genioglossus; FIS = Floor Intermuscular septum; GH = Geniohioide; MH: Miohioide.



Figura 5: collocazione della sonda a sinistra, immagine coronale della lingua al centro e sua rappresentazione schematica a destra adattata da Stone (2005): S = Tongue Surface; M = mucosa; N = Superior, Inferior longitudinalis, and Vertical, Horizontal network of fibres; GG = Genioglossus; FIS = Floor Intermuscular septum; GH = Geniohioide; MH: Miohioide.

MFS = Median Fibrous Septum, LM = Lateral Muscles, J = Jaw Inner Aspect, PS = Paramedian Septum, CF = Cervical Fascia.

I dati discussi in questo lavoro sono stati acquisiti con un sistema Aplio XV, costruito da Toshiba Medical System corp., che consente anche l'esportazione delle immagini ecografiche sotto forma di un flusso video continuo, ad una frequenza di 25 immagini al secondo, grazie all'apposito uscita S-Video. Il flusso video è stato acquisito in maniera sincrona con il segnale acustico per mezzo di una scheda esterna di acquisizione, e quindi registrato in tempo reale su un PC dedicato allo scopo (per maggiori dettagli cf. Grimaldi *et alii* 2007).

### 5.2 L'articulografo elettromagnetico 3D

Come il sistema ad ultrasuoni, anche l'articulografo elettromagnetico (EMA, ElectroMagnetic Articulograph) permette di ottenere importantissime informazioni cinematiche della lingua. Tuttavia sino a poco tempo fa questa tecnica era limitata da severe restrizioni imposte ai movimenti dei soggetti analizzati, compromettendo così la naturalezza delle loro produzioni. L'introduzione di un sistema di rilevamento tridimensionale all'interno di un cubo ha permesso di risolvere questi problemi. In particolare, il modello AG500, costruito da Carstens Medizinelektronik GmbH (Germania), consente di tracciare la posizione di un set di sensori. Tali sensori possono essere incollati su lingua, labbra e denti incisivi, nonché su altri punti presi come riferimento al fine di compensare gli eventuali movimenti della testa compiuti dal soggetto durante l'esperimento.

L'EMA è in grado di determinare le coordinate cartesiane  $x$ ,  $y$  e  $z$ , nonché azimuth ed elevazione di un massimo di 12 sensori di campo magnetico, ad una frequenza di campionamento pari a 200 Hz. I sensori di misura sono costituiti da piccole bobine a singolo asse: altre 6 bobine, posizionate in modo da formare un riferimento tridimensionale, emettono altrettanti campi magnetici a differenti e ben note frequenze, tra i 7500 e i 13750 Hz. L'idea principale che sta dietro il processo di determinazione delle coordinate spaziali dei sensori è che la distanza fra i sensori e le bobine di riferimento sia determinata dai valori assoluti delle correnti indotte. Il moto di un sensore e la direzione vettoriale di un sensore può essere così calcolato partendo dalle sei bobine di riferimento usando un metodo iterativo (Zierdt & Carstens, 1995).

Un limite dell'EMA è rappresentato dalla sua sensibilità ad oggetti che permettono la conduzione di elettricità o di campi magnetici. Se uno o più di questi oggetti sono collocati nelle vicinanze delle bobine, il segnale registrato viene distorto. L'EMA deve essere calibrato prima di ogni registrazione, in modo da sintonizzare le sei bobine di riferimento con i dodici sensori (o meno, a seconda del numero di dati richiesti dal protocollo sperimentale). I sensori sono collocati in un apposito caricatore installato su un disco (Circal) in alto al centro del cubo EMA e mosso da un motore elettrico. Il motore, grazie a un software sviluppato da Carstens Medizinelektronik GmbH, muove circolarmente ogni sensore sino a che non ne viene individuata la posizione precisa all'interno del campo magnetico e le corrispondenti ampiezze di corrente elettrica non sono state acquisite (cf. Tavella 2007).

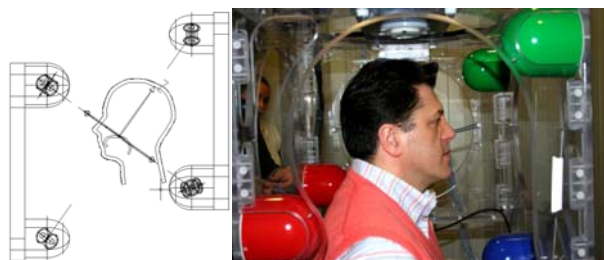


Figura 6: rappresentazione schematica del cubo EMA a sinistra, e immagine di una sessione di prova al CRIL a destra. Sono visibili cinque delle sei bobine di riferimento: due rosse, due verdi e una blu (la sesta, blu, si trova proprio dietro la testa del soggetto in cima la cubo). In alto sulla testa del soggetto è visibile il disco (bianco) su cui, nella fase della calibratura, è inserito il caricatore che contiene i sensori. Immagini adattate da Zierdt *et al*, 1999 e da Tavella, 2007.



Figura 7: immagine in primo piano del disco e del caricatore con quattro sensori a sinistra, e di cinque sensori EMA incollati sulla lingua di un soggetto prima di un esperimento a destra. Evidenziata dall'ellissi in rosso la collocazione dei sensori da cui sono stati tratti i dati discussi nel paragrafo in 6 (cf. Figure 6-14). Immagini adattate da Tavella, 2007.

Durante la sessione di registrazione, le correnti alternate indotte nei sensori dal campo magnetico generato dalle bobine di riferimento sono filtrate in frequenza, digitalizzate e trasmesse in tempo reale ad un PC di controllo attraverso una connessione Ethernet. Nel PC un software apposito, incluso in dotazione con l'AG500, registra su hard disk i valori ricevuti, rendendoli disponibili per la successiva fase di calcolo delle traiettorie seguite dai sensori mobili. Infine, un meccanismo di sincronizzazione (Sybox Opto-4) è connesso con l'EMA, permettendo la registrazione parallela di dati articolatori e del segnale vocale corrispondente (cf. Tavella 2007: 49-52).

## 6. PRELIMINARI PER UN'ANALISI ARTICOLATORIA DEI FONEMI DELL'ITALIANO

Pur con i limiti evidenziati in 2, i dati raccolti all'interno del progetto europeo CONTACT consentono di fare alcune riflessioni preliminari sulle potenzialità che gli strumenti sofisticati descritti in 5 offrono per lo studio articolatorio delle lingue naturali.<sup>4</sup> Tali limiti ci obbligano a prendere in considerazione le produzioni di un solo parlante dell'italiano salentino, visto che la scelta di non immobilizzare la testa dei soggetti registrati per il progetto CONTACT non rende possibile il confronto fra le stesse produzioni di parlanti diversi.<sup>5</sup> Tuttavia i dati di cui discuteremo fra poco, considerati alla stregua di

<sup>4</sup> Un altro limite è rappresentato dal fatto che i contesti segmentali del corpus non sono controllati, come in genere si richiede per l'osservazione sistematica di parametri sia acustici sia articolatori rilevanti.

<sup>5</sup> Sono stati acquistati i dati articolatori di parlato letto da 9 soggetti (6 di sesso maschile e 3 di sesso femminile), di età compresa fra i 23 e i 30 anni, tutti parlanti della varietà leccese di italiano. I soggetti hanno letto per tre volte 74 parole e 68 pseudo-parole, realizzate sia in forma dichiarativa che interrogativa. Parole e pseudo-parole sono state scelte in modo da rappresentare tutti i fonemi consonantici e vocalici attestati in italiano (esclusa l'opposizione /e/-/ɛ/ ed /o/-/ɔ/) in condizioni prosodiche paragonabili. Per questo motivo, le parole bersaglio hanno principalmente la prima sillaba accentata (ad esempio, /'matto, 'nome, 'strada/), il cui attacco è rappresentato da una delle consonanti bersaglio ed è seguito da diverse vocali (ad esempio, /'matto, 'muffa, 'moro/). Inoltre, per una maggiore rappresentatività del corpus, sono state anche inserite parole caratterizzate da una diversa posizione dell'accento (ad esempio, /mat'tone, pa'pa/). Le pseudo-parole, invece, sono monosillabi scelti in modo da rappresentare tutte le consonanti in posizione iniziale, seguite dalle vocali /a, u, i/ (ad esempio, /'na, 'nu, 'ni, 'la, 'lu, 'li/).

quelli provenienti da un esperimento pilota, si possono rilevare una buona base di partenza per calibrare in futuro protocolli meglio mirati allo studio di segmenti o gruppi di segmenti funzionali all'interno dei sistemi linguistici.

Per questo scopo mi pare utile provare ad analizzare le parti funzionali della lingua principalmente coinvolte nella realizzazione di un segmento fonologico complesso. Un buon esempio può essere rappresentato da un segmento fonologico affricato seguito da una vocale anteriore alta, come /tʃi/ (vedi Figura 8), prodotto da un parlante di sesso femminile. Come è abbondantemente noto:

- i suoni affricati sono realizzati da una prima fase occlusiva seguita da una breve fase fricativa (con un restringimento del tratto vocale).
- la /i/ è realizzata da un innalzamento ed avanzamento della lingua.

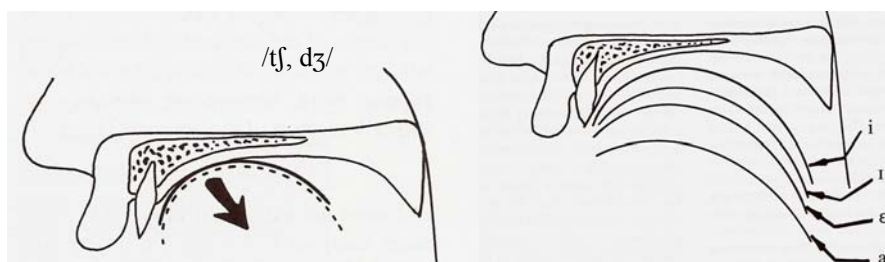


Figura 8: sulla sinistra rappresentazione schematica dell'articolazione di un segmento affricato. La freccia rappresenta il movimento che va dall'ostruzione (fase occlusiva) al restringimento del tratto vocale (fase fricativa). Sulla destra rappresentazione schematica della configurazione articolatoria relativa ad alcune vocali anteriori, fra cui /i/. Immagini adattate da Kent, 1997.

Utilizzando l'acquisizione sincronizzata di dati US e dati EMA, possiamo osservare le diverse fasi articolatorie attraverso cui la lingua dà forma a questo gruppo di fonemi, tentando, per quanto possibile, di focalizzare l'attenzione anche sui segmenti anatomici funzionali coinvolti.

Per visualizzare i risultati delle procedure di acquisizione dei dati, è stato implementato un agevole software di analisi. Il software è attualmente in fase di sviluppo da Francesco Sigona al CRIL e si basa sull'idea di integrare in un unico ambiente altri software già esistenti, specializzati ed ottimizzati nel compiere operazioni specifiche. Al momento l'ambiente si compone di PRAAT (realizzato da Boersma and Weenink, University of Amsterdam), Edgetrak (Li, *et alii*, 2005), il noto software open source Mplayer ed una interfaccia grafica (Graphic user interface), basata su Matlab e battezzata provvisoriamente soltanto GUI, in grado di interagire, entro certi limiti, con i sopra citati software, anche grazie alle funzionalità messe a disposizione dal linguaggio di scripting del sistema operativo e all'interfaccia a linea di comando di PRAAT e Mplayer (Cf. Grimaldi *et alii* 2007, Grimaldi *et alii* 2008 per maggiori dettagli).

Grazie a un algoritmo, anch'esso in fase di sperimentazione e perfezionamento presso il CRIL, i dati US a nostra disposizione possono essere correlati ai tre sensori EMA collocati sul piano sagittale della lingua e associati al suo contorno estratto dalle immagini US. come rappresentato nelle Figure 6-14 (la punta della lingua si trova nella parte sinistra di ogni figura, mentre la radice si trova sulla parte destra):

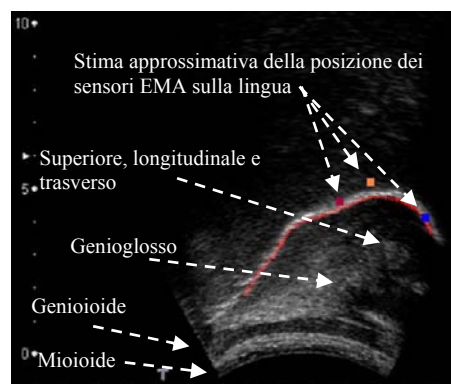


Figura 6: immagine sagittale del frame corrispondente alla fase di occlusione del segmento /tʃ/. La linea rossa individua il contorno della lingua. I quadrati colorati bordò, arancione e blu indicano la posizione approssimativa dei sensori EMA sulla lingua, rispettivamente sul dorso, sulla lamina e sulla punta.

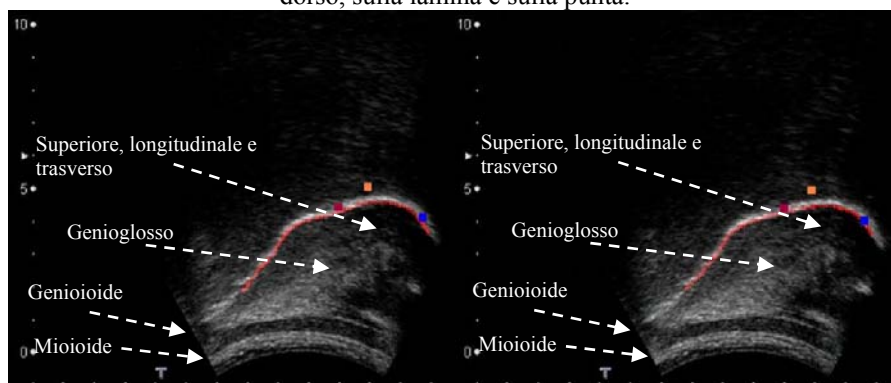


Figure 7-8: immagine sagittale del frame corrispondente alla fase di rilascio sulla sinistra ed alla fase di frizione del fonema /tʃ/ sulla destra.

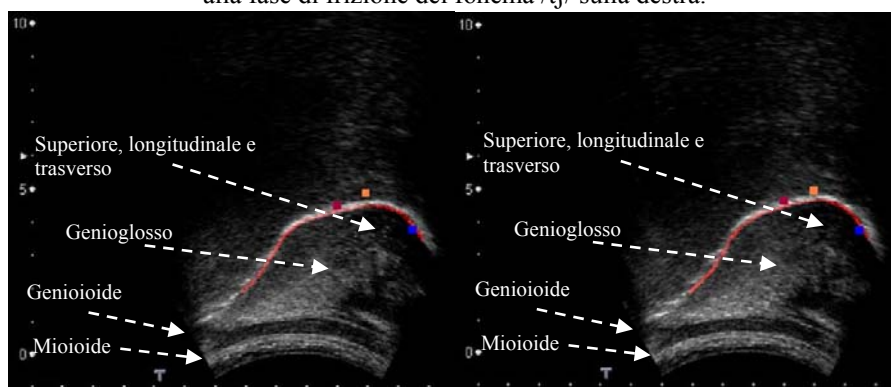


Figure 9-10: immagine sagittale del frame corrispondente alla fase interlocutoria di preparazione del processo di innalzamento nella realizzazione del fonema /i/ sulla sinistra e alla immediata fase di anteriorizzazione sulla destra.

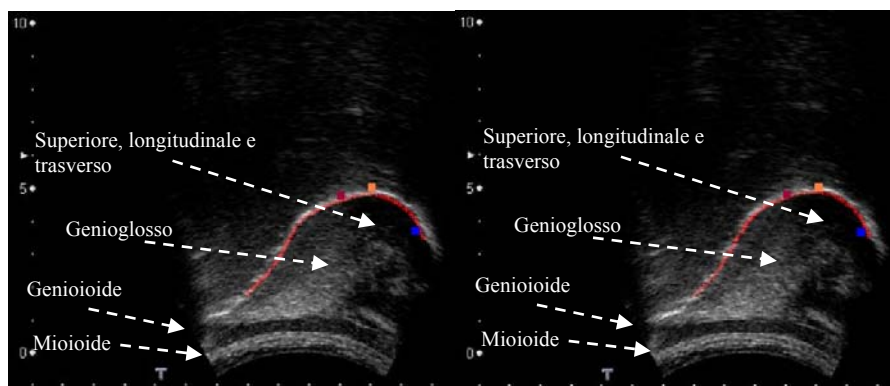


Figure 11-12: da sinistra a destra immagini sagittali dei frame corrispondenti al progressivo innalzamento ed anteriorizzazione della lingua durante la realizzazione del segmento /i/.

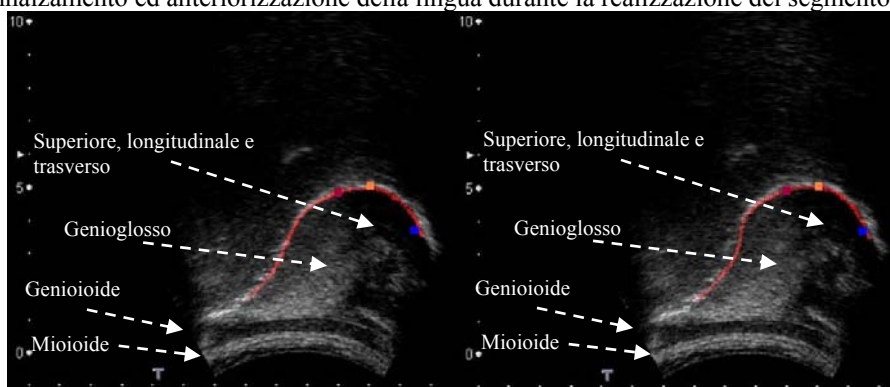


Figure 13-14: da sinistra a destra immagini sagittali dei frame corrispondenti alle fasi finali di innalzamento della lingua durante la realizzazione del segmento /i/.

Anche una semplice analisi dei frame in sequenza sembra offrire sufficienti dati per indagini più sistematiche in direzione dell'ipotesi secondo cui i goal articolatori della lingua, nella realizzazione dei segmenti fonologici, siano raggiunti attraverso l'attivazione differenziata di muscoli o di parti di muscoli di volta in volta diversi. In particolare, relativamente al caso preso in esame, si può ipotizzare che la realizzazione del segmento consonantico affricato e del segmento vocalico alto anteriore coinvolga un set di muscoli in modo differenziato. Come si può notare nelle Figure 6-8:

- nelle fasi di occlusione, di rilascio e di frizione sembrerebbe siano completamente coinvolti sia il muscolo genioglosso che i muscoli superiori longitudinali e trasversali.

Dalle Figure 9-14 emergerebbe invece che:

- nella realizzazione di /i/ l'anteriorizzazione precede la fase di innalzamento.
- le fasi di innalzamento e anteriorizzazione della lingua sono prodotte con l'espansione e la compressione solo della parte posteriore del muscolo genioglosso, unitamente ai muscoli superiori longitudinali e trasversali.



- l'espansione e la compressione della sola parte posteriore dei muscoli genioioide e mioioide paiono contribuire incisivamente nella realizzazione del tratto ATR che caratterizza il fonema /i/.

Dunque, limitatamente alla lingua – il principale organo dell'apparato vocale coinvolto nella fase cruciale in cui i suoni assumono la loro 'forma' definitiva – si potrebbe ipotizzare che un segmento fonologico complesso sia il risultato dell'attivazione differenziata e nello stesso tempo coordinata delle fibre di diversi muscoli linguali: in sostanza, il goal articolatorio finale sarebbe la somma di goal articolatori intermedi.

Questo complesso fenomeno può essere riconsiderato da un'altra prospettiva grazie alla possibilità di ottenere una rappresentazione grafica del tracciato dei sensori EMA e dei contorni sagittali della lingua nel tempo, che può essere relativa a un solo segmento fonologico, a una parola intera, o addirittura a una frase. Elaborando i dati US ed EMA precedentemente ottenuti, attraverso i punti x e y di ogni contorno e i valori t del tempo, possiamo infatti ricostruire la superficie della lingua in un ambiente 3D. Nel nostro caso il software GUI è in grado di estrarre automaticamente la rappresentazione 3D della lingua sia dal contorno sagittale ottenuto tramite US sia dal tracciato dei sensori EMA collocati sulla stessa.

Tuttavia le diverse tecniche di acquisizione dei dati articolatori offrono possibilità di rappresentazioni grafiche diverse. Abbiamo visto in 5.1 e 5.2 che il sistema US acquisisce ad una frequenza di 25 Hz (e quindi è in grado di registrare 25 immagini al secondo), mentre l'EMA ad una frequenza di 200 Hz (e quindi è in grado di registrare 200 posizioni al secondo per ogni sensore). Volendo ottenere una rappresentazione 3D di un segmento come /tʃi/, la cui durata si aggira intorno ai 0,6 secondi, risulta evidente che i dati ottenuti con l'EMA ci permettono di rappresentare proficuamente le diverse fasi articolatorie anche di segmenti che hanno breve durata, mentre l'US ci consente solo di rappresentare il segmento per intero.

Correlando con il software GUI il segnale acustico ai dati articolatori forniti dai sensori EMA (cf. Grimaldi *et alii*, 2007) e utilizzando la tecnica dell'interpolazione delle spline, possiamo ottenere rappresentazioni efficaci delle fasi di occlusione, rilascio, frizione, anteriorizzazione e innalzamento relativi alla produzione del segmento /tʃi/. Nelle Figure 15-20, possiamo infatti osservare le rappresentazioni grafiche 3D della lingua derivate dal tracciato dei sensori EMA collocati sul dorso, sulla lamina e sulla punta della lingua.

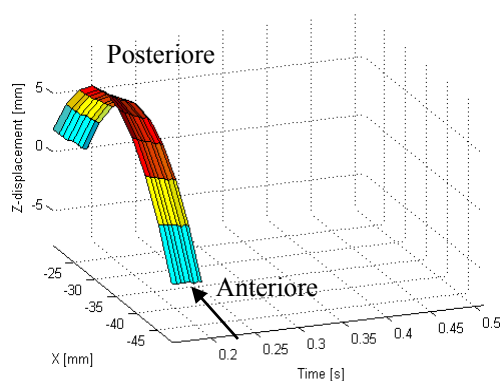


Figura 15: fase di occlusione. La freccia nera indica il probabile punto di contatto della punta della lingua con i denti.

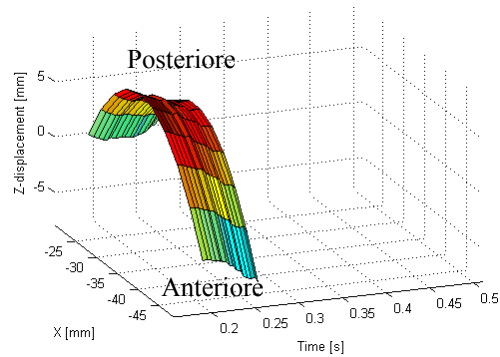


Figura 16: fasi di rilascio e frizione.

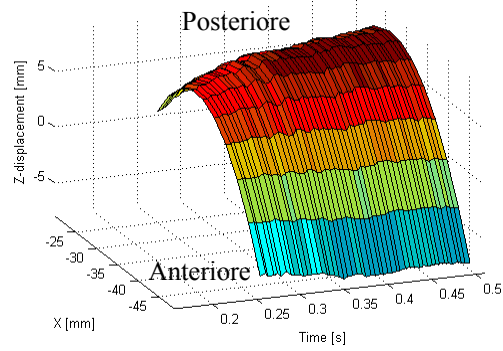


Figura 17: fasi di innalzamento e anteriorizzazione.

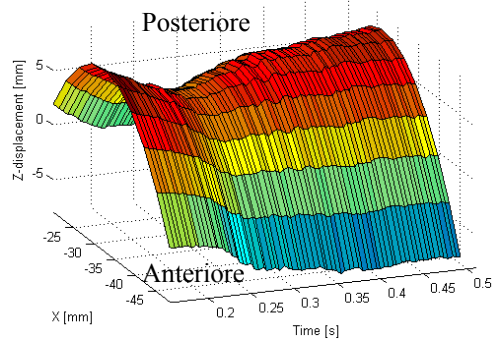


Figura 18: rappresentazione completa delle fasi di occlusione, rilascio, frizione, innalzamento e anteriorizzazione.

Allo stesso modo i contorni sagittali della lingua ricavati dalle immagini US (ed evidenziati nelle Figure 6-14), se proiettati nel tempo, possono essere elaborati per realizzare una rappresentazione 3D, come illustrato nella Figura 19:



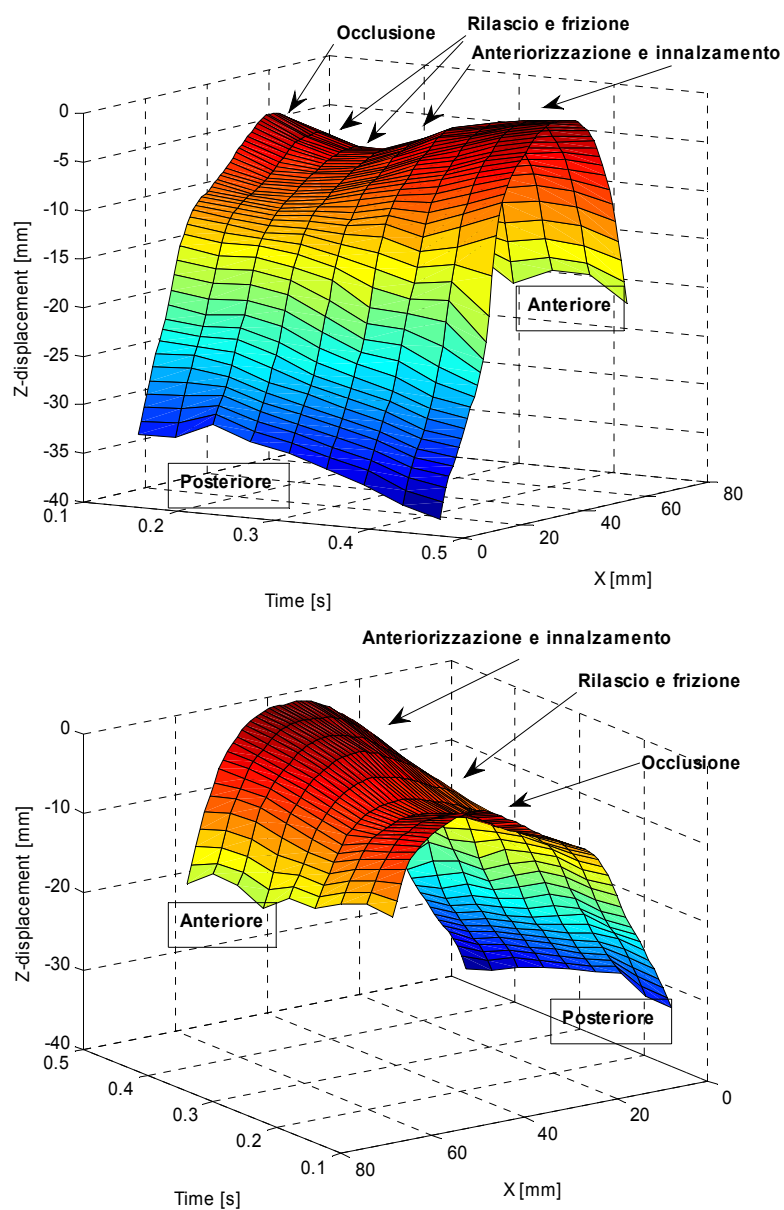


Figura 19: rappresentazione 3D da prospettive diverse dei contorni sagittali della lingua ottenuti tramite sistema US durante la realizzazione del segmento /tʃi/.

Se i dati EMA forniscono maggiori informazioni relativamente alla punta della lingua (che come abbiamo visto in 5.1 gli US non sempre riescono a monitorare), i dati US ci possono offrire informazioni più dettagliate in merito alla radice della lingua, in quanto i sensori EMA non possono essere collocati al di sotto del dorso.

Tentando di comparare, anche solo da un punto di vista impressionistico, le Figure 6-14 con le Figure 15-19 si può evincere come dalle rappresentazioni 3D relative ai dati EMA (e in particolare dalle Figure 15-18), è facile recuperare un particolare dato articolatorio inerente la punta della lingua: ovvero il graduale passaggio dalla posizione più elevata – in cui il contatto con gli incisivi superiori permette di creare la fase di occlusione – alla posizione di contatto con gli incisivi inferiori, attraverso il processo di interiorizzazione, innalzamento e di inarcamento della lingua, che permette il successivo allungamento e abbassamento della punta. Il fenomeno, nonostante i limiti del sistema US nel catturare i movimenti della punta della lingua, è sufficientemente visibile anche nel secondo grafico della Figura 19, che quindi conferma l’osservazione ricavata dai dati EMA.

## 7. CONCLUSIONI

In conclusione si può dire che mentre il sistema EMA – in quanto capace di decodificare in tempo reale la traiettoria di ogni sensore collocato su un punto della lingua – offre dati più robusti dal punto di vista temporale, il sistema US offre dati più efficaci da un punto di vista spaziale, in quanto l’onda sonora che si diffonde sotto il mento e si riflette cattura la maggior parte del corpo della lingua. Di queste potenzialità ci si può rendere meglio conto ricorrendo alle funzioni del software GUI che permette di assemblare in un unico ambiente i dati ottenuti durante la realizzazione del segmento /tʃi/ dal sistema US, insieme alla rappresentazione 3D di tutti i sensori collocati sulla lingua (cf. Figura 7), aggiungendo le informazioni ottenute dall’Elettroglottografo (EGG), insieme allo spettrogramma correlato e al filmato del soggetto durante l’esperimento, come si può vedere cliccando sulla [Figura 21](#):

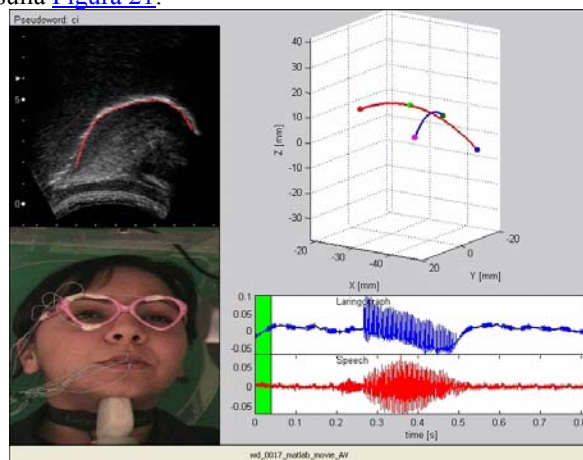


Figura 21: dati US, EMA, EGG insieme allo spettrogramma e al filmato del soggetto di sesso femminile che produce il segmento /tʃi/.

Al di là delle potenzialità delle singole tecnologie ora descritte, il perfezionamento del processo di acquisizione sincronizzata di dati EMA ed US, lo sviluppo e la verifica dell’algoritmo che permette di associare i sensori EMA all’immagine sagittale della lingua, e l’implementazione continua del software GUI, insieme alla stabilizzazione della testa dei soggetti monitorati (attualmente già possibile al CRIL), potrà sicuramente offrire possibilità di analisi articolatorie molto più complesse e mirate di quelle qui superficialmente

abbozzate a solo fine esemplificativo. Penso ovviamente a tutta la vasta gamma di processi coarticolatori che caratterizza le varietà romanze di area italiana, che si sono già rivelati di enorme interesse per le teorie fonologiche contemporanee, e che ora potrebbero essere guardati sotto una nuova luce e meglio interpretati anche in un prospettiva multidisciplinare.

## BIBLIOGRAFIA

- Cho, T. (2002), *The Effects of Prosody on Articulation in English*, Routledge, New York, NY.
- Gick, B., (2002), The use of ultrasound for linguistic phonetic fieldwork, *Journal of the International Phonetics Association*, 32(2): 113-122.
- Gick, B., Bernhardt B., Bacsfalvi P., Wilson I. (2008), Ultrasound imaging applications in second language acquisition, in J. G. Hansen Edwards & M. L. Zampini, *Phonology and Second Language Acquisition*, Benjamins, Amsterdam-Philadelphia, 309-322.
- Gray, H., (1825–1861), *Anatomy of the human body*, 20th ed., thoroughly rev. and re-edited by Warren H. Lewis, Philadelphia: Lea & Febiger, 1918.
- Grimaldi, M., Gili Fivela, B., Tavella, M., Sigona, F., Fitzpatrick, P., Metta, G., Craighero, L., Fadiga, L., Sandini, G., (2007), Synchronized acquisition of Italian speech articulatory data using ultrasound, 3D-articulograph and laryngograph: First results, in *Ultrafest VII*, New York, September 2007: <http://jerome.linguistics.fas.nyu.edu/ultrafest.html>.
- Grimaldi, M., Gili Fivela B., Sigona F., Tavella M., Fitzpatrick P., Metta G., Craighero L., Fadiga L., Sandini G., *New Technologies for Simultaneous Acquisition of Speech Articulatory Data: Ultrasound, 3D Articulograph and Electrolottograph*, in Cristina Delogu, Mauro Falcone (eds.), *LangTech 2008*, Fondazione Ugo Bordoni, Roma: 81-85.
- Guenther, F.H., Espy-Wilson, C., Boyce, S.E., Matthies, M.L., Zandipour, M., Perkell, J.S., (1999), Articulatory tradeoffs reduce acoustic variability during American English /r/ production, *Journal of Acoustical Society of America*, 105, 2854-2865.
- Guenther, F.H., Ghosh, S.S., Tourville, J.A., (2006), Neural modeling and imaging of the cortical interactions under-lying syllable production, *Brain and Language*, 96, 280-301.
- Hardcastle, W. J., (1976), *Physiology of Speech Production*, London, Academic Press.
- Hardcastle, W.J., (1985), Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences, *Speech Communication*, Vol. 4, 247-263.
- Hardcastle, W. J., (1999), Electromyography, in W. J. Hardcastle & N. Hewlett, *Coarticulation. Theory, Data and Techniques*, Cambridge University Press, 270-283.
- Hardcastle William J., Hewlett Nigel (editors), (1999), *Coarticulation, Theory, Data and Techniques*, Cambridge University Press, 1999.
- Hardcastle, W. J., Vaxelaire, B., Gibbon, F., Hoole, P., Nguyen, N., EMA / EPG study of lingual coarticulation in /kl/ clusters, in *1st ESCA Tutorial and Research Workshop on Speech Production Modeling*, Autrans, May 1996, 123-131.
- Hirose, H., (1997), Investigating the Physiology of Laryngeal Structures, in W. J. Hardcastle, J. Laver, *The Handbook of Phonetics Sciences*, Blackwell, Oxford, 11-32.

- Hoole, P. & Nguyen, N., *Electromagnetic articulography in coarticulation research*, in: W.H Hardcastle, *Instrumental Studies of Coarticulation*. CUP (in press).
- Hoole, P., Nguyen-Trong, N. & Hardcastle, W. (1993), A comparative investigation of coarticulation in fricatives: electropalatographic, electromagnetic and acoustic data, *Language and Speech*, 36:235-260.
- Kent R. D., (1997), *The Speech Sciences*, Singular Publishing, Group, San Diego-London.
- Jou, S.-C., Maier-Hein L., Schultz T, Waibel A., Articulatory Feature Classification Using Surface Electromyography, in *ICASSP 2006*, I, 605-608.
- Li, M., Kambhamettu, C. & Stone, M., (2005a), Automatic Contour Tracking in Ultrasound Images, *International Journal of Clinical Linguistics and Phonetics*, 19 (6-7), 545-554.
- Perkell, J. S., (2007), Sensory Goals and Control Mechanisms for Phonemic Articulations, in *ICPhS*, Saarbrücken, 6-10 August 2007, 169-174.
- Perkell, J. S., Matthies, M. L., Svirsky, M. A., Jordan, M. I., (1993), Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study, *Journal of Acoustical Society of America*, 2948-2961.
- Recasens, D., (2002), An EMA study of VCV coarticulatory direction, *Journal of the Acoustical Society of America*, 111, 2002, pp. 2828-2841.
- Rizzolati G., Craighero L., (2004), The Mirror-Neuron System, *Annual Review of Neuroscience*, 27, 169–192.
- Stone, M. & Lundberg A., (1996), Three-dimensional tongue surface shapes of English consonants and vowels, *Journal of the Acoustical Society of America*, 99, 3728–3737.
- Stone M., Epstein M. A., Iskarous K., (2004), Functional Segments in Tongue Movement, *Clinical Linguistics and Phonetics*, 18, 6-8, 507-521.
- Stone, M. (1997), Laboratory Techniques for Investigating Speech Articulation, in W. J. Hardcastle, J. Laver, *The Handbook of Phonetics Sciences*, Blackwell, Oxford, 11-32.
- Stone, M., (2005b), A Guide to Analyzing Tongue Motion from Ultrasound Images, *Clinical Linguistics and Phonetics*, 19, (6-7), 455-502.
- Takemoto, H. (2001), Morphological analyses of the human tongue musculature for three-dimensional modeling, *Journal of Speech and Language Hearing Research*, 44, 95–107.
- Tavella, M., (2007), *Simultaneous recording of phono-articulatory parameters during speech production*, unpublished Master Thesis, University of Genova, LiraLab (Laboratory for Integrated Advanced Robotics), 2006-2007.
- Wilhelms-Tricarico, R. (1995), Physiological modeling of speech production: Methods for modeling soft-tissue articulators, *Journal of the Acoustical Society of America*, 97, 3085–3098.
- Wrench, A. A., (2007), Advances in EPG palate design, *Advances in Speech-Language Pathology*, 99, Issue 1, March 2007 , 3 – 12.
- Zierdt, A., Carstens, B., Method and a device to determine the spatial arrangement of a directionally sensitive magnetic field sensor, *United States Patent Number*, 5, 453, 687, 1995.
- Zierdt, A., Hoole, P., Tillmann, H.G., (1999), Development of a System for Three-Dimensional Fleshpoint Measurement of Speech Movements”, in *Proceedings. ICPhS’99*, San Francisco. August.

## IL FRANCESE COME LINGUA STRANIERA: L'ASSIMILAZIONE DEL TRATTO DI SONORITÀ

Sonia d'Apolito, Barbara Gili Fivela  
Università del Salento & CRIL, Lecce  
[soniadap@libero.it](mailto:soniadap@libero.it), [barbara.gili@ateneo.unile.it](mailto:barbara.gili@ateneo.unile.it)

### 1. SOMMARIO

In questo studio\* prenderemo in considerazione la realizzazione dell'assimilazione del tratto di sonorità tra segmenti consonantici adiacenti. Il fenomeno è fortemente presente nella lingua francese e sarà indagato in un corpus di produzioni in francese di parlanti italofoeni che abbiano raggiunto un buon livello di competenza nella lingua straniera.

Il fenomeno osservato consiste nel cambiamento del tratto di sonorità di un segmento consonantico per la presenza di un altro segmento consonantico contiguo. Il quadro teorico di cui si è tenuto conto è quello della fonologia prosodica, sviluppata da Nespor e Vogel [1986]. Secondo la fonologia prosodica e le analisi della lingua francese che hanno preso in considerazione le ipotesi sviluppate all'interno di questo quadro teorico [Féry 2003], l'assimilazione si realizza all'interno del sintagma fonologico, mentre non si osserva al confine di questo costituente o di un costituente di livello superiore. Peraltro, in varie lingue, tra cui il francese e l'italiano, la focalizzazione rappresenta uno dei fattori in grado di indurre la presenza di confini prosodici, confini analizzati da alcuni studiosi come confini di sintagmi fonologici [Frascarelli, 1997]. Nel caso in cui sia presente, quindi, un confine fonologico tra due elementi, ad esempio per via della focalizzazione di uno dei due termini di un sintagma, la possibile assimilazione non si dovrebbe realizzare in quanto i due termini fanno parte di due sintagmi fonologici distinti.

Il nostro studio si compone di due esperimenti in cui è stata effettuata un'analisi uditiva delle produzioni di cinque parlanti (3 italofoeni e 2 francofoeni); inoltre, nel secondo esperimento, l'analisi uditiva è stata integrata con un'analisi acustica (sia del segnale verbale che del segnale glottico, registrati simultaneamente). Nel primo esperimento si è osservata la realizzazione di assimilazione del tratto di sonorità – sia all'interno che al confine di parola – al variare della posizione degli stimoli all'interno di frasi dichiarative (posizione iniziale e finale); nel secondo esperimento, la realizzazione di assimilazione è stata osservata in relazione ai contesti di focalizzazione neutra e contrastiva.

Dai risultati emerge che la realizzazione di assimilazione varia a seconda del contesto consonantico considerato e delle caratteristiche di focalizzazione. In base al fattore focalizzazione, in particolare, risulta che il maggior numero di assimilazioni si riscontra nei casi di focalizzazione neutra e, comunque, nei casi in cui non intervenga nessun confine prosodico. Il minor numero di assimilazioni, al contrario, si riscontra quando, all'interno di un sintagma, sia stato focalizzato il primo elemento piuttosto che il secondo, sia nelle produzioni dei parlanti italofoeni che nelle produzioni dei francofoeni.

---

\* Questo articolo è frutto del lavoro congiunto dei due autori. Tuttavia, ai soli fini accademici, 3, 4.1, 4.2.1, 4.2.2 vanno attribuiti a S.D. e 2, 4.2.3 a B.G.F.

## 2. INTRODUZIONE

L'assimilazione consiste nel cambiamento sistematico di un fono che, per uno o più tratti, assume le stesse caratteristiche fonetiche di un elemento adiacente o relativamente distante [Laver, 1994; Nespor, 1993]. Il fenomeno dell'assimilazione ha origine molto spesso dalla coarticolazione, dall'attenuazione delle distinzioni articolatorie, e quindi acustiche e percettive, di foni che appartengono allo stesso contesto.

In questo studio ci occuperemo dell'assimilazione del tratto di sonorità tra segmenti consonantici e prenderemo in esame la lingua francese, lingua in cui è molto frequente l'assimilazione regressiva del tratto di sonorità di consonanti adiacenti [De Dominicis, 1999]. Prendendo in considerazione la direzione di propagazione di uno o più tratti, si distingue l'assimilazione di tipo progressivo e regressivo. L'assimilazione è regressiva, o anticipatoria, quando un fono assimila caratteristiche di un elemento seguente poiché gli organi dell'apparato vocale ne anticipano (alcuni) aspetti articolatori; in francese si riscontrano sia sonorizzazioni (es. *iceberg* 'blocco di ghiaccio' [i'sbɛg], in cui la fricativa dentale acquisisce il tratto di sonorità per via dell'occlusiva sonora seguente) che desonorizzazioni (es. *médecin* 'medico' [mɛd'sɛ̃], in cui l'occlusiva dentale acquisisce il tratto di sordità della fricativa sorda seguente). Al contrario, l'assimilazione è progressiva quando un fono assimila alcune caratteristiche di un elemento che lo precede, in quanto sono mantenuti (alcuni) aspetti della sua articolazione; anche in questo caso, in francese si riscontrano sonorizzazioni (es. *pirogue partie* 'canoa partita' [piʁog paʁ'ti], in cui l'occlusiva bilabiale acquisisce il tratto di sonorità dell'occlusiva sonora precedente) e desonorizzazioni (es. *type gourde* 'tipo stupido' [tip' ɡuʁd], in cui l'occlusiva velare acquisisce il tratto di sordità dell'occlusiva sorda precedente). Inoltre, l'assimilazione può essere parziale e totale, in base al numero dei tratti propagati da un fono all'altro: è totale quando tutti i tratti vengono propagati e il segmento assimilato diventa identico al segmento assimilante (es. *adsorbant* 'assorbente' [as:oʁ'bã]) mentre è parziale quando l'assimilazione riguarda solo alcuni tratti (es. *anecdote* 'aneddoto' [anɛɡ'dot], *absolu* 'assoluto' [apso'ly], *iceberg* 'blocco di ghiaccio' [i'sbɛg] e *médecin* 'medico' [mɛd'sɛ̃]); gli esempi appena forniti evidenziano che nella lingua francese, l'assimilazione parziale può realizzarsi talvolta come assimilazione completa di un tratto (es. *absolu* 'assoluto' [apso'ly]) e talvolta come assimilazione non completa (es. *médecin* 'medico' [mɛd'sɛ̃]), anche in relazione allo stesso contesto (es. *médecin* realizzato come [mɛd'sɛ̃] e [mɛt'sɛ̃], rispettivamente). Infine, come si può notare dagli esempi citati, le assimilazioni del tratto di sonorità in francese possono verificarsi sia all'interno che al confine di parola.

Nel corso dello studio si è fatto riferimento al quadro teorico della fonologia prosodica sviluppata da Nespor e Vogel [Nespor e Vogel, 1986] secondo cui i costituenti prosodici sono organizzati gerarchicamente su vari livelli e sono motivati in quanto domini di fenomeni fonologici, sebbene siano 'costruiti' a partire da informazioni sintattiche. I sette livelli proposti da Nespor e Vogel sono: la sillaba, il piede, la parola fonologica, il gruppo clítico, il sintagma fonologico, il sintagma intonativo e l'enunciato. Secondo le analisi della lingua francese che hanno preso in considerazione le ipotesi sviluppate all'interno della fonologia prosodica [Nespor e Vogel, 1986; Féry, 2003], l'assimilazione si realizza all'interno di un costituente prosodico specifico, il sintagma fonologico.

In base alla definizione data da Nespor [1993], il dominio di un sintagma fonologico consiste di:

“un gruppo clítico che contiene una testa lessicale X più tutti i gruppi clitici sul suo lato non ricorsivo fino al primo gruppo clítico che contiene una testa lessicale al di fuori della proiezione massima di X”

Inoltre, il dominio può anche essere ristrutturato in base alla seguente regola facoltativa:

“un sintagma fonologico che non ramifica e che contiene il primo complemento o il primo modificatore di X dal lato ricorsivo viene incluso nel sintagma fonologico che contiene X”

Per fare un esempio, la frase *Il a utilisé une corde tendue pour le sauver*<sup>1</sup>, potrà essere composta dai seguenti sintagmi fonologici: [Il a utilisé]<sub>Φ</sub> [une corde tendue]<sub>Φ</sub> [pour le sauver]<sub>Φ</sub>; la presenza di *corde tendue* all'interno di un unico sintagma fonologico rende possibile l'assimilazione del tratto di sonorità al confine di parola, in relazione ai segmenti [d] e [t]. La frase italiana corrispondente, *Ha utilizzato una corda tesa per salvarlo*, sarebbe suddivisa in modo analogo: [Ha utilizzato]<sub>Φ</sub> [una corda tesa]<sub>Φ</sub> [per salvarlo]<sub>Φ</sub>.

Come già accennato, in base alla fonologia prosodica, anche la presenza di focalizzazione può interferire con l'assimilazione del tratto di sonorità, in quanto induce la modificazione dei domini fonologici. Diciamo, innanzitutto, che la focalizzazione può essere realizzata sintatticamente, lessicalmente o fonologicamente, e permette di evidenziare un elemento nuovo e/o contrastivo all'interno di una frase [Frascarelli, 1997]. Secondo molti studiosi (Frascarelli [1997], ma si veda anche Ladd [1996]), si distinguono due tipi di focalizzazione: focalizzazione neutra (o ampia) quando nessun elemento all'interno della frase è marcato e tutta l'informazione è nuova; focalizzazione ristretta<sup>2</sup>, quando essa è limitata ad alcuni costituenti di una frase [Frascarelli, 1997]. Secondo Frascarelli, quando si ha focalizzazione ristretta, dal punto di vista fonologico il costituente in focus coincide con uno specifico costituente prosodico: il sintagma fonologico [ ]<sub>Φ</sub>.

La frase italiana *Ha utilizzato una corda tesa per salvarlo*, in caso di focalizzazione ristretta su *corda* sarebbe suddivisa in [Ha utilizzato]<sub>Φ</sub> [una **corda**]<sub>Φ</sub> [tesa]<sub>Φ</sub> [per salvarlo]<sub>Φ</sub>. A questa rappresentazione è possibile applicare la regola del *Focus Restructuring*, secondo la quale, se un costituente è focalizzato, esso si ristruttura in un sintagma fonologico sul suo lato sinistro, formando un sintagma fonologico più ampio [Frascarelli, 1997], quindi la frase precedente potrebbe essere rappresentata come segue: [Ha utilizzato una **corda**]<sub>Φ</sub> [tesa]<sub>Φ</sub> [per salvarlo]<sub>Φ</sub>.

Frascarelli, inoltre, osserva che in italiano un elemento focalizzato crea un confine “rafforzato” di sintagma fonologico sul suo lato destro<sup>3</sup>.

<sup>1</sup> La frase è stata estratta dal corpus di analisi ed appartiene ai contesti previsti per la realizzazione di assimilazione regressiva ([dt] → [tt]) o progressiva ([dt] → [dd]) del tratto di sonorità a confine di parola.

<sup>2</sup> La focalizzazione ristretta si distingue in focus informativo, quando identifica un elemento all'interno di un insieme presupposto senza implicazione di esaustività, e focus contrastivo, quando l'identificazione esclude qualsiasi altro elemento dell'insieme, talvolta negando una precedente presupposizione [Frascarelli, 2005].

<sup>3</sup> Questo confine destro “rafforzato” comporta la non applicazione della regola della ristrutturazione facoltativa del sintagma fonologico [Frascarelli, 1997; Nespor, 1993]. Le due regole di ristrutturazione (regola del *Focus Restructuring* e regola della ristrutturazione facoltativa) non possono essere applicate, pertanto, all'interno dello stesso sintagma fonologico [Frascarelli, 1997]. Per questo motivo, non è possibile avere una frase ristrutturata nel modo seguente: [Ha utilizzato una **corda** tesa]<sub>Φ</sub> [per salvarlo]<sub>Φ</sub>.

Per quanto riguarda il francese, Féry [2001, 2003] mostra come l'elemento focalizzato formi un sintagma indipendente, poiché tale costituente è realizzato all'interno di un proprio sintagma fonologico con una breve pausa prima e/o dopo i confini del sintagma. L'autrice osserva che i fenomeni fonologici, e l'assimilazione del tratto di sonorità in particolare, hanno luogo all'interno del sintagma fonologico e la loro realizzazione viene bloccata in caso di focalizzazione ristretta. Tuttavia, a differenza di quanto avviene in italiano, in francese un costituente focalizzato introduce un confine fonologico prima dell'elemento focalizzato, sul suo confine sinistro [Féry, 2003]. Prendendo nuovamente in considerazione l'esempio precedente, sarà la presenza di focalizzazione ristretta sulla parola *tendue* (e non su *corde*) a indurre la presenza di un confine prosodico nel sintagma bersaglio ed a permettere quindi la seguente suddivisione in sintagmi fonologici: [Il a utilisé]<sub>Φ</sub> [une corde]<sub>Φ</sub> [**tendue**]<sub>Φ</sub> [pour le sauver]<sub>Φ</sub><sup>4</sup>.

I due esperimenti descritti in questo studio hanno lo scopo di osservare se studenti italofoni che parlino francese con un buon livello di competenza realizzino le assimilazioni del tratto di sonorità nei contesti nei quali le producono i parlanti francofoni; inoltre, sulla base di quanto appena discusso, lo scopo è anche quello di verificare la realizzazione di questo fenomeno in contesti di focalizzazione neutra e ristretta-contrastiva, data la diversa posizione dei confini prosodici attesa in italiano e in francese sulla base delle descrizioni in letteratura.

### 3. PRIMO ESPERIMENTO

#### 3.1 Corpus e metodo

Nel primo studio, l'assimilazione del tratto di sonorità è stata studiata rispetto alle consonanti /s, b, d e t/ che sono state considerate in tutte le loro possibili combinazioni – si veda la tabella 1. Il corpus è stato costruito in modo che fosse rappresentativo di tutte le combinazioni delle quattro consonanti e di contesti nei quali, in francese, fosse attesa la realizzazione di assimilazioni del tratto di sonorità regressive e progressive, totali e parziali - complete e non complete - sia all'interno che al confine di parola<sup>5</sup>. La posizione dei fonemi consonantici è stata anche controllata in relazione a quella dell'accento, considerando le consonanti bersaglio sia a confine tra sillabe non accentate che a confine tra sillabe con diversa specificazione in termini di accentuazione (non-accentata/accentata).

|      |      |      |      |
|------|------|------|------|
| /sd/ | /sb/ | /td/ | /tb/ |
| /ds/ | /bs/ | /dt/ | /bt/ |

Tabella 1: Contesti consonantici osservati.

<sup>4</sup> In questo esempio, tratto dal corpus di indagine, è stata indotta la realizzazione del focus contrastivo sul secondo dei due termini del sintagma (*tendue* 'tesa'), grazie al seguente contesto: *Pour le sauver il a utilisé une corde lâche. – ( Mais non,) il a utilisé une corde tendue pour le sauver* 'Per salvarlo ha utilizzato una corda lenta. – ( Ma no,) ha utilizzato una corda tesa per salvarlo'.

<sup>5</sup> Dalla consultazione del dizionario di francese [Petit Robert] si riscontra che le seguenti combinazioni danno luogo ad assimilazioni complete del tratto di sonorità all'interno di parola: /bs/→/p/; /bt/→/p/; /ds/→/t/ [Robert, Rey, Rey-Debove, Petit Robert, 2000].



Inoltre, al fine di osservare i fenomeni di assimilazione in relazione al variare della struttura prosodica, le parole target sono state inserite sia in posizione iniziale e sia in posizione finale di frasi dichiarative (ad esempio: *La **place barrée** est praticable seulement par les piétons* 'La piazza chiusa è transitabile solo dai pedoni'; *La circulation a été déviée par une **place barrée*** 'La circolazione è stata deviata a causa di una piazza chiusa). Il corpus, in cui la presentazione delle frasi è stata variata in modo casuale, è stato letto per tre volte ad una velocità di eloquio normale da cinque parlanti, di cui tre italofofoni e due francofofoni. I soggetti italofofoni, due di genere femminile e uno di genere maschile (età compresa tra i 23 e i 25 anni), provengono dalla provincia di Lecce ed hanno studiato lingue straniere e, soprattutto francese, presso la Facoltà di Lingue e Letterature Straniere del capoluogo. In particolare, due italofofoni (PI1 e PI2) hanno studiato la lingua francese a partire dalla scuola media, ed il parlante PI2 è stato anche in Francia (Nancy) per un soggiorno di 6 mesi, nell'ambito del programma Erasmus. I due soggetti francofofoni (PF4, PF5), entrambi di genere femminile (età: 21 e 22 anni), provengono da Nizza ed, al momento della registrazione, erano in Italia da cinque mesi come studenti Erasmus.

I parlanti hanno letto il corpus in un ambiente silenzioso e sono stati registrati con un MiniDisk portatile Sony MZN1 e microfono Plantronics. I materiali registrati sono stati analizzati uditivamente da parte del primo autore: tutte le frasi sono state ascoltate più volte al fine di verificare la realizzazione di assimilazione dei contesti in esame; il risultato dell'analisi è stato riportato sotto forma di trascrizione fonetica stretta dei contesti di studio (utilizzando l'alfabeto IPA - *International Phonetic Alphabet*). Successivamente, è stato effettuato un controllo acustico delle produzioni dei parlanti italofofoni, mediante ispezione del sonogramma. In questa prima indagine, l'analisi acustica è stata effettuata con il solo fine di ottenere una ulteriore fonte di informazione circa la presenza e tipologia (totale, parziale completa o non completa) del fenomeno di assimilazione.

### 3.2. Risultati del primo esperimento

Nelle tabelle 2-5, riportiamo i risultati dell'analisi circa la presenza di assimilazione regressiva delle produzioni in francese da parte di italofofoni (tabelle 2 e 3, relative all'analisi acustica dei contesti all'interno di parola ed al confine di parola, rispettivamente) e di francofofoni (tabelle 4 e 5, relative all'analisi uditiva dei contesti all'interno di parola ed al confine di parola, rispettivamente)<sup>6</sup>.

L'assimilazione del tratto di sonorità nelle produzioni dei parlanti, sia italofofoni che francofofoni, è prevalentemente parziale regressiva. I risultati complessivi indicano che le assimilazioni regressive sono realizzate dai parlanti italofofoni nel 60.28% dei contesti e dai francofofoni nell'82.29% dei casi<sup>7</sup>; di fatto, nessun parlante realizza assimilazioni progressive.

I contesti consonantici che gli italofofoni assimilano maggiormente sono la fricativa alveolare sorda /s/ seguita dalle occlusive bilabiale e dentale /b e d/, soprattutto all'interno

<sup>6</sup> Il simbolo “—” indica che, nella specifica posizione, il contesto non esiste; in particolare, non sono stati trovati stimoli in cui i contesti /tb/ e /td/ comparissero all'interno di parola. Il simbolo “Ø” indica la cancellazione del primo segmento in posizione finale di parola: in particolare, l'elisione è riscontrata per i contesti /tb/ e /td/, in cui il segmento /t/ non viene realizzato.

<sup>7</sup> Questi risultati si riferiscono alla media calcolata sulla base di tutte le realizzazioni di assimilazioni regressive, all'interno e al confine di parola, all'interno di tutte le produzioni realizzate da parte dei parlanti.

di parola, dove la percentuale è dell'88.88% (/sb/ 32 casi su 36; /sd/ 16 casi su 18), mentre si abbassa al 66.67% (24 casi su 36) nel caso di assimilazione al confine di parola (si vedano le tabelle 2 e 3, rispettivamente). La percentuale di realizzazione di assimilazione da parte dei francofoni è del 100% sia all'interno sia al confine di parola (si vedano le tabelle 4 e 5, rispettivamente)

All'interno di parola, l'occlusiva bilabiale viene assimilata dagli italofoeni in modo completo per il 100% dei casi (36 assimilazioni su 36) nel contesto /bs/ e per il 94.44% dei casi (34 su 36) nel contesto /bt/; al contrario le percentuali si abbassano notevolmente al confine di parola, 30.56% (11 casi su 36) e 19.44% (7 casi su 36) rispettivamente (cfr. tabelle 2 e 3). Anche per i parlanti francofoni si riscontra una diminuzione del numero di assimilazioni nel caso della presenza di un confine di parola (il contesto /bs/ è assimilato per il 45.83% (11 casi su 24) e il contesto /bt/ per il 50% (12 casi su 24)), mentre all'interno di parola le assimilazioni dei francofoni si riscontrano nel 100% dei casi indagati (24 casi su 24) – si vedano le tabelle 4 e 5. I contesti /dt/, /tb/ e /td/ non vengono mai assimilati, da nessun parlante, mentre emerge la tendenza a cancellare il segmento finale /t/ a confine di parola – si vedano le tabelle 3 e 5. Un minor numero di assimilazioni regressive riguarda il contesto /ds/ per gli italofoeni, mentre nel caso dei francofoni la diminuzione si riscontra solo nel confronto fra contesto interno ed a confine di parola. Si noti che il maggior numero di assimilazioni del tratto di sonorità all'interno piuttosto che al confine di parola, da parte di entrambi i gruppi di parlanti, dimostra l'integrità della parola come costituente.

Per quanto riguarda la posizione all'interno della frase, i parlanti tendono a realizzare l'assimilazione soprattutto in posizione iniziale, le percentuali sono del 59.09% per gli italofoeni e del 81.82% per i francofoni, contro il 55.56% e il 78.79% per italofoeni e francofoni, rispettivamente, in posizione finale<sup>8</sup>.

| Contesto:<br>Consonante-<br>interno<br>parola | Totale<br>Realizzazioni<br>Assimilazione |       | Interno Parola Inizio frase |       |              |       | Interno Parola Fine frase |       |              |       |
|-----------------------------------------------|------------------------------------------|-------|-----------------------------|-------|--------------|-------|---------------------------|-------|--------------|-------|
|                                               |                                          |       | Completa                    |       | Non Completa |       | Completa                  |       | Non Completa |       |
|                                               | N°                                       | %     | N°                          | %     | N°           | %     | N°                        | %     | N°           | %     |
| /bs/                                          | 34/36                                    | 94.44 | 18/18                       | 100   | 0            | 0     | 16/18                     | 88.88 | 0            | 0     |
| /bt/                                          | 36/36                                    | 100   | 18/18                       | 100   | 0            | 0     | 18/18                     | 100   | 0            | 0     |
| /ds/ <sup>9</sup>                             | 18/36                                    | 50    | 10/18                       | 55.56 | 0            | 0     | 8/18                      | 44.44 | 0            | 0     |
| /ds/                                          | 12/18                                    | 66.67 | 6/9                         | 66.66 |              |       | 6/9                       | 66.66 |              |       |
| /dt/                                          | —                                        | —     | —                           | —     | —            | —     | —                         | —     | —            | —     |
| /sb/                                          | 32/36                                    | 88.88 | 15/18                       | 83.33 | 3/18         | 16.67 | 9/18                      | 50    | 5/18         | 27.78 |
| /sd/-int                                      | 16/18                                    | 88.88 | 6/9                         | 66.66 | 2/9          | 22.22 | 6/9                       | 66.66 | 2/9          | 22.22 |
| /tb/                                          | —                                        | —     | —                           | —     | —            | —     | —                         | —     | —            | —     |
| /td/                                          | —                                        | —     | —                           | —     | —            | —     | —                         | —     | —            | —     |
| Tot                                           |                                          | 84.44 |                             | 78.70 |              | 7.78  |                           | 69.44 |              | 10    |

Tabella 2: Italofoeni. Risultati della verifica acustica e percentuali delle assimilazioni regressive all'interno di parola, complete/non complete in posizione iniziale/fine di frase.

<sup>8</sup> I risultati descritti si riferiscono alla media calcolata tenendo conto di tutte le realizzazioni di assimilazioni regressive per tutti stimoli (all'interno e al confine di parola) in base alla loro posizione all'interno della frase dichiarativa (posizione iniziale e finale di frase) all'interno delle produzioni.

<sup>9</sup> Il contesto /ds/ si riferisce agli stimoli in cui la realizzazione di assimilazione è completa (primo contesto; 'adstrat' 'adstrato' e 'adsorbant' 'adsorbente') ed agli stimoli la cui realizzazione è parziale non completa (secondo contesto; 'médecin' 'medico').

| Contesto:<br>Consonante-<br>confine<br>parola | Totale<br>Realizzazioni<br>Assimilazione |       | Confine Parola Inizio frase |       |              |       | Confine Parola Fine frase |       |              |       |
|-----------------------------------------------|------------------------------------------|-------|-----------------------------|-------|--------------|-------|---------------------------|-------|--------------|-------|
|                                               |                                          |       | Completa                    |       | Non Completa |       | Completa                  |       | Non Completa |       |
|                                               | N°                                       | %     | N°                          | %     | N°           | %     | N°                        | %     | N°           | %     |
| /bs/                                          | 11/36                                    | 30.56 | 0                           | 0     | 6/18         | 33.33 | 0                         | 0     | 5/18         | 27.78 |
| /bt/                                          | 7/36                                     | 19.44 | 0                           | 0     | 2/18         | 11.11 | 0                         | 0     | 5/18         | 27.78 |
| /ds/                                          | 12/36                                    | 33.33 | 6/18                        | 33.33 | 0            | 0     | 6/18                      | 33.33 | 0            | 0     |
| /dt/                                          | 0                                        | 0     | 0                           | 0     | 0            | 0     | 0                         | 0     | 0            | 0     |
| /sb/                                          | 24/36                                    | 66.67 | 10/18                       | 55.55 | 3/18         | 16.67 | 9/18                      | 50    | 3/18         | 16.67 |
| /sd/                                          | 24/36                                    | 66.67 | 9/18                        | 50    | 3/18         | 16.67 | 8/18                      | 44.44 | 4/18         | 33.33 |
| /tb/                                          | Ø                                        | Ø     | Ø                           | Ø     | Ø            | Ø     | Ø                         | Ø     | Ø            | Ø     |
| /td/                                          | Ø                                        | Ø     | Ø                           | Ø     | Ø            | Ø     | Ø                         | Ø     | Ø            | Ø     |
| Tot                                           |                                          | 36.11 |                             | 23.15 |              | 12.96 |                           | 21.30 |              | 17.59 |

Tabella 3: Italofofoni. Risultati della verifica acustica e percentuali delle assimilazioni regressive al confine di parola, complete/non complete in posizione iniziale/finale di frase.

| Contesto:<br>Consonante-<br>Interno<br>parola | Totale<br>Realizzazioni<br>Assimilazione |     | Interno Parola Inizio frase |     |              |   | Interno Parola Fine frase |     |              |   |
|-----------------------------------------------|------------------------------------------|-----|-----------------------------|-----|--------------|---|---------------------------|-----|--------------|---|
|                                               |                                          |     | Completa                    |     | Non Completa |   | Completa                  |     | Non Completa |   |
|                                               | N°                                       | %   | N°                          | %   | N°           | % | N°                        | %   | N°           | % |
| /bs/                                          | 24/24                                    | 100 | 12/12                       | 100 | 0            | 0 | 12/12                     | 100 | 0            | 0 |
| /bt/                                          | 24/24                                    | 100 | 12/12                       | 100 | 0            | 0 | 12/12                     | 100 | 0            | 0 |
| /ds/                                          | 24/24                                    | 100 | 12/12                       | 100 | 0            | 0 | 12/12                     | 100 | 0            | 0 |
| /ds/                                          | 12/12                                    | 100 | 6/6                         | 100 |              |   | 6/6                       | 100 |              |   |
| /dt/                                          | —                                        | —   | —                           | —   | —            | — | —                         | —   | —            | — |
| /sb/                                          | 24/24                                    | 100 | 12/12                       | 100 | 0            | 0 | 12/12                     | 100 | 0            | 0 |
| /sd/                                          | 12/12                                    | 100 | 6/6                         | 100 | 0            | 0 | 6/6                       | 100 | 0            | 0 |
| /tb/                                          | —                                        | —   | —                           | —   | —            | — | —                         | —   | —            | — |
| /td/                                          | —                                        | —   | —                           | —   | —            | — | —                         | —   | —            | — |
| Tot                                           |                                          | 100 |                             | 100 |              | 0 |                           | 100 |              | 0 |

Tabella 4: Francofoni. Risultati dell'analisi uditiva e percentuali delle assimilazioni regressive all'interno di parola, complete/non complete in posizione iniziale/finale di frase.

| Contesto:<br>Consonante-<br>confine<br>parola | Totale Realizzazioni<br>Assimilazione |       | Confine Parola Inizio frase |     |              |       | Confine Parola Fine frase |       |              |       |
|-----------------------------------------------|---------------------------------------|-------|-----------------------------|-----|--------------|-------|---------------------------|-------|--------------|-------|
|                                               |                                       |       | Completa                    |     | Non Completa |       | Completa                  |       | Non Completa |       |
|                                               | N°                                    | %     | N°                          | %   | N°           | %     | N°                        | %     | N°           | %     |
| /bs/                                          | 11/24                                 | 45.83 | 0                           | 0   | 6/12         | 50    | 0                         | 0     | 5/12         | 41.66 |
| /bt/                                          | 12/24                                 | 50    | 0                           | 0   | 6/12         | 50    | 0                         | 0     | 5/12         | 41.66 |
| /ds/                                          | 22/24                                 | 91.66 | 12/12                       | 100 | 0            | 0     | 10/12                     | 83.33 | 0            | 0     |
| /dt/                                          | 0                                     |       | 0                           | 0   | 0            | 0     | 0                         | 0     | 0            | 0     |
| /sb/                                          | 24/24                                 | 100   | 12/12                       | 100 | 0            | 0     | 12/12                     | 100   | 0            | 0     |
| /sd/                                          | 24/24                                 | 100   | 12/12                       | 100 | 0            | 0     | 12/12                     | 100   | 0            | 0     |
| /tb/                                          | Ø                                     | Ø     | Ø                           | Ø   | Ø            | Ø     | Ø                         | Ø     | Ø            | Ø     |
| /td/                                          | Ø                                     | Ø     | Ø                           | Ø   | Ø            | Ø     | Ø                         | Ø     | Ø            | Ø     |
| Tot                                           |                                       | 64.58 |                             | 50  |              | 16.67 |                           | 47.22 |              | 13.88 |

Tabella 5: Francofoni. Risultati dell'analisi uditiva e percentuali delle assimilazioni regressive al confine di parola, complete/non complete in posizione iniziale/finale di frase.

## 4. SECONDO ESPERIMENTO

### 4.1 Corpus e metodo

Nel secondo esperimento, oltre alle consonanti già considerate nella prima indagine, sono state prese in esame anche /p/, /k/, /g/, /f/ e /v/, ottenendo le combinazioni di contesti consonantici riportate nella tabella 6.<sup>10</sup>

|      |      |  |      |      |
|------|------|--|------|------|
| /sd/ | /ds/ |  | /kb/ | /bk/ |
| /sb/ | /bs/ |  | /kd/ | /dk/ |
| /sv/ | /vs/ |  | /kv/ | /vk/ |
| /sg/ | /gs/ |  | /fb/ | /bf/ |
| /tb/ | /bt/ |  | /fd/ | /df/ |
| /td/ | /dt/ |  | /pg/ | /gp/ |
| /tg/ | /gt/ |  | /pv/ | /vp/ |

Tabella 6: Contesti consonantici considerati.

Per la creazione del corpus sono state quindi scelte parole bersaglio che permettessero di osservare le combinazioni di consonanti suddette in relazione ai fenomeni di assimilazione regressiva e progressiva del tratto di sonorità, all'interno e al confine di parola; inoltre, come nel caso del primo corpus, si è tenuto conto della posizione reciproca dei fonemi consonantici e della posizione dell'accento – si veda la tabella 7 per alcuni esempi tratti dal corpus.

| Contesto | Interno Parola Complete  | Interno Parola Non complete | A confine di Parola Parziali   |                                    |
|----------|--------------------------|-----------------------------|--------------------------------|------------------------------------|
|          |                          |                             | A confine                      | Per confronto <sup>11</sup>        |
| /ds/     | Adstrat<br>[at'strat]    | Médecin<br>[mɛdɛ'sɛ̃]       | Amandes sèches<br>[amɑ̃d ʁsɛʃ] | Cul-de-sac<br>[ky d ʁsak]          |
|          | Adsorbant<br>[atsoʁ'bɑ̃] |                             | Code secret<br>[kɔd sɔʁkɛ]     | Coup de sonnette<br>[ku d ʁso'net] |

Tabella 7: Esempi di stimoli estratti dal corpus: contesto consonantico /ds/.

Le parole bersaglio, in base al loro significato sono state inserite in frasi che potessero rappresentare un contesto naturale per la loro produzione. La posizione degli stimoli all'interno della frase non è stata messa a fattore e tutti gli stimoli sono stati collocati in posizione iniziale di frase in frasi dichiarative. I risultati del primo esperimento avevano infatti mostrato che l'assimilazione tende a verificarsi maggiormente in questa posizione.

Inoltre, per studiare l'effetto della focalizzazione sui fenomeni di assimilazione - per via delle modifiche apportate alla struttura prosodica dell'enunciato - gli stimoli relativi all'assimilazione al confine di parola sono stati anche inseriti in contesti che favorissero la

<sup>10</sup> I contesti consonantici /bk/, /bs/, /bt/, /ds/, /gs/, /kd/ all'interno di parola danno luogo ad assimilazioni regressive complete del tratto di sonorità [Robert, Rey, Rey-Debove, Petit Robert, 2000].

<sup>11</sup> All'interno dei contesti relativi al segmento /d/ sono stati inseriti alcuni stimoli in cui il segmento è rappresentato dalla preposizione semplice “de”. Questi sintagmi sono stati inseriti in frasi dichiarative ed in posizione iniziale.

realizzazione di focalizzazione contrastiva: la focalizzazione è stata indotta, di volta in volta, sul primo o sul secondo membro del sintagma composto dalle due parole al cui confine si trovano le consonanti di studio.

Considerando, tra gli stimoli, la parola *médecin* 'medico' ed il sintagma *amandes fraîches* 'mandorle fresche', le frasi del corpus per il secondo esperimento sono così strutturate:

Fraasi dichiarative in cui lo stimolo è in posizione iniziale:

a) *Mon médecin de famille est toujours introuvable* 'Il mio medico di famiglia è sempre irreperibile';

b) *Les amandes fraîches sont très chères dans cette période* 'Le mandorle fresche sono molto costose in questo periodo';

Contesti per la realizzazione di focalizzazione contrastiva

b1) contesto: *Nous avons mangé des noix fraîches à déjeuner*

frase bersaglio: *Mais non, nous avons mangé des amandes fraîches à déjeuner;*

'Abbiamo mangiato qualche noce fresca per colazione ; Ma no, abbiamo mangiato qualche mandorla fresca per colazione'

b2) contesto: *Il a acheté des amandes sèches ce matin*

frase bersaglio: *Mais non, il a acheté des amandes fraîches ce matin.*

'Ha comprato delle mandorle secche questa mattina; Ma no, ha comprato della mandorle fresche questa mattina'.

Ricordiamo che in francese la focalizzazione sembra indurre un confine prosodico a sinistra dell'elemento in focus [Féry, 2003], mentre in italiano il confine dovrebbe essere realizzato sulla destra [Frascarelli, 1997]. L'ipotesi è quindi che 1) in francese l'assimilazione possa essere realizzata in b1 ma non in b2, a causa del confine prosodico; 2) gli italofoeni possano invece produrre più assimilazioni nel contesto b2 piuttosto che nel contesto b1.

Il *corpus*, nel quale le frasi ed i contesti sono stati disposti in ordine casuale, è stato sottoposto a cinque parlanti, tre italofoeni e due francofoeni. Gli italofoeni sono studenti di lingua francese presso la Facoltà di Lingue e Letterature Straniere dell'Università del Salento; due sono di genere femminile (età di 25 anni entrambe) ed hanno portato a termine il percorso di studi universitario - una parlante ha anche soggiornato in Francia (Nantes) per sei mesi, grazie al progetto Erasmus; un soggetto è di genere maschile (età di 23 anni) ed, al momento della registrazione, era iscritto al corso di laurea specialistica. I parlanti francofoeni, uno di sesso maschile e uno di sesso femminile, provengono da Metz ed, al momento della registrazione, erano in Italia da cinque mesi come studenti Erasmus. Tutti i parlanti hanno ripetuto la lettura del *corpus* per tre volte ad una velocità normale di eloquio.

Infine, ai soggetti hanno partecipato ad un Map-Task, un compito dialogico che ruota attorno alla ricostruzione di un tracciato tra icone disegnate su una mappa: nel caso specifico, un solo parlante partecipava al gioco e forniva informazioni circa il tracciato tra alcune icone i cui nomi contenevano stimoli previsti per lo studio dell'assimilazione all'intero di parola. L'obiettivo dell'acquisizione di materiali con il Map-Task è stato quello di ricavare dati per un confronto tra il parlato letto e semispontaneo.

In fase di registrazione, sono stati raccolti simultaneamente dati relativi al segnale verbale e glottico. L'acquisizione dei materiali, infatti, è stata fatta nella camera insonorizzata del Centro di Ricerche Interdisciplinari sul Linguaggio (CRIL) dove i materiali audio sono stati registrati con microfono Shure (SM58) e campionati direttamente su *hard-disk*, ed il segnale glottico è stato acquisito con l'Elettroglottografo prodotto dalla

Laryngograph (che prevede l'uso di due elettrodi posizionati sul collo dei parlanti, in corrispondenza della laringe) e grazie al programma software *Speech Studio*. L'uso dell'elettroglottografo ha permesso di verificare con sicurezza l'effettiva presenza o assenza di sonorità<sup>12</sup>.

I materiali sono stati analizzati sia uditivamente che acusticamente: l'analisi uditiva è stata effettuata al fine di osservare la realizzazione dei casi di assimilazione e la presenza o meno di focalizzazione nei contesti inseriti appositamente per indurla; l'analisi acustica è stata fatta per ricavare informazioni quantitative circa i contesti di assimilazione. Con il programma PRAAT (Boersma & Weenink, 2007) sono stati visualizzati sia il segnale verbale che il segnale glottico: la visualizzazione ha permesso di osservare le caratteristiche di sonorità tramite ispezione della forma d'onda del sonogramma e di effettuare la misurazione della durata dei segmenti consonantici rilevanti<sup>13</sup>.

Nelle figure 2-3 sono riportati alcuni esempi relativi al contesto /gp/, a confine di parola nel sintagma *pirogue partie* 'canoa partita'. Le figure riportano il sonogramma del segnale verbale, in alto, e glottico, in basso, per un caso di assimilazione parziale non completa (Figura 2) e per un caso di assimilazione parziale completa (Figura 3).

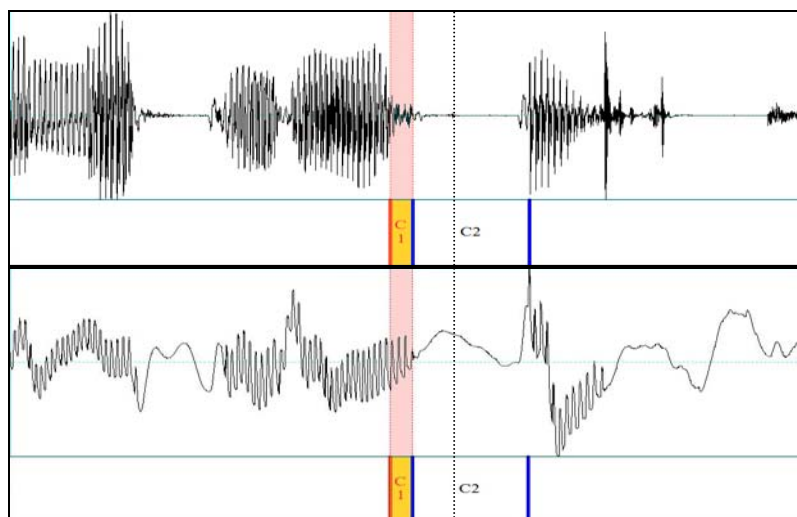


Figura 2: Segnale verbale (sopra) e segnale glottico (sotto) del contesto /gp/ al confine di parola in contesto di focalizzazione sul secondo termine. Il segmento /g/ risulta non completamente assimilato nella prima ripetizione del parlante italofono P2.

<sup>12</sup> Si veda Gendrot (2003) per uno studio relativo ai parametri EGG ricavati in relazione a segmenti vocalici, al variare della posizione all'interno del sintagma intonativo e dell'enunciato ed del contesto di focalizzazione.

<sup>13</sup> Sono stati segmentate ed etichettate anche le eventuali vocali epentetiche e pause realizzate tra i due segmenti consonantici, la nasale precedente i contesti /df/, /dt/, /gp/, /td/ e la vibrante precedente il contesto /dt/.

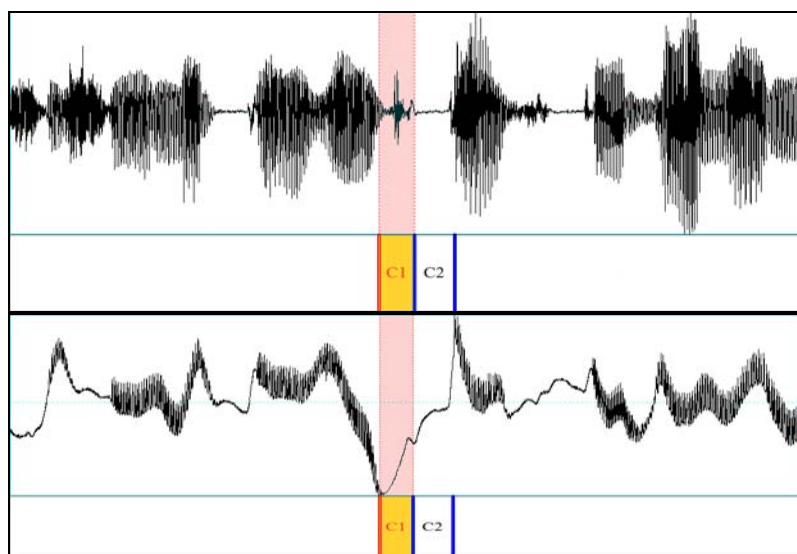


Figura 3: Segnale verbale (sopra) e segnale glottico (sotto) del contesto /gp/ al confine di parola in contesto di focalizzazione sul secondo termine. Il segmento /g/ risulta assimilato in modo completo nella seconda ripetizione del parlante italofono P1.

## 4.2 Risultati del secondo esperimento

### 4.2.1 Realizzazione di assimilazioni del tratto di sonorità

La realizzazione di assimilazione del tratto di sonorità, anche nel secondo corpus, è essenzialmente parziale regressiva. L'assimilazione regressiva si realizza, nel complesso di tutti i dati, per il 46.56% dei casi nelle produzioni degli italofoeni e per il 54.08% dei casi in quelle dei francofoeni<sup>14</sup>, al contrario, per gli italofoeni e i francofoeni l'assimilazione progressiva si realizza nell'11.88% e nel 19.8% dei casi, rispettivamente. In questa sede commenteremo approfonditamente i dati relativi alle sole assimilazioni regressive, le più frequenti nel nostro corpus, e commenteremo separatamente i dati relativi ai contesti nei quali era previsto che l'assimilazione del tratto di sonorità fosse completa (/bs/, /bt/, /ds/, /gs/ e /kd/ [Petit Robert, 2000]).

Nelle tabelle 8-11 sono riportati i dati relativi alle assimilazioni riscontrate nelle produzioni dei singoli parlanti sulla base dell'ispezione del segnale verbale e glottico. Per quanto riguarda i contesti nei quali era attesa l'assimilazione completa del tratto di sonorità all'interno di parola – si veda la tabella 8 – il segmento maggiormente assimilato dai parlanti italofoeni è l'occlusiva bilabiale /b/, soprattutto nella combinazione /bs/. Gli italofoeni non realizzano molte assimilazioni nei contesti /ds/, /kd/<sup>15</sup>, e /gs/ (quest'ultimo non è mai contesto di assimilazione, mentre /kd/ lo è solo per un parlante). I due parlanti francofoeni – si veda la tabella 9 – mostrano un comportamento linguistico eterogeneo: la parlante P4,

<sup>14</sup> I risultati sopra descritti si riferiscono alla media calcolata considerando tutte le assimilazioni regressive realizzate all'interno delle produzioni dei parlanti.

<sup>15</sup> Gli stimoli “anecdote” e “synecdoque”, in alcuni casi, sono oggetto di realizzazioni con assimilazione progressiva in cui si verifica la desonorizzazione del segmento /d/. In ogni caso, bisogna sottolineare che i parlanti italofoeni, tendono ad elidere il segmento /k/ (12 casi su 18, equivalenti al 66.67%).

infatti, effettua l'assimilazione nel 72.22% degli stimoli, considerati nel complesso, mentre il parlante P5 realizza solo il 19.45% delle assimilazioni. L'eterogeneità delle osservazioni ci sembra imputabile al fatto che il parlante P5, che presenta uno stile di eloquio molto fluente, tende a non pronunciare il primo segmento e ad annullare quindi il contesto di assimilazione. Si noti, comunque, che mentre i francofoni, in questi contesti, realizzano quasi sempre le assimilazioni del tratto di sonorità in modo completo (83.33%), negli stessi contesti i parlanti italo-foni realizzano le assimilazioni in modo non completo nel 47.48% dei casi.

Considerando le assimilazioni all'interno di parola attese in tutti gli altri casi (nei quali era attesa assimilazione, ma non necessariamente completa), tutti e tre i parlanti italo-foni mostrano di non avere alcuna difficoltà nel sonorizzare la fricativa alveolare /s/. Risultato poco sorprendente dal momento che anche in italiano si riscontrano casi di assimilazione della fricativa prima di consonante sonora. Quanto agli altri contesti consonantici, la parlante P2 desonorizza sempre l'occlusiva dentale /d/ seguita dalla fricativa alveolare /s/ all'interno dello stimolo *médecin*. Inoltre, in questi contesti gli italo-foni realizzano le assimilazioni in modo completo nel 70.95%.

Per quanto riguarda i risultati relativi alle produzioni semispontanee (Map-task), non emergono differenze per la realizzazione di assimilazione. I parlanti assimilano gli stessi contesti già evidenziati nel commento dei risultati del parlato letto. In particolare, nessuno dei parlanti italo-foni assimila il contesto /gs/ e il contesto /vs/; nel contesto /bk/, la bilabiale è desonorizzata solo da P1, mentre il contesto /ds/, relativo allo stimolo in cui è prevista un'assimilazione completa del tratto di sonorità, è luogo di assimilazione per due parlanti, P1 e P3; nel contesto /ds/ nel quale ci si aspettava assimilazione non completa nessun parlante italo-fono effettua assimilazione. Per quanto riguarda i francofoni, sono riscontrate le assimilazioni attese, in modo simile a quanto riportato per il parlato letto, salvo che nel caso del parlante P5 che tende a non pronunciare il primo segmento relativo ai contesti di assimilazione completa del tratto di sonorità.

| Contesti | P1  |       | P2  |       | P3  |       | Completa (%) | Non Completa (%) |
|----------|-----|-------|-----|-------|-----|-------|--------------|------------------|
|          | N°  | %     | N°  | %     | N°  | %     |              |                  |
| /bk/     | 3/6 | 50    | 2/6 | 33.33 | 1/6 | 16.67 | 0            | 100              |
| /bs/     | 6/6 | 100   | 3/6 | 50    | 4/6 | 66.67 | 61.54        | 38.46            |
| /bt/     | 3/6 | 50    | 3/6 | 50    | 1/6 | 16.67 | 28.57        | 71.43            |
| /ds/     | 2/6 | 33.33 | 0   | 0     | 2/6 | 33.33 | 25           | 75               |
| /gs/     | 0   | 0     | 0   | 0     | 0   | 0     | 0            | 0                |
| /kd/     | 0   | 0     | 0   | 0     | 1/6 | 16.67 | 100          | 0                |
| Tot      |     | 38.88 |     | 22.22 |     | 25    | 35.85        | 47.48            |

Tabella 8: Italo-foni. Contesti previsti per assimilazione completa all'interno di parola.

| Contesti | P4  |       | P5  |       | Completa (%) | Non Completa (%) |
|----------|-----|-------|-----|-------|--------------|------------------|
|          | N°  | %     | N°  | %     |              |                  |
| /bk/     | 3/6 | 50    | 0   | 0     | 33.33        | 66.67            |
| /bs/     | 6/6 | 100   | 0   | 0     | 100          | 0                |
| /bt/     | 2/6 | 33.33 | 1/6 | 16.67 | 66.67        | 33.33            |
| /ds/     | 4/6 | 66.67 | 1/6 | 16.67 | 100          | 0                |
| /gs/     | 5/6 | 83.33 | 0   | 0     | 100          | 0                |
| /kd/     | 6/6 | 100   | 5/6 | 83.33 | 100          | 0                |
| Tot      |     | 72.22 |     | 19.45 | 83.33        | 16.67            |

Tabella 9: Francofoni. Contesti previsti per assimilazione completa all'interno di parola.



| Contesti | P1  |       | P2  |       | P3  |       | Completa (%) | Non Completa (%) |
|----------|-----|-------|-----|-------|-----|-------|--------------|------------------|
|          | N°  | %     | N°  | %     | N°  | %     |              |                  |
| /sb/     | 4/6 | 66.67 | 5/6 | 83.33 | 5/6 | 83.33 | 100          | 0                |
| /sd/     | 3/3 | 100   | 3/3 | 100   | 3/3 | 100   | 100          | 0                |
| /sv/     | 4/6 | 66.67 | 6/6 | 100   | 4/6 | 66.67 | 21.42        | 78.58            |
| /vs/     | 0   | 0     | 1/3 | 33.33 | 0   | 0     | 100          | 0                |
| /ds/     | 0   | 0     | 3/3 | 100   | 0   | 0     | 33.33        | 66.67            |
| Tot      |     | 46.67 |     | 83.32 |     | 50    | 70.95        | 29.05            |

Tabella 10: Italofofoni. Contesti previsti per assimilazione non completa all'interno di parola.

| Contesti | P4  |       | P5  |       | Completa (%) | Non Completa (%) |
|----------|-----|-------|-----|-------|--------------|------------------|
|          | N°  | %     | N°  | %     |              |                  |
| /sb/     | 4/6 | 66.67 | 6/6 | 100   | 100          | 0                |
| /sd/     | 2/3 | 66.67 | 3/3 | 100   | 100          | 0                |
| /sv/     | 1/6 | 16.67 | 6/6 | 100   | 100          | 0                |
| /vs/     | 3/3 | 100   | 1/3 | 16.67 | 60           | 40               |
| /ds/     | 3/3 | 100   | 3/3 | 100   | 83.33        | 16.67            |
| Tot      |     | 80    |     | 88.33 | 88.67        | 11.33            |

Tabella 11: Francofofoni. Contesti previsti per assimilazione non completa all'interno di parola.

Per quanto riguarda le assimilazioni a confine di parola, i dati della tabella 12<sup>16</sup> mostrano il comportamento linguistico eterogeneo riscontrato nelle produzioni dei parlanti intervistati. Considerando i contesti in cui si prevedeva la sonorizzazione della fricativa alveolare /s/ - contesti nei quali ci si aspettava il maggior numero di realizzazioni in base ai risultati del primo esperimento e ai risultati delle rilevazioni all'interno di parola - si riscontra che il parlante P1 non sonorizza il segmento in nessun contesto, il parlante P2 lo sonorizza solo nel contesto /sb/ (con una percentuale dell'11.11% , cioè in 2 casi su 18), mentre il parlante P3 sonorizza la fricativa in tutti i contesti previsti (la percentuale più alta è riscontrata nel contesto /sv/ ed è pari al 55.56%, cioè la sonorizzazione avviene in 10 casi su 18). Nel caso di questi contesti consonantici, si potrebbe pensare che la differenza di comportamento all'interno ed al confine di parola dipenda dalla frequenza di occorrenza di contesti analoghi in italiano: il nesso /sb/, ad esempio, è frequente all'interno di parola piuttosto che a confine di parola. Tuttavia l'ipotesi che il diverso comportamento sia dovuto all'influenza della lingua materna non è coerente con i risultati del primo esperimento – in cui, tra interno e confine di parola, si era notata solo una flessione del numero di casi di assimilazione e non una differenza così ingente. Inoltre, i parlanti italofofoni effettuano l'assimilazione del tratto di sonorità in contesti anche molto differenti rispetto a quelli riscontrati in italiano. Ad esempio nel contesto /vp/ si ha assimilazione regressiva del tratto di sonorità nel 72.22% (13 casi su 18) e nel 61.11% (11 casi su 18) da parte dei parlanti P2 e P3, rispettivamente; nel contesto /gt/, la percentuale di assimilazioni è 55.56% (10 casi su

<sup>16</sup> Nelle tabelle relative alle assimilazioni al confine di parola (tabelle 12 e 13) ed alle assimilazioni al confine di parola nei contesti di focalizzazione (tabelle 14 e 15) riportiamo solo i casi più interessanti ai fini della nostra descrizione. I risultati nel loro complesso saranno descritti nel paragrafo relativo all'analisi statistica.

18) per i parlanti P1 e P3; infine, l'occlusiva bilabiale è desonorizzata nel 61.11% (11 casi su 18) nei contesti /bk/ e /bt/ da parte di P3.

Un comportamento comune nelle produzioni degli italofoeni è l'inserimento di una vocale schwa epentetica tra i due segmenti consonantici. In questi casi, l'eventuale assimilazione sembra assumere le caratteristiche di un'assimilazione a distanza, soprattutto nel caso di desonorizzazione; nel caso di sonorizzazione, infatti, non si può completamente escludere che il fenomeno sia dovuto alla posizione intervocalica acquisita per l'inserimento della vocale epentetica, benché nella varietà d'italiano parlata dai soggetti intervistati non si riscontri il fenomeno della sonorizzazione di /s/ in posizione intervocalica. In ogni caso, in tutte le realizzazioni del contesto /gt/ da parte di P1 si è verificato l'inserimento della vocale epentetica. Un altro fenomeno interessante già osservato nel primo esperimento, è la cancellazione del segmento finale di parola, riscontrato soprattutto per i contesti nei quali il segmento in questione è /t/ e per i contesti nei quali il segmento finale di parola è preceduto da una nasale. In particolare, per i contesti /td/ e /dt/ si registrano cancellazioni che superano il 40% (43.34% e 45.83% rispettivamente).

Infine, come si può osservare ancora nella tabella 12, gli italofoeni tendono a realizzare l'assimilazione sia come completa che come non completa, a seconda dei contesti consonantici e delle combinazioni. I segmenti realizzati con assimilazione completa in percentuale maggiore sono /b/ e /g/, in particolare quando sono seguiti dalla fricativa /s/ (infatti, per i contesti /bs/ e /gs/ le percentuali sono del 77.78% e del 75% rispettivamente). I segmenti realizzati con assimilazione non completa in percentuale maggiore sono /k/ (100%), /f/ (80%) e /s/ (80% nel contesto /sv/ e 75% in quello /sg/) e /d/ (75% nel contesto /ds/).

I pochi casi di assimilazioni progressive si riscontrano per i seguenti contesti: /fd/ (10 casi su 24, 41.67%), /sd/ (9 casi su 24, 44.44%) , /tb/ (9 casi su 18, 18.50%) per P3; /sv/ (7 casi su 18, 38.88%) e /sg/ (8 casi su 18, 44.44%) per P2; l'unico contesto riscontrato per P1 è /pg/ (5 casi su 27, 18.51%).

Per quanto riguarda i parlanti francofoeni, come si evince dalla tabella 13, la fricativa /s/ è sonorizzata in più del 50% dei casi: nei contesti /sd/ (70.83% per entrambi), /sv/ (72.22% per P4, 77.78% per P5), /sb/ (55.56% per P4, 66.67% per P5); si differenzia solo il contesto /sg/ in cui la percentuale è dell'11.11% e del 33.33%, per P4 e P5 rispettivamente. Considerando i contesti in cui si realizza la desonorizzazione del primo segmento, la consonante /b/ è assimilata nel 55.56% dei casi (10 su 18) da parte di P4, ma solo nel contesto /bs/; negli altri casi, infatti, il tratto è scarsamente assimilato. Il segmento /v/ è assimilato soprattutto quando è seguito da /p/ - infatti le percentuali sono del 100% per P4 (18 casi su 18) e dell'83.33% per P5 (15 casi su 18) - e nel contesto /gs/ il tratto di sonorità è assimilato da /g/ nel 94.44% da P4 (17 casi su 18) e nel 66.67% da P5 (12 casi su 18).

Nei casi di assimilazione a confine di parola, la realizzazione di una vocale epentetica è un fenomeno del tutto marginale nelle produzioni dei francofoeni. Al contrario, sono numerosi i casi di cancellazione del segmento finale di parola. In particolare, per il contesto /td/ si riscontrano cancellazioni nel 63.63% dei casi mentre nel il contesto /dt/ la percentuale è del 44.44%.

I parlanti francofoeni, al contrario degli italofoeni, tendono a realizzare assimilazioni soprattutto in modo completo: in contesti come /sv/, /sg/, /sb/, /sd/, /tb/, /kd/, /fb/, /bt/ il tratto di sonorità viene assimilato dal primo segmento con percentuali al di sopra del 90%.

Per quanto riguarda le assimilazioni progressive, i contesti più interessanti sono: /pg/ con il 37.03% (10 casi su 27) e il 77.78% (21 casi su 27) per P4 e P5 rispettivamente; /pv/ in cui l'assimilazione si riscontra nel 50% dei casi (9 casi su 18) da parte di P4; /sg/ con il

55.56% (10 casi su 18) da parte di P5; e, infine, /td/ e /tg/ in cui si ha assimilazione del tratto di sonorità per il parlante P5, nel 27.27% (9 casi su 33) e nel 77.78% (14 casi su 18), rispettivamente.

| Contesti | P1    |       | P2    |       | P3    |       | Completa (%) | Non Completa (%) |
|----------|-------|-------|-------|-------|-------|-------|--------------|------------------|
|          | N°    | %     | N°    | %     | N°    | %     |              |                  |
| /bk/     | 9/18  | 50    | 0/18  | 0     | 11/18 | 61.11 | 55           | 45               |
| /bs/     | 3/18  | 16.67 | 0/18  | 0     | 6/18  | 33.33 | 77.78        | 22.22            |
| /bt/     | 4/18  | 22.22 | 3/18  | 16.67 | 11/18 | 61.11 | 50           | 50               |
| /ds/     | 4/24  | 16.67 | 0/24  | 0     | 4/24  | 16.67 | 25           | 75               |
| /dt/     | 3/24  | 12.50 | 1/24  | 4.16  | 1/24  | 4.16  | 60           | 40               |
| /fb/     | 2/18  | 11.11 | 0/18  | 0     | 8/18  | 44.44 | 20           | 80               |
| /gp/     | 8/18  | 44.44 | 1/18  | 5.56  | 8/18  | 44.44 | 47.05        | 52.94            |
| /gs/     | 11/18 | 61.11 | 1/18  | 5.56  | 4/18  | 22.22 | 75           | 25               |
| /gt/     | 10/18 | 55.56 | 2/18  | 11.11 | 10/18 | 55.56 | 50           | 50               |
| /kd/     | 0/33  | 0     | 2/33  | 6.25  | 8/33  | 24.24 | 0            | 100              |
| /sb/     | 0/18  | 0     | 2/18  | 11.11 | 7/18  | 38.89 | 33.33        | 66.67            |
| /sd/     | 0/24  | 0     | 0/24  | 0     | 9/24  | 37.5  | 33.33        | 66.67            |
| /sg/     | 0/18  | 0     | 0/18  | 0     | 8/18  | 44.44 | 25           | 75               |
| /sv/     | 0/18  | 0     | 0/18  | 0     | 10/18 | 55.56 | 20           | 80               |
| /tb/     | 0/18  | 0     | 1/18  | 5.56  | 2/18  | 11.11 | 66.67        | 33.33            |
| /td/     | 1/33  | 3.03  | 0/18  | 0     | 0     | 0     | 0            | 100              |
| /vp/     | 0/18  | 0     | 13/18 | 72.22 | 11/18 | 61.11 | 21.74        | 78.26            |
| /vs/     | 0/18  | 0     | 1/18  | 5.56  | 4/18  | 22.22 | 100          | 0                |

Tabella 12: Italofoeni. Assimilazioni a confine di parola.

| Contesti | P4    |       | P5    |       | Completa (%) | Non Completa (%) |
|----------|-------|-------|-------|-------|--------------|------------------|
|          | N°    | %     | N°    | %     |              |                  |
| /bk/     | 0/18  | 0     | 2/18  | 11.11 | 0            | 100              |
| /bs/     | 10/18 | 55.56 | 6/18  | 33.33 | 18.75        | 81.25            |
| /bt/     | 1/18  | 5.56  | 0/18  | 0     | 100          | 0                |
| /ds/     | 16/24 | 66.67 | 8/24  | 33.33 | 62.5         | 37.5             |
| /dt/     | 0/24  | 0     | 0/24  | 0     | 0            | 0                |
| /fb/     | 8/18  | 44.44 | 14/18 | 77.78 | 95.45        | 4.54             |
| /gp/     | 3/18  | 16.67 | 2/18  | 11.11 | 40           | 60               |
| /gs/     | 17/18 | 94.44 | 12/18 | 66.67 | 58.62        | 41.38            |
| /gt/     | 4/18  | 22.22 | 3/18  | 16.67 | 28.57        | 71.43            |
| /kd/     | 9/33  | 27.27 | 4/33  | 12.12 | 100          | 0                |
| /sb/     | 10/18 | 55.56 | 12/18 | 66.67 | 100          | 0                |
| /sd/     | 17/24 | 70.83 | 17/24 | 70.83 | 97.06        | 2.94             |
| /sg/     | 2/18  | 11.11 | 6/18  | 33.33 | 100          | 0                |
| /sv/     | 13/18 | 72.22 | 14/18 | 77.78 | 100          | 0                |
| /tb/     | 2/18  | 11.11 | 4/18  | 22.22 | 100          | 0                |
| /td/     | 0/33  | 0     | 1/33  | 3.03  | 0            | 100              |
| /vp/     | 18/18 | 100   | 15/18 | 83.33 | 57.58        | 42.42            |
| /vs/     | 8/18  | 44.44 | 2/18  | 11.11 | 80           | 20               |

Tabella 13: Francofoeni. Assimilazioni a confine di parola.

#### 4.2.2 Assimilazione del tratto di sonorità e focalizzazione

I risultati finora descritti si riferiscono al numero di assimilazioni realizzate dai parlanti nei contesti di focalizzazione neutra. Tuttavia, nel caso di contesti di assimilazione a confine di parola, avevamo previsto anche casi di focalizzazione sul primo e sul secondo elemento del sintagma. La verifica uditiva circa la realizzazione di focalizzazione ristretta-contrastiva ha evidenziato la tendenza per i parlanti italofofoni a realizzare in modo più evidente la focalizzazione in corrispondenza del secondo elemento del sintagma, talvolta con uno spostamento della focalizzazione dal primo termine sul secondo. Per i francofoni, la distribuzione di realizzazione della focalizzazione è pressoché equivalente nei due contesti previsti<sup>17</sup>.

Per quanto riguarda la realizzazione di assimilazione in base al contesto di focalizzazione, dalla tabella 14 emerge che gli italofofoni non assimilano mai il tratto di sonorità in caso di focalizzazione sul primo termine del sintagma. Il maggior numero di casi assimilati viene riscontrato, infatti, nei contesti di focalizzazione neutra e comunque nei casi in cui non sia stato realizzato il focus previsto – si veda, in tabella, la colonna 'Focalizzazione neutra + casi non focalizzati'. Il parlante P1 assimila i contesti /bk/ e /gs/ solo in posizione neutra (9 casi su 9 per /bk/; 11 casi su 11 per /gs/). Il parlante P2 assimila il tratto di sonorità nel contesto /vp/ nel 76.92% dei casi in di focalizzazione neutra (10 casi su 13). Il parlante P3 effettua assimilazione nel 63.64% dei contesti /bk/ e /vp/ (7 casi su 11) e nel 90% dei contesti /gt/ (9 casi su 10). Al contrario di quanto accade per il contesto di focalizzazione sul primo termine, si riscontrano alcuni casi di assimilazioni in caso di focalizzazione sul secondo elemento del sintagma. Il parlante P3, ad esempio, effettua l'assimilazione nei contesti /bk/ e /vp/ nel 36.36% dei casi (4 su 11) e nel 45.45% dei casi per il contesto /bt/ (5 assimilazioni su 11).

Anche i parlanti francofoni realizzano assimilazioni maggiormente in caso di focalizzazione neutra. Come si evince dalla tabella 15, le percentuali superano il 50% per entrambi i parlanti. Il parlante P4, ad esempio, assimila nel contesto /gs/ per l'86.67% dei casi (13 su 15) e nel contesto /vp/ per il 77.78% (14 casi su 18); il parlante P5 assimila il tratto di sonorità nel contesto /gs/ per il 91.66% (11 casi su 12) e nel contesto /sb/ per l'83.33% dei casi (10 su 12). Per quanto riguarda le assimilazioni in contesto di focalizzazione sul secondo termine, i casi più interessanti si riscontrano nelle produzioni di entrambi i parlanti per il contesto /vp/ - 5 casi su 15 per P4; 3 casi su 15 per P5 - e del parlante P4, per il contesto /sb/ (4 casi su 10). Nelle produzioni dei francofoni si riscontrano anche alcuni casi di assimilazioni nel contesto di focalizzazione sul primo termine. In particolare, P4 effettua assimilazione del tratto di sonorità nel contesto /bs/ (3 casi su 10) e /gs/ (3 casi su 17); P5 nel solo contesto /fb/ (3 casi su 14).

I risultati sono coerenti rispetto all'attesa di un maggior numero di assimilazioni nel caso in cui non intervenga alcun confine fonologico, quindi nei casi di focalizzazione neutra o di non focalizzazione. Per gli italofofoni, l'assenza di assimilazioni nel caso di focalizzazione sul primo elemento del sintagma è coerente con il riferimento alle caratteristiche della lingua italiana, in cui si riporta la presenza di un confine prosodico a destra dell'elemento focalizzato, in grado di bloccare il fenomeno di assimilazione [Frascarelli, 1997]. La situazione riscontrata nelle produzioni dei francofoni sembra più eterogenea, con un certo numero, seppur limitato, di assimilazioni nel caso di focalizzazione sia sul primo che sul

<sup>17</sup> In ogni caso, gli italofofoni marcano l'elemento contrastivo in modo più evidente rispetto ai parlanti francofoni.

secondo membro del sintagma. L'ipotesi avanzata da Féry [2003] circa la presenza di un confine prosodico a sinistra dell'elemento in focus, in grado di bloccare l'assimilazione non è quindi completamente supportata dai nostri dati.

| Contesti | Focalizzazione neutra +<br>Casi non focalizzati |              |                 | Focalizzazione<br>parola 1 |    |    | Focalizzazione<br>parola 2 |             |                 |
|----------|-------------------------------------------------|--------------|-----------------|----------------------------|----|----|----------------------------|-------------|-----------------|
|          | P1                                              | P2           | P3              | P1                         | P2 | P3 | P1                         | P2          | P3              |
| /bk/     | 9/9<br>(100%)                                   |              | 7/11<br>(63.64) |                            |    |    |                            |             | 4/11(36.36)     |
| /bs/     | 3/3<br>(100%)                                   |              | 4/6 (66.67)     |                            |    |    |                            |             | 2/6(33.33)      |
| /bt/     | 3/4 (75)                                        | 3/3(100)     | 6/11<br>(54.55) |                            |    |    | 1/4 (25)                   |             | 5/11<br>(45.45) |
| /ds/     | 2/3*<br>(66.67)                                 |              | 3/3* (100)      |                            |    |    | 1/3(33.33)                 |             |                 |
| /fb/     | 2/2(100)                                        |              | 4/8 (50)        |                            |    |    |                            |             | 4/8 (50)        |
| /gp/     | 7/8(87.5)                                       | 1/1 (100)    | 5/8 (62.5)      |                            |    |    | 1/8(12.5)                  |             | 3/8 (37.5)      |
| /gs/     | 11/11(100)                                      |              | 2/4(50)         |                            |    |    |                            | 1/1(100)    | 2/4 (50)        |
| /gt/     | 8/10(80)                                        | 2/2 (100)    | 9/10 (90)       |                            |    |    | 2/10 (20)                  |             | 1/10 (10)       |
| /kd/     |                                                 | 1/2 (50)     | 3/4* (75)       |                            |    |    |                            | 1/2 (50)    | 1/4 (25)        |
| /sb/     |                                                 | 1/2 (50)     | 5/7 (71.43)     |                            |    |    |                            | 1/2 (50)    | 2/7(28.57)      |
| /sd/     |                                                 |              | 3/4* (75)       |                            |    |    |                            |             | 1/4 (25)        |
| /sg/     |                                                 |              | 6/8(75)         |                            |    |    |                            |             | 2/8 (25)        |
| /sv/     |                                                 |              | 6/10 (60)       |                            |    |    |                            |             | 4/10 (40)       |
| /vp/     |                                                 | 10/13(76.92) | 7/11(63.64)     |                            |    |    |                            | 3/13(23.08) | 4/11(36.36)     |
| /vs/     |                                                 | 1/1 (100)    | 3/4 (75 )       |                            |    |    |                            |             | 1/4 (25)        |

Tabella 14: Italofoeni. Assimilazioni regressive realizzate nei contesti di focalizzazione neutra e contrastiva<sup>18</sup>.

| Contesti | Focalizzazione neutra +<br>Casi non focalizzati |                | Focalizzazione<br>parola 1 |             | Focalizzazione<br>parola 2 |             |
|----------|-------------------------------------------------|----------------|----------------------------|-------------|----------------------------|-------------|
|          | P4                                              | P5             | P4                         | P5          | P4                         | P5          |
| /bk/     |                                                 | 2/2<br>(100)   |                            |             |                            |             |
| /bs/     | 7/10 (70)                                       | 5/6<br>(83.33) | 3/10 (30)                  |             |                            | 1/6 (16.67) |
| /bt/     | 1/1 (100)                                       |                |                            |             |                            |             |
| /ds/     | 13/15* (86.67)                                  | 4/5* (80)      |                            |             | 2/13(13.33)                | 1/5 (20)    |
| /fb/     | 6/8 (75)                                        | 10/14 (71.43)  | 1/8 (12.5)                 | 3/14(21.43) | 1/8(12.5)                  | 1/14 (7.14) |
| /gp/     | 3/3 (100)                                       | 2/2 (100)      |                            |             |                            |             |
| /gs/     | 12/17 (70.59)                                   | 11/12 (91.66)  | 3/1(17.64)                 | 1/12 (8.34) | 2/17(11.77)                |             |
| /gt/     | 4/4 (100)                                       | 3/3 (100)      |                            |             |                            |             |
| /kd/     | 5/5* (100)                                      | 2/2* (100)     |                            |             |                            |             |
| /sb/     | 6/10 (60)                                       | 10/12 (83.33)  |                            |             | 4/10 (40)                  | 2/12(16.67) |
| /sd/     | 7/8* (87.5)                                     | 12/12* (100)   |                            |             | 1/8 (25)                   |             |
| /sg/     | 1/2 (50)                                        | 5/6 (83.33)    | 1/2 (50)                   |             |                            | 1/6(16.67)  |
| /sv/     | 10/13 (76.92)                                   | 11/14 (78.57)  | 1/13 (7.7)                 | 1/14 (7.15) | 2/13(15.38)                | 2/14(14.28) |
| /vp/     | 14/18 (77.78)                                   | 10/15(66.67)   | 1/18(5.56)                 |             | 3/18(16.66)                | 5/15(33.33) |
| /vs/     | 8/8 (100)                                       | 2/2 (100)      |                            |             |                            |             |

Tabella 15: Francofoeni. Assimilazioni regressive realizzate nei contesti di focalizzazione neutra e contrastiva.

<sup>18</sup> L'asterisco indica una variazione del numero totale di assimilazioni dovuta al fatto che i sintagmi con la preposizione semplice "de" non rientrano nei contesti di focalizzazione.

#### 4.2.3 Analisi statistiche

Le analisi statistiche discusse in questi paragrafi riguardano i casi di assimilazione regressiva<sup>19</sup> e corrispondono ad analisi ANOVA multivariate, nelle quali il contesto consonantico è stato messo a fattore con la lingua materna del parlante (italiano/francese) e il fattore parlante (indipendentemente dalla sua lingua materna). In ogni caso, si è riscontrata la significatività dell'interazione tra i fattori. La posizione dell'accento, invece, non è stata messa a fattore nelle analisi statistiche ma è stata considerata nella costituzione del corpus per renderlo rappresentativo delle possibili posizioni prosodiche del contesto di studio, di fatto annullando il possibile effetto della posizione dell'accento sull'incidenza dei fenomeni di assimilazione.

##### 4.2.3.1 Assimilazioni all'interno di parola

Nel caso di assimilazioni regressive all'interno di parola, sia per gli italofofoni che per i francofoni, si riscontrano differenze significative in base ai tipi di segmenti coinvolti<sup>20</sup>. Test statistici ANOVA mostrano che il tipo di segmento risulta un fattore significativo in relazione al numero di realizzazioni di assimilazione del tratto di sonorità [ $F(216,9)=4.158;p<0.0001$ ], così come la lingua materna del soggetto [ $F(216,1)=21.078;p<0.0001$ ] (con un'interazione tra i due fattori  $F(216,9)=3.819;p=0.0002$ ). Come si può osservare dal grafico sottostante (Figura 4) e come confermato dal test post-hoc di Fisher<sup>21</sup>, nelle produzioni degli italofofoni i contesti in cui è presente la fricativa alveolare /s/ risultano, in media, maggiormente assimilati rispetto agli altri contesti consonantici (in particolare, si vedano /sb/, /sd/, /sv/ e /bs/).

Come si può osservare nel grafico, e come già accennato, i parlanti italofofoni non assimilano mai il fonema /g/ nelle loro produzioni ed assimilano molto raramente /k e /v/.

Si osserva un'alta variabilità inter-soggettiva [ $F(231,4)=5.004;p=0.0007$ ] – si veda la figura 5 – ma il post hoc di Fisher mostra che la differenza tra i due parlanti francofoni ed i parlanti italofofoni è significativa.

I risultati ottenuti dall'analisi delle produzioni nel Map-task - cfr. grafico in figura 6 – sono abbastanza coerenti con quelli ricavati dal *corpus* frasi: i contesti nei quali gli italofofoni resistono maggiormente al processo di assimilazione del tratto di sonorità sono /gs/ e /vs/. Una differenza tra il parlato letto e quello semispontaneo è data, invece, dall'assimilazione del contesto /sb/, assimilato maggiormente nel parlato letto. Infine, nel parlato semispontaneo, i parlanti assimilano sempre il contesto /bt/ che risulta significativamente diverso rispetto a /bk/, /sb/ e /vs/ ( $p=0.0222$ ).

---

<sup>19</sup> Esclusi, quindi, i contesti nei quali si è osservata la caduta del primo segmento consonantico.

<sup>20</sup> I grafici di riferimento per i risultati dei test ANOVA riportano la media dei casi nei quali è stata riscontrata assimilazione del tratto di sonorità sotto forma di valore numerico compreso tra '0' (assenza di assimilazione) ed '1' (presenza di assimilazione).

<sup>21</sup> Non riportiamo nel dettaglio i risultati di tutti i confronti effettuati nel test post-hoc di Fisher e ci limitiamo a sottolineare l'indicazione circa le differenze significative all'interno dell'insieme dei contesti, come supporto alla mera osservazione del grafico.

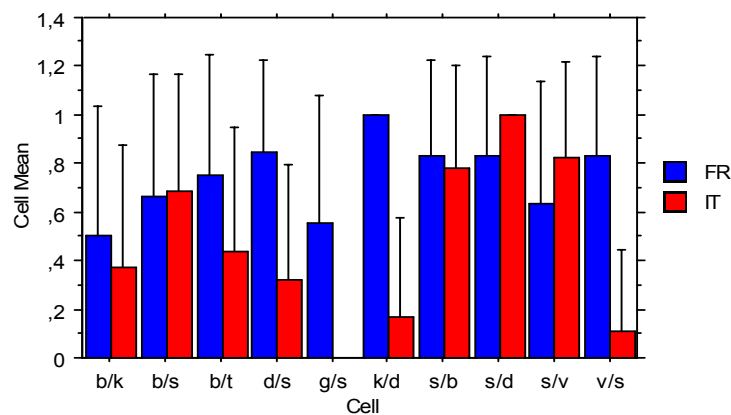


Figura 4: Italofofoni e francofofoni, parlato letto. Incidenza delle assimilazioni regressive (complete e non complete) all'interno di parola per contesto segmentale.

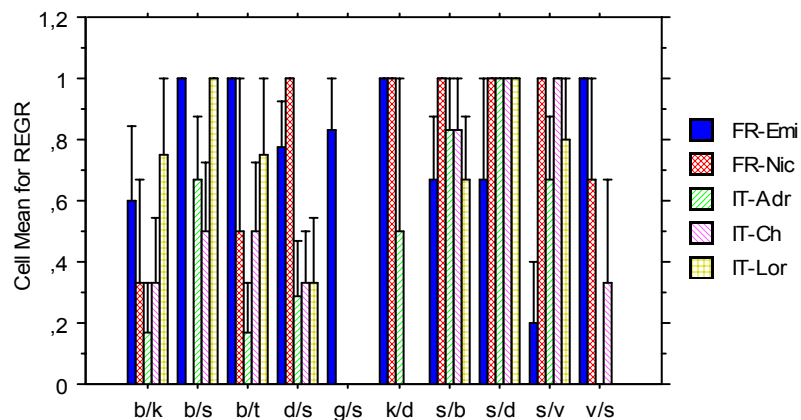


Figura 5: Italofofoni e francofofoni, parlato letto. Assimilazioni regressive (complete e non complete) all'interno di parola per soggetto (P1=IT-Lor; P2= IT-Ch; P3=IT-Adr; P4=FR-Emi; P5=FR-Nic).

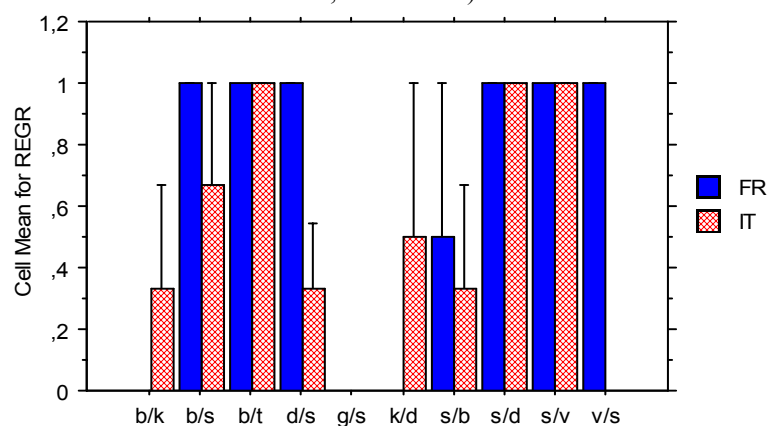


Figura 6: Italofofoni e francofofoni, parlato semispontaneo. Assimilazioni regressive (complete e non complete) all'interno di parola.

#### 4.2.3.2 Assimilazioni a confine di parola

##### *Focalizzazione neutra*

Nel caso di focalizzazione neutra, il test ANOVA mostra che il tipo di segmento considerato è significativo in relazione all'incidenza dei fenomeni di assimilazione del tratto di sonorità anche a confine di parola [ $F(868,27)=3.222; p<0.0001$ ], così come la lingua materna del soggetto [ $F(894,1)=330,655; p<0.0001$ ], con il maggior numero di assimilazioni registrate per i francofoni. Come si può osservare dal grafico sottostante (Figura 7), i parlanti italofofoni non assimilano mai i contesti /kv/, /td/, /tg/ e /sv/; relativamente alla assimilazione della fricativa, in particolare, gli stessi parlanti mostrano una maggiore difficoltà di assimilazione al confine di parola rispetto alla posizione all'interno di parola. Per quanto riguarda tutti gli altri contesti, le assimilazioni a confine di parola realizzate dagli italofofoni sono molte meno rispetto a quelle realizzate dai francofoni (che, comunque, non effettuano assimilazione nei contesti /dt/, /tb/, e /td/).

Osservando il comportamento linguistico degli italofofoni, sorprende l'assimilazione in alcuni contesti molto dissimili da quelli riscontrati nella lingua italiana, in particolare nei contesti /bk/, /gt/ e /vp/. Si consideri comunque che i parlanti differiscono significativamente gli uni dagli altri, come indicato dai risultati del test ANOVA [ $F(891,4)=111.042; p<0.0001$ ], e il post-hoc di Fisher mostra addirittura una differenza significativa tra tutti i parlanti tranne che tra i due italofofoni, P1 e P2.

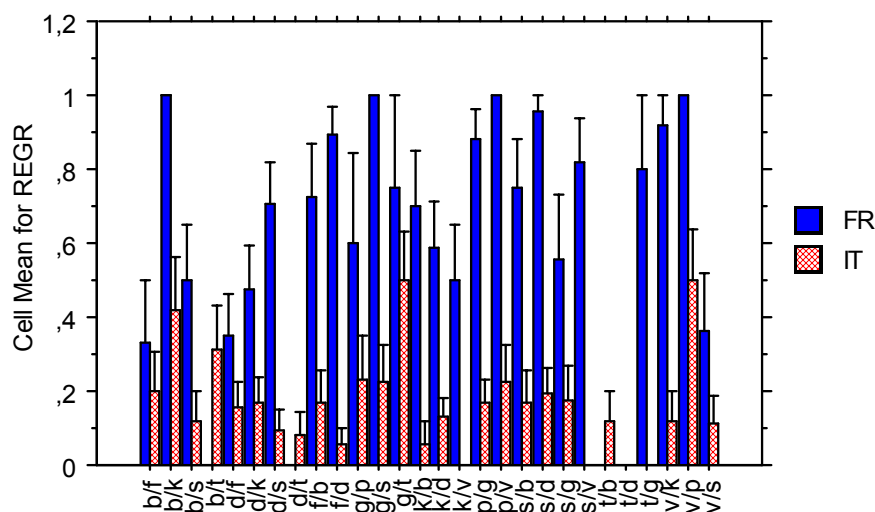


Figura 7: Italofofoni e francofoni, parlato letto. Assimilazioni regressive (complete e non complete) a confine di parola nei contesti di focalizzazione neutra.

##### *Focalizzazione contrastiva*

Passiamo adesso ad analizzare i risultati relativi al contesto di focalizzazione sul primo termine. Come si evince dal grafico sottostante (Figura 8, sinistra), i parlanti italofofoni realizzano assimilazioni regressive in due contesti consonantici: /dt/ e /tg/. L'ipotesi che esista quindi un confine prosodico profondo a destra dell'elemento focalizzato sembra quindi essere suffragata dai risultati con pochissime eccezioni (che riguardano in modo chiaro il contesto consonantico /dt/, l'unico in cui i segmenti sono caratterizzati dallo stesso modo e luogo di articolazione).



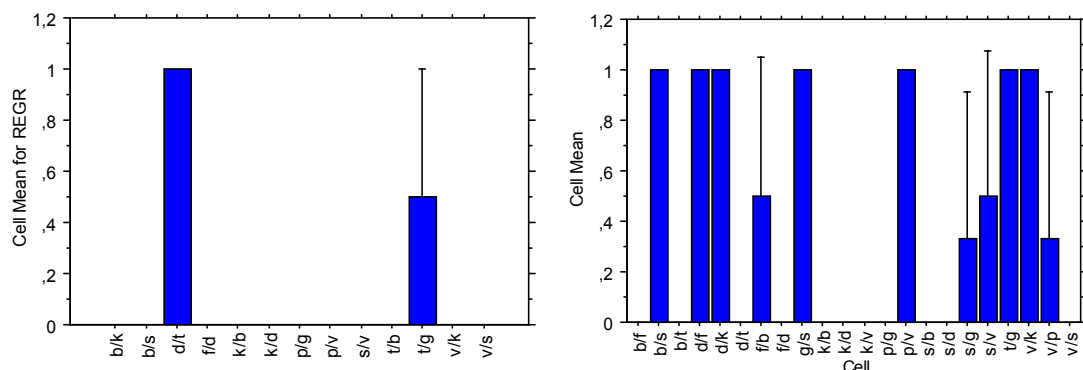


Figura 8: Italofoni (sinistra) e francofoni (destra), parlato letto. Assimilazioni regressive (complete e non complete) a confine di parola nei contesti di focalizzazione sul primo termine.

Per quanto riguarda i francofoni, si riscontrano molti più casi di assimilazione ed il test ANOVA indica differenze significative in relazione ai tipi di segmenti coinvolti [ $F(35,21)=34.375; p<0.0001$ ]. Come si può notare dal grafico in figura 8 (destra), i parlanti effettuano sempre assimilazione nei contesti /bs/, /df/, /dk/, /gs/, /pv/, /tg/, /vk/. Negli altri contesti l'assimilazione non si realizza, si realizza solo in alcuni casi, o, ancora, si riscontra la cancellazione della prima consonante del contesto (come in /bk/, /gp/, /td/, /tb/).

Nel caso di focalizzazione sulla seconda parola del sintagma, gli italofoni realizzano un maggior numero di assimilazioni del tratto di sonorità. L'analisi ANOVA dei risultati, rappresentati graficamente in figura 9 (sinistra), mostra che il comportamento linguistico cambia a seconda del contesto consonantico [ $F(277,27)=1.618; p=0.003$ ], sebbene, globalmente il numero di assimilazioni sia inferiore rispetto al contesto di focalizzazione neutra. Come si può notare dal grafico, i contesti consonantici nei quali si realizzano più assimilazioni sono /bk/, /dk/, /pv/, /vp/ e /bt/.

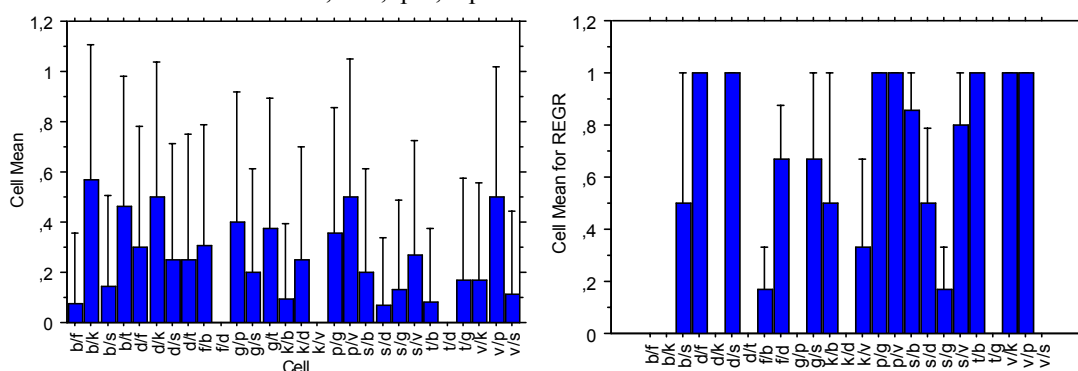


Figura 9: Italofoni (sinistra) e francofoni (destra), parlato letto. Assimilazioni regressive (complete e non complete) a confine di parola nei contesti di focalizzazione sul secondo termine.

Nel caso dei francofoni (figura 9, destra), doveva essere la focalizzazione sul secondo termine a indurre la presenza di un confine prosodico interno al sintagma considerato, in grado di interferire con l'assimilazione del tratto di sonorità. Al contrario, l'assimilazione

viene comunque realizzata a confine di parola e non si riscontra nemmeno una diminuzione del numero di casi osservati; anzi, i francofoni assimilano più contesti rispetto a quanto riscontrato alla presenza di focalizzazione sul primo termine (17 vs 11) e, salvo che per /fd/, nei casi in cui non effettuano assimilazione, cancellano il primo segmento consonantico. Il test ANOVA mostra che il tipo di segmento influenza in modo significativo la realizzazione di assimilazione [ $F(62,24)=4.403$ ,  $p<0.0001$ ] e il post-hoc di Fisher indica che esiste un maggior numero di assimilazioni in contesti che non coinvolgono sempre lo stesso segmento assimilato e che quindi dipendono dalla combinazione delle consonanti coinvolte.

#### 4.2.3.3 Misure di durata effettuate sul segnale verbale e sul segnale glottico

Come accennato in 4.1, sono state effettuate anche misurazioni di durata dei segmenti coinvolti, con riferimento al segnale verbale e al segnale glottico.

In generale, la segmentazione effettuata sulla base del segnale acustico ha permesso di individuare la presenza di segmenti di durata maggiore rispetto a quelli delimitati sulla base del segnale glottico (benché tra le durate ricavate con le due misurazioni ci sia un'elevata correlazione positiva - coefficiente di correlazione:  $r^2=0,63$  - grafico in figura 10, a sinistra). Il grafico in figura 10, a destra, riporta la differenza di durata tra le due misurazioni. Come si può osservare dal grafico, la differenza maggiore si riscontra per i contesti nei quali è stata realizzata assimilazione non completa del tratto di sonorità. Unica eccezione è rappresentata dalle misurazioni effettuate in contesto di focalizzazione sul primo elemento del sintagma, nella produzione degli italo-foni. In questi contesti, la differenza maggiore tra le misurazioni ricavate in base ai due segnali si riscontra nel caso di assenza di assimilazione.

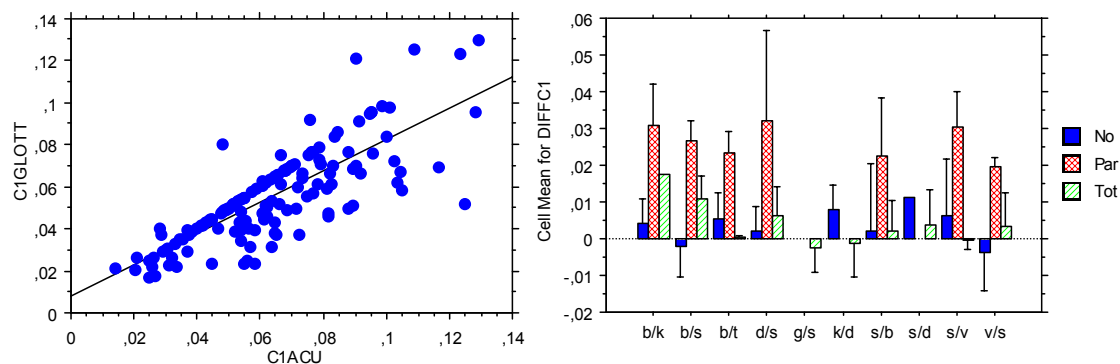


Figura 10: Prima consonante del contesto all'interno di parola, nel parlato letto dei soggetti italo-foni e francofoni. A sinistra: regressione lineare sulle durate ricavate in base al segnale verbale – ascisse – e glottico – ordinate; a destra: differenza di durata per le misurazioni effettuate sul segnale verbale e glottico.

Ci soffermeremo su alcune osservazioni relative all'assimilazione regressiva e alla durata del primo segmento consonantico, così come è stata ricavata sulla base del segnale verbale.

#### *Contesti all'interno di parola*

Nel caso di contesti di assimilazione in posizione interna di parola, l'ANOVA effettuata sui valori di durata del primo segmento consonantico mostra che il segmento ha in media una durata maggiore nelle produzioni degli italo-foni piuttosto che dei francofoni

[F(224,1)=4.405;p=0.037] e, comunque, ha una durata significativamente diversa a seconda del fatto che sia presente o meno un fenomeno di assimilazione del tratto di sonorità [F(224,2)=4,602;p=0,011], senza che i due aspetti interagiscano tra loro [F(224,2)=0,598;p=0,550]. In particolare, il post-hoc di Fisher mostra che la durata è maggiore nel caso di assimilazione parziale non completa – si veda la colonna 'Par' nel grafico in figura 11.

L'analisi dettagliata delle produzioni degli italofoeni mostra, come ci si poteva aspettare, una significativa differenza di durata a seconda del contesto consonantico considerato [F(136,9)=10,155;p<0.0001], e il post-hoc di Fisher mostra che i contesti differiscono tutti tra loro, tranne /bk/ e /ds/ - si veda la figura 12. Si noti che in questo caso non si riscontra una differenza significativa a seconda del parlante [F(143,2)=0.936;p=0.394].

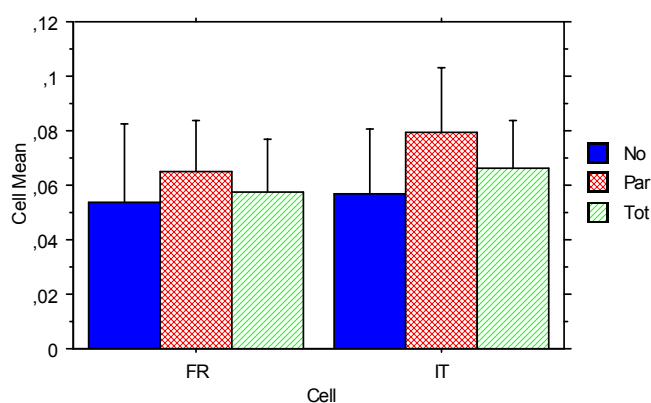


Figura 11: Assimilazioni all'interno di parola – analisi del segnale verbale di italofoeni e francofoeni. Durata del primo segmento consonantico del contesto bersaglio nei casi in cui non ci sia assimilazione (No) sia riscontrata assimilazione parziale non completa (Par) e completa (Tot)

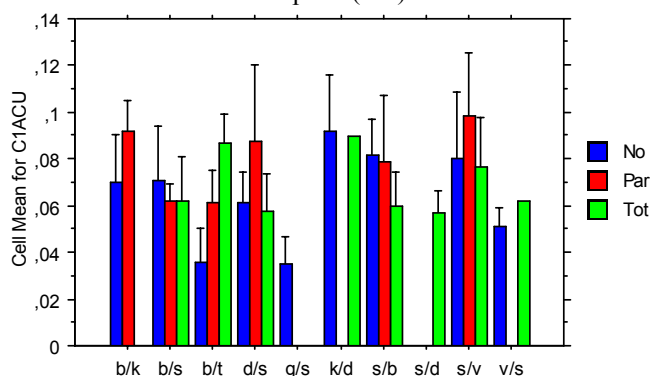


Figura 12: Assimilazioni all'interno di parola – analisi del segnale verbale degli italofoeni. Durata del primo segmento dei contesti consonantici esaminati nei casi in cui non ci sia assimilazione (No) sia riscontrata assimilazione parziale non completa (Par) e completa (Tot).

#### *Contesti a confine di parola*

Nel caso dei contesti di assimilazione a confine di parola e focalizzazione neutra, i risultati mostrano che il primo segmento consonantico ha durata maggiore nelle produzioni

degli italofofoni, e per entrambi i gruppi di parlanti la durata è leggermente minore nel caso di assimilazione non completa rispetto alla realizzazione senza assimilazione; nel caso di assimilazione completa la durata è minore solo per i francofoni, e non per gli italofofoni (che sembrano trattare il contesto in modo analogo nel caso di assimilazione completa e mancanza di assimilazione).

Nel caso di focalizzazione sul primo elemento del sintagma, non si riscontra differenza significativa nella durata del primo segmento consonantico del contesto bersaglio [F(88,1)=0.921;p=0,339]. Benché i valori di durata siano chiaramente diversi in caso di assimilazione parziale non completa e completa – si veda il grafico in figura 13, colonne 'Par' e 'Tot' (è bene ricordare che i parlanti italofofoni effettuano pochissime assimilazioni regressive in questo contesto - cfr paragrafo 4.2.3.2 - e lo scarso numero di osservazioni ha sicuramente influito sul risultato statistico).

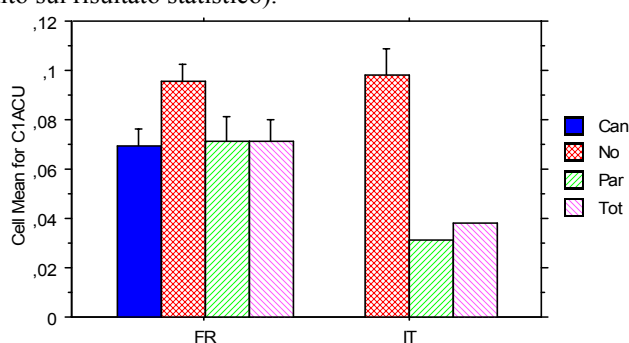


Figura 13: Assimilazioni a confine di parola, in contesto di focalizzazione sul primo elemento del sintagma – analisi del segnale verbale di italofofoni e francofoni. Durata del primo segmento consonantico del contesto bersaglio nei casi in cui non ci sia assimilazione (No), sia riscontrata assimilazione parziale non completa (Par) o completa (Tot), oppure cancellazione del segmento (Can<sup>22</sup>).

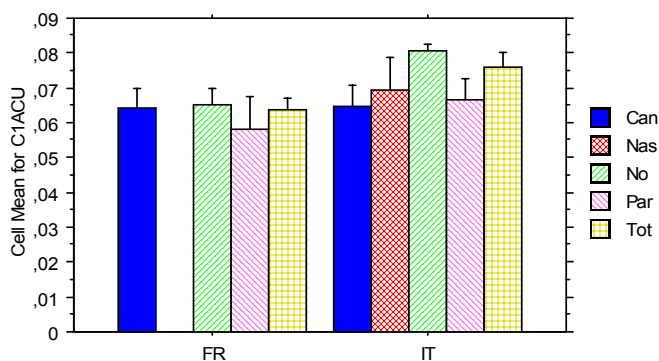


Figura 14: Assimilazioni a confine di parola, in contesto di focalizzazione sul secondo elemento del sintagma – analisi del segnale verbale di italofofoni e francofoni. Durata del primo segmento consonantico del contesto bersaglio nei casi in cui non ci sia assimilazione (No), sia riscontrata assimilazione parziale non completa (Par) o completa (Tot), cancellazione del segmento (Can) o eventuale allungamento della nasale precedente (Nas).

<sup>22</sup> In questi casi è stato misurato il segmento vocalico precedente.

Nel caso di focalizzazione sul secondo elemento del sintagma, al contrario, i risultati del test ANOVA mostrano che i parlanti italofofoni producono consonanti di durata significativamente maggiore rispetto ai francofoni [ $F(428,1)=17.549;p<0,0001$ ]. In ogni caso, la durata maggiore corrisponde ai segmenti non assimilati o con assimilazione completa. Si noti che in alcuni casi si riscontra la cancellazione del segmento consonantico: in questi casi è stato misurato il segmento vocalico precedente ('can') o, nel caso fosse presente, la nasale ('nas') – si veda la figura 14.

## 5. DISCUSSIONE E CONCLUSIONI

Dai risultati dell'indagine descritta nei paragrafi precedenti, emerge che le assimilazioni del tratto di sonorità si riscontrano sia all'interno che al confine di parola, sono principalmente parziali e regressive e, coerentemente con le descrizioni disponibili in letteratura, possono essere sia complete (e corrispondenti alla sonorizzazione o desonorizzazione dell'intero segmento) o non complete (perché è stata sonorizzata o desonorizzata solo parte del segmento).

Nei soggetti italofofoni, in alcuni casi non è stato possibile individuare né l'assenza di assimilazione né la sua presenza perché si è osservata piuttosto la caduta della prima consonante del contesto di studio: questo è accaduto soprattutto in posizione finale al confine della parola e nel caso dell'occlusiva dentale sorda, nei contesti /td/, /tb/, /tg/. Nel secondo esperimento, la cancellazione dell'elemento finale è stata osservata anche quando il bersaglio era preceduto da una nasale (per i contesti /df/, /dt/, /gp/, /gt/ e /bk/). In casi simili, quindi, il secondo segmento consonantico è stato realizzato in posizione intervocalica, o tra nasale e vocale.

Nelle loro produzioni verbali, talvolta i soggetti hanno inserito una vocale epentetica (all'interno ma soprattutto al confine di parola) tra i due segmenti consonantici, seguendo una classica strategia volta ad evitare nessi consonantici complessi. In casi simili, anche quando si sia osservata la sonorizzazione del primo segmento consonantico, non è stato possibile considerare la sonorizzazione come dovuta al segmento consonantico seguente poiché non si può escludere che si sia trattato semplicemente di assimilazione dovuta alla vocale epentetica (benché il fenomeno non sia probabile data la varietà di italiano parlata dai soggetti).

In ogni caso, la presenza ed entità dell'assimilazione varia a seconda del contesto consonantico considerato e delle caratteristiche di focalizzazione. Per quanto riguarda i contesti consonantici non è stato possibile osservare sistematicità molto chiare. Tuttavia dai risultati emerge che i parlanti italofofoni, seppur con comportamenti linguistici differenti, effettuano assimilazioni anche in contesti molto diversi rispetto a quelli nei quali si riscontra assimilazione del tratto di sonorità in italiano. Tra i contesti simili all'italiano, si osserva che, almeno per due parlanti su tre, la fricativa alveolare sorda /s/ è più frequentemente assimilata (diventando sonora) all'interno di parola piuttosto che al confine di parola. Osservando questi risultati, sembrerebbe ragionevole pensare ad un'influenza della lingua materna, in quanto, in italiano, è facile trovare esempi di nessi /sb/, /sd/, /sv/, e /sg/ all'interno di parola (piuttosto che nessi come /ds/, /vs/ e altri) nei quali la fricativa è sonorizzata. Tuttavia, soprattutto considerando gli stimoli al confine di parola, si osservano fenomeni di assimilazione in contesti dissimili rispetto a quelli riscontrati nella lingua italiana: ad esempio, nel caso della desonorizzazione del segmento /v/, soprattutto quando è seguito dall'occlusiva bilabiale sorda, del segmento /g/ nel contesto /gt/ e dell'occlusiva bilabiale nei contesti /bs/, /bf/ e /bt/.

Circa l'interferenza della focalizzazione rispetto alla presenza di assimilazioni, emerge che il maggior numero di assimilazioni si riscontra nei casi di focalizzazione neutra e, comunque, nei casi in cui non interviene nessun confine prosodico. Il minor numero di assimilazioni, al contrario, si riscontra quando, all'interno di un sintagma, sia stato focalizzato il primo elemento piuttosto che il secondo, sia nelle produzioni dei parlanti italofoeni che nelle produzioni dei francofoeni. Questo risultato conferma l'ipotesi di Frascarelli [1997] circa la presenza, almeno in italiano, di un confine prosodico a destra dell'elemento focalizzato che blocca la realizzazione di fenomeni fonologici come quello dell'assimilazione. Al contrario, il comportamento dei francofoeni non è coerente con l'ipotesi avanzata da Féry [2003] circa la presenza di un confine prosodico a sinistra dell'elemento in focus (ipotesi che implicava l'attesa di interferenza rispetto all'assimilazione nel caso di focalizzazione sul secondo elemento del sintagma).

Aspetto peculiare del nostro studio è stato anche l'aver effettuato un'indagine acustica sulla base sia di segnale verbale (acquisito tramite microfono) che di segnale glottico (acquisito tramite elettrodi posti sull'epidermide della parete anteriore del collo, in corrispondenza delle laringe). Lo studio ci ha permesso di osservare l'attesa correlazione tra i dati acustici acquisiti, ma anche possibili differenze nelle misurazioni effettuate. In particolare, i segmenti consonantici individuati sulla base del segnale verbale risultano mediamente più lunghi di quelli individuati sulla base del segnale glottico: per segmenti sonori, ad esempio, questo può essere dovuto alle risonanze nella cavità orale; per segmenti sordi, invece, può essere imputato al (seppur minimo) ritardo della comparsa dell'informazione circa la sonorità nel segnale verbale per via del passaggio dell'onda di pressione nella cavità epiglottidale. In ogni caso, le differenze risultano maggiori nei contesti di assimilazione non completa; l'unica eccezione è rappresentata dalle produzioni degli italofoeni in contesto di focalizzazione sul primo termine. In questi contesti la differenza maggiore tra segnale verbale e glottico si riscontra in corrispondenza dell'assenza di assimilazione, probabilmente perché le caratteristiche prosodiche dell'elemento in focus, legate anche ad un incremento di intensità (pressione sub-glottidale), determinano la presenza di un segnale verbale più dettagliato nei termini di presenza/assenza di sonorità e quindi più simile al segnale glottico.

Per quanto riguarda la durata del primo segmento consonantico in relazione alla presenza di assimilazione regressiva, i risultati dell'analisi effettuata sul segnale verbale hanno evidenziato che, in posizione interna di parola, il segmento ha in media una durata maggiore nelle produzioni degli italofoeni piuttosto che dei francofoeni e, comunque, ha durata significativamente maggiore nel caso di assimilazione non completa piuttosto che nel caso di assimilazione completa o assenza di assimilazione. Nel caso dei contesti di assimilazione a confine di parola e focalizzazione neutra, i risultati mostrano nuovamente che il primo segmento consonantico ha durata maggiore nelle produzioni degli italofoeni, e, per entrambi i gruppi di parlanti, la durata è leggermente minore nel caso di assimilazione non completa rispetto alla realizzazione senza assimilazione. Per la focalizzazione sul primo elemento del sintagma, non si riscontra differenza significativa nella durata del primo segmento consonantico del contesto bersaglio (benché questo possa essere dovuto al numero limitato di osservazioni disponibili per questa condizione, visto che gli italofoeni effettuano pochissime assimilazioni regressive in questo contesto). Nel caso di focalizzazione sul secondo elemento del sintagma, al contrario, i risultati mostrano che i parlanti italofoeni producono consonanti di durata significativamente maggiore rispetto ai francofoeni; in ogni caso, rispetto alla presenza/assenza di assimilazione, la durata maggiore

corrisponde ai segmenti non assimilati o con assimilazione completa, mentre nel caso di assimilazione non completa la durata è inferiore.

## BIBLIOGRAFIA

Boersma, P. & Weenink, D. (2007), *Praat: doing phonetics by computer (Version 5.0)*, University of Amsterdam, The Netherlands. Software liberamente scaricabile su <http://www.praat.org/>.

De Dominicis, A. (1999), *Fonologia comparata delle principali lingue europee moderne*, CLUEB, Bologna, 143-154.

Féry, C. (2001), Gradient prosodic correlates of phrasing in French, to appear in Meisenburg, *Nouveaux departs en phonologie*, Niemeger, 161-182.

Féry, C. (2003), Focus and phrasing in French, *Audiatur Vox Sapientiae: A fest schrift for AV Stechow*, Berlin, 1-29.

Frascarelli, M. (1997), The phonology of Focus and Topic in Italian, *The Linguistic Review*, 14, 221-248.

Gendrot, C. (2003), EGG and spectral investigations on final focalised positions in French, in *Proceedings of the 15<sup>th</sup> ICPHS*, Barcelona, 547-550.

Giannini, A., Pettorino, M. (1992), *La fonetica sperimentale*, Edizioni Scientifiche Italiane, Napoli, 75-81.

Laver, J., (1994), *Principles of phonetics*, Cambridge University Press, Cambridge.

Mereu, L., Frascarelli, M., (2005), L'interfaccia sintassi-fonologia. Interpretazione e implicazioni teoriche, in *Atti AISV*, Fisciano 30 novembre-2 dicembre, 256-284.

Nespor, M., Vogel, I. (1986), *Prosodic Phonology*, Foris Publications, Dordrecht, The Netherlands.

Robert P., Rey A., Rey-Debove J., (2000), *Le Petit Robert: Dictionnaire de la langue française*, Le Robert, Paris.

# SVILUPPO FONETICO E ACQUISIZIONE LESSICALE: L'EMERGERE DELLA FONOLOGIA DELL'ITALIANO IN UNA PROSPETTIVA INTERLINGUISTICA

°Claudio Zmarich, #Debora Stocco, Sara Giulivi, \*Mario Vayra

°Istituto di Scienze e Tecnologie della Cognizione del C.N.R., Sede di Padova; #Centro Medico di Foniatria di Padova, \*Dipartimento di Studi Linguistici e Orientali, Università di Bologna

°[zmarich@pd.istc.cnr.it](mailto:zmarich@pd.istc.cnr.it), # [dstocco@centrofoniatra.it](mailto:dstocco@centrofoniatra.it), [saragiulivi@hotmail.it](mailto:saragiulivi@hotmail.it), \* [vym@dns.unife.it](mailto:vym@dns.unife.it)

## 1. SOMMARIO

Nelle fasi iniziali dello sviluppo del linguaggio, l'insorgere dell'influenza linguo-specifica e le modalità con cui essa si manifesta costituiscono un aspetto aperto alla verifica sperimentale, in prospettiva interlinguistica. Infatti si può affermare che si è in presenza di influenze linguo-specifiche solo se le produzioni di babbling o delle prime parole di gruppi di bambini provenienti da ambienti linguistici diversi sono separate da differenze maggiori di quelle riscontrabili all'interno di ogni singolo gruppo, e se tali differenze riflettono i *patterns* caratteristici di ciascuna lingua di afferenza.

Il presente studio, che fa parte di un progetto di ricerca più ampio incentrato sulle produzioni linguistiche di bambini italiani dai 6 ai 27 mesi di vita, ri-analizza i dati sulla frequenza relativa dei tipi sillabici e dei foni consonantici calcolata dal primo autore in studi precedenti su un gruppo di 4 bambini (due maschi e due femmine, esaminati a 10, 12, 14, 16 e 18 mesi di vita, cfr. Zmarich e Miotti, 2003). In particolare, mentre negli studi precedenti i dati fonetici erano stati descritti in relazione all'età cronologica, nel presente studio sono messi in relazione a quattro stadi di produzione lessicale, riconosciuti in letteratura: lo stadio delle 0 parole (da 0 a 3 parole), quello delle 4 parole (da 4 a 14), delle 15 parole (da 15 a 24), e delle 25 parole (da 25 a 50), integrati con i dati di una bambina analizzata in relazione agli stadi lessicali da Giulivi, Zmarich, Vayra e Farnetani (2006). Al fine di identificare tempi e modalità del processo di "sintonizzazione" sulla fonologia dell'italiano adulto, i risultati sono stati confrontati sia con i dati relativi alle frequenze di occorrenza delle caratteristiche fonetiche del lessico italiano infantile e adulto, sia con i dati relativi alle frequenze di occorrenza dei segmenti, ordinati in riferimento agli stadi di parola, prodotti da quattro gruppi di bambini coetanei di quelli del presente studio e provenienti da comunità linguistiche diverse (francese, giapponese, inglese, svedese, cfr. Vihman, 1993).

I risultati preliminari qui presentati consentono di affermare che nello stadio finale delle 25 parole, le differenze fonetiche fra i gruppi linguistici nazionali sono maggiori delle differenze all'interno di ogni singolo gruppo, e nella maggior parte dei casi riflettono i *patterns* fonetico-fonologici caratteristici di ciascuna lingua.

## 2. INTRODUZIONE

È intuizione comune che alcuni aspetti fonetici del *babbling* e delle prime parole siano universali, cioè si trovino in tutti i bambini, a prescindere da diversità di razza e lingua, mentre altri tendano ad essere specifici, cioè presenti solo nei bambini di una data comunità linguistica, e che le specificità aumentino con la crescita del bambino. Tenendo presente che gli aspetti fonetici consistono nel tipo di foni e nel tipo di combinazione di questi foni in strutture fonotattiche minime, le sillabe, si è molte volte osservato, a partire da Locke



(1983), che i fonemi di gran lunga più frequenti sono le consonanti occlusive e le vocali medie, basse anteriori e centrali, e che questi due tipi di suoni si combinano prevalentemente in sillabe del tipo CV. Secondo alcuni autori (cfr. Davis e MacNeilage, 1995), questi aspetti fonetici universali sarebbero dovuti alle proprietà del sistema di produzione del parlato, poiché la tipologia dei fonemi e delle loro combinazioni deriverebbe dall'oscillazione verticale della sola mandibola durante la fonazione. Le particolarità linguospecifiche deriverebbero invece dall'apprendimento, da parte del bambino, di fonemi e combinazioni foniche salienti per l'udito e anche per la vista (nel caso delle consonanti bilabiali).

Alcune considerazioni di metodo. Perché si possa sostenere l'esistenza di un' influenza linguo-specifica, bisogna dimostrare : 1) che le differenze fonetiche fra i gruppi linguistici sono maggiori delle differenze presenti all'interno di ogni singolo gruppo; 2) che queste differenze riflettono i *patterns* fonetico-fonologici caratteristici di ciascuna lingua (Boysson-Bardies et al., 1992).

L'influenza delle caratteristiche fonetiche linguo-specifiche può essere indagata a livello di percezione (cfr. per esempio Best, 2002) e a livello di produzione, tramite analisi acustiche (Boysson-Bardies, Hallé, Sagart & Durand, 1989; Macken & Barton, 1980), cinematiche, e basate sulla trascrizione fonetica (cfr. Kent & Miolo, 1995). Poiché il presente studio si basa sulla trascrizione fonetica delle vocalizzazioni del bambino, la discussione sarà necessariamente limitata ai risultati ottenuti con tale metodo.

Inoltre, l'estensione dell'arco temporale che un'analisi di questo tipo deve considerare per cogliere i primi segni di un'influenza linguo-specifica e documentarne poi la progressione, ha condizionato il metodo qui adottato. Il periodo preso in esame va dai 6-8 mesi, età in cui tipicamente compaiono le prime produzioni di *babbling*, ai 16-18 mesi, età in cui viene prodotto un vocabolario di 25-50 parole. Per questo motivo l'analisi della produzione fonetica non può basarsi sullo studio degli inventari fonetici. Da Stoel-Gammon (1985) in poi essi vengono infatti calcolati in base alla presenza dello stesso *tipo* di fonema in almeno due parole diverse all'interno di un vocabolario che va da un minimo di 10 ad un massimo di 50 parole diverse. Tuttavia, l'assenza nel periodo iniziale di parole intese come associazioni stabili forma-significato impedisce di distinguere, in due produzioni uguali per forma fonica, la ripetizione di due occorrenze (*tokens*) di uno stesso tipo (*type*) dalla singola occorrenza di due tipi diversi. Nel nostro caso, l'unico modo per poter confrontare la produzione vocale lessicale e pre-lessicale consiste quindi nell'esaminare non i tipi lessicali, ma le occorrenze.

C'è inoltre da considerare il problema rappresentato dalla grande variabilità interlinguistica e intersoggettiva, che caratterizza lo sviluppo fonetico-fonologico del bambino. Si rinvia al riguardo a una serie di studi di Boysson-Bardies & Vihman (1991), Boysson-Bardies, Vihman, Roug-Hellichius, Durand, Landberg, Arao (1992), Vihman (1993), Vihman & Boysson-Bardies (1994), volti a far luce sui tempi e le modalità di insorgenza delle influenze linguo-specifiche nei primi stadi del linguaggio infantile. I risultati evidenziano come soggetti provenienti da ambienti linguistici differenti (francese, inglese, giapponese e svedese) raggiungano una determinata soglia nello sviluppo lessicale in età diverse. Da questi dati risulta evidente come una metodologia di analisi che consenta di mettere a confronto le produzioni di bambini di età cronologica differente, sulla base dell'appartenenza allo stesso stadio di sviluppo lessicale, permetta di ridurre al minimo gli effetti di quella grande variabilità intersoggettiva e interlinguistica che caratterizza il naturale sviluppo fonetico.

E' poi possibile affermare che il numero dei foni nativi aumenta al crescere non tanto dell'età, quanto del vocabolario. Infatti, studi diretti (per l'italiano cfr. Giulivi, Zmarich, Farnetani e Vayra, 2006) e indiretti (per l'italiano cfr. Zmarich et al., in questo volume) hanno trovato che a parità di età, il bambino che possiede un vocabolario più ampio produce anche un maggior numero di foni della lingua target.

Nel presente studio presenteremo i risultati relativi alle frequenze dei diversi tipi sillabici e segmentali in rapporto alla posizione nella parola, non in funzione dell'età, bensì in funzione dell'evoluzione lessicale del bambino. Come tappe significative dell'evoluzione lessicale prenderemo a riferimento gli stadi proposti da Boysson-Bardies & Vihman (1991): gli stadi cosiddetti delle 0 parole (da 0 a 3 parole), delle 4 parole (da 4 a 14), delle 15 parole (da 15 a 24) e delle 25 parole (da 25 a 50).

Le conclusioni a cui giungono questi autori sottolineano l'uso prevalente da parte di bambini inglesi, francesi, svedesi e giapponesi di consonanti occlusive, specialmente labiali e dentali/alveolari, a scapito delle altre categorie di foni; ciò però non esclude che ci siano differenze statisticamente significative tra i gruppi, già nella fase del *babbling*, dovute all'influenza linguospecifica. Il confronto ulteriore tra le forme foniche del *babbling* e delle prime parole rivela delle differenze che riflettono le conseguenze sul piano motorio delle restrizioni sulle sequenze lineari presenti nelle prime parole. Vihman & Boysson-Bardies (1994) riescono a precisare un'influenza positiva della lingua nativa sulle frequenze di occorrenza dei foni nel *babbling* già a 9-10 mesi, allorché i foni nativi aumentano, e un'influenza negativa a partire dai 12 mesi circa, quando i foni non nativi diminuiscono.

Gli unici studi per l'italiano di cui siamo a conoscenza, e che descrivono i risultati in base agli stadi lessicali, sono Bortolini (1993) e Giulivi, Zmarich, Vayra e Farnetani (2006).

Bortolini (1993) prende in esame 4 soggetti audio- e video-registrati ogni due settimane dai 10 ai 18 mesi di vita. I dati mostrano che il modo di articolazione prevalente è quello oclusivo. Le consonanti occlusive occorrono in generale con frequenza maggiore e tendono ad aumentare progressivamente dal primo all'ultimo stadio. Seguono le nasali, che registrano una crescita sensibile soprattutto negli ultimi due stadi. Riguardo al luogo di articolazione, è possibile osservare una prevalenza di labiali e dentali nel corso dei quattro stadi. Tuttavia, mentre le labiali decrescono progressivamente, le dentali tendono ad aumentare. In generale, i dati mostrano una forte correlazione tra il repertorio dei suoni presenti nel *babbling* e quello dei suoni presenti nelle prime parole.

In Giulivi et al. (2006), il soggetto indagato è una bambina audioregistrata e trascritta foneticamente a 6, 8, 10, 12, 14 e 16 mesi di età. Per questa bambina il tipo sillabico prevalente è il CV. Le sillabe di tipo CVC tendono ad aumentare progressivamente, mentre sillabe più complesse (come CCVC o CCCV) sono quasi del tutto assenti. Riguardo al modo di articolazione, le occlusive nasali sono inizialmente prevalenti (in posizione iniziale di sillaba), poi diminuiscono a favore delle occlusive di tipo orale. Riguardo al luogo di articolazione, occorrono con maggior frequenza (sempre in posizione iniziale di sillaba) le consonanti alveolari e bilabiali, che registrano rispettivamente una diminuzione e un aumento nel corso dei tre stadi lessicali considerati (0 parole, 4 parole e 15 parole).

Il presente studio ri-analizza i dati sulla frequenza relativa dei tipi sillabici e dei foni consonantici, raccolti dal primo autore (cfr. Zmarich e Miotti, 2003; Zmarich, Stocco, Minozzi, Bonifacio, 2005) in studi precedenti su un gruppo di 4 bambini (due maschi e due femmine, esaminati a 10, 12, 14, 16 e 18 mesi di vita). In particolare, mentre in questi ultimi studi i dati fonetici erano stati descritti in relazione all'età cronologica, in questo studio saranno messi in relazione ai quattro stadi di produzione lessicale prima descritti, e

saranno integrati con i dati di una bambina già analizzata in relazione agli stadi lessicali da Giulivi et al. (2006), qui arricchiti con l'analisi ex-novo della registrazione del 18° mese.

Al fine di identificare i tempi e le modalità del processo di sintonizzazione sulla fonologia dell'italiano adulto, i risultati saranno confrontanti sia con i dati relativi alle frequenze di occorrenza delle caratteristiche fonetiche del lessico infantile quale è rappresentato nel PVB (Caselli e Casadio, 1995; cfr. Zmarich, Dispaldro, Rinaldi, Caselli in questo volume), sia con i dati relativi alle frequenze di occorrenza dei segmenti, ordinati in riferimento agli stadi di parola, prodotti da quattro gruppi di bambini coetanei di quelli del presente studio e provenienti da comunità di lingua madre inglese, francese, svedese e giapponese (cinque bambini per gruppo) (Boysson-Bardies et al., 1992).

Questi dati sono stati pubblicati come valori delle medie e delle deviazioni standard (DS) del numero di fonemi, classificati per modo e luogo, prodotti o tentati (*target*) dai gruppi infantili nazionali (Boysson-Bardies & Vihman, 1991), o dai singoli bambini (Vihman, 1993). I valori dei *target* per queste lingue sono relativi ai fonemi presenti nelle parole adulte tentate dai bambini (Boysson-Bardies & Vihman, 1991).

### 3. METODO

#### 3.1 Soggetti

Il campione di bambini italiani è costituito da cinque soggetti dei quali si è osservato lo sviluppo fonetico al 10°, 12°, 14°, 16°, 18° mese d'età. Questi soggetti fanno parte di un database raccolto dal primo autore. Dei cinque soggetti, quattro sono già stati studiati da Zmarich e Miotti (2003) e da Zmarich, Minozzi, Stocco e Bonifacio (2005), ed il quinto da Giulivi, Zmarich, Vayra e Farnetani (2006), ma solo fino al 16° mese. I soggetti appartengono a una classe sociale di tipo medio e la loro partecipazione allo studio è stata a titolo gratuito. Si tratta di quattro femmine e un maschio, nati a termine, che hanno avuto uno sviluppo neuromotorio, cognitivo e linguistico nella norma rispetto all'età, e che sono esenti da patologie a carico dell'apparato uditivo e pneumo-fono-articolatorio.

I soggetti di Boysson-Bardies e Vihman (1991), a cui si rimanda per i dettagli, sono costituiti da 4 gruppi di 5 soggetti ciascuno, di madrelingua francese, inglese, giapponese e svedese, e sono stati registrati dai 9 mesi circa fino alla sessione di registrazione in cui hanno prodotto un vocabolario di almeno 25 parole (max. 17 mesi).

#### 3.2 Registrazione

I soggetti italiani sono stati registrati a partire da età diverse, e la prima tappa in cui è possibile considerarli tutti insieme è quella del 9°/10° mese. Le modalità di raccolta dei dati sono state descritte in Zmarich e Miotti (2003), a cui rinviamo per i dettagli; segnaliamo solamente che le sedute di registrazione, della durata di circa un'ora, sono state effettuate con frequenza bisettimanale e sono state eseguite dai familiari dei bambini, che erano stati dotati degli strumenti per la registrazione e delle istruzioni necessarie. Due dei bambini sono stati registrati con un registratore *Sony* di tipo *DAT* e un microfono *Sony* modello *ECM-T7*, mentre altri due bambini sono stati registrati con un registratore analogico semiprofessionale *AIWA* modello *TP-850* e un microfono *Sony* modello *ECM-T7*.

|                 | 10 mesi    | 12 mesi | 14 mesi | 16 mesi | 18 mesi |
|-----------------|------------|---------|---------|---------|---------|
| <b>Caterina</b> | 9;28       | 11;18   | 13;17   | 15;10   | 17;17   |
| <b>Giorgia</b>  | 9;05       | 11;12   | 13;14   | 15;14   | 17;24   |
| <b>Alessia</b>  | 9;08       | 11;22   | 13;21   | 15;19   | 18;10   |
| <b>Davide</b>   | 9;09+9, 28 | 11;21   | 13;07   | 15;16   | 17;23   |
| <b>*Arianna</b> | 8;09       | 10;19   | 12;09   | 14;19   | 18;10   |

Tab. 1. Età dei bambini in mesi e giorni a seconda delle tappe. \* Arianna è stata registrata una prima volta a 6;15

I soggetti di Boysson-Bardies e Vihman (1991), a cui si rimanda per i dettagli, sono stati registrati in 6 sessioni della durata di mezz'ora a partire dai 9-10 mesi di vita, fino all'età in cui venivano prodotte 25 parole per sessione.

|               | <i>Stadi lessicali<br/>(n. parole)</i> | 0     | 4     | 15    | 25    |
|---------------|----------------------------------------|-------|-------|-------|-------|
| <i>Lingua</i> |                                        |       |       |       |       |
| francese      |                                        | 10    | 12    | 15    | 16-17 |
| inglese       |                                        | 9-10  | 11-12 | 14    | 16    |
| giapponese    |                                        | 13    | 14    | 17-18 | 19    |
| svedese       |                                        | 9     | 11-12 | 15-16 | 16-17 |
| *italiano     |                                        | 10-12 | 12-14 | 14-16 | 18    |

Tab. 2. Età media dei soggetti in mesi alle 4 tappe principali. \*Vedi sotto

Ci sembra giusto sottolineare una differenza importante nel modo in cui sono stati raccolti i dati dei bambini italiani rispetto a quelli di Boysson-Bardies e Vihman (1991), visto che comporta una diretta conseguenza sulle età considerate. I dati dei soggetti italiani sono stati raccolti a tappe di età prefissate (appunto ogni 2 mesi a partire al massimo dai 10 fino ai 18) mentre quelle di Boysson-Bardies e Vihman (1991) partono dai 9 mesi e includono la prima registrazione in cui i soggetti producevano il numero minimo di parole necessarie a stabilire il raggiungimento di un determinato stadio lessicale e cioè 4, 15 e 25 parole.

### 3.3 Indice di concordanza

Per i bambini italiani, l'attendibilità delle trascrizioni è stata controllata attraverso due diverse procedure. I trascrittori di uno dei soggetti (Davide) hanno trascritto in modo congiunto circa il 25% dell'intero prodotto verbale del bambino, discutendo i dubbi e giungendo ad una trascrizione consensuale, con un accordo percentuale del 98,3% (calcolato con la formula *point by point*, Shriberg & Lof, 1991) sui simboli IPA. Per gli altri soggetti, due trascrittori hanno eseguito in modo indipendente la trascrizione di 10 enunciati consecutivi di ciascuna registrazione, e alla fine è risultato un accordo del 62,7% su tutti i simboli IPA.

Dal momento che nello studio di Davis & MacNeilage (1995), come d'altronde nella grande maggioranza delle ricerche basate sulla trascrizione fonetica del percolato uditivo del parlato infantile, l'accordo era del 76,8% (63%-83%) per le consonanti, e del 44,8% per le vocali (33%-69%), e che la stima sul tasso medio di accordo inter- e intra-trascrittore per un

tipo di trascrizione stretta (*narrow*) eseguite da Shriberg & Lof (1991) risulta essere del 74% circa, valutiamo le nostre percentuali di accordo come sufficienti.

### 3.4 Criteri di selezione e analisi

Nel presente studio sono stati adottati i seguenti criteri di selezione e analisi delle produzioni dei soggetti presi in esame:

- Sono state escluse tutte le produzioni disturbate da rumore, oltre ai suoni di tipo riflesso e vegetativo, e a quelli classificabili come appartenenti allo stadio *pre-babbling* del *vocal play*. Sono state inoltre eliminate le vocalizzazioni prodotte con vocalizzazione non modale, cioè eccessivamente laringalizzate o in falsetto.
- Sono state accettate le sillabe di tipo CV isolate. Sono state considerate come appartenenti a enunciati (cioè episodi di babbling o singole parole) differenti due sillabe successive che fossero separate da più di 250 ms.
- Per quanto riguarda le produzioni di *babbling* sono state esaminate tutte le produzioni non escluse in base alle motivazioni sopra elencate; per quanto riguarda i foni delle prime parole sono stati considerati solo quelli prodotti all'interno di parole che potessero essere identificate come tali in base ai criteri esposti in Vihman & McCune (1994):
  - *la somiglianza fonetica con la forma adulta*, cioè la presenza di almeno due suoni, preferibilmente consonantici, che siano in comune alla forma adulta e a quella infantile
  - *il contesto d'uso* di una determinata produzione e quindi la sua *funzione comunicativa*
  - *l'identificazione, da parte del genitore*, delle produzioni linguistiche del bambino, cioè il riconoscimento che determinate forme fonetiche vengono coerentemente associate a determinati significati
- Sono state analizzate solamente le denominazioni e le produzioni spontanee, mentre le ripetizioni immediate (cioè successive alla produzione adulta) sono state escluse, così come le onomatopее (ma solo se elicitate dall'adulto, cioè solo se prodotte in risposta a domande del tipo: "Come fa X?")

Rimandiamo a Boysson-Bardies e Vihman (1991) per i dettagli relativi ai criteri di trascrizione e selezione delle produzioni ivi esaminate, peraltro simili ai nostri.

### 3.5 Codifica

Ogni parola è stata suddivisa in sillabe, sulla base della gerarchia di sonorità e dei principi di sillabificazione (Nespor, 1993; Blevins, 1995). I foni [j, w] sono stati codificati come consonanti in base alla considerazione che non costituiscono apice di sillaba. Il fono [s] in posizione iniziale di nesso consonantico intervocalico è stato attribuito alla sillaba successiva (per es.: 'a.sta' e non 'as.ta'). Sono state codificate, come casi di una matrice di *Systat*, le sillabe della produzione infantile i cui foni hanno occupato colonne separate; inoltre sono stati codificati, per ogni produzione, il numero progressivo delle sillabe, il numero corrispondente alla numerazione delle sillabe nella trascrizione, il numero progressivo di appartenenza di una data sillaba a un dato enunciato, la posizione della sillaba nell'enunciato (iniziale, mediana, finale), il *target* corrispettivo, il tipo sillabico. Si è usato il programma statistico *Systat* per codificare i foni e per attribuire ad ognuno la classe articolatoria appropriata, secondo i principi IPA di classificazione delle consonanti in termini di modo e luogo di articolazione.

### 3.6 Misure quantitative

In modo analogo a Boysson-Bardies e Vihman (1991), per ciascun soggetto italiano sono stati eseguiti, ad ogni tappa, conteggi di frequenza delle consonanti e dei tipi sillabici. Sono state contate e sommate tutte le consonanti presenti nelle diverse posizioni di un episodio di *babbling* o di una parola, fatta eccezione per la posizione finale. Sono state considerate, per la presente analisi, tutte le consonanti, fatta eccezione per le approssimanti, [h] e occlusive glottidali. In modo analogo a Boysson-Bardies e Vihman (1991), le consonanti sono state classificate per modo e luogo di articolazione. Per quanto riguarda il luogo, le consonanti sono state classificate in tre categorie: labiali (comprese le labiodentali), dentali (comprese alveolari, palatali e retroflesse), e velari (comprese le uvulari). Per il modo di articolazione sono state considerate solo cinque categorie: occlusive, affricate, fricative, nasali e liquide. Affricate e fricative sono state in seguito raggruppate, allo scopo di eliminare la categoria delle affricate, per le quali si è registrata una frequenza inferiore al 4% in tutte e cinque le lingue.

## 4. RISULTATI

Per ciascuno dei soggetti italiani abbiamo calcolato la distribuzione % delle consonanti non finali nelle categorie di luogo (tab. 3) e modo (tab. 4), e li abbiamo uniti ai dati individuali presentati in Vihman (1993), allo stadio iniziale delle “0 parole” e allo stadio finale delle “25 parole”.

Nonostante la notevole differenza tra i soggetti all'interno dei gruppi, si può osservare la comune tendenza alla riduzione della variabilità, misurata dalla deviazione standard (DS) con la produzione delle prime parole, particolarmente per le categorie più favorite nel *babbling* (labiali, dentali, occlusive orali e nasali).

Per testare l'ipotesi che la provenienza linguistica influenzi la distribuzione delle consonanti nelle categorie di modo e luogo delle produzioni infantili, così riflettendo le tendenze principali trovate nelle parole *target* di ciascun gruppo, è stata eseguita un'analisi della varianza sulla distribuzione delle consonanti nelle produzioni dello stadio delle 0 parole, cioè quando si presume che i bambini esibiscano più caratteristiche fonetiche di tipo universale che linguo-specifico. E' stato considerato come fattore *between-subjects* il gruppo linguistico (italiano, francese, inglese, giapponese e svedese) e come fattore *within-subjects* le categorie di luogo (labiale, dentale, velare). Lo stesso tipo di analisi è stato eseguito per le categorie di modo. I due tipi di analisi sono stati poi applicati anche alle produzioni dello stadio delle 25 parole, quando si presume che i bambini esibiscano un minor numero di caratteristiche di tipo universale che linguo-specifico. Se questa ipotesi fosse corretta, l'analisi statistica dovrebbe confermare che le differenze fonetiche tra i gruppi nazionali, che emergono dall'interazione dei fattori “lingua” e “categoria fonetica”, sono maggiori delle differenze all'interno dei gruppi.

|              |            | <i>LABIALI</i> |              | <i>DENTALI</i> |              | <i>VELARI</i> |              | <i>Freq.Ass.Tot</i> |           |
|--------------|------------|----------------|--------------|----------------|--------------|---------------|--------------|---------------------|-----------|
|              |            | <i>0</i>       | <i>25</i>    | <i>0</i>       | <i>25</i>    | <i>0</i>      | <i>25</i>    | <i>0</i>            | <i>25</i> |
| Caterina     | italiano   | 38             | 27           | 47             | 40           | 15            | 33           | 387                 | 451       |
| Arianna      | italiano   | 57             | 47           | 32             | 44           | 12            | 9            | 1015                | 825       |
| Giorgia      | italiano   | 30             | 21           | 62             | 65           | 8             | 14           | 204                 | 1153      |
| Alessia      | italiano   | 14             | 34           | 81             | 64           | 5             | 3            | 397                 | 439       |
| Davide       | italiano   | 34             | 23           | 54             | 53           | 13            | 24           | 280                 | 479       |
| <b>MEDIA</b> |            | <b>34</b>      | <b>30</b>    | <b>55</b>      | <b>53</b>    | <b>10</b>     | <b>16</b>    |                     |           |
| <b>D.S.</b>  |            | <b>15.35</b>   | <b>10.65</b> | <b>18.45</b>   | <b>11.28</b> | <b>4.12</b>   | <b>12.21</b> |                     |           |
| Deborah      | inglese    | 30             | 46           | 62             | 39           | 8             | 15           | 218                 | 167       |
| Emily        | inglese    | 90             | 33           | 0              | 47           | 10            | 20           | 20                  | 202       |
| Molly        | inglese    | 62             | 57           | 15             | 28           | 23            | 15           | 47                  | 500       |
| Sean         | inglese    | 37             | 36           | 57             | 54           | 6             | 10           | 98                  | 203       |
| Timmy        | inglese    | 63             | 44           | 25             | 31           | 13            | 25           | 94                  | 249       |
| <b>MEDIA</b> |            | <b>56</b>      | <b>43</b>    | <b>32</b>      | <b>40</b>    | <b>12</b>     | <b>17</b>    |                     |           |
| <b>D.S.</b>  |            | <b>23.9</b>    | <b>9.4</b>   | <b>26.9</b>    | <b>10.8</b>  | <b>6.7</b>    | <b>5.7</b>   |                     |           |
| Carole       | francese   | 51             | 56           | 22             | 23           | 27            | 21           | 109                 | 222       |
| Charles      | francese   | 31             | 53           | 58             | 44           | 11            | 3            | 75                  | 253       |
| Laurent      | francese   | 24             | 45           | 71             | 39           | 3             | 16           | 34                  | 374       |
| Marie        | francese   | 66             | 44           | 34             | 54           | 1             | 3            | 271                 | 758       |
| Noel         | francese   | 87             | 63           | 7              | 36           | 6             | 1            | 68                  | 329       |
| <b>MEDIA</b> |            | <b>52</b>      | <b>52</b>    | <b>39</b>      | <b>39</b>    | <b>10</b>     | <b>9</b>     |                     |           |
| <b>D.S.</b>  |            | <b>25.7</b>    | <b>7.9</b>   | <b>26.1</b>    | <b>11.3</b>  | <b>10.4</b>   | <b>9.1</b>   |                     |           |
| Emi          | giapponese | 70             |              | 22             |              | 9             |              | 23                  |           |
| Haruo        | giapponese | 61             | 13           | 31             | 55           | 7             | 31           | 163                 | 531       |
| Kazuko       | giapponese | 6              | 46           | 44             | 29           | 50            | 25           | 18                  | 686       |
| Kenji        | giapponese | 19             | 24           | 59             | 47           | 23            | 29           | 263                 | 125       |
| Taro         | giapponese | 33             | 14           | 45             | 55           | 22            | 31           | 82                  | 212       |
| <b>MEDIA</b> |            | <b>38</b>      | <b>24</b>    | <b>40</b>      | <b>47</b>    | <b>22</b>     | <b>29</b>    |                     |           |
| <b>D.S.</b>  |            | <b>27.2</b>    | <b>15.3</b>  | <b>14.2</b>    | <b>12.3</b>  | <b>17.2</b>   | <b>2.8</b>   |                     |           |
| Didrik       | svedese    | 65             | 26           | 12             | 51           | 22            | 24           | 49                  | 333       |
| Hanna        | svedese    | 70             | 7            | 20             | 70           | 10            | 23           | 10                  | 404       |
| Kurt         | svedese    |                | 53           |                | 23           |               | 24           |                     | 135       |
| Lina         | svedese    | 22             | 20           | 47             | 57           | 31            | 22           | 363                 | 393       |
| Stig         | svedese    | 10             | 18           | 39             | 59           | 52            | 22           | 211                 | 274       |
| <b>MEDIA</b> |            | <b>42</b>      | <b>25</b>    | <b>30</b>      | <b>52</b>    | <b>29</b>     | <b>23</b>    |                     |           |
| <b>D.S.</b>  |            | <b>30.2</b>    | <b>17.2</b>  | <b>16.3</b>    | <b>17.6</b>  | <b>17.7</b>   | <b>1.0</b>   |                     |           |

Tab. 3. Distribuzione delle frequenze percentuali per le consonanti non finali nelle categorie di luogo nello stadio delle 0 parole e nello stadio delle 25 parole

|              |            | OCCLUSIVE   |             | FRICATIVE   |             | NASALI      |             | LIQUIDE     |             | Freq.Ass.Tot |      |
|--------------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|------|
|              |            | 0           | 25          | 0           | 25          | 0           | 25          | 0           | 25          | 0            | 25   |
|              |            |             |             |             |             |             |             |             |             |              |      |
| Caterina     | italiano   | 63          | 68          | 25          | 8           | 11          | 14          | 1           | 10          | 388          | 440  |
| Arianna      | italiano   | 31          | 54          | 4           | 2           | 56          | 39          | 8           | 5           | 1002         | 828  |
| Giorgia      | italiano   | 25          | 55          | 12          | 7           | 53          | 13          | 10          | 26          | 187          | 1142 |
| Alessia      | italiano   | 50          | 46          | 8           | 3           | 37          | 28          | 5           | 22          | 391          | 427  |
| Davide       | italiano   | 58          | 60          | 16          | 8           | 21          | 31          | 5           | 0           | 277          | 474  |
| <b>MEDIA</b> |            | <b>45</b>   | <b>57</b>   | <b>13</b>   | <b>5</b>    | <b>36</b>   | <b>25</b>   | <b>6</b>    | <b>13</b>   |              |      |
| <b>D.S.</b>  |            | <b>16.6</b> | <b>8.1</b>  | <b>7.8</b>  | <b>2.9</b>  | <b>19.6</b> | <b>11.5</b> | <b>3.4</b>  | <b>10.9</b> |              |      |
|              |            |             |             |             |             |             |             |             |             |              |      |
| Deborah      | inglese    | 75          | 62          | 5           | 25          | 19          | 11          | 1           | 1           | 211          | 169  |
| Emily        | inglese    | 43          | 62          | 0           | 8           | 52          | 28          | 5           | 1           | 21           | 205  |
| Molly        | inglese    | 86          | 80          | 2           | 1           | 8           | 18          | 4           | 2           | 49           | 508  |
| Sean         | inglese    | 55          | 63          | 24          | 8           | 2           | 17          | 19          | 13          | 123          | 232  |
| Timmy        | inglese    | 78          | 78          | 16          | 4           | 2           | 13          | 4           | 6           | 98           | 264  |
| <b>MEDIA</b> |            | <b>67</b>   | <b>69</b>   | <b>9</b>    | <b>9</b>    | <b>17</b>   | <b>17</b>   | <b>7</b>    | <b>5</b>    |              |      |
| <b>D.S.</b>  |            | <b>17.8</b> | <b>9.2</b>  | <b>10.2</b> | <b>9.3</b>  | <b>21.0</b> | <b>6.6</b>  | <b>7.1</b>  | <b>5.1</b>  |              |      |
|              |            |             |             |             |             |             |             |             |             |              |      |
| Carole       | francese   | 42          | 64          | 5           | 6           | 37          | 29          | 16          | 0           | 130          | 223  |
| Charles      | francese   | 36          | 46          | 12          | 30          | 40          | 19          | 13          | 5           | 86           | 267  |
| Laurent      | francese   | 46          | 44          | 5           | 6           | 27          | 26          | 23          | 24          | 44           | 499  |
| Marie        | francese   | 71          | 34          | 7           | 16          | 17          | 26          | 6           | 25          | 289          | 1005 |
| Noel         | francese   | 30          | 64          | 9           | 7           | 59          | 24          | 3           | 5           | 70           | 347  |
| <b>MEDIA</b> |            | <b>45</b>   | <b>50</b>   | <b>8</b>    | <b>13</b>   | <b>36</b>   | <b>25</b>   | <b>12</b>   | <b>12</b>   |              |      |
| <b>D.S.</b>  |            | <b>15.7</b> | <b>13.2</b> | <b>3.0</b>  | <b>10.4</b> | <b>15.7</b> | <b>3.7</b>  | <b>11.8</b> | <b>11.8</b> |              |      |
|              |            |             |             |             |             |             |             |             |             |              |      |
| Emi          | giapponese | 61          |             | 22          |             | 17          |             | 0           |             | 23           |      |
| Haruo        | giapponese | 59          | 59          | 18          | 8           | 22          | 25          | 1           | 7           | 165          | 574  |
| Kazuko       | giapponese | 47          | 59          | 0           | 11          | 47          | 25          | 5           | 5           | 19           | 719  |
| Kenji        | giapponese | 71          | 60          | 9           | 11          | 18          | 13          | 3           | 17          | 271          | 150  |
| Taro         | giapponese | 55          | 67          | 14          | 10          | 29          | 21          | 1           | 2           | 83           | 216  |
| <b>MEDIA</b> |            | <b>59</b>   | <b>61</b>   | <b>13</b>   | <b>10</b>   | <b>27</b>   | <b>21</b>   | <b>2</b>    | <b>8</b>    |              |      |
| <b>D.S.</b>  |            | <b>8.8</b>  | <b>3.9</b>  | <b>8.5</b>  | <b>1.4</b>  | <b>5.7</b>  | <b>5.7</b>  | <b>2.0</b>  | <b>6.5</b>  |              |      |
|              |            |             |             |             |             |             |             |             |             |              |      |
| Didrik       | svedese    | 33          | 61          | 10          | 4           | 21          | 5           | 6           | 29          | 52           | 469  |
| Hanna        | svedese    | 90          | 92          | 0           | 1           | 10          | 7           | 0           | 0           | 10           | 404  |
| Kurt         | svedese    |             | 73          |             | 5           |             | 20          |             | 2           |              | 137  |
| Lina         | svedese    | 64          | 72          | 1           | 8           | 21          | 5           | 14          | 15          | 420          | 462  |
| Stig         | svedese    | 86          | 66          | 4           | 8           | 1           | 11          | 9           | 15          | 330          | 324  |
| <b>MEDIA</b> |            | <b>68</b>   | <b>73</b>   | <b>11</b>   | <b>5</b>    | <b>13</b>   | <b>10</b>   | <b>7</b>    | <b>12</b>   |              |      |
| <b>D.S.</b>  |            | <b>26.1</b> | <b>11.8</b> | <b>19.2</b> | <b>2.9</b>  | <b>9.7</b>  | <b>6.3</b>  | <b>5.9</b>  | <b>11.7</b> |              |      |

Tab.4. Distribuzione delle frequenze % per le consonanti non finali nelle categorie di modo nello stadio delle 0 parole e nello stadio delle 25 parole.

Per quanto riguarda lo stadio delle 0 parole, l'ANOVA mostra che l'effetto principale del luogo di articolazione è significativo:  $F(2,38) = 8.61$ ,  $p = 0.001$ . Una serie di t-test evidenzia che le consonanti labiali contrastano significativamente con le velari,  $t(23) =$



4.31,  $p = 0.000$ , e le dentali con le velari  $t(23) = 4.35$ ,  $p = 0.000$ . L'interazione tra luogo e gruppo linguistico non è però significativa.

Sempre in relazione allo stadio delle 0 parole, per quanto riguarda l'effetto principale dell'ANOVA relativo al modo di articolazione, esso è significativo:  $F(3,57) = 51.42$ ,  $p = 0.000$ . Una serie di t-test evidenzia che le consonanti occlusive orali contrastano significativamente con le fricative,  $t(23) = 9.65$ ,  $p = 0.000$ , con le nasali,  $t(23) = 15.52$ ,  $p = 0.000$  e con le liquide,  $t(23) = 11.16$ ,  $p = 0.000$ , che le consonanti nasali contrastano significativamente con le fricative  $t(23) = 3.33$ ,  $p = 0.003$ , e le nasali con le liquide,  $t(23) = 5.16$ ,  $p = 0.000$ . Anche in questo caso l'interazione tra luogo e gruppo linguistico non è significativa.

Per quanto riguarda lo stadio delle 25 parole, l'ANOVA mostra che l'effetto principale del luogo di articolazione è significativo:  $F(2,38) = 23.626$ ,  $p = 0.000$ . Una serie di t-test evidenzia che le consonanti labiali contrastano significativamente con le velari,  $t(23) = 3.66$ ,  $p = 0.001$ , e le dentali con le velari  $t(23) = 8.00$ ,  $p = 0.000$ . L'interazione tra luogo e gruppo linguistico è questa volta significativa:  $F(8,38) = 3.17$ ,  $p = 0.007$ . Questo effetto d'interazione è principalmente dovuto al fatto che i gruppi linguistici differiscono per le labiali:  $F(4,19) = 4.68$ ,  $p = 0.008$ , e per le velari  $F(4,19) = 4.58$ ,  $p = 0.009$ . Una serie di confronti post-hoc a coppie tra i gruppi linguistici eseguiti col metodo Bonferroni mostra come, nelle labiali, il francese contrasti significativamente con il giapponese ( $p = 0.034$ ) e con lo svedese ( $p = 0.025$ ), mentre per le velari il francese contrasta con il giapponese ( $p = 0.008$ ).

Per quanto riguarda il modo di articolazione, nello stadio delle 25 parole l'ANOVA mostra che l'effetto principale del modo di articolazione è significativo:  $F(3,57) = 154.01$ ,  $p = 0.000$ . Una serie di t-test evidenzia che le consonanti occlusive orali contrastano significativamente con le fricative,  $t(23) = 15.87$ ,  $p = 0.000$ , con le nasali,  $t(23) = 11.29$ ,  $p = 0.000$  e con le liquide,  $t(23) = 13.30$ ,  $p = 0.000$ , che le consonanti nasali contrastano significativamente con le fricative  $t(23) = 4.60$ ,  $p = 0.000$  e le nasali con le liquide,  $t(23) = 3.23$ ,  $p = 0.004$ . Anche questa volta l'interazione tra luogo e gruppo linguistico è significativa:  $F(3,57) = 2.47$ ,  $p = 0.011$ . Questo effetto d'interazione è principalmente dovuto al fatto che i gruppi linguistici differiscono per le occlusive:  $F(4,19) = 4.08$ ,  $p = 0.015$ , e per le nasali:  $F(4,19) = 3.77$ ,  $p = 0.020$ . Una serie di confronti post-hoc a coppie tra i gruppi linguistici eseguiti col metodo Bonferroni mostra come, nelle occlusive, il francese contrasti significativamente con lo svedese ( $p = 0.017$ ), mentre per le nasali quest'ultimo contrasta significativamente con il francese ( $p = 0.028$ ) e l'italiano ( $p = 0.027$ ).

Abbiamo fin qui dimostrato che l'influenza linguo-specifica è assente nello stadio delle 0 parole, ma presente in quello delle 25 parole, quando compare la significatività dell'interazione tra gruppi linguistici e, rispettivamente tra modi e luoghi di articolazione. Ci proponiamo ora di indagare se queste differenze riflettono i *patterns* caratteristici di ciascuna lingua (Boysson-Bardies et al., 1992).

I grafici 1-3 illustrano la distribuzione percentuale delle consonanti non finali di parola nelle categorie di luogo di articolazione organizzate per gruppi linguistici e riferite allo stadio delle 0 parole, a quello delle 25 parole, e alle lingue *target* (le frequenze fanno riferimento rispettivamente alla distribuzione generale e indipendente dalla posizione e a quella legata alla consonante iniziale). I dati per i bambini non italiani sono tratti come al solito da Boysson-Bardies & Vihman (1991), e Vihman (1993). I dati relativi ai *target* che le autrici presentano sono calcolati sulle forme lessicali adulte che i bambini avevano tentato di produrre. I dati relativi ai *target* dell'italiano sono stati calcolati sulla base delle

frequenze consonantiche nelle parole del PVB (Caselli e Casadio, 1995) relative alla fascia di età 18-23 mesi (cfr. Zmarich, Dispaldro, Rinaldi e Caselli, questo volume). Per simulare le frequenze delle occorrenze (*token*) e non dei tipi lessicali (*types*), abbiamo moltiplicato le frequenze ricavate dalle parole del quarto quartile (cioè le più frequenti perché prodotte dal 75 al 100% dei bambini) per quattro, quelle relative al terzo quartile per tre, quelle relative al secondo quartile per due, e le ultime, quelle relativamente rare del primo quartile, le abbiamo lasciate invariate.

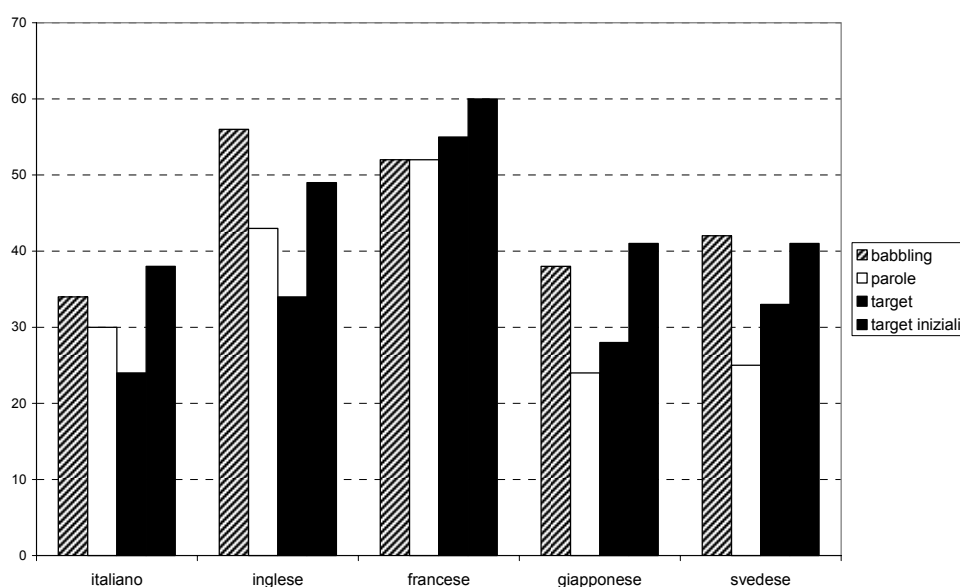


Fig.1. Distribuzione delle frequenze percentuali delle consonanti labiali organizzate per gruppi linguistici e riferite allo stadio delle 0 parole, quello delle 25 parole, e alle lingue target (rispettivamente generali e limitate alla consonante iniziale).

La produzione delle labiali (fig. 1), basata sulla distribuzione generale o in posizione iniziale nei target, predice lo stesso ordinamento gerarchico: francesi > inglesi > svedesi = giapponesi > italiani. E in effetti lo stesso ordinamento si trova anche per le produzioni infantili, che seguono in particolare la distribuzione delle consonanti iniziali.

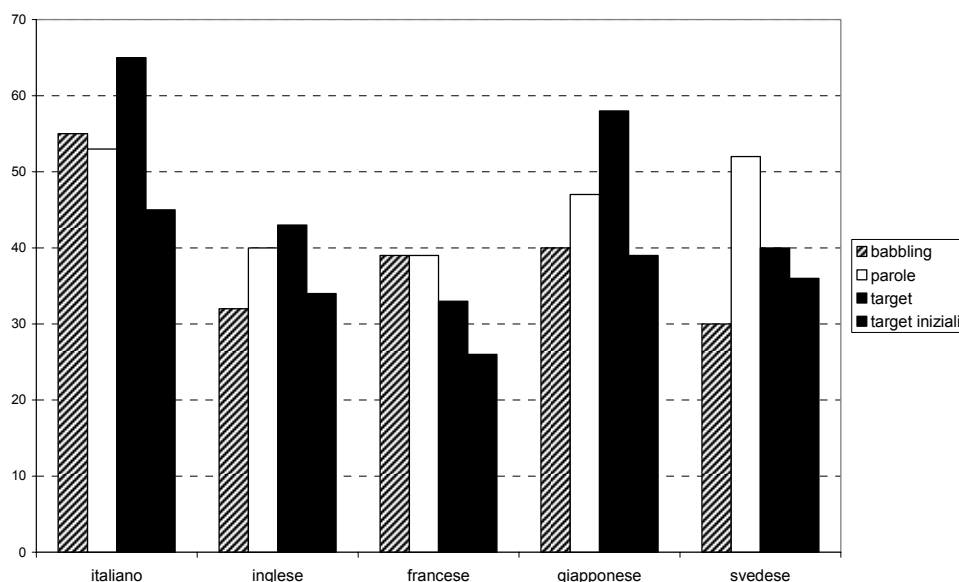


Fig.2. Distribuzione delle frequenze percentuali delle consonanti dentali organizzate per gruppi linguistici e riferite allo stadio delle 0 parole, quello delle 25 parole, e alle lingue target (rispettivamente generali e limitate alla consonante iniziale).

Nella produzione delle dentali (fig. 2), considerando le distribuzioni delle consonanti nei *target* adulti ci si aspetta che i bambini italiani producano più dentali dei giapponesi e degli svedesi, e i francesi e gli americani il numero minore, e in effetti accade esattamente questo, soprattutto nello stadio delle 25 parole.

Per le velari (fig. 3), le distribuzioni generali e in posizione iniziale nei target adulti predicono esattamente l'ordinamento gerarchico che troviamo nelle produzioni infantili: svedesi > inglesi = giapponesi > italiani > francesi.

Per le occlusive (fig. 4), le distribuzioni generali e in posizione iniziale nei target adulti seguono l'ordinamento gerarchico svedesi > inglesi > giapponesi = francesi > italiani, mentre le produzioni infantili ribaltano l'ordinamento degli ultimi tre posti: svedesi > giapponesi > italiani > francesi. Le percentuali di produzione delle occlusive da parte dei bambini sono quelle che cambiano di meno, tra tutte le categorie considerate, tra i due stadi delle 0 e delle 25 parole.

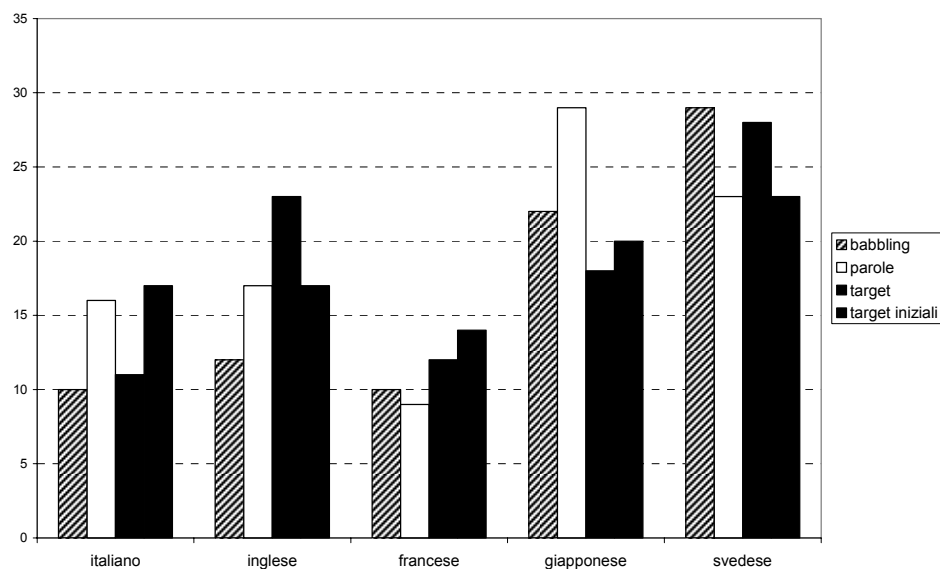


Fig.3. Distribuzione delle frequenze percentuali delle consonanti velari organizzate per gruppi linguistici e riferite allo stadio delle 0 parole, quello delle 25 parole, e alle lingue target (rispettivamente generali e limitate alla consonante iniziale).

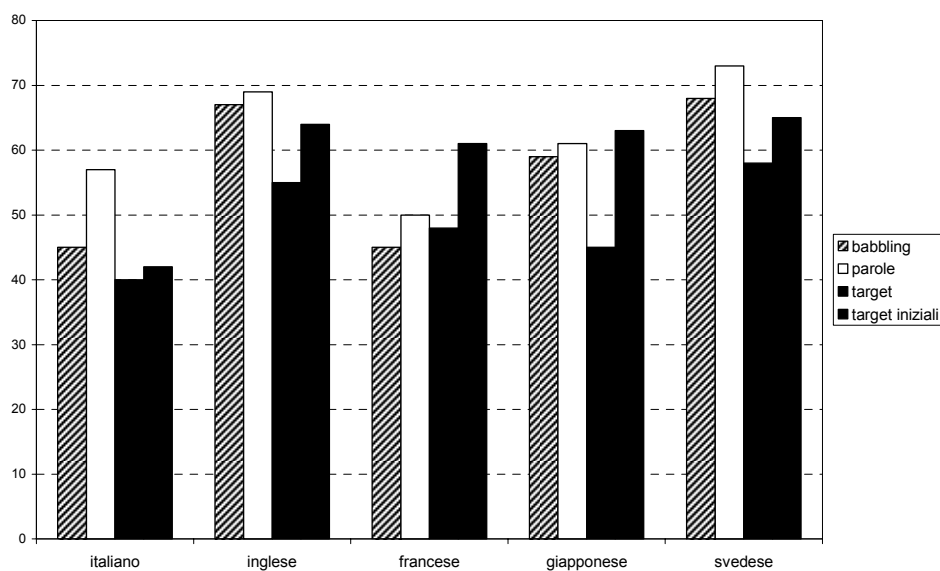


Fig.4. Distribuzione delle frequenze percentuali delle consonanti occlusive organizzate per gruppi linguistici e riferite allo stadio delle 0 parole, quello delle 25 parole, e alle lingue target (rispettivamente generali e limitate alla consonante iniziale).

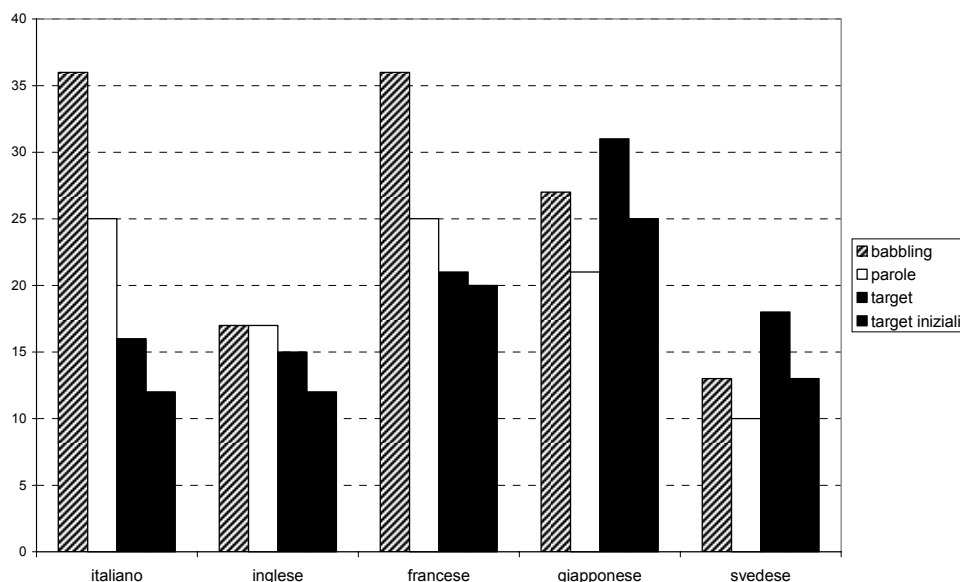


Fig.5. Distribuzione delle frequenze percentuali delle consonanti nasali organizzate per gruppi linguistici e riferite allo stadio delle 0 parole, quello delle 25 parole, e alle lingue target (rispettivamente generali e limitate alla consonante iniziale)

Per le nasali, le distribuzioni generali e in posizione iniziale nei target adulti (giapponese > francese > svedese > italiano = inglese) non sono molto rispettate nelle produzioni infantili, che invece seguono l'ordinamento: italiani = francesi > giapponesi > inglesi > svedesi.

Per le fricative e le liquide, le percentuali sono troppo ridotte per essere informative e attendibili.

## 5. DISCUSSIONE E CONCLUSIONE

L'analisi, esposta al paragrafo precedente, delle frequenze dei foni consonantici in posizione non finale, in bambini di lingua madre italiana (cfr. anche Zmarich & Miotti, 2003; Giulivi et al., 2005), francese, inglese, giapponese e svedese (cfr. Boysson-Bardies & Vihman, 1991; Vihman, 1993), ci consente di trarre la seguente generalizzazione. Il cambiamento più evidente in ciascun bambino quando passa dal *babbling* (stadio delle 0 parole) alla produzione del primo vocabolario (stadio delle 25 parole) consiste nella riduzione del numero dei foni prodotti (indicato dalla DS). Mentre nel *babbling* tutti i bambini, indipendentemente dall'appartenza al gruppo linguistico, mostrano una preferenza relativa per le consonanti occlusive (orali e nasali) articolate nel luogo labiale e dentale, nello stadio delle 25 parole si assiste all'influenza dei *pattern* delle singole lingue (vedi la significatività dell'interazione gruppo linguistico X modo/luogo di articolazione). È questo, per esempio, il caso dei bambini italiani che riducono grandemente le percentuali delle occlusive nasali passando dal *babbling* allo stadio delle 25 parole, adeguandosi ai livelli richiesti dalla lingua *target*.

Per quanto riguarda la prevalenza delle occlusive orali nello stadio delle prime parole, prevalenza che caratterizza anche i bambini italiani, rimandiamo all'estesa discussione in Boysson-Bardies & Vihman (1991), che sottolineano come i bambini cerchino se possibile di riprodurre nelle prime parole gli schemi motori ben rodati nel babbling, come quelli che caratterizzano le occlusive, in ciò favoriti dalla relativa frequenza delle occlusive orali nelle lingue del mondo. Le nasali, invece, pur essendo altrettanto frequenti delle occlusive orali nel babbling, lo sono molto meno nelle parole, soprattutto in alcune lingue, e infatti i dati sperimentali documentano una riduzione generalizzata delle occlusive nasali allo stadio delle 25 parole. Inoltre queste due tendenze sono meglio spiegabili se immaginiamo che i bambini si stiano rapportando alle consonanti iniziali dei target lessicali, dove le occlusive orali prevalgono più nettamente sulle occlusive nasali di quanto accada in posizione non iniziale.

Osserviamo inoltre come ci siano notevoli differenze tra i gruppi linguistici nazionali nelle percentuali di frequenza di occorrenza di ciascuna categoria di modo e luogo di articolazione già dallo stadio iniziale di questa ricerca, quello delle 0 parole, relativo alle produzioni di solo *babbling*. Il fatto che i diversi gruppi nazionali non si uniformino alla stessa percentuale per ciascuna categoria, e che le differenze non siano casuali, può dimostrare che, pur in presenza di differenze notevoli tra i gruppi nazionali per l'età di campionamento dello stadio delle 0 parole, nell'intervallo temporale tra i 9 e i 13 mesi questi bambini in generale strutturano già le caratteristiche fonetiche del loro *babbling* in funzione delle caratteristiche fonetiche della lingua nativa.

In questa sede comunque ci preme di sottolineare le differenze dei risultati qui presentati rispetto ai precedenti. Per esempio, Boysson-Bardies & Vihman (1991) enfatizzano l'aumento di frequenza delle consonanti labiali nelle prime parole rispetto al *babbling*, giustificandolo con la salienza percettiva e visiva offerta dalla labiali, soprattutto occlusive, rispetto alle altre consonanti. I nostri dati invece documentano una riduzione delle labiali per tutti i gruppi linguistici, salvo il francese, nel passaggio dal *babbling* allo stadio delle 25 parole (vedi fig. 1). L'apparente contraddizione si potrebbe spiegare tenendo presente che mentre i dati discussi da Boysson-Bardies & Vihman (1991) riguardano solo le produzioni infantili che allo stadio delle 25 parole sono lessicali, i dati presentati in questo lavoro (sia quelli, originali, relativi ai bambini italiani, sia quelli dei bambini francesi, inglesi, svedesi e giapponesi analizzati in Boysson-Bardies & Vihman, 1991) presentano le frequenze consonantiche allo stadio delle 25 parole senza distinguere tra produzioni lessicali o di *babbling*. Se ne evince che le frequenze dei fonemi presenti negli episodi di *babbling* allo stadio delle 25 parole sono ancora molto consistenti e molto diverse, almeno per quanto riguarda la preferenza labiale, rispetto a quelle delle prime parole. Queste considerazioni comunque restano allo stato di ipotesi finché non verrà condotta un'analisi statistica separata sulle produzioni di *babbling* e quelle lessicali allo stadio delle 25 parole.

Un altro dato leggermente diverso da quanto trovato da Boysson-Bardies & Vihman (1991) riguarda la questione della somiglianza delle frequenze di occorrenza delle categorie prodotte dai gruppi nazionali di bambini allo stadio delle 25 parole con le frequenze iniziali dei target lessicali o con le frequenze generali. Se le frequenze consonantiche allo stadio delle 25 parole vengono considerate come nel presente lavoro in modo indifferenziato rispetto all'appartenenza agli episodi di *babbling* o alle parole, si vede che si può parlare, come fanno Boysson-Bardies & Vihman (1991), di un adeguamento alle frequenze delle consonanti iniziali dei *target*, piuttosto che alle frequenze generali, solo per le consonanti occlusive e per le consonanti velari. Per quanto riguarda invece le consonanti dentali e le

consonanti nasali, le frequenze di tutti i gruppi linguistici sono più simili a quelle relative alla distribuzione generale.

Un altro dato forse non sufficientemente colto da Boysson-Bardies & Vihman (1991) riguarda le occlusive nasali, che calano drammaticamente nello stadio delle 25 parole rispetto agli alti livelli del *babbling*, soprattutto per i bambini francesi e italiani, apparentemente per adeguarsi ai livelli richiesti nelle lingue target.

Per concludere, pensiamo che questi risultati possano passare il vaglio del criterio generale di Boysson-Bardies et al. (1992), esposto nell'introduzione, per attestare l'esistenza di un'influenza linguo-specifica, poiché ci troviamo effettivamente di fronte, nello stadio delle 25 parole, a differenze fonetiche fra i gruppi linguistici che sono maggiori delle differenze presenti all'interno di ogni singolo gruppo, e queste differenze nella maggior parte dei casi riflettono i *patterns* fonetico-fonologici caratteristici di ciascuna lingua.

## BIBLIOGRAFIA

Best C. T. (2002), Revealing the mother tongue's nurturing effects on the infant ear, *Infant Behavior and Development*, 25, 134-139

Blevins, J. (1995), The syllable in phonological theory. In *The Handbook of Phonological Theory* (J. A. Goldsmith, editor), Cambridge (MA): Basil Blackwell, 206-244

Bortolini U., Continuità fonetica tra "babbling" e prime parole, in *Ricerche sull'acquisizione dell'italiano* (a cura di E. Cresti e M. Moneglia), Bulzoni, Roma, 1993.

Boysson Bardies B. de, Hallé P., Sagart L. & Durand C. (1989), A crosslinguistic investigation of vowels formants in babbling, *Journal of Child Language*, 16, 1-17.

Boysson-Bardies B. de, Vihman M.M., (1991) Adaptation to language: evidence from babbling and first words in four languages, *Language* 67, 297-319

Boysson-Bardies B. de, Vihman, M. M., Roug-Hellichius, L., Durand, C., Landberg, I., Arao, F. (1992), Material evidence of infant selection from the target language, in *Phonological Development. Models, Research, Implications* (C. A. Ferguson, L. Menn & C. Stoel-Gammon, editors), Timonium: York Press, 369-391.

Caselli M.C. e Casadio P. (1995), *Il primo vocabolario del bambino*, Franco Angeli. Milano.

Davis B.L., MacNeilage P.F. (1995), The articulatory basis of babbling. *J. Speech Hearing Res.*, 38, 1199-1211.

Giulivi S., Zmarich C., Vayra M., Farnetani E. (2006), Lo sviluppo fonetico in relazione agli stadi di produzione della parola: studio pilota di una bambina italiana, in *Atti del II° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV)* (a cura di R. Savy & C. Crocco), Salerno, 30/11-2/12/2005, EDK Editore, Brescia, 472-490.

Kent R.D. & Miolo G. (1995), Phonetic abilities in the first year of life, in *The Handbook of child language* (P. Fletcher & B. MacWhinney, editors), Blackwell.

Locke, J.L. (1983) *Phonological acquisition and change*. New York: Academic Press

- Macken M. A. & Barton, D. (1980), A longitudinal study of the acquisition of the voicing contrast in American-English word-initial stops, as measured by Voice Onset Time, *Journal of Child language*, 7, 41-74.
- Nespor, M. (1993), *Fonologia*, Il Mulino, Bologna.
- Shriberg, L.D. & G.L., Lof (1991), Reliability studies in broad and narrow phonetic transcription, *Clinical Linguistics & Phonetics*, 5, 225-279
- Stoel-Gammon, C. (1985), Phonetic inventories, 15-24 months: a longitudinal study, *Journal of Speech and Hearing Research*, 28, 505-512.
- Vihman, M.M. (1993), Variable paths to early word production, *Journal of Phonetics*, 21, 61-82
- Vihman, M. M., de Boysson-Bardies B. (1994), The Nature and Origin of Ambient Language Influence on Infant Vocal Production and Early Words. *Phonetica*, 51, 159-169.
- Vihman, M. & McCune, L. (1994) When is a word a word?, *Journal Child Language*, 21, 517-542.
- Zmarich C., Dispaldro M, Caselli M. C., Rinaldi P., La composizione fonetica del primo vocabolario del bambino, in questo volume.
- Zmarich C., Miotti R. (2003), The frequency of consonants and vowels and their co-occurrences in the babbling and early speech Italian children, in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 4-9 August 2003, 1947-1950.
- Zmarich C., Stocco D., Minozzi M., Bonifacio S. (2005), La frequenza di occorrenza di consonanti e vocali e delle loro combinazioni nella sillabe del babbling e delle prime parole dai 10 ai 27 mesi di età, in *Misura dei parametri, Atti del 1° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV)* (a cura di P. Cosi), Padova, 2-4 dicembre, 2004, EDK Editore, Brescia, 481-510.



# L'INTONAZIONE NELL'ITALIANO DELL'AREA LECCESE: PRIME OSSERVAZIONI DAL PUNTO DI VISTA AUTOSEGMENTALE-METRICO

Antonio Stella, Barbara Gili Fivela  
Università del Salento & CRIL, Lecce

[antonio.stella@ateneo.unile.it](mailto:antonio.stella@ateneo.unile.it), [barbara.gili@ateneo.unile.it](mailto:barbara.gili@ateneo.unile.it)

## 1. SOMMARIO

In questo contributo\*, proponiamo una prima analisi fonologica di tipo autosegmentale-metrico delle caratteristiche intonative della varietà leccese di italiano, con particolare attenzione agli enunciati corrispondenti alla modalità interrogativa.

Il corpus d'indagine è stato elicitato con il metodo del *Map Task* (Anderson *et alii*, 1991), un gioco dialogico che consiste nella ricostruzione di un percorso su una mappa, e con quello del *Test delle differenze*, che permette di acquisire materiale dialogico prodotto nel tentativo di trovare le differenze tra due vignette apparentemente uguali. I soggetti analizzati sono tutti studenti dell'Università del Salento, nati e cresciuti a Lecce o in paesi limitrofi. I dialoghi sono stati trascritti ortograficamente delimitando i turni dialogici ad ognuno dei quali è stata attribuita un'etichetta relativa al tipo di mossa conversazionale (Carletta *et alii*, 1996). I materiali sono stati poi analizzati uditiivamente ed acusticamente al fine di individuare le caratteristiche fonologiche del sistema intonativo dei parlanti, secondo il quadro teorico della fonologia autosegmentale-metrica (Ladd, 1996) e con l'uso del modello ToBI (*Tones and Break Indices*, cfr. Beckman & Ayers, 1997).

Si è rilevato che le mosse Query-Wh, corrispondenti alle domande aperte, e le mosse Query-YN, corrispondenti a domande polari, sono prevalentemente realizzate con un contorno nucleare discendente-ascendente, analizzato come H+L\*...L-H%; nel caso delle domande aperte, si è riscontrato, inoltre, l'uso di un *pattern* nucleare discendente, cioè con un tono di confine L-L%.

Nelle mosse conversazionali di tipo Check, che corrispondono a richieste di conferma, il contorno nucleare non differisce da quello individuato per le domande polari; tuttavia, tra le strategie seguite dai parlanti è stato anche riscontrato l'utilizzo di un andamento discendente, caratterizzato da toni di confine bassi L-L% in luogo di una specificazione di confine L-H%.

Le mosse Object, invece, sono realizzate con un accento tonale L+H\* caratterizzato da un allineamento particolarmente ritardato rispetto alla sillaba, oppure con una categoria accentuale differente, etichettata come H\*+L.

Infine, grazie all'analisi dei contesti di focalizzazione, anche in relazione a diverse costruzioni sintattiche finalizzate alla focalizzazione sintattica, topicalizzazione e dislocazione, si è fornita una prima analisi di altri eventi tonali. Ad esempio, è stato individuato l'accento tonale L+H\* sembra mettere in evidenza un elemento focalizzato, e, quando sia seguito da un tono di confine, solitamente L-, individua una frase intermedia, soprattutto in casi di dislocazione.

---

\* Questo articolo è frutto del lavoro congiunto dei due autori. Tuttavia, ai soli fini accademici, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6 vanno attribuiti a A.S. e 2, 3, 4, 5.7, 5.8, 5.9 a B.G.F.

Nell'articolo sono inoltre discussi fenomeni di associazione secondaria dei toni di confine, ipotizzata nel caso di enunciati con focus non finale.

## 2. INTRODUZIONE

In questo contributo proponiamo una prima analisi fonologica di tipo autosegmentale-metrico (d'ora in avanti, AM; Pierrehumbert (1980); si veda Ladd (1996) per una panoramica) dell'intonazione nella varietà leccese dell'italiano - in particolare in quella utilizzata nella città di Lecce e nei paesi limitrofi - per gli enunciati interrogativi e nei contesti di focalizzazione e topicalizzazione.

Obiettivo della teoria AM è quello di caratterizzare adeguatamente i contorni intonativi in termini fonologici, con il riferimento ad una serie di elementi categoricamente distinti, corrispondenti a parametri acustici continui; l'analisi AM assume che la struttura tonale sia lineare, ossia formata da una concatenazione di eventi locali associati a determinati punti della gerarchia metrica: tali eventi consistono in accenti tonali (*pitch accents*), associati a sillabe metricamente forti dell'enunciato, e toni di confine (*edge tones*), associati invece ai confini di costituenti prosodici.

Ai fini dell'analisi delle lingue intonative, come l'italiano, la teoria AM distingue tra l'accentuazione lessicale, anche dovuta alla modulazione della frequenza fondamentale (F0) e correlata alla prominenza delle sillabe all'interno di parola, e la modulazione relativa a domini più ampi della parola, di carattere postlessicale, dovuta principalmente a modificazioni tonali in corrispondenza di sillabe accentate e di confini prosodici. L'altezza tonale viene indicata attraverso due soli indici, corrispondenti al livello alto, indicato con 'H', e al livello basso, indicato con 'L'. Una caratteristica fondamentale della teoria AM, inoltre, è quella che riguarda gli aspetti fonetici delle produzioni intonative, i quali dipendono da una varietà di fattori come il grado di enfasi o la posizione dell'evento nell'enunciato: l'identità fonetica del tono H o L è quindi determinata tenendo conto di tali fattori, ma la sua identità fonologica, in buona sostanza, prescinde da essi. La variazione fonetica può interessare il singolo evento tonale o l'andamento intonativo dell'intero sintagma.

A differenza di un approccio puramente fonetico, l'approccio fonologico utilizzato nella teoria AM differenzia gli aspetti fonetici e quelli fonologici e permette di catturare anche aspetti funzionali dell'intonazione, cercando di individuare l'uso che il parlante fa di determinate categorie fonologiche tonali e il significato pragmatico che vuole trasmettere producendole.

L'ausilio di un metodo fonologico per la descrizione dell'intonazione differenzia profondamente la nostra indagine da altri studi sperimentali condotti sugli aspetti prosodici nelle varietà salentine, lavori che, in base alle nostre conoscenze, sono tutti ad opera di Romano (1998, 1999, 2000). L'autore, infatti, utilizzando un approccio di tipo fonetico-strumentale, descrive le caratteristiche dell'intonazione affermativa ed interrogativa del dialetto e dell'italiano regionale parlati nella penisola salentina<sup>1</sup>. Tuttavia, a causa delle

---

<sup>1</sup> All'interno di questa zona, egli individua due aree differenti dal punto di vista delle caratteristiche prosodiche: l'area del "salentino meridionale estremo", comprendente l'area meridionale del Salento, per il quale si riscontrano proprietà intonative che lo accomunano con i dialetti meridionali estremi quali quello siciliano (cfr. Grice, 1995), e l'area del "salentino leccese" che comprende la zona centro-settentrionale estendendosi anche verso l'area orientale di Gallipoli. Romano effettua una descrizione dei dialetti e delle varietà di

divergenze metodologiche, i risultati dei lavori appena menzionati non sono facilmente comparabili con il tipo di studio che è descritto in questa sede. L'analisi dell'intonazione qui fornita, infatti, si pone l'obiettivo di considerare solo le variazioni fonetiche che corrispondano a scelte fonologiche del parlante, cercando quindi di astrarre dalle differenze fonetiche dell'enunciato e delle unità prosodiche per individuare unità fonologiche di riferimento. Nel prosieguo del presente lavoro, ad ogni modo, saranno indicate eventuali analogie che possano emergere confrontando i risultati dell'indagine qui descritta con quelli già documentati in letteratura.

### 3. METODO DI ANALISI

Per la notazione degli eventi tonali che caratterizzano un contorno intonativo è stato utilizzato il metodo di trascrizione ToBI, acronimo di *Tones and Break Indices*, (Silverman *et alii*, 1992; Beckman & Hirschberg, 1994; Pitrelli *et alii*, 1994; Beckman & Ayers, 1997), il quale si propone come uno standard per etichettare le caratteristiche prosodiche di *corpora* digitalizzati di parlato; esso fu inizialmente pensato per la lingua inglese, ma la sua adattabilità ne ha permesso l'uso per la descrizione di una grande varietà di lingue. In questa sede si utilizzerà una versione di ToBI opportunamente modificata e adattata alle caratteristiche intonative dell'italiano.

Una trascrizione ToBI completa comprende quattro differenti livelli (*tiers*), i quali consistono in stringhe di etichette disposte sull'asse temporale in corrispondenza di determinati punti della forma d'onda relativa ad un enunciato (Beckman & Ayers, 1997):

- LIVELLO TONALE (*tonal tier*): consiste in etichette che rappresentano livelli tonali distinti, trascritti come sequenze di toni alti (H) e toni bassi (L) e marcati con diacritici indicanti la loro funzione all'interno dell'unità prosodica.
- LIVELLO ORTOGRAFICO (*orthographic tier*): riporta la diretta trascrizione di tutte le parole dell'enunciato.
- LIVELLO DEGLI INDICI DI DISGIUNZIONE (*break index tier*): riporta i raggruppamenti prosodici delle parole di un enunciato; la fine di ogni parola viene marcata con un'etichetta che indica la forza soggettiva della sua associazione con la parola successiva, su una scala da "0" a "4" che rappresenta la variazione dal minor al maggior grado di disgiunzione.
- LIVELLO MISTO (*miscellaneous tier*): riporta i commenti del trascrittore, i quali indicano disfluenze quali, ad esempio, i colpi di tosse.

Il nucleo fondamentale dell'analisi fonologica è rappresentato dal livello tonale, al quale sarà riservata l'attenzione in questa sede. Il livello tonale comprende una serie di etichette posizionate in corrispondenza dei punti della stringa segmentale che corrispondono alle unità metriche con i quali un dato evento tonale si associa. Le etichette che indicano la presenza di un accento tonale sono contrassegnate dal diacritico '\*', il quale indica un tono associato con una sillaba prominente (*starred tone*), di solito lessicalmente accentata (ad es. H\* o L\*); gli accenti, inoltre, possono essere bitonali, cioè caratterizzati dalla presenza di entrambi i livelli tonali in un unico evento: in tal caso, i due toni sono uniti tra loro mediante il diacritico '+', ma solo uno dei due sarà quello associato alla sillaba, indicato con il diacritico '\*', mentre l'altro potrà essere un tono precedente (*leading tone*), come nel caso di L+H\*, oppure un tono successivo (*trailing tone*) rispetto al tono asteriscato, come nel caso di L\*+H. L'accento associato alla sillaba metricamente più forte (normalmente

italiano con l'obiettivo di compararle e determinare le caratteristiche che differenziano le due aree.

individuata in base alla posizione nell'albero metrico) è riconosciuto come accento nucleare.

Per i toni di confine, *phrase accent* e *boundary tone*, si utilizzano rispettivamente i diacritici '!' e '%': il primo indica la presenza di un confine di sintagma intermedio (ip), cioè un costituente prosodico minore, mentre il secondo indica il confine di sintagma intonativo (IP), ossia il costituente prosodico maggiore; il confine di sintagma intonativo, però, essendo contemporaneamente anche un confine di sintagma intermedio, conterrà sempre entrambe le etichette dei toni di confine (come ad esempio in L-H%).

#### 4. CORPUS DI INDAGINE

##### 4.1 Metodi di elicitazione dei dialoghi

Il corpus di indagine è stato elicitato grazie a due metodi già ampiamente utilizzati negli studi sull'intonazione: il *Map Task* e il *Test delle differenze*.

Il Map Task (Anderson *et alii*, 1991) consiste in un gioco dialogico che coinvolge due parlanti a ciascuno dei quali viene consegnata una mappa; le due mappe non sono identiche: una contiene un tracciato disegnato tra alcune icone (*landmarks*), con un punto di partenza e uno d'arrivo; l'altra mappa contiene solo le icone e il punto di partenza. Il compito dei partecipanti al gioco consiste nell'interagire verbalmente per riprodurre, nella maniera più accurata possibile, il tracciato sulla mappa in cui esso non è indicato. Il ruolo dei due interlocutori è perciò stabilito a priori: vi sarà un *instruction giver* (IG), ossia colui che dà indicazioni e istruzioni, e un *instruction follower* (IF), cioè colui che segue le indicazioni e le istruzioni. Tra le due mappe, inoltre, vi sono delle incongruenze riguardo alla numero, alla natura e alla posizione delle icone, in modo da rendere possibili fraintendimenti che rendano più variegato e spontaneo lo scambio comunicativo; per questo motivo, prima di ogni sessione di dialogo, nessun parlante è stato informato che le due mappe non erano identiche. Per l'elicitazione dei dialoghi Map Task sono state utilizzate 2 coppie di mappe, rese disponibili nell'ambito del progetto AVIP (Archivio delle Varietà dell'Italiano Parlato; cfr. Bertinetto, 2001).

Il *Test delle differenze* è, invece, un gioco dialogico effettuato da due partecipanti ad ognuno dei quali viene data una vignetta; le due vignette sono apparentemente uguali, ma si differenziano per alcuni particolari che gli interlocutori hanno lo scopo di scoprire, interagendo verbalmente. A differenza del Map Task, in questo gioco non vi è un ruolo prestabilito ed asimmetrico per gli interlocutori<sup>2</sup>. Per l'elicitazione dei dialoghi si è utilizzata una coppia di vignette che fa parte del materiale del progetto CLIPS (Corpora e Lessici dell'Italiano Parlato e Scritto; cfr. Albano Leoni, 2003).

---

<sup>2</sup> In realtà questa identità di ruoli nel Test delle differenze non è sempre garantita e ciò dipende dalla condizione psicologica dei parlanti e dal ruolo che si instaura tra i due durante la pianificazione dei compiti; ad esempio, in un dialogo del corpus sembra quasi che tra i due attori si instauri un rapporto IG-IF simile a quello pre-determinato del Map Task. Ciò porta a considerazioni riguardanti la validità di etichette attribuite a mosse che in alcuni casi possono sembrare parafrasi di informazione già condivisa ma in realtà sono delle richieste di conferma, per le quali in effetti il parlante aspetta un feedback assumendo un ruolo simile a quello di un IF. In questi casi, allora, il valore pragmatico che si attribuisce ad una mossa potrebbe non essere quello corretto e di ciò bisognerebbe tener conto in fase di etichettatura e di analisi fonologica.

#### 4.2 Criteri di scelta dei parlanti

La raccolta dei materiali effettuata per la descrizione dell'intonazione della varietà leccese dell'italiano proposta in questo studio costituisce solo una parte di un'indagine più ampia, volta a confrontare l'intonazione delle produzioni in italiano e in tedesco da parte di parlanti della varietà leccese dell'italiano. La selezione degli informatori è stata quindi influenzata da questo ulteriore obiettivo<sup>3</sup>.

Il primo corpus di indagine (Stella, 2007), costituito da 4 dialoghi Map Task, comprendeva 8 parlanti (4 maschi e 4 femmine) provenienti dalla città di Lecce e da paesi limitrofi (Arnesano, Copertino e Salice Salentino). I parlanti avevano in comune le seguenti caratteristiche:

- nati e cresciuti nel Salento;
- età compresa tra i 22 e i 32 anni;
- studenti presso differenti Facoltà dell'Università del Salento.

Il corpus è stato poi ampliato e, attualmente, include altri 4 dialoghi prodotti con il metodo del Map Task e 2 dialoghi prodotti effettuando il Test delle differenze; tali dialoghi sono stati prodotti da 8 parlanti (1 maschio e 7 femmine) provenienti dalla città di Lecce e da paesi limitrofi (Copertino e Veglie), i quali avevano in comune le seguenti caratteristiche:

- nati e cresciuti nel Salento;
- età compresa tra i 18 e i 20 anni;
- studenti al I anno del Corso di Laurea in Traduzione e Interpretariato presso la Facoltà di Lingue e Letterature Straniere dell'Università del Salento.

#### 4.3 Registrazione dei dialoghi

Tutti i dialoghi sono stati acquisiti nello studio di registrazione del Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) dell'Università del Salento, in un ambiente insonorizzato e sufficientemente spazioso. Secondo le stesse modalità seguite per l'acquisizione dei materiali durante il progetto AVIP (si veda, in particolare, Savino, 2000), nello studio sono state sistemate due scrivanie separate da un pannello divisorio per impedire il contatto visivo tra i parlanti durante il dialogo e massimizzare l'interazione verbale. Alle registrazioni era presente solo uno degli sperimentatori, il quale ha fornito le spiegazioni necessarie ai partecipanti ma non è intervenuto in alcun modo durante lo svolgimento del dialogo.

La strumentazione utilizzata per la registrazione è la seguente:

- n° 2 microfoni SHURE mod. SM58
- scheda di acquisizione audio EDIROL mod. UA-5
- computer portatile HP Pavilion ze4943ea

I dialoghi sono stati registrati direttamente sul disco fisso del computer utilizzando il software Goldwave (versione 5.17) su sistema operativo Windows XP; la registrazione è avvenuta su due canali, uno per parlante, con una frequenza di campionamento di 24 kHz per canale a 16-bit. L'analisi intonativa e l'estrazione dell'andamento di F0 sono state effettuate con l'ausilio del software Praat (Boersma & Weenink, 2007).

---

<sup>3</sup> Nel presente contributo si tralascerà di elencare le caratteristiche funzionali all'analisi della lingua tedesca le quali hanno determinato la scelta dei parlanti; inoltre, il numero dei dialoghi indicati nel presente paragrafo non comprende i dialoghi effettuati in tedesco.

#### 4.4 Trascrizione ed etichettatura dei dialoghi

La fase di trascrizione dei dialoghi è stata effettuata seguendo le convenzioni riportate in Savy (2001), già utilizzate per la trascrizione dei dialoghi del corpus AVIP: i dialoghi sono suddivisi in *turni dialogici*, i quali sono delimitati dalla presa di parola da parte di uno dei due interlocutori; quelli del primo corpus sono stati tutti trascritti ortograficamente<sup>4</sup>, mentre per i dialoghi aggiunti è stata effettuata la trascrizione dei soli turni giudicati rilevanti ai fini dell'analisi qui descritta.

Una prima fase di etichettatura è consistita nella determinazione dello scopo e della funzione degli enunciati in relazione al contesto dialogico, in base alle convenzioni di etichettatura proposte in Carletta *et alii* (1996), nelle quali l'unità minima è definita *mossa conversazionale*. Le mosse conversazionali prese in considerazione nel presente studio sono principalmente quelle che hanno la funzione di richiesta di informazioni o di richiesta di conferma, ossia:

- CHECK: richiede all'interlocutore di confermare un'informazione che il parlante ha motivo di credere che sia vera, senza esserne completamente certo. In genere questa mossa è la richiesta di conferma di qualcosa che il partner ha cercato di comunicare esplicitamente o qualcosa che il parlante crede che possa, o debba, essere dedotta da quanto detto in precedenza (ma non necessariamente in eventi dialogici immediatamente precedenti).
- ALIGN: controlla l'attenzione o l'accordo del partner e consente di verificare che l'informazione sia stata trasferita con successo (ad es. "Ok?", "Ci sei?" "No?", ecc.).
- QUERY-YN: è una domanda di tipo polare che richiede una risposta del tipo "Sì" oppure "No" e che non rientra nei casi di mossa *Check* o *Align*. Essa focalizza l'attenzione dell'interlocutore su un determinato punto della mappa o su un dominio o su un compito, ma comunque su informazione che non può essere direttamente dedotta dal contesto dialogico.
- QUERY-WH: è una domanda che non appartiene alle altre categorie e che richiede una risposta aperta, che consiste nella scelta di un elemento informativo sull'asse paradigmatico (con l'esclusione di 'sì' e 'no') o, talvolta, in un insieme esplicitato.
- OBJECT: questa mossa solitamente esprime il rifiuto di assumere lo scopo preposto, normalmente perché il parlante nota che esiste un modo migliore di portare a termine il compito oppure perché non condivide l'obiettivo dell'interlocutore; spesso questa mossa si risolve con una domanda eco prodotta con un'intonazione di sorpresa particolarmente evidente (Grice & Savino, 1995)<sup>5</sup>.

---

<sup>4</sup> Nella trascrizione ortografica vengono riportati inoltre tutti i fenomeni rilevanti ai fini dell'analisi intonativa, quali pause, sovrapposizioni, disfluenze, allungamenti, fenomeni vocali, ecc.

<sup>5</sup> La mossa Object può comunque essere realizzata anche con un'intonazione differente da quella interrogativa, ad esempio con espressioni come "*No, parliamo di...*" oppure "*Cosa dici?*". Per questo studio sono state prese in considerazione esclusivamente le mosse Object che presentavano una chiara intonazione interrogativa.

## 5. ANALISI FONOLOGICA

### 5.1 Query-Wh: domande aperte

Le domande aperte sono solitamente realizzate con un contorno che presenta un accento nucleare discendente H+L\* seguito da un tono di confine ascendente L-H%. In figura 1 è riportato un esempio di domanda con focus largo, nella quale l'accento tonale viene realizzato con un andamento di F0 che raggiunge una frequenza molto alta sulla sillaba pretonica *ti* e che inizia la sua discesa dall'incipit della sillaba tonica *chia-* di 'chiami'<sup>6</sup>. Lo stesso parlante realizza un andamento melodico identico nell'enunciato in figura 2 nel quale però il focus è ristretto sulla parte iniziale dell'enunciato: l'ascesa che precede la realizzazione del *pitch accent* si estende sulla sillaba prenucleare *Dov-*, mentre la discesa si realizza su *-v'è*. Il movimento tonale, da valori alti a valori bassi, è realizzato quindi sull'intero nucleo della sillaba tonica, motivo per il quale si propone la trascrizione dell'accento con H+L\*.

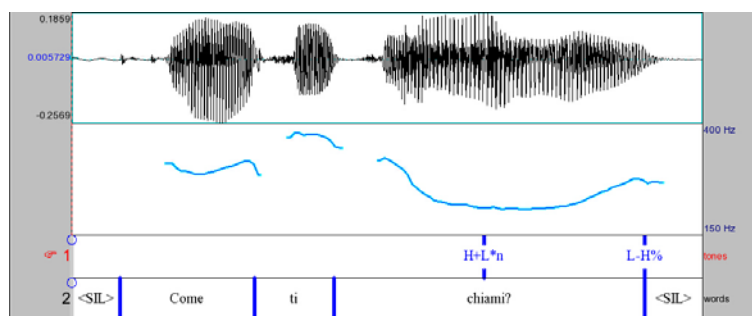


Figura 1: mossa Query-Wh del parlante 'Ba': "Come ti chiami?".

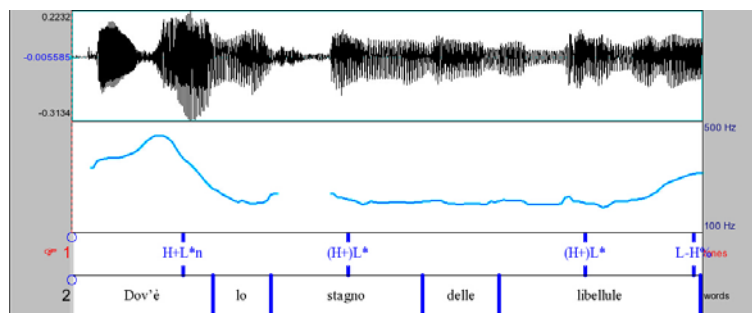


Figura 2: mossa Query-Wh del parlante 'Ba': "Dov'è lo stagno delle libellule?".

Osservando le due figure, inoltre, si nota come, nonostante l'aumento del numero di sillabe che separano l'accento tonale dal tono di confine, i valori di F0 rimangono costanti e corrispondano ad un lungo *plateau* basso. Nel *plateau* dell'enunciato in figura 2 si individuano due prominenze accentuali postnucleari: l'andamento di F0 è leggermente discendente e potrebbe corrispondere alla presenza di un accento tonale L\* o di un accento bitonale H+L\* nel quale, per motivi legati al contesto tonale, il bersaglio alto non è

<sup>6</sup> Nelle figure riportate in questo lavoro, l'accento tonale nucleare è indicato apponendo il diacritico 'n' alla destra dell'etichetta che identifica il tipo di accento (Grice *et alii*, 2005).

implementato in modo evidente. In posizione postnucleare, gli accenti di questo tipo sembrano essere caratterizzati da una minima escursione di F0 (si veda anche la figura 4) e presentano caratteristiche compatibili con la presenza di un solo bersaglio basso. Tuttavia la possibilità che si tratti della posizione postnucleare a determinare la compressione dell'intervallo di frequenze e che, quindi, sia possibile trovare anche altre categorie accentuali in cui i toni alti non siano implementati con valori alti di F0 (ad esempio, H\*+L), ci porta a preferire l'analisi (H+)L\* in questa prima fase di analisi. Gili Fivela (2004), infatti, propone l'etichetta (H+)L\* per casi analoghi, con il *trailing tone* indicato tra parentesi perché non chiaramente visibile in termini di escursione di F0, ad esempio, per via della compressione del *pitch range*<sup>7</sup>.

In entrambi gli enunciati il tono di confine è realizzato con un'ascesa che investe solo parzialmente la sillaba finale: l'attacco dell'ultima sillaba è infatti caratterizzato da un valore basso di F0 e l'ascesa che si forma sull'ultima sillaba raggiunge valori medi del campo di frequenze del parlante. Su tali basi si è quindi optato per una notazione L-H%.

Nelle figure 3 e 4 sono riportate due realizzazioni che differiscono dalle precedenti per la combinazione dei toni di confine: entrambi i contorni presentano un tono nucleare H+L\*, implementato con una discesa di F0 sulla sillaba accentata, seguito da toni di confine bassi (L-L%) anziché da una sequenza basso-alto che determina una fase ascendente (L-H%). Inoltre, così come è stato evidenziato dal confronto tra il contorno melodico in figura 1 e quello in figura 2, anche confrontando le figure 3 e 4 si può rilevare che l'aumento del numero di sillabe interposte tra l'accento tonale nucleare e il tono di confine porta alla realizzazione di un *plateau* basso tra i due eventi tonali.

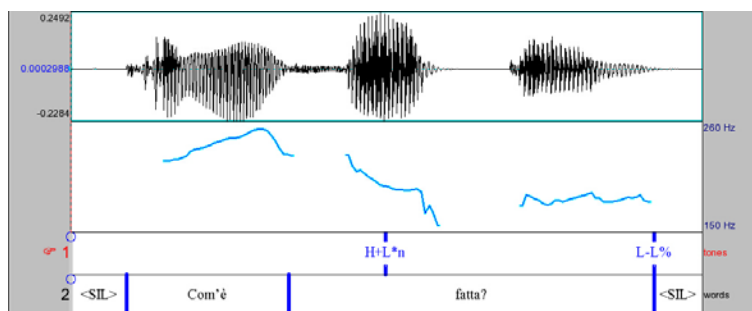


Figura 3: mossa Query-Wh del parlante 'Be': "Com'è fatta?".

<sup>7</sup> Gili Fivela (2004) rileva che il target H di un accento nucleare H+L\* può non essere chiaramente visibile. Tra i fattori rilevanti, si individuano quelli legati alla scelta della *range* di frequenze o al fatto che l'implementazione fonetica di un *trailing tone* alto sia localizzata sulla linea di interpolazione tra un accento tonale prenucleare H\* e il target basso dell'accento nucleare H+L\*.



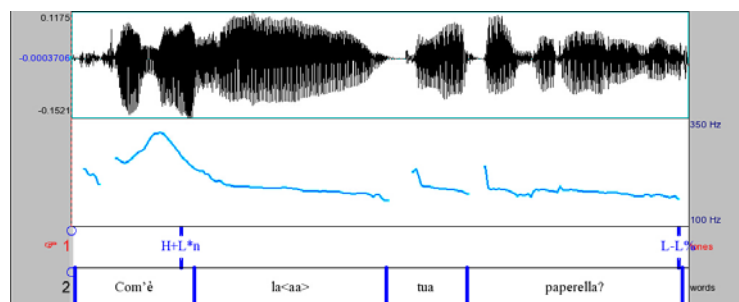


Figura 4: mosca Query-Wh del parlante 'Be': “Com'è la tua paperella?”.

L'inserimento di un elemento lessicale che rappresenta una informazione particolarmente rilevante nello scambio dialogico (tipicamente, informazione nuova) porta alla formazione di un contorno differente, il quale contiene un accento tonale la cui funzione sembra proprio essere quella di mettere in rilievo tale informazione. Un esempio si può riscontrare nell'enunciato riportato in figura 5, dove sulla parola 'capelli' il parlante produce un accento tonale ascendente, etichettato come L+H\*: questo accento corrisponde ad un bersaglio tonale alto nella parte finale della sillaba accentata *-pel-*, o persino nella sillaba postonica, e ad un bersaglio tonale basso allineato ad una delle sillabe pretoniche. Nel nostro corpus di indagine, la presenza di un accento tonale L+H\* è stata più volte rilevata come la categoria fonologica che il parlante usa per focalizzare un elemento all'interno dell'enunciato (cfr. 5.8 e 5.9 per una descrizione più approfondita).

L'accento nucleare della domanda si realizza comunque attraverso un accento tonale H+L\* e, nel caso in esame, i due eventi tonali L+H\* e H+L\* sono legati tra loro da un *plateau* che interpola i due target alti degli eventi tonali. Il contorno continua poi su un livello basso di F0 fino alla fine dell'enunciato, indicando la presenza di toni di confine bassi, come in corrispondenza di una specificazione L-L%.

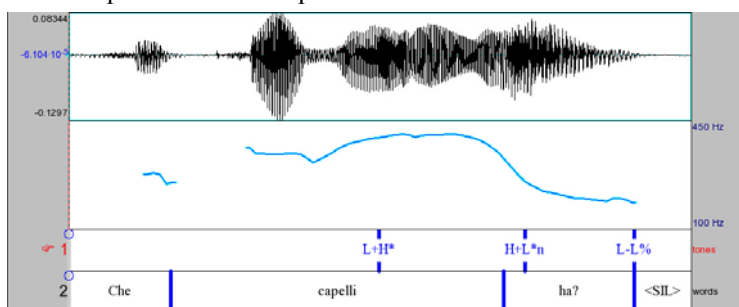


Figura 5: mosca Query-Wh del parlante 'El': “Che capelli ha?”

Il caso riportato in figura 6 rappresenta un esempio di Query-Wh con una struttura sintattica marcata per molte varietà di italiano, nella quale la costruzione è quella delle frasi scisse<sup>8</sup>: la frase principale è composta dall'elemento interrogativo e dalla copula ed è

<sup>8</sup> Le interrogative con struttura scissa rappresentano la scelta non marcata in alcune varietà settentrionali (Poletto e Vanelli, 1995). Altrimenti rientrano tra i fenomeni di segmentazione dell'enunciato e corrispondono a frasi sintatticamente marcate (Salvi e Vanelli, 2004).

seguita da una subordinata dalla quale proviene l'elemento *wh* (Poletto e Vanelli, 1995). Gli spostamenti rispetto all'assetto sintattico non marcato sono correlati alla presenza di differenti accenti tonali ascendenti L+H\* che si susseguono e che sono realizzati a diversi livelli di F0 per effetto della normale declinazione o di graduali compressioni dell'intervallo di frequenze utilizzato. Come si descriverà anche più avanti (cfr. 5.8 e 5.9) l'accento L+H\* sembra essere utilizzato dal parlante per porre in evidenza un elemento all'interno della frase; in questo caso la presenza di questa serie di accenti è determinata dalla struttura sintattica marcata, ma non si può parlare di formazione di frasi intermedie differenti: non è infatti possibile distinguere la presenza di alcun fenomeno fonetico (ad esempio una pausa o un allungamento finale) associabile con la presenza di un tono di confine di frase intermedia dopo ogni istanza di L+H\*. E' comunque possibile affermare che gli accenti si susseguono l'uno all'altro come in una serie di riaccentuazioni, così come avviene in strutture marcate quali quelle topicalizzate (Grice e Savino, 1997). In questi contesti, l'accento tonale può essere “copiato” nel costituente successivo, benché sia realizzato in un campo di frequenze compresso. Nel nostro caso, non sembra essere presente alcun tono di confine e la compressione sembra corrispondere ad una discesa costante, caratteristica della declinazione fonetica osservata durante la fonazione, all'interno dei costituenti prosodici<sup>9</sup>.

L'unico punto in cui sembra possa essere presente un tono di confine intermedio ci sembra essere dopo la parola 'diciannove': sebbene non vi siano allungamenti (o, ovviamente, pause), infatti, si nota una differenza di ampiezza del *range* di frequenze usato nella parte successiva. L'andamento che identifica l'interrogativa si realizza proprio nella parte successiva dell'enunciato, sulla parola 'tu', ma, data la brevità della parola sulla quale si realizza la modulazione tonale, non si può dire se la mancanza di compressione sia dovuta ad una caratteristica intrinseca nella realizzazione del *pattern* melodico (che prevede anche toni di confine L-H%) oppure se ciò sia dovuto alla presenza di un tono confine precedente e ad una scelta di *range* diversa che dipende dal *phrasing* più che dal *pattern* tonale specifico. Ad ogni modo, l'unico punto nell'enunciato nel quale si possa ipotizzare la presenza di un tono di confine di frase intermedia è proprio prima di 'tu', dove peraltro si potrebbe realizzare solo un confine basso L- (in linea con le convenzioni ToBI l'incertezza nella trascrizione è resa esplicita tramite il simbolo '?')<sup>10</sup>.

<sup>9</sup> Non ci sembra quindi possibile ipotizzare la presenza di *downstep*, cioè di un fenomeno fonologico che provoca un abbassamento del valore di F0 su specifici accenti tonali (per una panoramica, cfr. Ladd, 1996): non viene quindi utilizzato il diacritico '!' per indicare l'occorrenza del fenomeno (Beckman & Ayers, 1997). Al contrario, il graduale restringimento del campo di frequenze durante la produzione di un enunciato può essere ascrivito ad una diminuzione graduale della pressione subglottale durante la fonazione con un conseguente abbassamento della frequenza di vibrazione delle corde vocali (Gussenhoven, 2004).

<sup>10</sup> L'accento iniziale del sintagma è particolarmente prominente. Un enunciato di questo tipo è indicativo di come sia problematico individuare l'accento nucleare, e di come, sulla base della prominenza, più di un elemento possa svolgere questo ruolo nonostante sia seguito da altri accenti (cfr. 5.3 per una discussione sulla possibilità di accenti postnucleari).

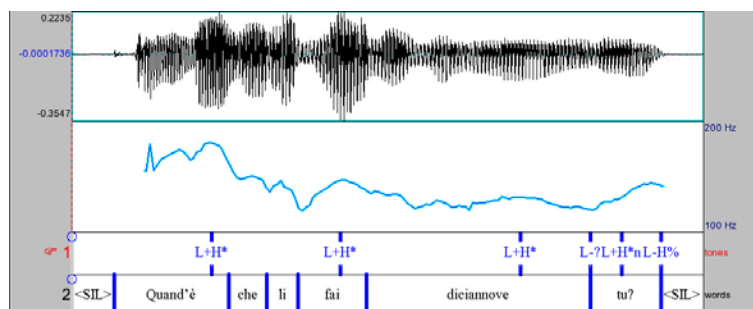


Figura 6: mossa Query-Wh del parlante 'Ma': "Quand'è che li fai diciannove tu?"

## 5.2 Query-Wh: proposta tra due alternative

Alcune delle domande aperte si configurano come proposte tra due alternative che il parlante espone esplicitamente all'interlocutore perché le ritiene egualmente plausibili nella situazione in cui si trova; queste vengono proposte, quindi, come elementi tra i quali operare una scelta.

Le figure 7 e 8 mostrano due enunciati identici dal punto di vista sintattico e molto simili a livello melodico: essi infatti differiscono solamente per la presenza del tono di confine discendente L-L% nel primo e ascendente L-H% nel secondo. In entrambi gli enunciati la prima delle due alternative è messa in rilievo attraverso un aumento repentino di F0 sulla sillaba accentata (la sillaba *de-* di 'destra') che porta alla configurazione della categoria L+H\*, mentre sulla seconda alternativa si crea un accento discendente H+L\* per la realizzazione del quale i valori di F0 iniziano ad abbassarsi subito dopo l'accento in corrispondenza della prima alternativa.

A seconda della distanza che intercorre tra i due accenti tonali, la discesa di F0 può essere identificata o meno sulla sillaba tonica della seconda alternativa: nell'esempio della figura seguente infatti, l'aumento del numero di sillabe che intercorrono tra i due accenti tonali porta ad una discesa molto più graduale rispetto a quanto visibile nelle figure 7 e 8. Nella figura 9 infatti il *pitch accent* sulla seconda alternativa potrebbe anche essere identificato come L\* dato che non si presenta alcuna discesa evidente all'interno del nucleo sillabico della sillaba tonica.

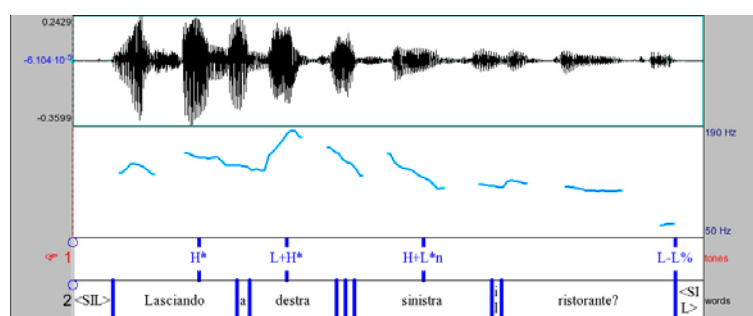


Figura 7: mossa Query-Wh del parlante 'Ma': "Lasciando a destra o a sinistra il ristorante?"

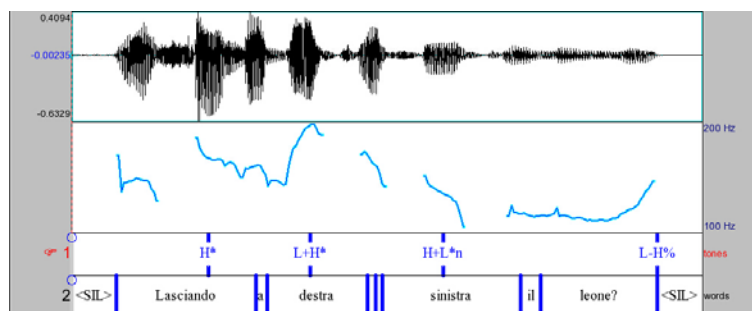


Figura 8: mosca Query-Wh del parlante 'Ma': “Lasciando a destra o a sinistra il leone?”.

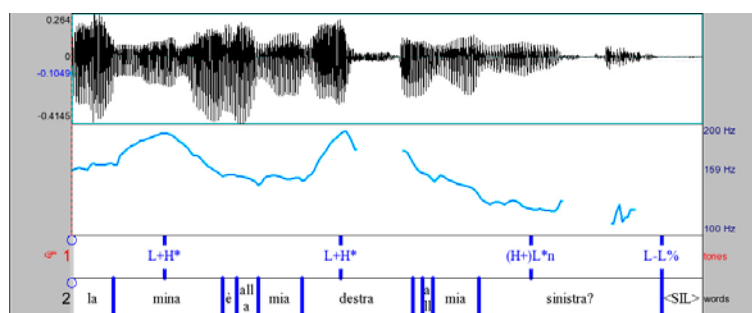


Figura 9: mosca Query-Wh del parlante 'Ma': “La mina è alla mia destra o alla mia sinistra?”

### 5.3 Query-YN: domande polari

Le domande polari presentano un contorno nucleare con caratteristiche simili a quello evidenziato per le domande aperte: esso è infatti composto da un accento tonale discendente  $H+L^*$  ed un tono di confine ascendente  $L-H\%$ <sup>11</sup>. La figura 10 mostra una mosca conversazionale nella quale si realizza tale andamento melodico: F0 ha un livello costante sulle sillabe pretoniche e discende in corrispondenza del nucleo della sillaba *li-* di 'linee', alla quale l'accento è associato; il tono tende poi verso l'alto solo sulla parte finale dell'enunciato, indicando la presenza dei toni di confine  $L-H\%$ .

<sup>11</sup> Secondo Romano (1998), la domanda totale, in locutori provenienti da un'area contigua a quella in esame in questa sede, è caratterizzata da “un andamento ondulato con picco opzionale sulla vocale tonica del verbo, da un andamento piatto nella parte seguente (vocale tonica compresa) e da un picco con massimo sulla parte finale dell'ultima vocale della frase.” (Romano, 1998: 36-37); il confronto fra gli andamenti intonativi riportati dall'autore e quelli proposti da noi in questa sede mostra che vi è una fortissima somiglianza tra i *pattern* descritti. Nonostante l'impossibilità di effettuare un confronto dal punto di vista dell'analisi fonologica, possiamo comunque dire che l'ascesa (verso valori medi), che in termini autosegmentali-metrici abbiamo qui correlato alla presenza di toni di confine  $L-H\%$ , sembra essere una caratteristica tipica delle domande polari in informatori provenienti dalla zona centro-settentrionale del Salento.

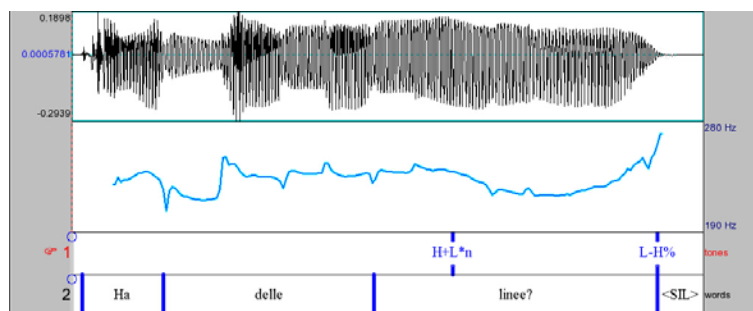


Figura 10: mosca Query-YN del parlante 'Ba': "Ha delle linee?".

In figura 11 è riportato invece un esempio nel quale l'oggetto della domanda, ossia il sintagma nominale *'la casa del bigné'*, viene dislocato a sinistra e richiamato nell'enunciato attraverso un pronome clitico. Tale sintagma forma una frase intermedia indipendente, composta da un accento tonale ascendente  $L+H^*$  particolarmente prominente, associato alla sillaba accentata *-gnè* di *'bigné'*; l'ascesa del *pitch accent* è particolarmente evidente e si realizza con un'escursione di  $F_0$  molto ampia, che pone in particolare evidenza l'elemento all'interno della produzione. L'elemento può essere considerato nucleare all'interno del sintagma intermedio, delimitato in contesti analoghi da un tono basso (cfr. anche 5.9 per altri casi di topicalizzazione). Tuttavia, nell'esempio in esame, la presenza di una parola tronca può indurre particolari fenomeni di "aggiustamento" circa l'allineamento degli eventi tonali: in particolare, il picco accentuale risulta leggermente anticipato rispetto alla fine della sillaba ed un apprezzabile abbassamento di  $F_0$  è visibile solo dopo il confine sillabico. In questo caso, quindi si potrebbe sostenere che il tono basso può essere stato troncato (come ad esempio, suggerisce la trascrizione in figura) o che la sua realizzazione fonetica si estenda dopo il confine stesso.

Nella parte restante dell'enunciato il campo di frequenze del parlante si comprime e si realizza il *pattern* nucleare di una domanda polare, analogo a quello già descritto in figura 10: il contorno di  $F_0$  discende sulla sillaba tonica di *'destra'*, coerentemente con la presenza di un accento tonale  $H+L^*$ , e risale poi verso valori medi sull'ultima parte dell'ultima sillaba, per via della specificazione tonale di confine  $L-H\%$ .

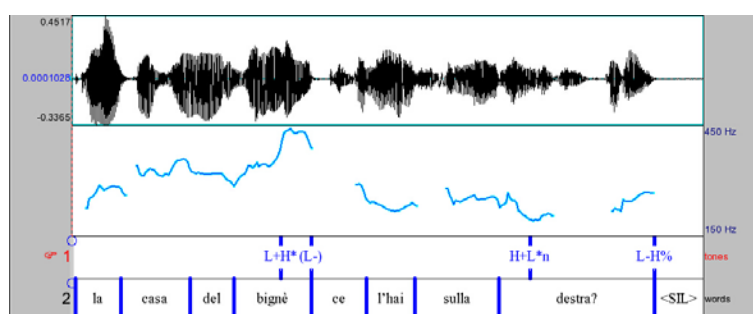


Figura 11: mosca Query-YN del parlante 'Ba': "La casa del bigné ce l'hai sulla destra?".

L'andamento melodico dell'enunciato riportato in figura 12, pur essendo frutto dell'implementazione delle stesse categorie accentuali dell'enunciato precedente, si realizza in maniera foneticamente differente. L'accento nucleare sul verbo *'vede'* corrisponde a una

discesa di F0 nella sillaba tonica (cioè ad una categoria fonologica come H+L\*), ma la presenza di un bersaglio tonale alto, allineato alla sillaba prenucleare, risulta mascherata dal graduale abbassamento di F0 che si realizza a partire dal picco precedente. Per questo motivo si è optato per l'etichetta (H+)L\* per l'accento nucleare, con la quale si identifica un accento che fonologicamente si realizza in maniera discendente, ma per il quale il tono alto di provenienza non è foneticamente evidente sulla sillaba prenucleare. L'andamento al confine è ascendente, verso valori medi di F0 come già osservato in precedenza, ed è etichettato come L-H%.

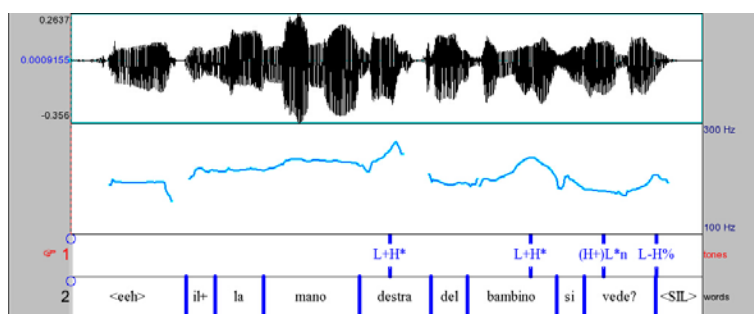


Figura 12: mosca Query-YN del parlante 'Be': “La mano destra del bambino si vede?”.

Nelle domande polari si è inoltre riscontrata la realizzazione di un *pattern* intonativo con accento nucleare ascendente L+H\*, analogamente a quanto accade nelle domande aperte (cfr. figura 6). Nell'esempio in figura 13, infatti, il parlante produce una domanda polare con innalzamento del valore di F0 sulla sillaba nucleare, configurando un accento L+H\* seguito da una valle che inizia in corrispondenza del confine destro del nucleo sillabico; il *pattern* termina poi con un tono di confine ascendente L-H% che occorre sulla parte finale dell'ultima sillaba dell'enunciato.

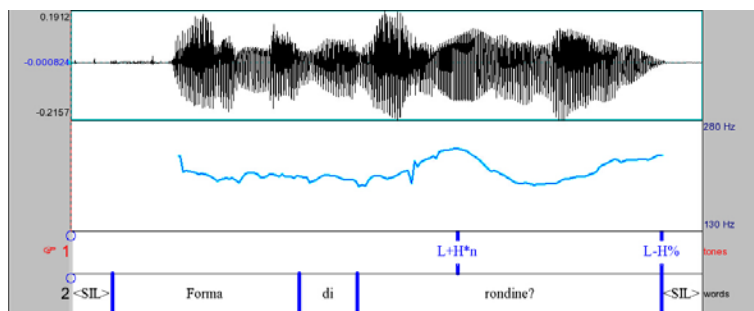


Figura 13: mosca Query-YN del parlante 'Be': “Forma di rondine?”.

Riassumendo, quindi, l'analisi di domande polari su informazione nuova ha permesso di individuare un accento nucleare discendente H+L\* seguito da un innalzamento dei valori di frequenza fondamentale in corrispondenza della parte finale del sintagma, inteso qui come dovuto alla presenza di una sequenza L-H% di toni di confine. In alcuni casi, si è anche osservata la presenza di un accento ascendente L+H\*, seguito dagli stessi toni di confine.

Nel caso di focalizzazione su un elemento in posizione non finale, l'andamento melodico si presenta con caratteristiche in buona parte analoghe, ma permette anche di

osservare interessanti eventi tonali in posizione postnucleare. Nelle figure 14, 15 e 16 sono riportati tre esempi di domande polari in cui l'accento nucleare non si trova sull'ultima parola dell'enunciato. Come si può osservare da questi esempi, il focus della domanda è comunque realizzato con un accento H+L\* ed è seguito, in tutti i casi, da un evento tonale ascendente associato all'ultima sillaba accentata del sintagma; negli esempi, l'innalzamento si osserva sulla sillaba *an-* di 'anche' (figura 14), su *l'om-* di 'ombra' (figura 15) e su *-ma-* di 'immagine' (figura 16).

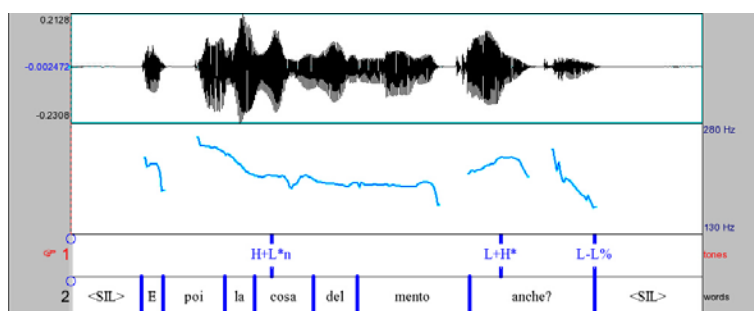


Figura 14: mossa Query-YN del parlante 'Be': “E poi la cosa del mento anche?”.

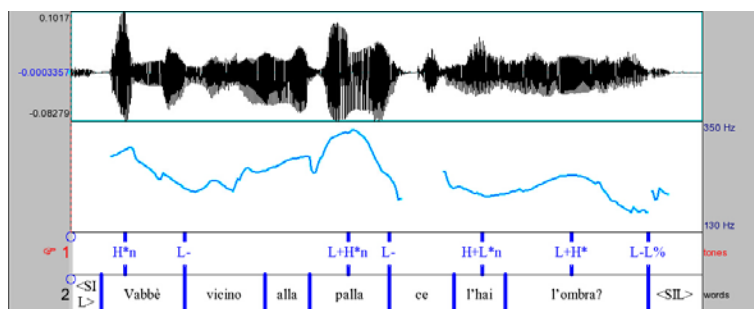


Figura 15: mossa Query-YN del parlante 'El': “Vabbè, vicino alla palla ce l'hai l'ombra?”.

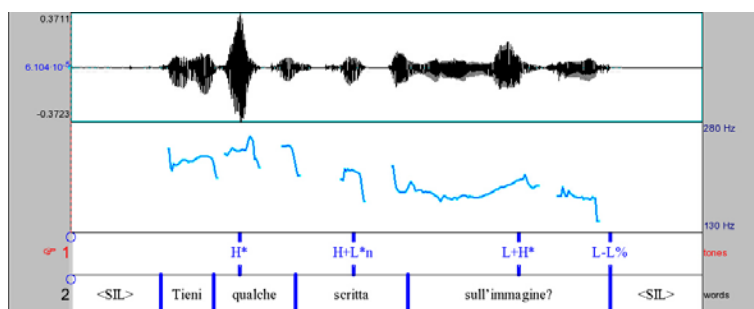


Figura 16: mossa Query-YN del parlante 'Be': “Tieni qualche scritta sull'immagine?”.

Una possibile analisi di contorni di questo tipo consiste nell'ipotizzare che, nel caso di focus non finale, si realizzi un accento tonale ascendente postnucleare e i toni di confine siano bassi. In base ad un'analisi di questo tipo, i parlanti utilizzerebbero quindi l'accento nucleare discendente in posizione di focus e marcherebbero la prominza postnucleare con

un accento ascendente, utilizzato anche in altri contesti per assegnare particolare prominenza – si vedano le trascrizioni fornite nelle figure. Tuttavia, contorni melodici di questo tipo, fanno riflettere sulla natura di eventi tonali ascendenti in posizione diversa da quella solitamente riscontrata, cioè la posizione finale assoluta. Gussenhoven (2002) ha infatti osservato che un tratto tipico delle domande polari consiste nella presenza di toni alti-ascendenti, sotto forma di specificazione tonale di confine o di accento. In questo senso, l'evento tonale LH, oltre ad essere analizzato come semplice accento ascendente, potrebbe anche essere analizzato come evento tonale dovuto ad un fenomeno di associazione secondaria delle specificazioni di confine tipiche delle domande sino ad ora analizzate (L-H%)<sup>12</sup> – per una discussione più approfondita, si veda il paragrafo 5.8.

#### 5.4 *Align: controllo dell'attenzione del partner*

La mossa Align consiste in un controllo dell'attenzione del partner ed ha l'unico obiettivo di allineare le posizioni tra i due interlocutori e di capire se il partner si trova nel punto atteso dall'interlocutore per poter iniziare un nuovo gioco dialogico. Nei dialoghi presi in esame, le mosse Align si riducono solitamente ad un'espressione quale 'giusto' o 'okay'<sup>13</sup>.

La mossa Align riportata in figura 17 presenta un andamento complessivamente ascendente: il contorno melodico è composto infatti da un accento alto L+H\* che si estende sulla prima sillaba senza significative variazioni di F0<sup>14</sup> ed un tono di confine che sale verso valori di F0 ancora più alti - senza formare alcuna valle tra i due eventi - motivando la scelta di un'etichetta H-H%.

<sup>12</sup> Sino ad ora, è stato riscontrato un unico esempio di accento postnucleare discendente. A parte il fatto che non sembra si tratti di un contorno tipico, è interessante notare che, coerentemente con l'ipotesi di Gussenhoven (2002), in questo caso i toni di confine sono L-H% e, quindi, una fase ascendente di F0 è comunque garantita.

<sup>13</sup> Si è notato che vi è una certa ambiguità nell'interpretazione di un contorno intonativo quale: “*Tu sei alla mina, giusto?*”: il problema è capire se tale produzione sia composta da due mosse differenti, ossia una mossa Check e una mossa Align, oppure se sia da considerare una mossa Align nel suo complesso. Tecnicamente la mossa Align comprende solo l'allineamento vero e proprio, quindi solo quella parte di enunciato che serve per controllare l'attenzione del partner e ricevere il suo feedback; viene però naturale pensare che la richiesta di conferma della posizione sia necessaria in casi simili a quello in esame per la corretta verifica della posizione. Nel presente paragrafo analizziamo quella che è indicata in letteratura come mossa Align, rimandando al paragrafo 5.5, sulle richieste di conferma, la discussione riguardante la prima parte dell'enunciato.

<sup>14</sup> Sulla base delle caratteristiche descritte, il tono potrebbe anche essere analizzato come appartenente ad una diversa categoria fonetica, H\*. Tuttavia, in questo stadio di analisi, non abbiamo dati sufficienti per ipotizzare l'esistenza di due categorie diverse, L+H\* ed H\*. Utilizziamo quindi la prima etichetta per l'analisi fonologica, ascrivendo a fattori fonetici la minima fase di innalzamento osservata negli esempi di mosse Align.



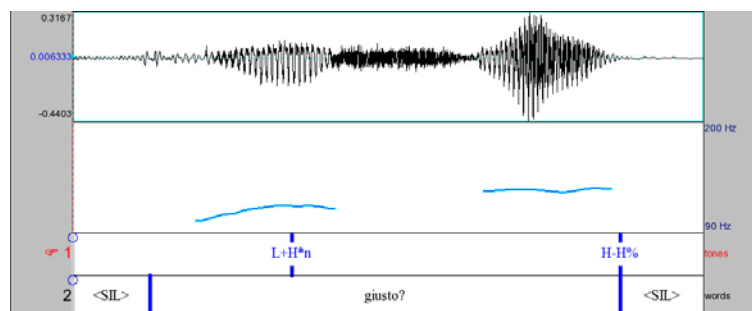


Figura 17: mosca Align del parlante 'Ba': "Giusto?".

### 5.5 Check: richiesta di conferma

La mosca Check è una mosca molto particolare e a volte anche ostica da riconoscere. La difficoltà sta nel fatto che, per sua natura, essa serve a chiedere conferma riguardo ad un'informazione condivisa; la differenza sommaria tra una mosca Check e una mosca Query-YN è proprio nella novità, o meno, dell'informazione all'interno dello scambio dialogico: l'etichetta Check si attribuisce prevalentemente a domande incentrate su argomenti per i quali uno scambio di informazioni è già avvenuto nei passaggi precedenti del dialogo. Inoltre una mosca Check, proprio perché rappresenta una richiesta di conferma, può presentarsi anche come parte integrante di una mosca Align, configurandosi quindi come una richiesta di conferma della posizione nella quale l'interlocutore si trova – si veda nota 12.

Le mosse Check presentano contorni intonativi nucleari molto simili a quelli delle domande polari, cioè formati da un accento tonale discendente  $H+L^*$  e un tono di confine ascendente  $L-H\%$ . Questa somiglianza si può notare dall'esempio riportato in figura 18: il *pattern* melodico di questo enunciato è costituito da un accento prenucleare ascendente  $L+H^*$  posizionato sulla sillaba accentata *-va-* di 'davanti'; il *pattern* nucleare si trova poi sull'ultima parola dell'enunciato e si realizza con una discesa del tono allineata con la sillaba *-bi-* di 'bambino', sulla quale si realizza un accento tonale  $H+L^*$ , e un leggero innalzamento finale che interessa la parte finale dell'ultima sillaba e che si configura come dovuta ad una sequenza di toni. Come si può notare, tale realizzazione è perfettamente compatibile con quella già riscontrata per le domande polari.

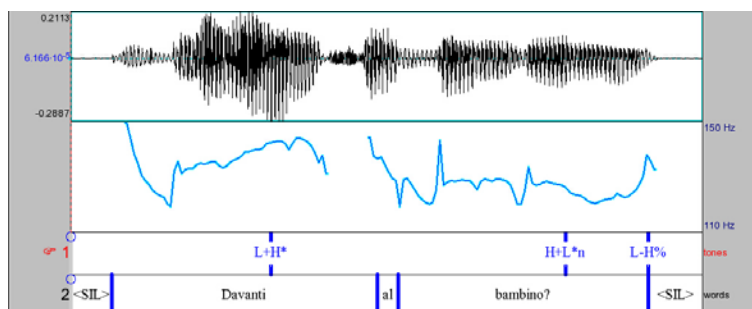


Figura 18: mosca Check del parlante 'Ma': "Davanti al bambino?".

Anche nel caso delle mosse Check, quindi, l'accento nucleare è analizzato come  $H+L^*$  e seguito da una specificazione di confine  $L-H\%$ . Come già osservato per le Query-YN, in

caso di focalizzazione su una parola in posizione non finale, l'accento nucleare può essere seguito da un fenomeno postnucleare ascendente associato ad una sillaba accentata e da toni di confine bassi – si veda l'esempio nella figura seguente e la discussione in 5.3 e 5.8.

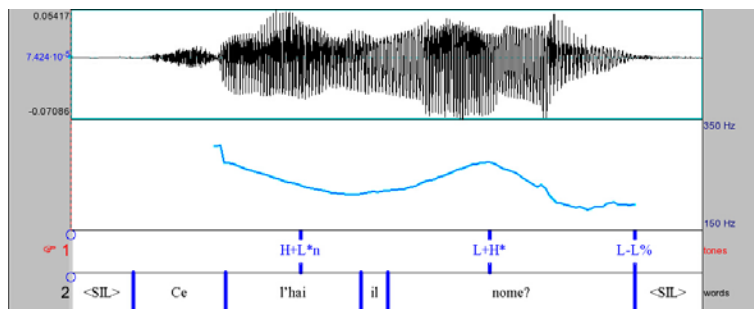


Figura 19: mossa Check del parlante 'El': “Ce l'hai il nome?”.

Nella classificazione degli enunciati di tipo Check è da tenere comunque in considerazione soprattutto un aspetto funzionale, legato a quanto il parlante ritiene che l'informazione per la quale chiede conferma sia effettivamente corretta. All'interno della categoria Check si impone infatti un'ulteriore distinzione legata al grado di sicurezza del parlante, come già è stato rilevato nell'analisi intonativa di altre lingue e varietà di italiano (cfr. Grice *et alii*, 1995). Distinguiamo perciò i *confident checks*, ossia le richieste di conferma con le quali il parlante trasmette un alto grado di sicurezza che l'informazione per la quale richiede una conferma sia corretta, e i *tentative checks*, ossia le richieste che corrispondono ad un basso grado di sicurezza. Le differenze tra i due tipi sono comunque molto sottili e legate a caratteristiche sia fonetiche, cioè di implementazione delle categorie, che fonologiche, ossia di uso di categorie differenti. Si osservino a tal proposito i due enunciati nelle figure seguenti: l'enunciato in figura 20 corrisponde ad un *confident check*, mentre quello in figura 21 corrisponde ad un *tentative check*; entrambi gli enunciati sono stati prodotti dallo stesso parlante in due turni susseguenti dello stesso dialogo.

Confrontando le due produzioni si può notare come abbiano molto in comune, in particolar modo nella parte iniziale dell'enunciato: l'accento prenucleare è posto sulla prima sillaba della produzione, la quale inizia con un valore alto di F0 nell'intervallo di frequenze del parlante e continua declinando verso il basso. Nella produzione in figura 20 questa declinazione è più ripida rispetto all'enunciato in figura 21: ciò porta alla realizzazione di un accento tonale discendente H+L\* sulla sillaba accentata *mam-* di 'mammole' e di un tono di confine basso-discendente L-L%, in modo molto simile a quanto si osserva per enunciati affermativi (cfr. per la varietà leccese, Romano, 1999; Stella, 2007). L'enunciato in figura 21, invece, mantiene un tono più sostenuto creando un contorno più simile ad un *plateau*, con la conseguenza che la discesa dovuta alla presenza dell'accento tonale discendente sulla sillaba nucleare *-re-* di 'amarene' non è così evidente: per questo motivo si è attribuita l'etichetta (H+)L\* all'accento. Infine, il tono di confine è ascendente L-H%, cosa che contraddistingue nettamente questa produzione rispetto alla precedente. Ci sembra anche rilevante sottolineare una differenza nel parametro intensità: una caduta dei valori di intensità si verifica sulla parola 'mammole' nel *confident check*, mentre si può notare una maggiore intensità sulla parola 'amarene' nel *tentative check*.

Le osservazioni evidenziate nel confronto tra i due elementi portano alla definizione di alcune caratteristiche melodiche che sembrano poter rendere conto delle differenze tra un

*tentative check* ed un *confident check*: sebbene entrambe abbiano la configurazione propria di una domanda polare, specialmente nella realizzazione del tono nucleare, il quale appartiene sempre alla stessa categoria fonologica H+L\*, la presenza di un tono di confine ascendente in luogo del tono discendente, percettivamente, trasmette un maggior senso interrogativo, così come un minor grado di sicurezza da parte del parlante; per contro il tono di confine discendente - il quale, anche se riscontrato in produzioni interrogative (cfr. Query-Wh), è tipico delle dichiarative - trasmette un maggior senso di ovvietà all'enunciato.

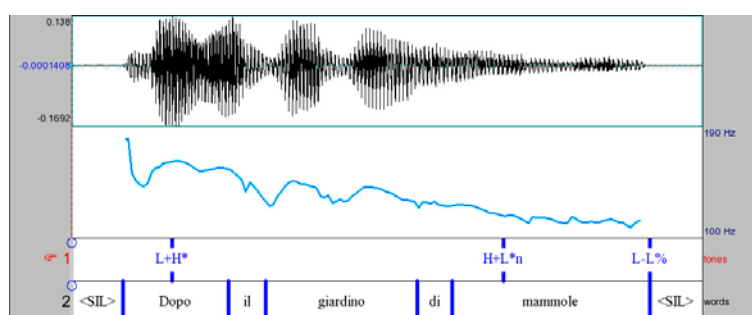


Figura 20: mosca Check del parlante 'Ma': "Dopo il giardino di mambole".

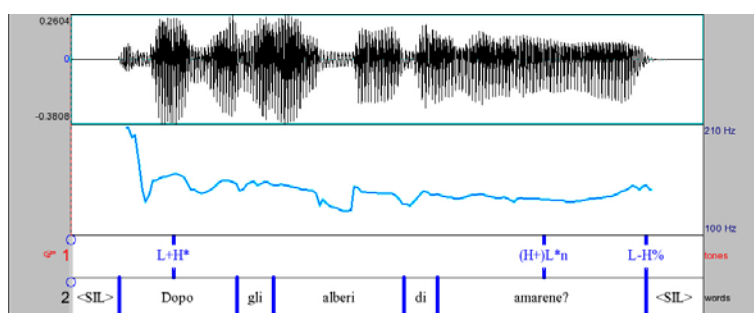


Figura 21: mosca Check prodotta dal parlante 'Ma': "Dopo gli alberi di amarene?".

In figura 22 è riportato un altro esempio di mosca Check che il parlante produce con una certa enfasi, focalizzando l'elemento verbale. In tal caso il parlante produce un andamento complessivamente discendente che parte da un valore molto alto del campo di frequenze<sup>15</sup> e discende fino a raggiungere i valori più bassi. L'andamento melodico è prodotto inoltre con una leggera ascesa che copre lo spazio della sillaba prenucleare -sa- di 'risalire' e una discesa che è parte caratterizzante dell'accento nucleare ed è perfettamente allineata con la sillaba nucleare -li-; tale realizzazione, dal punto di vista percettivo, conferisce una leggera enfasi alla richiesta, soprattutto grazie alla leggera ascesa prenucleare. L'effetto è quindi di

<sup>15</sup> Si noti che il parlante 'Ba', che ha prodotto la mosca in esame, ha un tono di voce molto basso e nel parlato non-enfatico non supera quasi mai i 120-150 Hz. In questa mosca, inoltre, la produzione inizia da un valore molto alto di F0 sulle prime sillabe dell'enunciato, dove sembra configurarsi un accento tonale H\*, categoria che però non è stata riscontrata in contesti simili e della cui esistenza non si hanno chiare evidenze nella varietà leccese dell'italiano.

una richiesta di conferma veicolata dalla presenza di un focus ristretto, o addirittura contrastivo<sup>16</sup>, sull'elemento bersaglio della richiesta stessa.

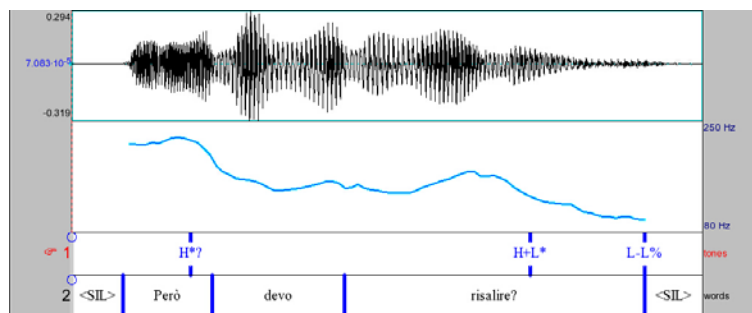


Figura 22: mossa Check del parlante 'Ma': "Però devo risalire?".

Nella varietà leccese dell'italiano, inoltre, si è attestato l'uso del focus contrastivo in enunciati di tipo Check, con i quali il parlante chiede all'interlocutore una conferma riguardo qualcosa che risulta evidente dallo stato del gioco o che crede esatta per motivi a lui ovvi; tale uso del focus contrastivo è già stato individuato in altre varietà di italiano, quali il barese (Savino, 1997) e il pisano (Gili Fivela, 2004)<sup>17</sup>. La produzione in figura 23 riporta proprio un caso simile. In linea generale si può dire che, sovente, l'occorrenza di questo tipo di focalizzazione, in relazione anche alle caratteristiche pragmatiche dell'enunciato, mostra un campo di frequenze espanso, che porta il parlante a produrre accenti tonali raggiungendo picchi molto alti e trasmettendo, oltre al contrasto, anche una certa enfasi<sup>18</sup>. La categoria fonologica utilizzata per identificare l'accento contrastivo

<sup>16</sup> Per un chiaro accento contrastivo si veda l'esempio in figura 23. Tuttavia, l'accento riportato in figura 22 presenta caratteristiche fonetiche che lo rendono particolarmente enfatico. Confrontando l'andamento tonale dell'esempio in figura 22 con quello in figura 20, il quale rappresenta invece un caso di focus ampio, si può notare come l'accento nucleare venga prodotto in maniera differente: mentre nella produzione con focus ampio l'intero contorno melodico, che include l'accento nucleare, è discendente, nella produzione con focus ristretto il contorno intonativo risale in modo evidente prima della discesa in corrispondenza della sillaba nucleare. Tale risalita sembra funzionale ad aumentare la percezione del picco alto di F0 e conseguentemente la percezione della discesa sulla sillaba accentata. L'analisi uditiva, infatti, porta a considerare l'accento come un evento tonale che possa anche indicare contrasto. In questo caso, la realizzazione di un *trailing tone* H+ a livelli alti di F0 sortirebbe, nella varietà leccese, effetti percettivi simili a quelli osservati in pisano. Nella varietà di Pisa, infatti, la percezione di stimoli nei quali l'altezza del tono H+ sia stata incrementata si avvicina alla percezione di accenti contrastivi piuttosto che di accenti neutri (Gili Fivela, 2006).

<sup>17</sup> Nella varietà di Pisa, in particolare, è stato osservato l'uso del contorno contrastivo (H+L\* L-L%) da parte dell'*instruction follower*, per via del suo ruolo nello scambio dialogico (nel caso del *giver*, infatti, l'accento contrastivo sarebbe seguito da una diversa combinazione di toni di confine). Nella varietà di Lecce sembra valere la stessa correlazione (si vedano gli esempi nelle figure 22 e 23).

<sup>18</sup> Nella notazione ToBI applicata all'inglese (Beckman & Ayers, 1997), l'etichetta %H è usata per indicare la presenza di un tono di confine iniziale alto, il quale è utilizzato dal

nell'enunciato in figura 23 è  $H^*+L$  e consiste in una sequenza tonale ascendente-discendente che investe l'intero nucleo della sillaba accentata relativa all'elemento posto in contrasto (si rimanda a 5.8 per la discussione relativa alla scelta di questa etichetta accentuale, ma si veda anche la discussione circa la figura 28). La parte restante del contorno dell'esempio è prodotta realizzando un accento tonale discendente  $H+L^*$  ed un tono di confine basso-discendente  $L-L\%$ ; entrambi questi eventi tonali occorrono sull'ultima sillaba dell'enunciato e, complice la declinazione globale dell'andamento melodico, risultano fusi in un unico e più ripido abbassamento del valore di  $F_0$  e non possono essere chiaramente differenziati.

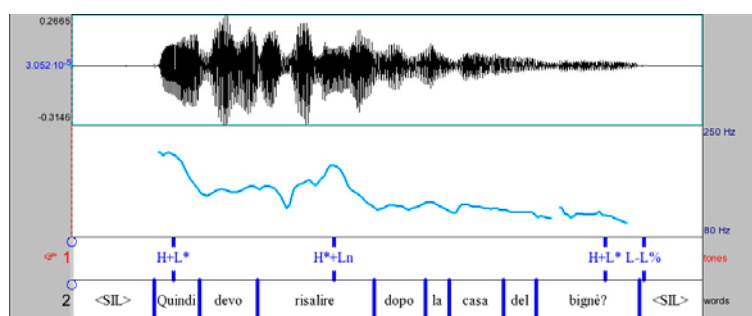


Figura 23: mossa Check del parlante 'Ma': “Quindi devo risalire dopo la casa del bigné?”.

Un altro andamento melodico riscontrato nel leccese per la realizzazione della mossa Check prevede la realizzazione di un accento nucleare in *downstep* o prodotto in un intervallo di frequenze compresso. Come si può osservare nell'enunciato in figura 24, infatti, il *pattern* accentuale è formato da un accento prenucleare ascendente  $L+H^*$  - che si associa alla sillaba *-là* di 'Liola' e che mette in evidenza l'elemento topicalizzato - mentre l'accento nucleare si realizza come una sorta di *plateau* o come un leggero innalzamento in un intervallo di frequenze compresso sulla sillaba *-no-* di 'anonima': il fatto che il nucleo di tale sillaba corrisponda ad un andamento piatto giustificerebbe la scelta di un'etichetta  $!H^*$  per indicare un *pitch accent* nucleare che viene compresso, indipendentemente dalla presenza di declinazione. Tuttavia la compressione dell'intervallo di frequenze può essere dovuta alla posizione dell'accento, come già osservato in precedenza. Nella parte di contorno che segue la realizzazione dell'accento nucleare,  $F_0$  scende ulteriormente fino a raggiungere i valori più bassi dell'intervallo di frequenze, indicando la presenza di un tono di confine  $L-L\%$ .

Non è infrequente nelle mosse Check, inoltre, l'uso di *tag questions*. Nell'esempio riportato in figura 25,  $F_0$  inizia la sua ascesa subito dopo il tono di confine di sintagma intermedio, per via della presenza di un accento tonale ascendente  $L+H^*$  e di un tono di confine  $H-H\%$ .

---

parlante per trasmettere ridondanza o sorpresa; anche in questo caso, si è notato che la presenza di un valore di  $F_0$  iniziale così alto è usato con lo stesso fine. Si potrebbe quindi usare l'etichetta ‘%H’ sia per l'esempio in figura 22, che per gli esempi nelle figure 23 e 34.

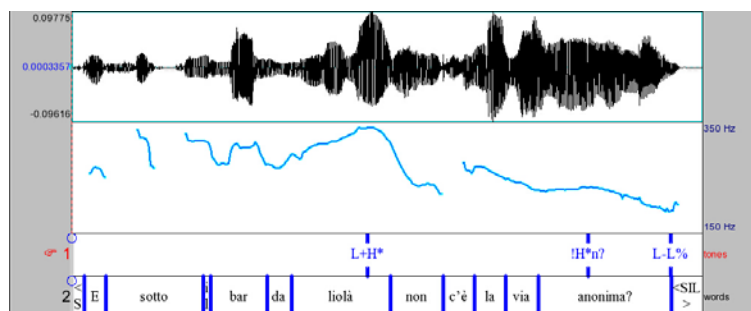


Figura 24: mosca Check del parlante 'El': “*E sotto il bar da Liolà non c'è la via anonima?*”.

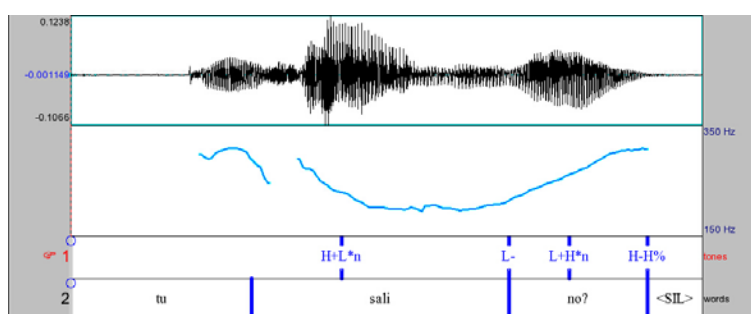


Figura 25: mosca Check del parlante 'El': “*Tu sali, no?*”.

Infine, come già anticipato in 5.4, una mosca Check può precedere un Align. Nell'enunciato riportato in figura 26, la parte che precede la mosca Align vera e propria consiste in una richiesta di conferma della posizione. Questa richiesta di conferma ricalca in tutto e per tutto l'andamento intonativo già riscontrato nelle domande polari per le quali si è individuato un accento nucleare discendente H+L\* ed un tono di confine basso, che in questo caso si realizza come tono di confine di frase intermedia L-, il quale, comunque, trasmette un certo grado di sicurezza del parlante allo stesso modo dei *tentative checks* terminanti con un tono di confine L-L%..

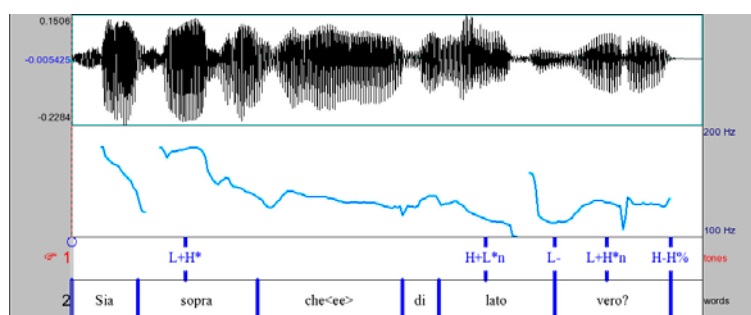


Figura 26: mosca Align del parlante 'Ma': “*Sia sopra che di lato, vero?*”.

### 5.6 Object: obiezione

Le mosse Object occorrono prevalentemente nei dialoghi Map Task, quando il parlante si rende conto dell'incongruenza tra le mappe, nella maggior parte dei casi manifestando la propria sorpresa attraverso una domanda eco.

Un esempio di domanda eco è riportato in figura 27; in questo tipo di domanda il senso di incredulità è trasmesso attraverso un accento nucleare ascendente L+H\* che subisce però un forte ritardo nell'allineamento degli eventi tonali, tant'è che il picco dell'ascesa si trova oltre il confine destro della sillaba accentata -no- di 'anonima' (tale ritardo è evidenziato attraverso il diacritico '>', previsto nella trascrizione ToBI per indicare questo tipo di fenomeno). Il tono di confine è etichettato come L-L%.

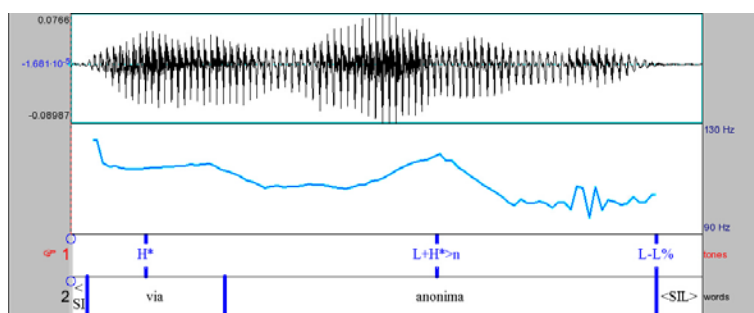


Figura 27: mossa Object del parlante 'Lu': “Via anonima?!”.

In figura 28 è riportato un altro tipo di mossa Object che sembra essere molto più vicina ad una produzione con focus contrastivo; in questo caso il valore di F0 sulla sillaba nucleare raggiunge valori molto alti su *stia-* di 'stiamo' per scendere poi sulle sillabe successive. Percettivamente, infatti, il parlante crea un focus contrastivo sul verbo, tant'è che l'enunciato può essere interpretato come “stiamo o non siamo nello stesso paese?”; il valore di F0 nel nucleo della sillaba accentata è molto alto e scende rapidamente all'interno del nucleo sillabico piuttosto che gradualmente, come invece succede nel caso delle realizzazioni dell'accento H+L\* riscontrato nella maggior parte delle domande polari. Per questa maggior salienza del tono alto all'interno del nucleo sillabico, la categoria fonologica alla quale appartiene tale realizzazione è stata identificata come H\*+L – si veda anche l'esempio in figura 23. L'altezza tonale diminuisce poi bruscamente e sulle sillabe post-toniche corrisponde ad un *plateau* basso con una leggera prominenzia posizionata su -e- di 'paese' ed un tono di confine discendente L-L%.

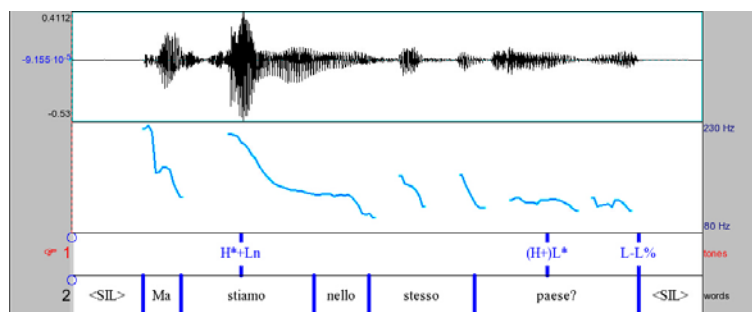


Figura 28: mossa Object del parlante 'Ma': “Ma stiamo nello stesso paese?”.

### 5.7 Continuation rise e domande ellittiche

Come già discusso in 5.2 i parlanti possono formulare domande anche proponendo due elementi tra i quali deve essere effettuata una scelta. La proposta di alternative e, in generale, l'articolazione di una domanda può anche corrispondere alla realizzazione di un contorno di continuazione. Nell'enunciato riportato in figura 29, ad esempio, il parlante presenta due possibilità di scelta all'interlocutore, producendo due differenti sintagmi intermedi, uno per ogni elemento: ogni sintagma contiene un accento ascendente L+H\* sulla sillaba tonica della parola che rappresenta l'alternativa proposta. In particolare, la prima termina con un tono di confine alto H-, tipico di un enunciato non concluso, che rimane in sospeso e che si suppone sia seguito da altro materiale; la seconda alternativa, finale di enunciato, termina con un tono di confine H-H%.

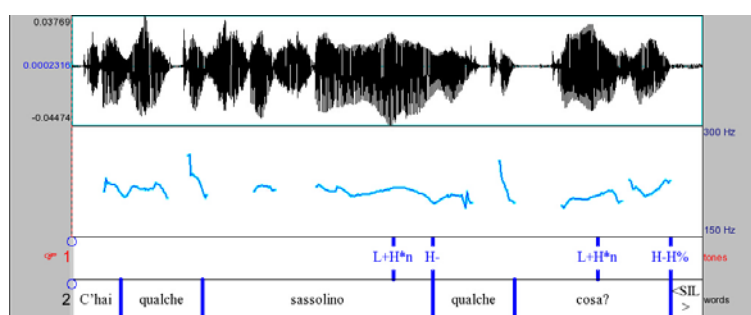


Figura 29: mosca Query-YN del parlante 'El': "C'hai qualche sassolino, qualche cosa?".

Nelle figure 30 e 31 sono riportati due esempi di domande ellittiche, le quali si configurano come delle proposte che il parlante fa al partner riguardo alla presenza o meno di un oggetto: inserendole nel loro contesto ed in relazione al gioco conversazionale in atto, si può ipotizzare che esse siano state prodotte come domande aperte, mentre, considerandole in isolamento, non è possibile stabilire se il parlante intenda formulare una domanda polare come "Hai le nuvole?" oppure una domanda aperta del tipo "Come sono le nuvole?".

I due contorni melodici sono composti da un accento tonale ascendente L+H\*, il quale investe l'intero nucleo della sillaba tonica partendo da un livello basso di F0, mentre differiscono per quanto riguarda i toni di confine. Nell'enunciato in figura 30 il livello di F0 raggiunto dall'accento L+H\* è mantenuto sulle ultime due sillabe dell'enunciato, formando un *plateau* alto che configura un andamento di confine H-H%; in figura 31, invece, dopo la realizzazione dell'ascesa, F0 scende progressivamente fino a raggiungere i valori bassi del campo di frequenze, indicando così la presenza di toni di confine L-L%. Anche se l'obiettivo di questo lavoro non è di fornire dati quantitativi, nel corpus di indagine abbiamo osservato una tendenza a preferire i toni di confine H-H% nella produzione di domande ellittiche.



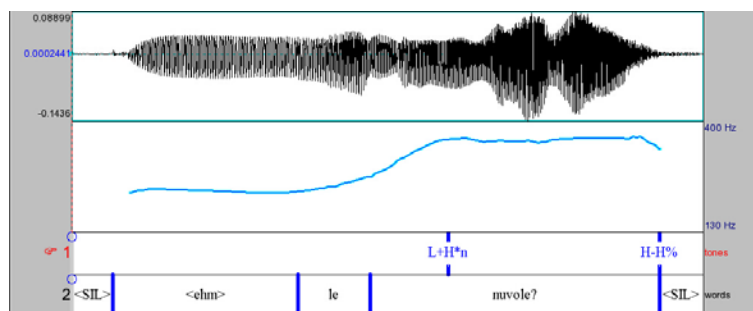


Figura 30: mosca Query-Wh del parlante 'El': "Ehm, le nuvole?".

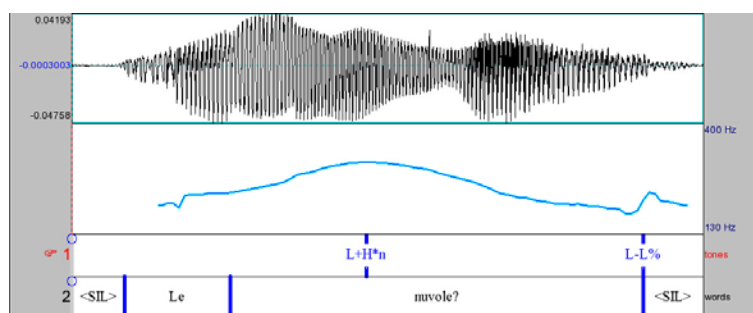


Figura 31: mosca Query-Wh del parlante 'El': "Le nuvole?".

### 5.8 Focalizzazione

L'osservazione dei contorni melodici delle mosse conversazionali che abbiamo considerato ci ha permesso di evidenziare alcune strategie utilizzate dai parlanti nei contesti di focalizzazione.

La presenza di un accento tonale ascendente su un determinato elemento permette al parlante di mettere in rilievo l'elemento stesso, sia quando esso rappresenti l'oggetto diretto della focalizzazione, sia quando segua un elemento focalizzato. Ad esempio, nella figura 6, qui ripetuta per comodità, è rappresentata una mosca Query-Wh realizzata come frase scissa: il costituente focalizzato sintatticamente è messo in evidenza grazie ad un accento ascendente L+H\*, seguito peraltro da altri accenti simili che rendono evidente la presenza di molteplici picchi di prominenza nell'enunciato.

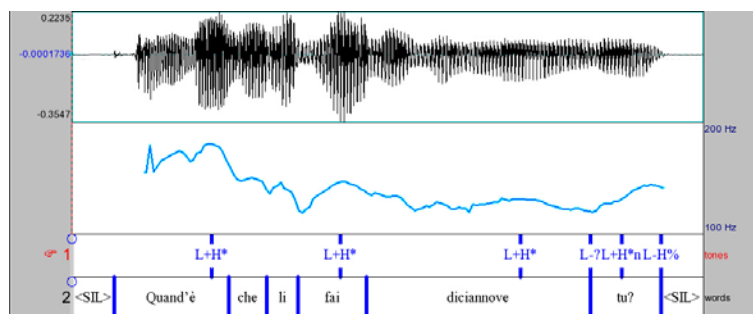


Figura 6: mosca Query-Wh del parlante 'Ma': "Quand'è che li fai diciannove tu?"

Inoltre, come descritto per le figure 14, 15 e 16, (per comodità riportiamo nuovamente la figura 14), la presenza di focalizzazione nel caso di domande sì-no (sia Query-YN che Check) corrisponde alla realizzazione dell'accento nucleare (H+L\*) in corrispondenza dell'elemento focalizzato, seguito da un fenomeno postnucleare ascendente associato all'ultima sillaba accentata e seguito da toni di confine bassi (L-L%).

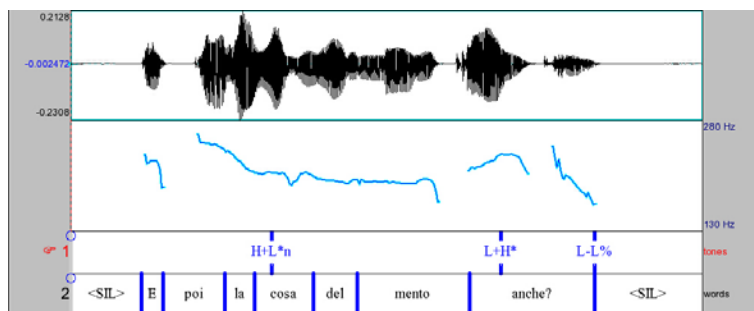


Figura 14: mossa Query-YN del parlante 'Be': "E poi la cosa del mento anche?".

Come già accennato, contorni di questo tipo possono essere analizzati come dovuti alla presenza di un accentto postnucleare ascendente e di toni di confine bassi, ma possono anche essere attribuiti alla presenza di un fenomeno di associazione secondaria delle specificazioni di confine riscontrate in domande funzionalmente equivalenti (L-H%).

Il meccanismo dell'associazione secondaria è stato inizialmente proposto nell'analisi del giapponese da Pierrehumbert & Beckman (1988). In giapponese, secondo le autrici, un tono alto è associato al margine sinistro del sintagma accentuale ed è anche legato da associazione secondaria ad una mora. A partire da questa proposta, l'esistenza di associazione secondaria è stata ipotizzata nell'analisi di diverse lingue (Grice, 1995; D'Imperio, 1999; Grice *et alii*, 2000; Grice *et alii*, 2005). Di particolare interesse in questa sede, è la proposta avanzata da Grice *et alii* (2000) in un'importante indagine sull'ungherese (standard e nella varietà della Transilvania), sul romeno, sul greco (standard e cipriota), sull'inglese e sul tedesco. E' stato infatti osservato che in queste lingue il tono di sintagma intermedio è realizzato in concomitanza con sillabe - accentate o non accentate - che si trovino nelle vicinanze della periferia del sintagma, per via di un meccanismo di associazione secondaria. In romeno, ad esempio, sono riportati due diversi andamenti melodici per la realizzazione delle domande polari con accentto nucleare in posizione non finale: in un caso (percepito come più marcato), l'accentto nucleare basso (L\*) è seguito da un *phrase accent* alto (H-) che, per via di associazione secondaria, occorre in corrispondenza di una sillaba accentata postnucleare ed è seguito da un tono di confine basso (L%); nell'altro caso (percepito come meno marcato), l'accentto nucleare basso (L\*) è seguito da un *phrase accent* basso (L-) e da un tono di confine alto (H%) che determina un repentino innalzamento in posizione finale assoluta. Nel caso del romeno, quindi, l'associazione secondaria dell'accentto di sintagma intermedio alto determina un picco di F0 allineato con l'ultima sillaba accentata. Questo fenomeno di associazione secondaria del *phrase accent*, come è stato anche osservato per il tedesco da Grice *et alii* (2005), genera delle prominenze che non indicano focalizzazione. In tedesco, ad esempio, il contorno H\*...L-L% può essere realizzato come H\*...L(\*)...L-L% per via dell'associazione secondaria dell'accentto di sintagma intermedio alla sillaba accentata precedente. La parola realizzata con accentto L\* è prominente ma non focalizzata.

Gli ultimi esempi descritti per il leccese sembrano compatibili con un fenomeno di questo tipo, in base al quale, però, l'intera specificazione L-H% del confine verrebbe associata all'ultima sillaba accentata del sintagma. Tale associazione renderebbe la parte finale del sintagma prominente, ma non interferirebbe con la realizzazione del focus in posizione 'iniziale' all'interno del sintagma. Nel caso del leccese, quindi, l'associazione secondaria coinvolgerebbe entrambi i toni di confine, e lascerebbe il posto ad un specificazione finale corrispondente a toni di confine bassi, che dovrebbe quindi essere prevista come soggiacente o sotto specificata. All'interno del sistema leccese, la scelta di associazione secondaria assolverebbe la funzione di evitare ambiguità all'interno del sistema, in quanto permetterebbe di non mantenere una specificazione tonale finale L-H% che, dopo un accento nucleare ascendente L+H\*, coinciderebbe con uno dei *pattern* tonali nucleari osservati per le domande sì-no (cfr. figura 13). Ulteriori osservazioni sono necessarie per poter indicare chiaramente quale sia l'ipotesi più adatta a spiegare i fenomeni in questione, fermo restando che un evento tonale ascendente con la caratteristica principale del *pitch accent*, cioè quella di fornire prominenza, è individuato in posizione postnucleare.

In ogni caso, nella varietà leccese di italiano, la prominenza tonale è spesso veicolata da un accento tonale ascendente come L+H\*. In relazione alla Query-YN riportata in figura 12, erano state osservate due prominenze accentuali ascendenti L+H\* che precedevano il *pattern* nucleare. In modo analogo, nella figura seguente è rappresentata una Query-Wh (“*e la mina è alla mia destra o alla mia sinistra?*”), preceduta da una mossa che somiglia molto ad un Align, anche se non rientra tra i comuni casi di allineamento tra i parlanti (“*Arrivo alla mina*”)<sup>19</sup>. L'esempio qui riportato permette di osservare la realizzazione di un accento L+H\* in corrispondenza del soggetto dell'enunciato ed anche all'interno di un contorno di continuazione, caratterizzato dalla presenza di un tono di confine alto (si veda anche 4.7).

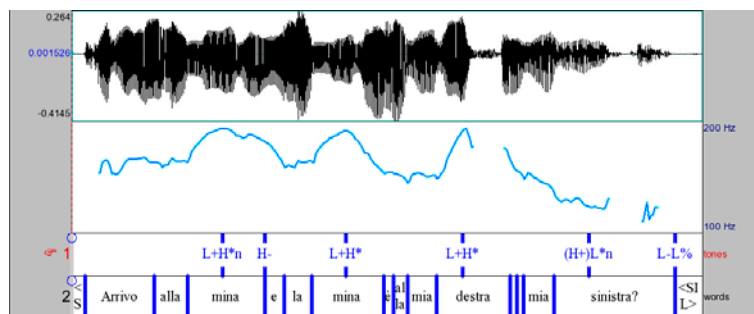


Figura 32: mossa Query-Wh del parlante 'Ma': “*Arrivo alla mina e la mina è alla mia destra o alla mia sinistra?*”.

Nel caso, invece, di realizzazione di richieste di conferma è stato osservato l'uso di un accento contrastivo, analizzato fonologicamente come H\*+L – si veda la figura 23. L'accento contrastivo è stato riscontrato anche in produzioni di frasi dichiarative, come si può notare dall'esempio in figura 33. Osservando il tracciato di F0, dal punto di vista fonetico l'accento è caratterizzato dalla presenza di tre bersagli tonali (basso-alto-basso) e presenta quindi un andamento globale di tipo ascendente-discendente. In attesa di effettuare

<sup>19</sup> Gili Fivela (2004) propone un'analisi dell'andamento melodico dello *start assessment*, sintagma che include informazione data che tuttavia deve essere evidenziata per permettere il corretto proseguimento dello scambio di informazioni.

misurazioni circa la latenza dei bersagli tonali che possano gettare luce sulla struttura interna dell'accento (ad esempio indicando se esso si componga di due o di tre bersagli tonali che sono realizzati ad una distanza stabile tra loro o rispetto a punti specifici della catena fonica), si propone di trascrivere l'accento come  $H^*+L$ , etichetta sufficientemente trasparente circa le caratteristiche fonetiche dell'accento e utile a distinguerlo chiaramente dagli altri eventi tonali riscontrati nel sistema fonologico sino a questa fase di analisi<sup>20</sup>.

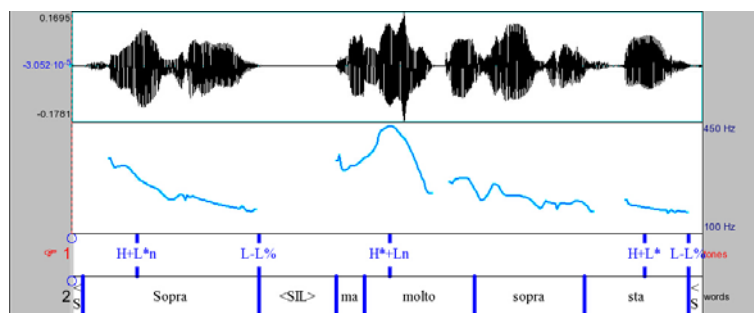


Figura 33: mosca Explain del parlante 'Be': "Sopra, ma molto sopra sta!".

Quindi, come già osservato per altre varietà di italiano (Savino, 1998; Gili Fivela 2004), nello scambio dialogico, l'uso di questo andamento melodico è riscontrato in enunciati che svolgono la funzione di frasi di modalità interrogativa e che sono interpretati come richieste di conferma. In leccese, come già osservato per il pisano (si veda la discussione in 5.5), questo *pattern* è riscontrato nelle mosse dei parlanti che, per via del loro ruolo nel dialogo Map Task, non sono in possesso di tutte le informazioni e non potrebbero quindi opporre con decisione il loro punto di vista circa l'argomento di discussione – ad esempio la posizione di un'icona rispetto al tracciato. In questi casi, quindi, l'intonazione induce l'ascoltatore ad effettuare un'inferenza simile a quella attesa nel caso di violazione delle massime conversazionali e di realizzazione di implicature conversazionali (Grice, 1975).

Come già osservato, vi sono anche casi in cui il parlante utilizza un accento tonale discendente come  $H+L^*$  per mettere in rilievo un elemento. Nell'enunciato in figura 34, infatti,  $H+L^*$  è realizzato con un'escursione di  $F_0$  molto elevata, per realizzare la quale il parlante aumenta la frequenza fondamentale della voce sulle ultime tre sillabe pretoniche, raggiunge il picco di  $F_0$  sull'ultima pretonica per poi scendere repentinamente ed

<sup>20</sup> Per una discussione circa la relazione tra gli aspetti fonetici e fonologici nella codifica degli eventi tonali di tipo accentuale si veda Gili Fivela (2006). Analizzando la varietà pisana di italiano, l'autrice prende spunto proprio dalle caratteristiche dell'accento contrastivo. Nella varietà pisana, il bersaglio basso che precede il picco tonale è allineato stabilmente rispetto al picco, così come riscontrato anche per il *trailing tone* basso, che lo segue. L'osservazione di caratteristiche di stabilità analoghe in relazione ad altri accenti caratterizzati da un picco tonale alto ( $H^*$ ) ed altre considerazioni per le quali si rimanda al lavoro suddetto, portano l'autrice a ritenere superflua, in una codifica fonologica, la specificazione del *leading tone* basso. L'evento tonale è invece considerato rilevante da un punto di vista fonetico e, per questo motivo, la sua presenza è codificata in modo specifico all'interno della trascrizione, con l'utilizzo di parentesi quadre (nel caso specifico,  $[L+][H^*+L]$ ).

ampiamente sulla sillaba tonica e sulle post-toniche, realizzando anche un tono di confine L- (si veda anche quanto osservato alla nota 17).

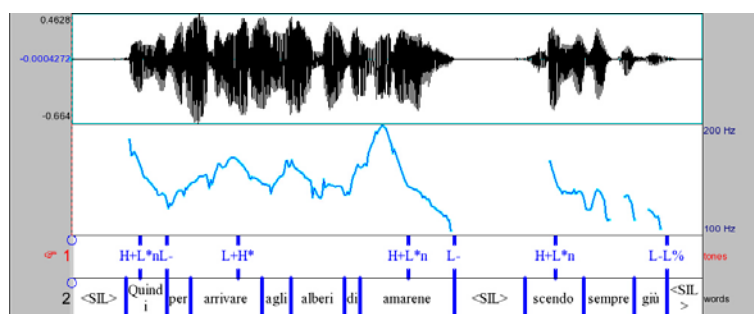


Figura 34: mossa Check del parlante 'Ma': “*Quindi, per arrivare agli alberi di amarene, scendo sempre giù?*”.

### 5.9 Topicalizzazione

Nei dialoghi raccolti per questo studio si è notata una forte tendenza alla topicalizzazione di alcuni elementi dell'enunciato. Come già rilevato in altri studi sulla correlazione tra il livello sintattico e il livello fonologico (ad esempio, Frascarelli (1997), Gili Fivela (1999), Avesani e Vayra (2000), Crocco (in stampa)), la dislocazione a sinistra di un elemento porta alla formazione di un gruppo intonativo separato, che corrisponde ad un sintagma intermedio. Nel caso della varietà di italiano parlata a Lecce, coerentemente con le descrizioni fornite in letteratura, si riscontra la presenza di specificazioni tonali relative alla presenza di un sintagma intermedio. Per quanto riguarda la scelta dell'accento tonale, nel corpus analizzato è stata riscontrata la presenza dell'accento ascendente L+H\*, già identificato come indicativo della presenza di una forte prominenzza e di un particolare rilievo dell'informazione relativa.

Nelle figure 35 e 36 sono riportati esempi di enunciati nei quali è presente un costituente dislocato a sinistra che rappresenta il topic della frase. In entrambi gli esempi è possibile osservare la presenza di un'ascesa dei valori di F0 e di un picco alto realizzato nel nucleo sillabico (la -o- di 'leone' e la -a- di 'mammole'), seguito da una discesa dei valori di F0 sino al raggiungimento della fine del costituente dislocato (si veda in particolare l'esempio in figura 36, dove sono presenti due sillabe atone dopo l'accento).

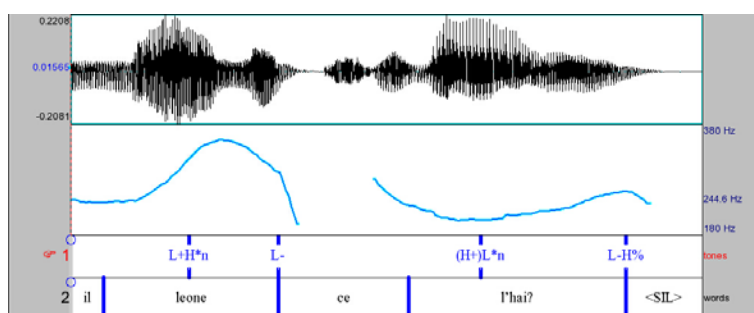
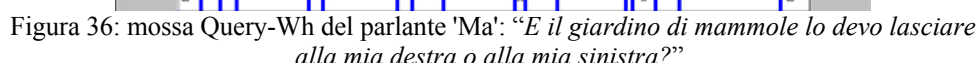
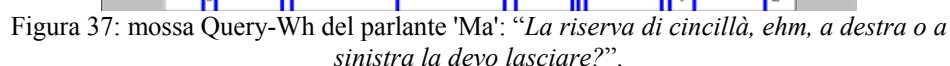


Figura 35: mossa Query-YN del parlante 'El': “*Il leone ce l'hai?*”.



Da una prima analisi degli enunciati con topicalizzazione contenuti nei dialoghi del corpus, si è notato che gli elementi che costituiscono il topic dell'enunciato sono sempre delimitati a destra da un tono di confine discendente L-, che sembra essere utilizzato con sistematicità per questo tipo di *pattern*. Alcuni casi problematici corrispondono alla realizzazione dell'accento tonale e del tono di confine in corrispondenza di parole tronche. Come già osservato nell'esempio riportato in figura 11, e come visibile in figura 37, l'andamento tonale al confine è discendente, ma implementato ad un livello molto alto di F0. Tuttavia l'implementazione fonetica del tono di confine può essere dovuta a troncamento, come già osservato per altre lingue (si veda Ladd (1996) per una panoramica), per via della compresenza di più eventi tonali che risultano associati ad una sillaba in posizione finale (nel caso in figura 37, infatti, l'accento tonale e il tono di confine occorrono entrambi sulla sillaba *-l'à* di 'cincillà'). In linea con le convenzioni di trascrizione utilizzate in casi analoghi, per il tono di confine dell'elemento dislocato si propone l'etichetta '(L-)', nella quale le parentesi tonde indicano che l'elemento non è realizzato foneticamente. Ulteriori evidenze e prove sistematiche potrebbero comunque portare ad affermare che il parlante della varietà leccese dell'italiano può utilizzare anche il tono di confine H- nei casi di dislocazione a sinistra, tono peraltro già osservato in relazione ai contorni di continuazione (cfr. paragrafi 5.7).



## 6. CONCLUSIONI

In questo articolo, si è presentata una prima analisi autosegmentale-metrica di alcune caratteristiche intonative della varietà leccese di italiano. L'analisi è stata condotta considerando un corpus di dialoghi semispontanei - elicitati grazie al Map Task ed al Test delle Differenze - ed è stata principalmente volta a delineare le caratteristiche intonative di enunciati interrogativi.

I dialoghi sono stati analizzati in termini di mosse conversazionali, in modo da poter ottenere una caratterizzazione funzionale dei materiali che fosse indipendente dalle caratteristiche prosodiche degli enunciati. Tramite l'analisi prosodica dei materiali si è poi accertato che i parlanti leccesi fanno uso prevalentemente di accenti nucleari discendenti  $H+L^*$  ed ascendenti  $L+H^*$ : i primi svolgono un ruolo fondamentale per la realizzazione di domande polari e domande aperte, in unione a specifiche combinazioni di toni di confine, mentre i secondi vengono utilizzati soprattutto per evidenziare degli elementi all'interno dell'enunciato, in caso di costruzioni sintattiche marcate e non. Tuttavia anche l'accento discendente  $H+L^*$  può essere utilizzato per assegnare prominenza ad un elemento della frase, così come l'accento  $L+H^*$  è realizzato in alcune occorrenze di domande polari o di domande aperte. L'accento discendente ad esempio è stato osservato in contesti enfatici, dove è caratterizzato da valori particolarmente alti in corrispondenza del *leading tone*; la connotazione conferita all'enunciato da queste scelte prosodico-intonative ricorda quella dei contesti contrastivi peraltro espressi in modo evidente dall'accento  $H^*+L$ . Questo accento è stato riscontrato in mosse Object, Check ed Align; in queste ultime, sono state inoltre osservate occorrenze di un accento che sembra rappresentare la categoria  $H^*$ .

I toni di confine sono sia alti che bassi e, nelle domande, compaiono in contorni discendenti ( $L-L\%$ ), ascendenti verso valori medi di  $F_0$  ( $L-H\%$ ), ed in *plateau* alti ( $H-H\%$ ), soprattutto nel caso di domande ellittiche o mosse di tipo Align. Una specificazione tonale bassa ( $L-$ ) è stata osservata in sintagmi prosodici corrispondenti a costituenti dislocati a sinistra ed una alta ( $H-$ ) in relazione ai contorni di continuazione. Sono inoltre stati riscontrati fenomeni di troncamento del tono di confine basso nel caso della associazione del tono e di un accento tonale alla sillaba metricamente forte di parole tronche in posizione finale.

Nella tabella seguente, riassumiamo i principali contorni nucleari rilevati in relazione alle mosse conversazionali considerate.

|          |                                                |
|----------|------------------------------------------------|
| Query-Wh | $H+L^*n...L-H\%$ o $L-L\%$<br>$L+H^*n...L-H\%$ |
| Query-YN | $H+L^*n...L-H\%$<br>$L+H^*n...L-H\%$           |
| Check    | $H+L^*n...L-H\%$ o $L-L\%$<br>$H^*+Ln...L-L\%$ |
| Align    | $H^*n...H-H\%$                                 |
| Object   | $L+H^*>n...L-L\%$<br>$H^*+Ln...L-H\%$          |

Tabella 1: schema riassuntivo dei principali contorni nucleari di enunciati intonativi della varietà leccese dell'italiano.

Infine, in caso di domande polari con focalizzazione stretta in posizione non finale, sono stati osservati accenti postnucleari molto prominenti. L'analisi di questi eventi tonali postnucleari deve essere ulteriormente approfondita in quanto sembra poter corrispondere ad almeno due possibili scelte fonologiche operate dai parlanti: è possibile che i parlanti realizzino una variante melodica dell'andamento utilizzato in caso di focus finale, producendo un accento tonale ascendente postnucleare seguito da toni di confine bassi; oppure l'evento tonale ascendente può essere dovuto ad un fenomeno di associazione secondaria delle specificazioni di confine tipiche delle domande polari con posizione finale del focus. Quest'ultima analisi presenta alcuni aspetti problematici all'interno di un quadro autosegmentale-metrico, in quanto prevede che entrambi i toni di confine subiscano associazione secondaria, lasciando il posto ad una specificazione tonale bassa che dovrebbe quindi essere prevista come soggiacente o sotto specificata. Peraltro, l'analisi sarebbe coerente con quanto riscontrato in altre lingue, nelle quali l'associazione secondaria di un tono di confine crea una prominenza accentuale che non indica però focalizzazione.

## BIBLIOGRAFIA

Albano Leoni, F. (2003), Tre progetti per l'italiano parlato: AVIP, API e CLIPS, in Maraschio, N. & Poggi Salani, T. (eds.), *Italia linguistica anno mille – Italia linguistica anno duemila*, Roma: Bulzoni, 675-683.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G.M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S. & Weinert, R. (1991), The HCRC Map Task Corpus, *Language and Speech*, 34, 351-366.

Avesani, C., Vayra, M. (2000), Costruzioni marcate e non marcate in italiano: Il ruolo dell'intonazione, in *Atti delle X giornate di studio del Gruppo di Fonetica Sperimentale (GFS)*, 1999, Napoli, Italy: 1-14.

Beckman, M. & Ayers, G. (1997), *Guidelines for ToBI labelling, Version 3.0*, Ms. and accompanying speech materials, Ohio State University, Ohio.

Beckman, M. & Hirschberg, J. (1994), *The ToBI annotation conventions*, Ms. and accompanying speech materials, Ohio State University, Ohio.

Bertinetto P.M. (2001), AVIP (*Archivio delle Varietà dell'Italiano Parlato*, CD-ROM, distribuito da Scuola Normale Superiore di Pisa).

Boersma, P. & Weenink, D. (2007), *Praat: doing phonetics by computer (Version 5.0)*, University of Amsterdam, The Netherlands. Software liberamente scaricabile su <http://www.praat.org/>.

Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G. & Anderson, A. (1996), *HCRC Dialogue Structure Coding Manual*, Technical Report 82, Human Communication Research Centre, University of Edinburgh, Edinburgh.

Crocco, C., (in stampa), Topic accent and prosodic structure, in Lunella Mereu e Giorgio Banti (a cura di) *Information structure and its interfaces. Studies on standard and regional Italian (and other languages)*, Berlin/NewYork: Mouton De Gruyter.

D'Imperio, M. (1999), Tonal structure and pitch targets in Italian focus constituents, in *Proceedings of the 1999 International Congress of Phonetic Sciences*, San Francisco, 1757-1760.



Gili Fivela, B. (1999), The prosody of left-dislocated topicalized constituents in Italian read speech, in *Proceedings of Eurospeech '99*, Budapest, Hungary, September 5-9, vol. 1, 531-534.

Gili Fivela, B. (2004), *The phonetics and phonology of intonation: the case of Pisa Italian*, Tesi di perfezionamento, Scuola Normale Superiore, Pisa, Italy.

Gili Fivela, B. (2006), The coding of target alignment and scaling in pitch accent transcription, in *Italian Journal of Linguistics*, 18.1, pp. 189-221.

Grice, H. P. (1975), Logic and conversation, in Cole, P. & Morgan, J. L. (eds.), *Syntax and Semantics 3: Speech Acts*, New York: Academy Press, 1975, pp. 41-58.

Grice, M. (1995), *The intonation of interrogation in Palermo Italian; implication for intonational theory*, Tuebingen: Niemeyer, L.A. Series.

Grice, M., Benzmueller, R., Savino, M., Andreeva, B. (1995), The intonation of queries and checks across languages: data from Map Task dialogues, in *Proceedings of the 1995 International Congress of Phonetics Sciences*, Stockholm, Sweden, 3, pp. 648-651.

Grice, M. & Savino, M. (1995), Intonation and communicative function in a regional variety of Italian, in Barry, W. J. & Koreman, J. (eds.) *PHONUS*, 1, pp. 19-32.

Grice, M. & Savino, M. (1997), Can pitch accent type convey information status in yes-no questions?, in Alter K., Pirker H., Finkler W., (eds.), *Proceedings of the ACL 97 Workshop on Concept to Speech Generation System*, Madrid, Spain, pp. 29-38.

Grice, M., Ladd, D. R. & Arvaniti, A. (2000), On the place of phrase accent in intonational phonology, in *Phonology*, 17, pp. 143-185.

Grice, M., D'Imperio, M., Savino, M., Avesani, C. (2005), Strategies for intonation labelling across varieties of Italian, in S.-A. Jun (eds.), *Prosodic Typology and Transcription. A Unified Approach*, Oxford: Oxford University Press, 362-389.

Gussenhoven, C. (2002), Intonation and interpretation: phonetics and phonology, in *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 11-13 April,

Gussenhoven, C. (2004), *The phonology of tone and intonation*, Cambridge: Cambridge University Press.

Ladd, D.R. (1996), *Intonational Phonology*, Cambridge: Cambridge University Press.

Pierrehumbert, J. (1980), *The phonology and phonetics of English intonation*, PhD Thesis, MIT, published 1988 by IULC

Pierrehumbert, J. & Beckman, M. (1988), *Japanese Tone Structure*, Massachussets Institute of Technology.

Pitrelli, J., Beckman, M. & Hirschberg, J. (1994), Evaluation of prosodic transcription labeling in the ToBI framework, in *Proceedings of the 1994 International Conference on Spoken Language Processing*, Yokohama, Japan, 1, 123-126.

Romano, A. (1998), Analisi comparata di alcuni schemi intonativi di base (frasi affermative e interrogative) di due varietà salentine (salentino meridionale e centrale), in *Studi Linguistici Salentini*, 23, 23-46.

Romano, A. (1999), *Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento: approche linguistique et instrumentale*, Tesi di Dottorato, Université Stendhal, Grenoble, France.

Romano, A. (2000), Variabilità degli schemi intonativi dialettali e persistenza di tratti prosodici nell'italiano regionale: considerazioni sulle varietà salentine, in A. Zamboni *et alii* (a cura di), *"La dialettologia oggi fra tradizione e nuove tecnologie"*, *Atti del Convegno Internazionale*, Università di Pisa, Pisa, Italia, 10-12 Febbraio, 73-91.

Salvi, G. & Vanelli, L. (2004). *Nuova grammatica italiana*. Il Mulino, Bologna.

Savino, M. (1997), *Il ruolo dell'intonazione nell'interazione comunicativa. Analisi strumentale delle domande polari in un corpus di dialoghi spontanei (varietà di Bari)*, Tesi di Dottorato, Politecnico di Bari.

Savino, M. (2000), Modalità di acquisizione dei dialoghi del corpus AVIP. Il metodo del Map Task, in Crocco, C., Savy, R. & Cutugno, F. (2003), *API – Archivio Parlato Italiano*, DVD-Rom distribuito da CIRASS, Università degli Studi di Napoli “Federico II”, Napoli, Italia.

Savy, R. (2001), Documento di specifiche per la rappresentazione, analisi e codifica dei dati, in Crocco, C., Savy, R. & Cutugno, F. (2003), *API – Archivio Parlato Italiano*, DVD-Rom distribuito da CIRASS, Università degli Studi di Napoli “Federico II”, Napoli, Italia.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992), ToBI: a standard for labeling English prosody, in *Proceedings of the 1992 International Conference on Spoken Language Processing*, Banff, Canada, 2, 867-870.

Stella, A. (2007), *La fonologia dell'intonazione: l'italiano e il tedesco nelle produzioni di parlanti dell'area leccese*, Tesi di laurea, Università del Salento, Lecce, Italy.

## EMOVO, DATABASE DI PARLATO EMOTIVO PER L'ITALIANO

Iacopo Iadarola  
Dipartimento di Studi Glottoantropologici e Discipline musicali  
dell'Università di Roma "La Sapienza"  
Fondazione Ugo Bordonì  
*iacopo.iadarola@gmail.com*

### 1. SOMMARIO

EMOVO è il primo database di parlato emotivo per la lingua italiana, ed è stato realizzato per conto del Dipartimento di Studi Glottoantropologici e Discipline musicali dell'Università di Roma "La Sapienza" e della Fondazione Ugo Bordonì, sotto la supervisione del Prof. Paolo Di Giovine e dell'Ing. Andrea Paoloni.

Sono stati convocati 6 attori, 3 maschi e 3 femmine, di provata professionalità, e si sono fatte loro recitare 14 frasi (assertive, interrogative, elenchi) in base ai 6 stati emotivi di base (disgusto, paura, rabbia, gioia, sorpresa, tristezza) più lo stato neutro. Tali emozioni sono le ben note "big six" riscontrabili in molta della letteratura inerente al parlato emotivo, nonché quelle prese in considerazione dal CNR di Padova, che ha condotto i principali studi sul parlato emotivo italiano. Le registrazioni sono state effettuate, con adeguate strumentazioni professionali, presso i laboratori della Fondazione Ugo Bordonì, dove è stata garantita agli attori la possibilità di autoindursi lo stato emotivo desiderato in base alle proprie tecniche professionali. Si è optato per una frequenza di campionamento di 48 kHz, 16 bit stereo, formato wav. Ogni singola elicitazione delle frasi è stata etichettata ed archiviata in un database di tipo Access, in modo da consentirne una rapida individuazione. In totale sono stati archiviati 588 record: 6 attori X 14 frasi X 7 stati. Il totale del materiale registrato, escluse le pause, è all'incirca di 10 minuti per attore, quindi un'ora per l'intero database, che verrà distribuito su cd-rom a chi ne farà [richiesta](#). Nel presente articolo è possibile averne qualche saggio (cf. § 2.2.1, § 2.2.3).

Inoltre, EMOVO è stato validato con un test di discriminazione delle emozioni su due frasi ('la casa forte vuole col pane', 'il gatto sta scorrendo nella pera'), condotto, parallelamente e separatamente, dall'autore e dall'Università della Calabria. Entrambi i test hanno avuto un campione di 12 soggetti riconoscitori, i quali hanno dovuto di volta in volta indicare, scegliendo fra due possibili risposte, lo stato emotivo delle frasi ascoltate. Il test ha avuto esito positivo in quanto è risultata un'accuratezza complessiva dei riconoscimenti pari all'80%, senza significative discrepanze fra i risultati dei due test. E' stata valutata inoltre la scala di riconoscibilità delle emozioni, e si è riscontrata una sostanziale concordanza con la scala di riconoscibilità desunta dalla letteratura al riguardo.

### 2. I DATABASE DI PARLATO EMOTIVO E EMOVO

#### *2.1 I corpora del parlato emotivo: quadro d'insieme, problematiche e prospettive*

Nel campo degli studi sul parlato emotivo c'è ad oggi urgente bisogno non solo di ulteriori e più approfondite analisi, ma soprattutto di standardizzazioni e convenzioni condivise dalla comunità scientifica al fine di rendere tali analisi comparabili e convergenti.

Per troppo tempo la nostra conoscenza del parlato emotivo è progredita a rilento a causa della mancanza di sforzi congiunti da parte degli studiosi: e questa è stata una conseguenza, più che di problemi logistici e organizzativi, della mancanza di un modello teorico (nella

concettualizzazione delle emozioni) e di prassi comuni (soprattutto nella rilevazione dei parametri acustici). Ora, sia dal punto di vista teorico sia dal punto di vista pratico, siamo lontani dall'auspicabile risultato di potersi riferire a modelli unanimemente accettati. A cavaliere fra i due versanti si trova una delle questioni fondamentali da affrontare per il progresso degli studi del parlato emotivo: la creazione di database linguistici emotivamente connotati. Dopo che nella comunità scientifica si sono effettuati centinaia di esperimenti isolati su materiale linguistico estremamente limitato, e dopo che si è assodato che l'enorme mole dei risultati così ottenuta è lungi dall'essere soddisfacente, si è cominciato a sentire l'urgenza di partire da altre basi: di passare dai *dataset* ai database. Infatti, se uno degli obiettivi della comunità scientifica è la comparazione e la convergenza dei risultati, la prima ovvia cosa da fare è cercare di partire da basi comuni di dati, oltre che da premesse teoriche comuni.

Un caso illuminante in questo senso è la creazione del network europeo HUMAINE, lanciato nel 2004 con l'adesione di 33 partner da 14 paesi europei, con l'obiettivo di sviluppare sistemi capaci di registrare, modellare e influenzare stati emotivi umani:

Tali sistemi potrebbero essere di importanza centrale nelle interfacce future, ma i loro presupposti teorici non sono abbastanza avanzati da renderci sicuri sul loro effettivo potenziale e sulle modalità per il loro sviluppo. Uno dei motivi è che le conoscenze al riguardo sono disperse fra varie discipline. HUMAINE intende raggruppare i leader del settore delle discipline-chiave in un programma finalizzato ad ottenere maggiore integrazione intellettuale.<sup>1</sup>

Il programma prevede l'organizzazione di aree di ricerca con relativi laboratori, nonché la creazione delle infrastrutture necessarie, quali appunto "databases recognising cultural and gender diversity." Sul sito del network si può trovare molto materiale interessante da un punto di vista teorico e metodologico, nonché utili software di analisi, ma un database HUMAINE ancora non è disponibile.

E' ora utile sottoporre a disamina le caratteristiche e le problematiche di cui lo sviluppo di un database deve tener conto. Innanzitutto va definito il numero, il sesso dei parlanti, la lingua, le emozioni, le occorrenze di una data emozione, eventuali configurazioni funzionali/sociali: la diversità dei risultati ottenuti nei vari studi può avere valore euristico soltanto se è ben chiaro l'assetto di tali parametri, dal momento che l'espressione vocale di un'emozione può cambiare notevolmente di parlante in parlante, di cultura in cultura, o ancora in base al sesso o alla situazione sociologica. Per quanto riguarda le emozioni, la loro selezione è fondamentale: se di solito la scelta "di default" nei database è quella di utilizzare le emozioni di base (fra cui le più consuete: paura, rabbia, tristezza, gioia), bisogna anche tener conto del fatto che nella vita di tutti i giorni le emozioni sono fortemente condizionate dalle *display rules* (i condizionamenti socioculturali all'atto espressivo, studiati in particolar modo da Paul Ekman), e da altre profonde influenze culturali (cf. Harré 1986): il risultato è che nella realtà difficilmente si hanno emozioni pure, ben più spesso svariati stati emotivi sfumati, la cui codifica presenta difficoltà ancora maggiori di quante già ne comporti la codifica delle emozioni di base. Un database ideale dovrebbe contemplare entrambe le tipologie di stati emotivi, il che renderebbe la sua portata di gran lunga più estesa dei database e dei *dataset* attualmente esistenti.

---

<sup>1</sup> Cf. il sito: [www.emotion-research.net](http://www.emotion-research.net)

Altro elemento qualificante di un database è costituito dal fatto di essere *naturale* o meno. Si apre qui una questione dibattutissima sui *pro* e sui *contro* riguardo al materiale raccolto naturalmente rispetto al materiale ottenuto artificialmente. Nel primo caso, ovviamente, i vantaggi sono dati da una maggiore rappresentatività, da parte del materiale analizzato, della reale espressione delle emozioni (ma si tenga presente quanto detto poc'anzi sulla scarsa frequenza di emozioni pure nella vita di tutti i giorni). Gli svantaggi sono di ordine pratico, deontologico e teorico: registrare del parlato emotivo in situazioni reali comporta inevitabilmente che il suono non sia pulito e che possa difficilmente essere confrontato con altro materiale raccolto in modo analogo. Inoltre, la registrazione di parlato emotivo dal vero comporta spinose questioni di *copyright*, e conseguentemente di accessibilità ai dati da parte di altri studiosi. Infine, come etichettare il materiale così ottenuto? Non c'è nessuna garanzia sull'attendibilità della classificazione fornita per lo stato emotivo "catturato", e ciò ovviamente comporta seri problemi di sistemazione e comparabilità dei dati. La comparabilità dei dati così ottenuti è ovviamente inficiata anche dalla estrema diversità delle situazioni da cui sono sorte le emozioni analizzate. Il materiale raccolto finora dal vivo, infatti, proviene dalle fonti più disparate: registrazioni di piloti di aerei prima che questi si schiantassero al suolo, giornalisti che commentano eventi disastrosi o che intervistano persone appena uscite da traumi fisici o psichici, *talkshow* (Johannes et al. 2000, Cowie – Douglas-Cowie 1996, Eldred – Price 1958, Hargreaves et al. 1965, Frolov et al. 1999, Huttar 1968, Kuroda et al. 1979, Niwa 1971, Roessler – Lester 1976, 1979, Simonov – Frolov 1973, Sulc 1977, Utsuki – Okamura 1976, Williams – Stevens 1969, 1972, Zuberbier 1957, Zwirner 1930). Se sulla variabilità di questi dati si può soprassedere all'interno di un singolo *dataset* per un esperimento, il problema certo non è di facile risoluzione nel caso di un intero database. La soluzione di compromesso è spesso quella di limitare la gamma di emozioni studiate, come nel caso del database *Vanderbilt II* (usato in France et al. 2000), che contempla solo la depressione e gli stati emotivi tendenti al suicidio, o del *SUSAS* (Hansen – Bou-Ghazale 1997), incentrato sullo stress. Ugualmente limitato a pochi stati emotivi è ovviamente il *Geneva airport lost luggage study* (Scherer – Ceschi 1997, 2000), in cui sono archiviate le reazioni emotive dei passeggeri nell'aeroporto di Ginevra, ai quali si era fatto credere che i loro bagagli fossero andati perduti, come in una sorta di "*candid camera*", sia pure con fini scientifici.

Le difficoltà su elencate non si pongono, invece, nel caso di materiale elicitato in laboratorio, sebbene anche qui si presentino varie questioni metodologiche di non facile soluzione, *in primis* sui soggetti elicitanti: attori professionisti o non-professionisti? Nel *Berlin database* (Kienast – Sandlmeier 2000, Paeschke – Sandlmeier 2000) le emozioni sono registrate in laboratorio dalla bocca di parlanti non attori, in quanto i professionisti produrrebbero, secondo gli autori, emozioni stereotipate. A ciò si potrebbe obiettare, pirandellianamente, che non solo gli attori, ma qualsiasi persona in pubblico recita, in misura minore o maggiore. Inoltre, a giudicare dai buoni risultati di riconoscimento delle emozioni elicitate da attori professionisti,<sup>2</sup> si può tenere per acquisita, in linea generale, la

---

<sup>2</sup> Banse – Scherer 1996, Bortz 1966, Coleman – Williams 1979, Costanzo et al. 1969, Cosmides 1983, Davitz 1964, Fairbanks – Pronovost 1939, Fairbanks – Hoaglin 1941, Fonagy 1978, 1983, Fonagy – Magdics 1963, Green – Cliff 1975, Höffe 1969, Kaiser 1962, Kienast et al. 1999, Klasmeyer – Meier 1999, Klasmeyer – Sandlmeier 1997, Klasmeyer – Sandlmeier 1999, Klasmeyer 1999, Kotlyar – Morozov 1976, Levin – Lord 1975, Paeschke et al. 1999, Scherer et al. 1972a,b, 1973, Sedlacek – Sychra 1963, Sobin – Alpert 1999,

capacità di un attore di rispecchiare la realtà dell'espressione emotiva, per quanto delle differenze sicuramente possano permanere. In particolar modo si è convenuto che le emozioni simulate dagli attori ci diano informazioni sulle componenti socioculturali delle espressioni emotive, mentre le ricorrenze naturali di un'emozione ci permettono di indagare con più facilità le sue caratterizzazioni psico-fisiologiche: e va puntualizzato che l'aspetto socioculturale non è certo meno importante di quello psico-fisiologico, dal momento che il le emozioni e il parlato emotivo trovano gran parte della propria ragione d'essere per l'appunto nella funzione comunicativa e di interazione sociale. Altri svantaggi dell'uso di attori in laboratorio sono legati al tipo di materiale elicitato: per lo più testi scritti e monologhi. Ciò ovviamente preclude l'analisi di importanti aspetti dell'espressione emotiva legati all'interrelazione e all'improvvisazione.

Come si può vedere gli argomenti a favore e a sfavore sono numerosi gli uni e gli altri, e la scelta di una soluzione o di un'altra in ultima analisi dipende ancora dalle idiosincrasie dei singoli studiosi. Sembrerebbe auspicabile utilizzare database naturali, ma resta ancora da vedere se le attuali tecnologie ci possono permettere di far fronte al problema in modo efficace. Per compendiare il doppio corno del problema, riporto infine un'affermazione di Scherer 1981: "Se i campioni di parlato usati nelle ricerche di simulazione non sono abbastanza naturali, i campioni costituiti da parlato spontaneo non sono abbastanza emotivi."

Vanno menzionati, ancora, tentativi di mediare fra le due posizioni: sono i casi di materiale "seminaturale", che prevede l'induzione di emozioni in laboratorio. I principali metodi di induzione emotiva sono stati l'uso di droghe (Helfrich et al. 1984), l'assegnazione di compiti particolarmente difficili per generare stress, la presentazione di film o diapositive atti a indurre date emozioni (Alpert et al. 1963, Bachorowski – Owren 1995, Bonner 1943, Brown 1980, Fernandez – Picard 2000, Havrdova – Moravek 1979, Hicks 1979, Karlsson et al. 1998, Markel et al. 1973, Plaikner 1970, Roessler – Lester 1979, Scherer 1977, 1979, Scherer et al. 1985, Skinner 1935, Tolkmitt – Scherer 1986). I problemi che si accompagnano a queste procedure sono in primo luogo metodologici (di fronte a uno stesso stimolo un soggetto può reagire con emozioni diverse, e anche qui sorge il problema di come si possano etichettare inequivocabilmente le emozioni elicitate), e in secondo luogo di efficienza: "these procedures often produce only weak affect", come sottolinea Scherer 2003, p. 232. Un'altra soluzione di tipo "seminaturale" è invece il "bootstrapping": l'uso di materiale naturale per guidare l'elicitazione di materiale recitato. I risultati di questa procedura sono tuttavia a mio parere inversamente proporzionali alla bravura di un attore: nel senso che un vero attore professionista dovrebbe essere in grado da sé di fare del "bootstrapping" mentale prima della propria *performance*, non così nel caso di attori meno competenti, che si affidano a moduli emotivi stereotipati. Ma sulla questione degli attori ritorneremo in seguito.

Infine, di un database è importantissima l'*accessibilità*: sua vera e propria ragione d'essere, se si considera che l'istanza di creare database nasce proprio da un bisogno sempre più urgente di condivisione dei saperi. L'*accessibilità* comporta senz'altro l'uso di formati simili per i descrittori e per il materiale raccolto (ad esempio, la questione se archiviare i file audio in formato WAV o MPEG: quest'ultimo è indubbiamente un formato più versatile per l'archiviazione e la trasmissione di dati, ma comporta minori informazioni sul segnale). Un altro problema legato all'*accessibilità* è quello già accennato del *copyright*:

---

Tischer 1993, Van Bezooijen 1984, Wallbott – Scherer 1986, Whiteside 1999, Williams – Stevens 1972.

il materiale raccolto dal vivo abbisogna del consenso da parte dei soggetti, ma anche degli enti televisivi di cui si sono eventualmente registrate le trasmissioni: questi ultimi possono dare il nulla osta per un singolo esperimento, ma possono negarlo per una libera diffusione del database (è il caso dei *Belfast* e *Reading-Leeds database*).

## 2.2 Il database EMOVO

Un *corpus* emotivo per l'italiano ancora non è stato realizzato, e anche dal punto di vista dei *dataset* il materiale raccolto non è molto. Ricordo in particolare i seguenti studi:

- AA. VV. 2008  
Svariati casi di studio contenuti negli Atti del 1° convegno nazionale GSCP (gruppo di studio della comunicazione parlata), a cura di Emanuela Magno Caldognetto, Federica Cavicchio e Piero Così. Segnalo in particolar modo Moneglia 2008, che contiene un sondaggio sulle manifestazioni acustiche delle emozioni nel *corpus* di parlato spontaneo C-ORAL-ROM (Cresti – Moneglia 2005)
- Anolli – Ciceri 1992  
1 frase pronunciata da 2 parlanti non attori. Stati emotivi: collera, gioia, paura, tristezza, disprezzo, tenerezza.
- De Dominicis 1992  
2 frasi, una sintetizzata ed una naturale, pronunciate da 1 attore. Più che stati emotivi sono stati presi in esame atti linguistici, con finalità di analisi del profilo intonativo: richiesta, provocazione, ordine, concessione di facoltatività, informazione, valutazione, rivelazione, domanda sì/no.
- Kori – Magno Caldognetto 1986  
3 vocali pronunciate da 1 attore. Stati emotivi: collera, disgusto, gioia, paura, sorpresa, tristezza.
- Kori – Magno Caldognetto 1990  
1 vocale sintetizzata. Stati emotivi: collera, disgusto, gioia, paura, sorpresa, tristezza.
- Kori – Magno Caldognetto 2003  
2 frasi, una in giapponese e una in italiano, per 2 attori, un giapponese e un italiano. Stati emotivi: collera, disgusto, gioia, paura, sorpresa, tristezza.
- Magno Caldognetto et al. 1998  
1 parola pronunciata da 3 parlanti non attori. Stati emotivi: collera, disgusto, gioia, paura, sorpresa, tristezza.
- Magno Caldognetto – Ferrero 1996  
1 brano e 1 frase pronunciati da 1 attore. Stati emotivi: collera, disgusto, paura, tristezza.

Constatata pertanto la mancanza, per il parlato emotivo, di un *corpus* di dati più esteso, non ho ritenuto inutile la creazione del database di parlato emotivo EMOVO, pur non potendo rispondere a tutte le istanze presentate nello scorso paragrafo, dati i limiti in termini di tempo e risorse a disposizione. Ho tentato comunque di gettare le basi in questa direzione, nella quale ulteriori sviluppi saranno sicuramente necessari, tanto in termini di analisi di dati quanto di raccolta e sistemazione degli stessi.

Descriverò ora la struttura di EMOVO:

### 2.2.1 Scelta delle emozioni

Come già anticipato nel primo capitolo, si è deciso di focalizzare l'attenzione sulle “*big six*”. Pertanto le *performance* emotive espresse nel database ineriscono ai seguenti stati (cliccando sui link su potrà sentire l'emozione corrispondente):

- [disgusto](#)
- [gioia](#)
- [paura](#)
- [rabbia](#)
- [sorpresa](#)
- [tristezza](#)
- (stato emotivo neutro)

Sono queste le emozioni che hanno ricevuto più attenzione dalla comunità scientifica italiana e pertanto si è ritenuto opportuno mantenersi entro questo assortimento di emozioni di base piuttosto che analizzarne altre, le quali non avrebbero potuto beneficiare di termini di paragone derivanti dai precedenti studi. Per evitare interpretazioni troppo diverse di queste emozioni, si sono date apposite spiegazioni almeno per le emozioni di più problematica definizione. In particolar modo, si è spiegato ad ogni attore che la rabbia da esprimere è quella identificata da Scherer come *calda*. Mentre, per quel che attiene al disgusto, si è sottolineato come si intendesse un disgusto fisico più che morale<sup>3</sup>. Per la sorpresa ho invitato gli attori a mantenere questo stato emotivo per l'intera durata della frase, e di non usare a tal fine una tonia interrogativa. Per le altre emozioni non ho ritenuto necessario fornire ulteriori spiegazioni. Infine, è stato richiesto agli attori di non esprimere indicatori emotivi espliciti quali risate, singhiozzi ecc., che avrebbero falsato i test di riconoscimento.

### 2.2.2 Attori

Sebbene convenga con quanti sostengono la necessità di analisi su materiale naturale, ho ritenuto opportuno, data l'esigua mole di studi per la lingua italiana, di rafforzare le nostre conoscenze innanzitutto su un *corpus* di materiale recitato, il quale sarà utile se non altro come termine di paragone per eventuali studi sul parlato emotivo naturale. Abbiamo inoltre già sottolineato come si sia convenuto da parte di molti studiosi sul fatto che gli attori possono darci informazioni probanti sulle componenti socioculturali delle emozioni, mentre il materiale non recitato ce ne segnalerebbe gli aspetti psico-fisiologici. Infatti, anche in altri paesi dove sono stati elaborati molti più studi, come in Germania, l'uso di attori per l'elicitazione delle emozioni a tutt'oggi non è affatto disdegnato.

A differenza di alcuni studiosi tedeschi, si è ritenuto non irrilevante l'impiego di attori *professionisti*. Come già accennato nel primo paragrafo, c'è chi sostiene che gli attori

---

<sup>3</sup> E' stato riscontrato come negli studi relativi a questo stato emotivo i valori della media del pitch dipendano dalle modalità di elicitazione dell'emozione: là dove il disgusto è stato indotto tramite materiale audiovisivo, si è rilevata una media del pitch maggiore rispetto al parlato non emotivo; laddove è stato simulato, una media minore. A mio parere, ciò rivela come questa etichetta si presti ad essere rianalizzata in due stati emotivi abbastanza distinti: un disgusto “fisico”, caratterizzato da una media maggiore del pitch e da una maggiore attivazione; e un disgusto “morale”, culturale, etico o che dir si voglia: questo è caratterizzato da una media minore del pitch, suscitando un *arousal* minore (eventualmente attenuato da *display rules*), in quanto si tratta ovviamente di un'emozione meno fisica.



professionisti produrrebbero emozioni stereotipate, a differenza dei parlanti non professionisti cui si è richiesto di esprimere una data emozione.<sup>4</sup> Non si capisce su quali basi si sostenga ciò, e ritengo invece la situazione esattamente ribaltata. E' proprio la ragion d'essere dell'attore quella di riprodurre le emozioni della vita reale, e una delle prime cose che un attore impara è di non essere impostato, ma "naturale". In generale, si ricorre invece agli stereotipi quando non si ha perizia performativa, esattamente come nel caso dei disegni dei bambini, o appunto in quello di parlanti non professionisti cui si richiede di "fare la rabbia" o altre emozioni: da alcune prove preliminari in laboratorio, con parlanti non attori, ho potuto verificare ciò ottenendo *performance* pessime, che sono migliorate solo nel momento in cui si è chiesto ai parlanti di essere più naturali, cioè impartendo loro un primo ovvio rudimento di recitazione.

Pertanto, ho scelto 6 attori professionisti, 3 maschi e 3 femmine, e mi son premurato di verificare i loro trascorsi professionali, in particolar modo la loro conoscenza del famoso metodo Stanislavski, consistente nell'autoindursi una data emozione ricordando le situazioni della propria vita in cui si è provata intensamente la medesima emozione. La loro età media è di 27 anni: non mi sono preoccupato di scegliere attori di età molto diverse in quanto allo stadio attuale degli studi rimane ancora indeterminata e del tutto negletta la correlazione fra età ed espressione delle emozioni. Ho dunque preferito uniformarmi agli standard di ricerca, piuttosto che disperdere energie in nuove direzioni.

Gli attori sono stati classificati come [M1](#), [M2](#), [M3](#), [F1](#), [F2](#), [F3](#). Ecco le loro referenze professionali:

**M1:**

Francesco Maiorca. 30 anni, palermitano. Formazione presso: laboratorio teatrale a cura Bibi Bianca; Accademia di formazione per attori di musical "MASTMASTER" a cura di Sofia Amendolea; Seminario di Giancarlo Sepe "La Musica in Movimento" presso il Teatro La Comunità; Seminario di educazione vocale e canto a cura di Lena Biolcati e Patrizia Troiani; Seminario di uso della voce "La Vita Bestia" a cura di Filippo Timi; Laboratorio teatrale "Masculi e Fimmini" a cura di Davide Enia; Laboratorio teatrale "Il contatto" a cura di Marco Maltauro. Partecipazione come attore in vari spettacoli teatrali e cortometraggi, nonché del film "Ho ammazzato Berlusconi" di Daniele Gionetti e Gianluca Rossi, attualmente in lavorazione.

**M2:**

Simone Arcese. 27 anni, romano. Conseguimento dei titoli di attore rilasciati dalla Scuola di recitazione "RIBALTE", di Enzo Garinei, e dal Centro Sperimentale di Cinematografia. Partecipazione a svariati film (di registi quali Michele Soavi, Daniela Stroppa, Pupi Avati; ruolo di protagonista in *Saremo film* di Ludovica Marinese) e sceneggiati televisivi ("Incantesimo", "Distretto di polizia").

**M3:**

Filippo Sandon. 30 anni, padovano. Ha studiato con Marina De Luca, attrice e regista padovana, per 2 anni. Con lei ha iniziato a lavorare in teatro, allestendo due spettacoli teatrali. Ha proseguito gli studi a Roma con Beatrice Bracco e Michael Margotta dell'Actor's centre. Ha lavorato in film per cinema e televisione con registi come Luchetti, Soavi, Vanzina e Perelli, anche in parti di coprotagonista.

---

<sup>4</sup> Cf. ad esempio Burkhardt et al. 2005.

**F1:**

Simona Augelli. 28 anni, gaetana. Diplomata presso la Scuola di Teatro “Auroville” diretta da Augusto Zucchi; laboratori teatrali con: Tullio Solenghi, Giuliana De Sio, Massimo Dapporto, Paolo Giuranna. Svariate rappresentazioni teatrali e partecipazione a spot televisivi e alla fiction “Un posto al sole”.

**F2:**

Maria Cristina Papeo. 23 anni, foggiana. Laureata in DAMS nel curriculum “regista e coordinatore di gruppi di teatro”; formatasi come attrice con un Master Class in clown-attore tenuto da Vladimir Olshansky (artista ospite del Cirque du Soleil) e con due laboratori teatrali tenuti rispettivamente da Paolo Alessandri e Pino Casolaro. Attrice in due rappresentazioni teatrali.

**F3:**

Chiara Fiorelli. 25 anni, romana. Qualifiche di "attore" e "aiuto regista teatrale" conseguite presso la Scuola Internazionale di Teatro "Circo a Vapore" di Roma, riconosciuta dalla Regione Lazio. Esperienze lavorative: otto messe in scena, presso i teatri "Euclide", "Anfitrione" e "In Portico" di Roma, nell'ambito della didattica triennale della scuola di teatro; tournèe estive di teatro di strada in numerose piazze italiane; partecipazione a vari festival tra cui Caorle, Aosta, Carpineto Romano e la Notte Bianca di Roma; interprete di *Trincee* allo "Spazio Uno" di Roma; partecipazione come attrice a vari cortometraggi.

### 2.2.3 *Materiale linguistico*

Riporto qui l’elenco delle frasi da elicitare:

- 1 [Gli operai si alzano presto.](#)
- 2 [I vigili sono muniti di pistola.](#)
- 3 [La cascata fa molto rumore.](#)
- 4 [L’autunno prossimo Tony partirà per la Spagna: nella prima metà di ottobre.](#)
- 5 [Ora prendo la felpa di là ed esco per fare una passeggiata.](#)
- 6 [Un attimo dopo s’è incamminato...ed è inciampato.](#)
- 7 [Vorrei il numero telefonico del Signor Piatti.](#)
- 8 [La casa forte vuole col pane.](#)
- 9 [La forza trova il passo e l’aglio rosso.](#)
- 10 [Il gatto sta scorrendo nella pera](#)
- 11 [Insalata pastasciutta coscia d’agnello limoncello.](#)
- 12 [Uno quarantatré dieci mille cinquantasette venti.](#)
- 13 [Sabato sera cosa farà?](#)
- 14 [Porti con te quella cosa?](#)

Ho ritenuto funzionale ai fini proposti l’uso di frasi isolate. Nella maggior parte dei database si è optato per una scelta simile e l’unica questione che si potrebbe sollevare è quella relativa alla maggiore o minore efficacia, ai fini dell’elicitazione e dell’analisi di un’emozione, della lettura di brani anziché frasi. A mio parere l’impiego di brani da colorare emotivamente nella *performance* presenta numerose questioni metodologiche: se da una parte è vero che l’impiego di un brano sufficientemente lungo permette meglio di una frase di cogliere lo sviluppo di un’emozione nel tempo, dall’altra ciò complica di molto

la già spinosa questione dell'etichettamento dell'emozione<sup>5</sup>. In un brano si può cambiare facilmente lo stato emotivo: si può passare dal neutro all'emotivo e viceversa (e di una stessa emozione si possono rendere stati più o meno intensi, in modo del tutto “*speaker-specific*”) se non passare addirittura ad un'altra emozione: come si classificherà tutto ciò? Nello studio di Anolli – Ciceri 1992 si è risolto il problema isolando una frase al centro del brano e limitando la funzione e l'impiego del brano ad una semplice preparazione dell'antecedente emotivo del parlante. Tutto ciò mi è sembrato superfluo nel caso di attori professionisti, la cui preparazione interiore supplisce di per sé alla contestualizzazione emotiva della frase.

Per quanto riguarda il contenuto delle frasi, ai fini di non falsare i test di riconoscimento dell'emozione, una condizione necessaria è che il valore semantico di tale contenuto sia emotivamente neutro. A tal fine è possibile utilizzare frasi che non si riferiscano specificatamente a una data espressione emotiva, e frasi “*nonsense*”. Se le prime possono porre qualche problema all'attore per la loro collocazione nella giusta situazione emotiva,<sup>6</sup> le seconde comportano il rischio di essere recitate in una maniera stereotipata in quanto l'attore può non riuscire a “sentirle” come naturali.<sup>7</sup> Consapevoli di queste problematiche, ho deciso di usare entrambe le categorie (frasi 1-7, 13, 14 semanticamente neutre, frasi 8-10 *nonsense*), aggiungendo due frasi che si situano in una posizione intermedia rispetto alle frasi nonsense e a quelle semanticamente non emotive: un elenco di numeri e un elenco di alimenti (frasi 11,12).

Delle frasi semanticamente neutre ho inoltre formulato delle versioni brevi (1-3) e lunghe (4-7).

Ai fini delle analisi spettroacustiche sono state soddisfatte le seguenti condizioni basilari:<sup>8</sup>

- presenza nelle frasi di tutti i fonemi della lingua italiana
- presenza in ogni frase di un giusto bilanciamento fra consonanti sorde e sonore
- assenza di consonanti occlusive iniziali (in quanto in tale posizione non è possibile misurare la durata della fase di occlusione).

Per quanto riguarda eventuali analisi prosodiche si è tenuto conto dei quattro tipi di tonia descritti da Canepari 1985. La *tonia conclusiva* è analizzabile nel database nelle frasi 1-5 e 7-13. Di norma la tonia conclusiva viene impiegata quando il parlante comunica di aver terminato il suo enunciato con un significato semplicemente denotativo: è ovvio pertanto che la sua ricorrenza può cambiare in base allo stato emotivo in questione. Con la *tonia sospensiva* invece il parlante avverte il suo ascoltatore che sta seguendo un'informazione importante, avvalendosi di una pausa per fini di *suspence*: è quanto è stato richiesto di fare agli attori per la frase 6. La *tonia divisiva* è la tonia non marcata, la cui funzione è la semplice suddivisione dell'enunciato in diversi segmenti fono-sintattici, senza particolari sfumature semantiche: agli attori è stato spiegato che è questo il significato dei due punti nella frase 4. Infine la *tonia interrogativa* è rappresentata dalle frasi interrogativa

<sup>5</sup> Per cui si rimanda a D'Urso – Cavicchio – Magno Caldognetto 2008.

<sup>6</sup> Frasi emotivamente neutre in assoluto in realtà non esistono, ed è del tutto dipendente dalle idiosincrasie e dai vissuti personali dei singoli interpreti il fatto che una frase “neutra” possa addirsi ad un'emozione piuttosto che ad un'altra.

<sup>7</sup> Cf. ad esempio Scherer 1981.

<sup>8</sup> Colgo qui l'occasione per ringraziare la Dottoressa Magno Caldognetto per i preziosi suggerimenti datimi al riguardo.

parziale 14, che si distingue dalla frase 13, interrogativa totale (o a risposta sì/no, che è invece caratterizzata di norma da una tonia conclusiva). Questa distinzione è importante perché, come rilevano Scherer et al. 1984, il profilo intonativo di una domanda può cambiare in base a questi due casi.<sup>9</sup> Come specificato per la tonia conclusiva, anche per gli altri tipi di tonia la loro effettiva realizzazione, secondo i criteri descrittivi di Canepari o di altri studiosi, può cambiare notevolmente di emozione in emozione: in alcune emozioni date tonie possono essere annullate o modificate a fini espressivi più specifici, e secondo quali modalità ciò avvenga sarà precipuamente il risultato delle analisi prosodiche che si potranno effettuare sul materiale offerto da questo database.

#### 2.2.4 Registrazione e archiviazione dei dati

La *performance* degli attori è stata registrata in una stanza dei laboratori della Fondazione Ugo Bordoni a Roma, in orari che permettessero adeguate condizioni di quiete acustica. Sono stati impiegati due microfoni professionali marca SHURE modello SM58LC ed un registratore digitale MARANTZ modello PMD670. Le registrazioni sono state eseguite con una frequenza di campionamento di 48 kHz, 16 bit stereo, formato wav. Gli attori hanno avuto la possibilità di muoversi liberamente e ciò ha ovviamente inficiato i dati relativi all'intensità assoluta del segnale, dipendente dalla lontananza della bocca dal microfono. In altri casi si è invece dovuto provvedere all'abbassamento manuale del volume di registrazione, onde evitare la saturazione del segnale nei casi di emozioni che generano livelli di energia elevati, quali la rabbia.

Si è consentito agli attori di avere tutto il tempo necessario per prepararsi all'autoinduzione dell'emozione, concedendo loro le pause richieste e consentendo la ripetizione delle frasi la cui esecuzione non li convinceva.

Ogni sessione di registrazione è durata all'incirca un'ora. Successivamente le frasi registrate sono state isolate, una per *file*, ed archiviate nel database in 6 cartelle diverse corrispondenti ai 6 attori. Ogni *file* è stato etichettato nel seguente modo:

##### ***neu-m1-b1.wav***

Nella prima parte del nome è indicato lo stato emotivo (**neutro**, **disgusto**, **gioia**, **paura**, **rabbia**, **sorpresa**, **tristezza**).

Nella seconda parte del nome è identificato l'attore/attrice (**m1**, **m2**, **m3**, **f1**, **f2**, **f3**) che recita la frase.

Nella terza parte del nome è indicato il tipo di frase:

le frasi 1-3 sono state classificate rispettivamente. b1, b2, b3 (**brevi**)

le frasi 4-7 sono state classificate risp. l1, l2, l3, l4 (**lunghe**)

le frasi 8-12 sono state classificate risp. n1, n2, n3, n4, n5 (**nonsense**)

le frasi 13,14 sono state classificate risp. d1, d2 (**domande**).

Infine, tutte le frasi sono state inserite in un database elaborato col programma Access, grazie al quale è possibile ordinare i *record* a piacere in base all'emozione, al testo della frase, al tipo di frase o all'attore/attrice, nonché procedere all'ascolto della frase cliccando su un apposito *link*.

In totale sono stati archiviati 588 *record*: per ognuno dei 6 attori sono state registrate 98 frasi, corrispondenti alle 14 frasi pronunciate nei 6 stati emotivi più lo stato neutro.

---

<sup>9</sup> In particolare rilevano come uno contorno del pitch discendente venga percepito come neutro in una "WH-question", come aggressivo in una "Y/N question."

Il totale del materiale registrato, escluse le pause, è all'incirca di 10 minuti per attore, quindi un'ora per l'intero database. EMOVO è attualmente disponibile su un singolo cd-rom: sarà distribuito a che ne farà [richiesta](#).

### 3. VALIDAZIONE DEL DATABASE EMOVO

#### 3.1 *Il decoding del parlato emotivo*

La capacità umana di riconoscere le emozioni in base a meri stimoli vocali ha interessato psicologi e psichiatri sin dall'inizio del secolo passato, in special modo fra gli anni '50 e '80. Gli studi al riguardo si sono configurati per lo più in questo modo: si è richiesto a dei soggetti ascoltatori di indicare lo stato emotivo di stimoli vocali fatti loro ascoltare (nella maggior parte dei casi, elicitati da attori), facendo loro scegliere fra una lista più o meno lunga di alternative; successivamente sono state calcolate le percentuali di stimoli correttamente interpretati ed è stato confrontato tale risultato con le percentuali di risposte che ci si sarebbe aspettati se il soggetto avesse deciso a caso. In alcuni studi sono state inoltre elaborate delle "matrici di confusione", per inquadrare meglio i tipi di errori e di false attribuzioni cui sono andati incontro i soggetti. Secondo Scherer – Johnstone – Klasmeyer 2003, a tirare le somme di tutti questi studi (all'incirca una trentina) si può quantificare la percentuale di accuratezza umana nel riconoscere le emozioni, in base a soli stimoli vocali, intorno al 60%: una percentuale cinque volte maggiore di quanto ci si aspetterebbe per effetto del caso. E' un livello di accuratezza notevole, se si tiene in considerazione che in molti di questi studi gli stimoli vocali erano assai brevi e che gli stati emotivi con cui li si caratterizzava erano rari o poco studiati (come l'orgoglio o la gelosia), se paragonati alle più comuni emozioni di base. Anche altri studiosi riportano valori analoghi di percentuale di accuratezza: 65% per Van Bezooijen 1984; 56% per Scherer – Banse – Wallbott – Goldbeck 1991.

Questi valori sono inferiori, in media, di un 15% rispetto alla capacità umana di riconoscere emozioni in base alla mimica facciale (cf. Ekman 1994): in realtà, la differenza fra riconoscimento a partire dalla voce e riconoscimento dalla mimica facciale cambia alquanto di emozione in emozione, come si può vedere nella seguente tabella (fonti: Ekman 1994 per la prima riga, Scherer – Banse – Wallbott 2001 per la seconda):

|                        | neutro | rabbia | paura | gioia | tristezza | disgusto | sorpresa | media      |
|------------------------|--------|--------|-------|-------|-----------|----------|----------|------------|
| <b>mimica facciale</b> | N.P.   | 78%    | 77%   | 95%   | 79%       | 80%      | 88%      | <b>78%</b> |
| <b>voce</b>            | 74%    | 77%    | 61%   | 57%   | 71%       | 31%      | N.P.     | <b>62%</b> |

A falsare i valori medi sono soprattutto i valori relativi a gioia e disgusto, fin troppo riconoscibili nel volto di una persona. Per quanto riguarda il riconoscimento vocale possiamo rilevare come vengano identificate meglio la tristezza e la rabbia, seguite dalla paura. E' interessante notare come questi primati di riconoscimento siano da inquadrare per Johnstone – Scherer 2000 nel contesto dell'evoluzione darwinianamente intesa: essa avrebbe selezionato e favorito, per la rabbia e la paura, sviluppi nell'espressione vocale, in modo che gli antenati dell'uomo potessero avvertirsi e minacciarsi in modo esplicito, anche a lunghe distanze. Avrebbe invece selezionato e favorito sviluppi biologici relativi alla mimica facciale per emozioni quali disgusto e gioia, di modo che fra compagni, nel senso etimologico del termine, ci si potesse capire al volo. Ciò presuppone, naturalmente, che l'espressione facciale sia in qualche modo indipendente dalla volontà dell'individuo, cosa nel del tutto ovvia.

Per quanto riguarda invece le attribuzioni erranee, dalle matrici di confusione dei vari studi si possono evincere dati importanti sul processo di *decoding* del soggetto, e di riflesso, sulla natura del parlato emotivo. Dallo studio di Banse – Scherer 1996 è emerso, ad esempio, che le confusioni non avvenivano casualmente fra tutte le emozioni, ma tra famiglie di emozioni: segno che non si trattava di veri e propri errori, ma di approssimazioni (dallo studio delle quali emergono rapporti di contiguità e prossimità fra le emozioni, secondo il modello dimensionale di cui si è parlato nel primo capitolo: resterebbe da indagare più in dettaglio quali dimensioni siano maggiormente coinvolte nel generare confusione). Per citare degli esempi, dallo studio sopra ricordato è emerso che il disgusto è stato confuso solo con emozioni negative, che la rabbia calda è stata confusa quasi esclusivamente con la rabbia fredda e la riprovazione, che l'interessamento è stato confuso con l'orgoglio e la gioia molto più frequentemente che con le altre emozioni. Da altri studi riporto ulteriori correlazioni: tristezza confusa con amore; vergogna confusa con tristezza; disprezzo confuso con collera o disgusto; gioia e sorpresa reciprocamente confuse (Fairbanks – Pronovost 1939, Davitz – Davitz 1959, Davitz 1964, Dimitrovsky 1964, Albas et al. 1976, Van Bezooijen 1984). Tuttavia, per quanto si possano trovare correlazioni sistematiche nelle matrici di confusione all'interno dei singoli studi, scarseggiano correlazioni fra le matrici di confusioni dei vari studi. Ciò è dovuto senz'altro ai seguenti motivi:

- diverso assortimento e/o etichettamento delle emozioni impiegate
- diversità nelle impostazioni dei test, in particolar modo riguardo al tipo di domande
- diverse moli di dati raccolte, in particolar modo riguardo al numero dei soggetti e alla quantità del materiale sottoposto ai test, se vocali, frasi, brani o altro.

### 3.2 Descrizione del test di validazione

Gli studi citati nel paragrafo precedente sono stati spesso criticati per aver usato un *set* ristretto e limitato di emozioni da far scegliere ai soggetti ascoltatori, caratterizzandosi pertanto come test di discriminazione più che test di riconoscimento. Nel nostro caso, invece, il problema non si pone, essendo il fine principale del test non il riconoscimento né la discriminazione, ma la validazione del database EMOVO, ovvero garantire la bontà delle *performance* emotive degli attori coinvolti. Tale test è stato somministrato a due gruppi di soggetti ascoltatori: i risultati verranno presentati separatamente e congiuntamente.

#### 3.2.1 Campione

- Primo gruppo:  
12 soggetti<sup>10</sup>

---

<sup>10</sup> Il primo gruppo è costituito da 6 maschi e 6 femmine, di età compresa fra i 21 e i 32 anni. Riporto qui i loro nomi non solo a mo' di ringraziamento, ma anche per indicarne il sesso, la cui influenza sulla performance riconoscitiva verrà presa in considerazione nella discussione dei risultati. **Soggetto 1:** Stefano V. Talone, 24 anni. **Soggetto 2:** Nicola Villa, 23 anni. **Soggetto 3:** Elisabetta Lopalco, 24 anni. **Soggetto 4:** Erica Pulvirenti, 22 anni. **Soggetto 5:** Beatrice Bertini, 29 anni. **Soggetto 6:** Michelangelo Iadarola, 26 anni. **Soggetto 7:** Marina Sforza, 22 anni. **Soggetto 8:** Valerio Friello, 21 anni. **Soggetto 9:** Chiara Carmosino, 21 anni. **Soggetto 10:** Luigina Ruffolo, 23 anni. **Soggetto 11:** Valerio Pescatore, 32 anni. **Soggetto 12:** Rio Olmedo, 22 anni.

- Secondo gruppo:

12 soggetti<sup>11</sup>

### 3.2.2 Procedura

Ai soggetti è stata fornita oralmente una sintetica descrizione dei fini del test. Si è successivamente somministrato loro un pre-test per acclimatarli all'esperimento e al software impiegato a tal fine. Il test preliminare è consistito nell'ascolto di 6 frasi, una per ogni stato emotivo, una per attore: di frase in frase il soggetto doveva scegliere fra una coppia di risposte possibili (ex: rabbia/gioia, paura/tristezza).

In seguito si è somministrato al soggetto ascoltatore il test di validazione completo, così costituito:

Per ogni attore, fra le 14 tipologie di frasi del database, sono state scelte due frasi *nonsense*:

- la casa forte vuole col pane
- il gatto sta scorrendo nella pera

ciò per ridurre al minimo l'eventualità che il contenuto semantico delle frasi potesse indirizzare o sviare l'attribuzione emotiva degli ascoltatori. Di frasi ne sono state scelte due in modo che, nel caso di due domande con diverse emozioni da riconoscere ma con una medesima coppia di risposte, si potesse usare ora una frase, ora l'altra. Ovviamente, di tali tipologie di frasi sono state selezionate all'interno del database tutte le coloriture emotive, per poterle presentare al soggetto ascoltatore pronunciate nei sette stati emotivi: neutro, disgusto, gioia, paura, rabbia, sorpresa, tristezza.

Quindi la frase di un attore, per ogni stato emotivo, è stata presentata al soggetto ascoltatore sei volte, fornendo ogni volta una coppia diversa di risposte alternative: una risposta giusta per lo stato emotivo in questione, una risposta sbagliata relativa a un altro stato emotivo scelto fra gli altri sei.

Riporto, a mo' di esemplificazione, la parte iniziale della strutturazione del test:

#### Domanda 1

in quale stato emotivo pensi sia pronunciata la domanda? (frase da ascoltare nello stato emotivo **neutro**).

- neutro
- disgusto

#### Domanda 2

in quale stato emotivo pensi sia pronunciata la domanda? (frase da ascoltare nello stato emotivo **neutro**).

- neutro
- gioia

[...]

#### Domanda 6

in quale stato emotivo pensi sia pronunciata la domanda? (frase da ascoltare nello stato emotivo **neutro**).

- neutro
- tristezza

#### Domanda 7

---

<sup>11</sup> Il secondo gruppo è costituito da studenti della Facoltà di Lettere e Filosofia dell'Università della Calabria, cui è stato somministrato il mio test grazie alla gentile collaborazione del Prof. Luciano Romito e del Dott. Vincenzo Galatà, su invito dell'Ing. Andrea Paoloni della Fondazione Ugo Bordoni.

in quale stato emotivo pensi sia pronunciata la domanda? (frase da ascoltare nello stato emotivo di **disgusto**).

- disgusto
- neutro

(In quest'ultima domanda, essendosi presentata la stessa coppia di risposte della domanda 1, oltre allo stato emotivo della frase da sentire è stato cambiato anche il testo, come spiegato prima)

E così via, per un totale di 42 domande (sei domande per sette stati emotivi). L'ordine delle domande, così come quello delle risposte, è stato opportunamente reso casuale tramite un'apposita funzione del software di *test-making*. Lo stesso software ha fornito l'interfaccia da presentare al soggetto ascoltatore, contenente il testo della domanda, le due caselle delle risposte da segnare, un pulsante per sentire il file audio abbinato alla domanda, un pulsante per procedere alla domanda successiva. E' stato concesso al soggetto di poter riascoltare la frase della domanda tutte le volte che ritenesse necessario. Non è stato concesso al soggetto di poter controllare le risposte precedenti. Orientativamente, tutti i soggetti hanno impiegato dieci minuti di tempo per rispondere alla batteria di 42 domande.

Ovviamente, la batteria delle 42 domande così strutturate, è stata approntata per ogni attore e per ogni attrice. A ogni soggetto ascoltatore sono state somministrate due batterie di domande, una per un attore maschio e una per un'attrice, secondo questo schema.

Batteria relativa a M1

Somministrata ai soggetti 1 2 3 4 di entrambi i gruppi di ascolto

Batteria relativa a F1

Somministrata ai soggetti 1 2 3 4 id.

Batteria relativa a M2

Somministrata ai soggetti 5 6 7 8 id.

Batteria relativa a F2

Somministrata ai soggetti 5 6 7 8 id.

Batteria relativa a M3

Somministrata ai soggetti 9 10 11 12 id.

Batteria relativa a F3

Somministrata ai soggetti 9 10 11 12 id.

### 3.3 *Presentazione dei risultati del test*

Mi sono premurato di calcolare l'accuratezza del riconoscimento correggendo i dati percentuali delle risposte secondo questa formula:

$$\text{accuratezza} = [(\text{risposte corrette} - \text{risposte sbagliate}) / \text{numero delle domande}] \cdot 100$$

Vale a dire si è sottratto il numero delle risposte sbagliate da quello delle risposte corrette, e si è divisa la differenza per il numero totale delle domande.

Con tale formula, che si è applicata in tutti i casi, si è voluto eliminare l'effetto della casualità nelle risposte date dai soggetti ascoltatori. I risultati sono stati qui approssimati per chiarezza espositiva.



### 3.3.1 Risultati relativi al primo gruppo

#### 3.3.1.1 Risultati relativi agli attori

| <b>RIQUADRO 1</b>             |                   |                    |                |                           |
|-------------------------------|-------------------|--------------------|----------------|---------------------------|
| <b>BATTERIA RELATIVA A M1</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 1                             | 37                | 5                  | 42             | 76%                       |
| 2                             | 40                | 2                  | 42             | 90%                       |
| 3                             | 36                | 6                  | 42             | 71%                       |
| 4                             | 40                | 2                  | 42             | 90%                       |
|                               |                   |                    |                | <b>accuratezza per M1</b> |
| 1,2,3,4                       | 153               | 15                 | 168            | <b>82%</b>                |
| <b>BATTERIA RELATIVA A F1</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 1                             | 36                | 6                  | 42             | 71%                       |
| 2                             | 39                | 3                  | 42             | 86%                       |
| 3                             | 37                | 5                  | 42             | 76%                       |
| 4                             | 36                | 6                  | 42             | 71%                       |
|                               |                   |                    |                | <b>accuratezza per F1</b> |
| 1,2,3,4                       | 148               | 20                 | 168            | <b>76%</b>                |
| <b>BATTERIA RELATIVA A M2</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 5                             | 40                | 2                  | 42             | 90%                       |
| 6                             | 37                | 5                  | 42             | 76%                       |
| 7                             | 37                | 5                  | 42             | 76%                       |
| 8                             | 38                | 4                  | 42             | 81%                       |
|                               |                   |                    |                | <b>accuratezza per M2</b> |
| 5,6,7,8                       | 152               | 16                 | 168            | <b>81%</b>                |
| <b>BATTERIA RELATIVA A F2</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 5                             | 41                | 1                  | 42             | 95%                       |
| 6                             | 38                | 4                  | 42             | 81%                       |
| 7                             | 40                | 2                  | 42             | 90%                       |
| 8                             | 40                | 2                  | 42             | 90%                       |
|                               |                   |                    |                | <b>accuratezza</b>        |

|                         |                   |                    |                |                          |
|-------------------------|-------------------|--------------------|----------------|--------------------------|
|                         |                   |                    |                | per F2                   |
| 5,6,7,8                 | 159               | 9                  | 168            | 89%                      |
| BATTERIA RELATIVA A M3  |                   |                    |                |                          |
| soggetto ascoltatore    | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto |
| 9                       | 37                | 5                  | 42             | 76%                      |
| 10                      | 40                | 2                  | 42             | 90%                      |
| 11                      | 37                | 5                  | 42             | 76%                      |
| 12                      | 36                | 6                  | 42             | 71%                      |
|                         |                   |                    |                | accuratezza per M3       |
| 9,10,11,12              | 150               | 18                 | 168            | 79%                      |
| BATTERIA RELATIVA A F3  |                   |                    |                |                          |
| soggetto ascoltatore    | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto |
| 9                       | 40                | 2                  | 42             | 90%                      |
| 10                      | 38                | 4                  | 42             | 81%                      |
| 11                      | 40                | 2                  | 42             | 90%                      |
| 12                      | 41                | 1                  | 42             | 95%                      |
|                         |                   |                    |                | accuratezza per F3       |
| 9,10,11,12              | 159               | 9                  | 168            | 89%                      |
| ACCURATEZZA RELATIVA A: |                   |                    |                |                          |
| ATTRICI                 | risposte corrette | risposte sbagliate | numero domande |                          |
| 85%                     | 466               | 38                 | 504            |                          |
| ATTORI                  |                   |                    |                |                          |
| 81%                     | 455               | 49                 | 504            |                          |
| COMPLESSIVA             |                   |                    |                |                          |
| 83%                     | 921               | 87                 | 1008           |                          |

3.3.1.2 Risultati relativi ai soggetti ascoltatori del primo gruppo:

| <b>RIQUADRO 2</b>            |                        |                           |                       |                    |
|------------------------------|------------------------|---------------------------|-----------------------|--------------------|
| <b>soggetto</b>              | <b>risposte giuste</b> | <b>risposte sbagliate</b> | <b>numero domande</b> | <b>Accuratezza</b> |
| 1                            | 73                     | 11                        | 84                    | 74%                |
| 2                            | 79                     | 5                         | 84                    | 88%                |
| <b>3</b>                     | <b>73</b>              | <b>11</b>                 | <b>84</b>             | <b>74%</b>         |
| <b>4</b>                     | <b>76</b>              | <b>8</b>                  | <b>84</b>             | <b>81%</b>         |
| <b>5</b>                     | <b>81</b>              | <b>3</b>                  | <b>84</b>             | <b>93%</b>         |
| 6                            | 75                     | 9                         | 84                    | 79%                |
| 7                            | 77                     | 7                         | 84                    | 83%                |
| 8                            | 78                     | 6                         | 84                    | 86%                |
| 9                            | 77                     | 7                         | 84                    | 83%                |
| <b>10</b>                    | <b>78</b>              | <b>6</b>                  | <b>84</b>             | <b>86%</b>         |
| 11                           | 77                     | 7                         | 84                    | 83%                |
| 12                           | 77                     | 7                         | 84                    | 83%                |
| soggetti maschi              | 459                    | 45                        | 504                   | 82%                |
| <b>soggetti femmine</b>      | <b>462</b>             | <b>42</b>                 | <b>504</b>            | <b>83%</b>         |
| <b>risultati complessivi</b> | <b>921</b>             | <b>87</b>                 | <b>1008</b>           | <b>83%</b>         |

(in neretto sono evidenziati i soggetti di sesso femminile)

3.3.1.3 Risultati relativi alle emozioni del primo gruppo

Tale tabella va letta nel seguente modo:

- emozioni nelle ordinate confuse con emozioni nelle ascisse nel X% dei casi
- quando l'emozione dei due assi è la stessa il valore nella casella indica l'accuratezza di riconoscimento di quell'emozione per tutti i soggetti del gruppo

| <b>EMOZIONE ELICITATA</b> | <b>RIQUADRO 3</b> | <b>EMOZIONE PERCEPITA</b> |                 |              |              |               |                 |                  |
|---------------------------|-------------------|---------------------------|-----------------|--------------|--------------|---------------|-----------------|------------------|
|                           |                   | <b>Neutro</b>             | <b>Disgusto</b> | <b>Gioia</b> | <b>Paura</b> | <b>Rabbia</b> | <b>Sorpresa</b> | <b>Tristezza</b> |
|                           | <b>Neutro</b>     | <b>96%</b>                | 1,3%            | 0%           | 0%           | 1,3%          | 0%              | 1,3%             |
|                           | <b>Disgusto</b>   | 3%                        | <b>76%</b>      | 3%           | 6%           | 6%            | 2%              | 6%               |
|                           | <b>Gioia</b>      | 1%                        | 3%              | <b>64%</b>   | 6%           | 8%            | 15%             | 3%               |
|                           | <b>Paura</b>      | 1%                        | 10%             | 0%           | <b>72%</b>   | 4%            | 6%              | 7%               |
|                           | <b>Rabbia</b>     | 0%                        | 3%              | 0%           | 3%           | <b>93%</b>    | 0%              | 1%               |
|                           | <b>Sorpresa</b>   | 1,5%                      | 3%              | 3%           | 1,5%         | 0%            | <b>83%</b>      | 8%               |
|                           | <b>Tristezza</b>  | 1,3%                      | 0%              | 1,3%         | 1,3%         | 0%            | 0%              | <b>96%</b>       |

### 3.3.2 Risultati relativi al secondo gruppo

#### 3.3.2.1 Risultati relativi agli attori

| <b>RIQUADRO 4</b>             |                   |                    |                |                           |
|-------------------------------|-------------------|--------------------|----------------|---------------------------|
| <b>BATTERIA RELATIVA A M1</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 1                             | 35                | 7                  | 42             | 67%                       |
| 2                             | 37                | 5                  | 42             | 76%                       |
| 3                             | 39                | 3                  | 42             | 86%                       |
| 4                             | 41                | 1                  | 42             | 95%                       |
|                               |                   |                    |                | <b>accuratezza per M1</b> |
| 1,2,3,4                       | 152               | 16                 | 168            | <b>81%</b>                |
| <b>BATTERIA RELATIVA A F1</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 1                             | 37                | 5                  | 42             | 76%                       |
| 2                             | 35                | 7                  | 42             | 67%                       |
| 3                             | 37                | 5                  | 42             | 76%                       |
| 4                             | 41                | 1                  | 42             | 95%                       |
|                               |                   |                    |                | <b>accuratezza per F1</b> |
| 1,2,3,4                       | 150               | 18                 | 168            | <b>79%</b>                |
| <b>BATTERIA RELATIVA A M2</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 5                             | 39                | 3                  | 42             | 86%                       |
| 6                             | 34                | 8                  | 42             | 62%                       |
| 7                             | 39                | 3                  | 42             | 86%                       |
| 8                             | 36                | 6                  | 42             | 71%                       |
|                               |                   |                    |                | <b>accuratezza per M2</b> |
| 5,6,7,8                       | 148               | 20                 | 168            | <b>76%</b>                |
| <b>BATTERIA RELATIVA A F2</b> |                   |                    |                |                           |
| soggetto ascoltatore          | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 5                             | 38                | 4                  | 42             | 81%                       |
| 6                             | 31                | 11                 | 42             | 48%                       |
| 7                             | 40                | 2                  | 42             | 90%                       |
| 8                             | 41                | 1                  | 42             | 95%                       |
|                               |                   |                    |                | <b>accuratezza per F2</b> |
| 5,6,7,8                       | 150               | 18                 | 168            | <b>79%</b>                |
| <b>BATTERIA RELATIVA A M3</b> |                   |                    |                |                           |

| soggetto ascoltatore           | risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
|--------------------------------|-------------------|--------------------|----------------|---------------------------|
| 9                              | 35                | 7                  | 42             | 67%                       |
| 10                             | 39                | 3                  | 42             | 86%                       |
| 11                             | 36                | 6                  | 42             | 71%                       |
| 12                             | 38                | 4                  | 42             | 81%                       |
|                                |                   |                    |                | <b>accuratezza per M3</b> |
| 9,10,11,12                     | 148               | 20                 | 168            | <b>76%</b>                |
| <b>BATTERIA RELATIVA A F3</b>  |                   |                    |                |                           |
| soggetto ascoltatore           | Risposte corrette | risposte sbagliate | numero domande | accuratezza per soggetto  |
| 9                              | 36                | 6                  | 42             | 71%                       |
| 10                             | 39                | 3                  | 42             | 86%                       |
| 11                             | 35                | 7                  | 42             | 67%                       |
| 12                             | 40                | 2                  | 42             | 90%                       |
|                                |                   |                    |                | <b>accuratezza per F3</b> |
| 9,10,11,12                     | 150               | 18                 | 168            | <b>79%</b>                |
| <b>ACCURATEZZA RELATIVA A:</b> |                   |                    |                |                           |
| <b>ATTRICI</b>                 | risposte corrette | risposte sbagliate | numero domande |                           |
| <b>79%</b>                     | 450               | 54                 | 504            |                           |
| <b>ATTORI</b>                  |                   |                    |                |                           |
| <b>78%</b>                     | 448               | 56                 | 504            |                           |
| <b>COMPLESSIVA</b>             |                   |                    |                |                           |
| <b>78%</b>                     | 898               | 110                | 1008           |                           |

### 3.3.2.2 Risultati relativi ai soggetti ascoltatori del secondo gruppo

| <b>RIQUADRO 5</b>            |                          |                           |                       |                    |
|------------------------------|--------------------------|---------------------------|-----------------------|--------------------|
| <b>soggetto</b>              | <b>risposte corrette</b> | <b>risposte sbagliate</b> | <b>numero domande</b> | <b>Accuratezza</b> |
| 1                            | 72                       | 12                        | 84                    | 71%                |
| 2                            | 72                       | 12                        | 84                    | 71%                |
| 3                            | 76                       | 8                         | 84                    | 81%                |
| 4                            | 82                       | 2                         | 84                    | 95%                |
| 5                            | 77                       | 7                         | 84                    | 83%                |
| 6                            | 65                       | 19                        | 84                    | 55%                |
| 7                            | 79                       | 5                         | 84                    | 88%                |
| 8                            | 77                       | 7                         | 84                    | 83%                |
| 9                            | 71                       | 13                        | 84                    | 69%                |
| 10                           | 78                       | 6                         | 84                    | 86%                |
| 11                           | 71                       | 13                        | 84                    | 69%                |
| 12                           | 78                       | 6                         | 84                    | 86%                |
| <b>risultati complessivi</b> | <b>898</b>               | <b>110</b>                | <b>1008</b>           | <b>78%</b>         |

### 3.3.2.3 Risultati relativi alle emozioni del secondo gruppo

Tale tabella va letta nel seguente modo:

- emozioni nelle ordinate confuse con emozioni nelle ascisse nel X% dei casi
- quando l'emozione dei due assi è la stessa il valore nella casella indica l'accuratezza di riconoscimento di quell'emozione per tutti i soggetti del gruppo

| <b>EMOZIONE ELICITATA</b> | <b>RIQUADRO 6</b> | <b>EMOZIONE PERCEPITA</b> |                 |              |              |               |                 |                  |
|---------------------------|-------------------|---------------------------|-----------------|--------------|--------------|---------------|-----------------|------------------|
|                           |                   | <b>Neutro</b>             | <b>Disgusto</b> | <b>Gioia</b> | <b>Paura</b> | <b>Rabbia</b> | <b>Sorpresa</b> | <b>Tristezza</b> |
|                           | <b>Neutro</b>     | <b>90%</b>                | 0%              | 0%           | 0%           | 7%            | 0%              | 3%               |
|                           | <b>Disgusto</b>   | 3%                        | <b>60%</b>      | 1%           | 5,5%         | 14%           | 11%             | 5,5%             |
|                           | <b>Gioia</b>      | 3%                        | 5,5%            | <b>67%</b>   | 8%           | 5,5%          | 5,5%            | 5,5%             |
|                           | <b>Paura</b>      | 3%                        | 4%              | 4%           | <b>74%</b>   | 1%            | 1%              | 13%              |
|                           | <b>Rabbia</b>     | 1,3%                      | 0%              | 1,3%         | 3%           | <b>90%</b>    | 3%              | 1,3%             |
|                           | <b>Sorpresa</b>   | 1%                        | 3%              | 6%           | 1%           | 1%            | <b>78%</b>      | 10%              |
|                           | <b>Tristezza</b>  | 3%                        | 4%              | 0%           | 4%           | 0%            | 0%              | <b>89%</b>       |

### 3.3.3 Risultati relativi a tutto il campione

#### 3.3.3.1 Risultati relativi agli attori

| <b>RIQUADRO 7</b>              |                   |                    |                |                           |
|--------------------------------|-------------------|--------------------|----------------|---------------------------|
| <b>BATTERIA RELATIVA A M1</b>  |                   |                    |                |                           |
| soggetti ascoltatori           | risposte corrette | Risposte sbagliate | numero domande | accuratezza <sup>12</sup> |
| 1,2,3,4 <sup>13</sup>          | 305               | 31                 | 336            | <b>82%</b>                |
| <b>BATTERIA RELATIVA A F1</b>  |                   |                    |                |                           |
| soggetti ascoltatori           |                   |                    |                |                           |
| 1,2,3,4                        | 298               | 38                 | 336            | <b>77%</b>                |
| <b>BATTERIA RELATIVA A M2</b>  |                   |                    |                |                           |
| soggetti ascoltatori           |                   |                    |                |                           |
| 5,6,7,8                        | 300               | 36                 | 336            | <b>79%</b>                |
| <b>BATTERIA RELATIVA A F2</b>  |                   |                    |                |                           |
| soggetti ascoltatori           |                   |                    |                |                           |
| 5,6,7,8                        | 309               | 27                 | 336            | <b>84%</b>                |
| <b>BATTERIA RELATIVA A M3</b>  |                   |                    |                |                           |
| soggetti ascoltatori           |                   |                    |                |                           |
| 9,10,11,12                     | 298               | 38                 | 336            | <b>77%</b>                |
| <b>BATTERIA RELATIVA A F3</b>  |                   |                    |                |                           |
| soggetti ascoltatori           |                   |                    |                |                           |
| 9,10,11,12                     | 309               | 27                 | 336            | <b>84%</b>                |
| <b>ACCURATEZZA RELATIVA A:</b> |                   |                    |                |                           |
| ATTRICI                        | risposte corrette | risposte sbagliate | numero domande |                           |
| <b>82%</b>                     | 916               | 92                 | 1008           |                           |
| ATTORI                         |                   |                    |                |                           |
| <b>79%</b>                     | 903               | 105                | 1008           |                           |
| <b>COMPLESSIVA</b>             |                   |                    |                |                           |
| <b>80%</b>                     | 1819              | 197                | 2016           |                           |

<sup>12</sup> Relativa all'attore/attrice per tutti i soggetti di entrambi i campioni

<sup>13</sup> Qui ovviamente ci si riferisce, come nelle celle sottostanti, ai soggetti 1,2,3,4 di entrambi i campioni

### 3.3.3.2 Risultati relativi ai soggetti ascoltatori

Come si evince dal riquadro precedente, il campione complessivo dei 24 soggetti ascoltatori ha dimostrato un'accuratezza dell'**80%**.

### 3.3.3.3 Risultati relativi alle emozioni

Tale tabella va letta nel seguente modo:

- emozioni nelle ordinate confuse con emozioni nelle ascisse nel X% dei casi
- quando l'emozione dei due assi è la stessa il valore nella casella indica l'accuratezza di riconoscimento di quell'emozione per tutti i soggetti del campione

| EMOZIONE ELICITATA | RIQUADRO 8 | EMOZIONE PERCEPITA |          |       |       |        |          |           |
|--------------------|------------|--------------------|----------|-------|-------|--------|----------|-----------|
|                    |            | Neutro             | Disgusto | Gioia | Paura | Rabbia | Sorpresa | Tristezza |
|                    | Neutro     | 93%                | 1%       | 0%    | 0%    | 4%     | 0%       | 2%        |
|                    | Disgusto   | 3%                 | 67%      | 2%    | 6%    | 10%    | 6%       | 6%        |
|                    | Gioia      | 2%                 | 4,3%     | 65%   | 7%    | 7%     | 10,3%    | 4,3%      |
|                    | Paura      | 2%                 | 7%       | 2%    | 74%   | 3%     | 3%       | 9%        |
|                    | Rabbia     | 1%                 | 1%       | 1%    | 3%    | 92%    | 1%       | 1%        |
|                    | Sorpresa   | 1%                 | 3%       | 4%    | 1%    | 1%     | 81%      | 9%        |
|                    | Tristezza  | 2%                 | 2%       | 1%    | 3%    | 0%     | 0%       | 92%       |

## 4. CONCLUSIONI

### 4.1 Discussione dei risultati del test

Si comincerà col prendere in considerazione i dati relativi alla riconoscibilità delle *performance* degli attori: per il primo gruppo l'accuratezza complessiva è dell'**83%**, per il secondo del **78%**, per ambedue dell'**80%**. Entrambi i gruppi di ascoltatori hanno riconosciuto meglio le attrici (**82%**) che gli attori (**79%**): per un così lieve scarto, e tenuto conto dell'esiguità del campione di ascolto, sarebbe azzardato fare generalizzazioni sulla migliore capacità delle attrici di esprimere emozioni.<sup>14</sup> Tenuto invece presente che l'obiettivo del test è precipuamente quello di garantire la bontà delle *performance* degli attori coinvolti nella creazione del database EMOVO, ci si può ritenere più che soddisfatti: un'accuratezza di riconoscimento dell'**80%** è di molto al di sopra dei valori medi desunti dalla letteratura sul riconoscimento delle emozioni, come riportati in § 3.1, che si collocano intorno al 62%: tuttavia non si dovrebbe dare troppa importanza a questo scarto, dal

<sup>14</sup> Per il Prof. Paolo Di Giovine, su comunicazione personale, la migliore riconoscibilità delle emozioni espresse dalle attrici potrebbe esser dovuta, *ceteris paribus*, al fatto che la voce femminile ricopre un gamma di frequenze maggiore rispetto a quella maschile. Ciò potrebbe senz'altro darsi, considerato il ruolo preminente che i livelli di pitch giocano nell'espressione dell'emozione, ma necessiterebbe di una verifica sperimentale che purtroppo esula dal presente lavoro.



momento che quest'ultimo valore si riferisce a test di riconoscimento senza dubbio più difficili del test sopra presentato, con risposte aperte o con molte più alternative di scelta.

Prendiamo ora invece in considerazione le *performance* dei singoli attori, partendo dai risultati di tutto il campione di ascolto: come è visibile dal riquadro 7, gli attori più riconoscibili risultano essere stati F2 e F3 (**84%**), a seguire M1 (**82%**), M2 (**79%**), e infine F1 e M3 (**77%**). Quel che è più importante rilevare è però il fatto che fra gli attori/attrici non ci sono scarti significativi che possano pregiudicare l'impiego, per eventuali analisi, del materiale elicitato da un attore piuttosto che da un altro: la riconoscibilità è sufficientemente garantita per tutti gli attori e le attrici.

Un'osservazione analoga va fatta per l'abilità dei soggetti ascoltatori: a parte il soggetto 6 del secondo gruppo di ascolto, non si sono rilevate significative devianze dall'indice di accuratezza dell'80%, relativo alla riconoscibilità delle emozioni da parte di tutti i soggetti. Un'altra differenza è quella relativa alle percentuali di riconoscimento, nel primo gruppo di ascolto, dei soggetti di sesso femminile rispetto a quelli di sesso maschile (cf. riquadro 2): uno scarto così lieve su un campione così esiguo può bastare, se non altro, a farci riflettere in rapporto ai casi in cui si è registrata una maggiore abilità di riconoscimento da parte delle donne (Dusenberry – Knower 1939), e, per converso, ai casi in cui si è registrata una sostanziale omogeneità fra i sessi nelle abilità riconoscitive (Hall 1978, Van Bezooijen 1984, Anolli – Ciceri 1992).

Restano da considerare ora i risultati pertinenti alle emozioni. Come si può vedere dal riquadro 8, nella scala di riconoscibilità delle emozioni c'è una sostanziale concordanza degli indici di riconoscibilità del campione del nostro test con gli indici di riconoscibilità delle emozioni desunti dalla letteratura (cf. § 3.1)<sup>15</sup>:

|                                           |                 |                    |                    |                |                   |                   |
|-------------------------------------------|-----------------|--------------------|--------------------|----------------|-------------------|-------------------|
| <b>Emozioni<br/>nel nostro<br/>test</b>   | Neutro<br>(93%) | Tristezza<br>(92%) | Rabbia<br>(91%)    | Paura<br>(73%) | Disgusto<br>(67%) | Gioia<br>(65%)    |
| <b>Emozioni<br/>nella<br/>letteratura</b> | Rabbia<br>(77%) | Neutro<br>(74%)    | Tristezza<br>(71%) | Paura<br>(61%) | Gioia<br>(57%)    | Disgusto<br>(31%) |

Se non si dà troppa importanza ai singoli livelli, che differiscono fra loro solo per pochi punti percentuale, si nota come la concordanza sia molto significativa: il fattore di correlazione fra le due serie di dati è di **0.87**. Per quanto riguarda, invece, il livello inferiore dei valori della letteratura rispetto ai nostri, ciò è del tutto normale. Ricordiamo quanto affermato poc'anzi: essendo il fine del test la validazione della qualità del materiale registrato del database EMOVO, e non la valutazione della capacità umana di riconoscere le emozioni, si è progettato il test in modo meno complesso che nel secondo caso, con un conseguente livello di difficoltà minore.

Ma quel che è più importante notare è come siano nettamente più riconoscibili la rabbia e la tristezza rispetto alla gioia e al disgusto, la paura avendo una riconoscibilità media. Ciò concorda perfettamente con quanto affermano Scherer – Johnstone – Klasmeyer 2003, p. 444, in una loro sintesi teorica sul parlato emotivo: “Sadness and anger are generally best

<sup>15</sup> La sorpresa, di cui non dispongo una media degli indici di riconoscibilità degli studi in cui è stata presa in considerazione, è risultata riconoscibile a un livello medio-alto (81%), fra la paura e la rabbia/tristezza

recognized vocally, followed by fear. Joy has rather mixed accuracy percentages in different studies, possibly due to differences with respect to quiet happiness versus elated joy being portrayed by the actors.” Senz’altro è quel che è avvenuto nel nostro caso: in italiano il termine “gioia” si presta molto di più ad essere inteso come contentezza esaltata che come serena contentezza: sono memore del fatto che a tre attori su sei, di quelli da me convocati, la prima idea che è venuta in mente per autoindursi lo stato di gioia è stato “l’aver vinto la lotteria” (di fronte a tale associazione io non mi espressi per evitare di condizionare la *performance*). Il risultato è quindi stato una gioia presumibilmente più riconoscibile rispetto alla gioia nelle sue rimanenti accezioni. Un discorso analogo va fatto per il disgusto, il cui basso indice di riconoscibilità desunto dalla letteratura è dovuto indubbiamente all’ambiguità semantica su cui mi sono soffermato nella nota 3, per cui nei vari studi sono stati persino rilevati due profili acustici del disgusto molto diversi, l’uno con una media del pitch maggiore al parlato non emotivo, l’altro con una media del pitch minore. Essendomi invece premurato di chiarire bene agli attori che si richiedeva il disgusto “fisico” piuttosto che quello “morale”, ne è conseguita una resa senza dubbio più chiaramente riconoscibile.

Infine, qualche considerazione sulle matrici di confusione. Come si vede dal riquadro 8, le confusioni fra due emozioni che si sono verificate maggiormente (stabilendo una soglia d’interesse del 5% dei casi) sono:

- 10,3 % gioia confusa con sorpresa
- 10% disgusto con rabbia
- 9% sorpresa con tristezza
- 9% paura con tristezza
- 9% gioia con paura
- 7% gioia con rabbia
- 7% paura con disgusto
- 6% disgusto con paura
- 6% disgusto con sorpresa
- 6% disgusto con tristezza

Queste confusioni sono tutte asimmetriche, a parte quella fra disgusto e paura: una confusione simmetrica fra due emozioni attesta che esse sono molto più vicine fra loro che non emozioni confuse asimmetricamente, e, in effetti, non è difficile scorgere una forte analogia fra il disgusto e la paura, emozioni volte entrambe all’avversare e/o evitare lo stimolo che ha generato l’emozione.

Per quanto riguarda le altre confusioni, nella letteratura sono state date due distinte interpretazioni che ne rendano conto: per Van Bezooijen 1984 le emozioni si confondono in base al livello di attivazione, per cui le emozioni ad alto livello di attivazione si confondono con altre emozioni ad alto livello, e così anche per le emozioni a basso livello. Per Dawes – Kramer 1966 invece le confusioni dovrebbero imputarsi alle contiguità semantiche fra le emozioni, ad es. la rabbia col disprezzo, o, come scritto poc’anzi, la paura col disgusto. Per le confusioni che sono risultate dal nostro test, la prima spiegazione potrebbe rendere conto della gioia confusa con la paura, e di disgusto confuso con sorpresa; la seconda spiegazione di sorpresa con tristezza, paura con tristezza, disgusto con tristezza (se accettiamo le

contiguità semantiche del cerchio di Ekman<sup>16</sup>). Gli altri casi potrebbero invece essere spiegati in entrambe le direzioni.

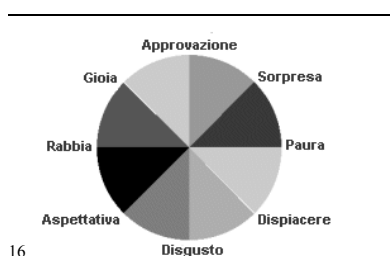
Un'altra questione, alquanto negletta nella letteratura, è quella relativa alla capacità di un'emozione non di essere confusa, ma di confondere. Per renderci conto di questo fenomeno, riporto qui una tabella col conteggio di tutte le volte che uno stato emotivo è stato erroneamente attribuito allo stato emotivo ascoltato, dato non direttamente evincibile dal riquadro 8 (dove, colonna per colonna, è invece possibile vedere le percentuali dei casi di erronea attribuzione di un'emozione sbagliata a quella ascoltata):

| neutro | disgusto | gioia | Paura | rabbia | Sorpresa | tristezza |
|--------|----------|-------|-------|--------|----------|-----------|
| 16     | 26       | 14    | 28    | 35     | 31       | 47        |

E' molto interessante notare come l'emozione che è stata attribuita erroneamente più volte, la tristezza, è stata anche una delle emozioni più riconoscibili (riconosciuta con un'accuratezza del 92%!). Non stupisce invece che lo stato emotivo neutro, riconoscibile con un'accuratezza del 93%, sia stato attribuito erroneamente molte volte di meno. Caso diametralmente analogo alla tristezza è invece quello della gioia, emozione fra le più difficilmente riconoscibili (accuratezza del 65%), eppure attribuita erroneamente pochissime volte. Tali correlazioni sembrano avallare il pensiero di Leopardi: è proprio come se l'orecchio umano non sapesse bene cos'è la gioia, anche se capirebbe bene cosa non lo è. Parimenti, saprebbe bene cosa è la tristezza, eppure, non riuscirebbe a capire cosa non lo è.

Detto questo, non posso spingermi in ulteriori speculazioni sulle confusioni del nostro test, dal momento che questo non è stato progettato per tal fine (bisognerebbe disporre di un campione più ampio, e con domande diversamente impostate) e dal momento che non sono state fatte analisi elettroacustiche delle frasi, indispensabili, fra le altre cose, per stabilire con certezza il livello di attivazione delle ricorrenze delle emozioni (in particolar modo quelle relative alla sorpresa e al disgusto): accertato il livello di attivazione e i valori degli altri parametri acustici, si potrà avere una migliore comprensione dei motivi delle confusione delle emozioni, individuando, ad esempio, valori simili di parametri che soggiacciono alle emozioni confuse fra di loro nei test di riconoscimento.

Fatte queste precisazioni, possiamo ritenere le *performance* degli attori, il riconoscimento dei soggetti e la resa delle emozioni pienamente soddisfacenti: la validazione del database di parlato emotivo EMOVO ha avuto esito positivo.



16

#### 4.2 Possibili impieghi di EMOVO

Il database EMOVO può prestarsi ai seguenti impieghi:

- **Analisi spettroacustiche:** assodato, grazie al test di validazione, che le emozioni che abbiamo registrato sono soddisfacentemente rappresentative delle emozioni percepite, il materiale del *corpus* può essere sottoposto a numerose ulteriori misurazioni, come analisi macroprosodiche di pitch, intensità e durata, analisi microprosodiche di detti indici e delle formanti vocaliche, analisi relative alla *voice quality*, analisi combinate, in vario modo, di tutti o alcuni di questi parametri. Ma più che gli indici macroprosodici, che presumibilmente confermeranno i risultati degli altri esperimenti, sarà interessante misurare gli andamenti del profilo intonativo in base alle diverse strutture semantiche e morfosintattiche offerte dal database, territorio ancora manchevole di generalizzazioni, soprattutto per l'italiano. In particolar modo, sarà interessante osservare il co-variare del contorno del pitch in relazione da una parte con la tonia conclusiva, sospensiva, interrogativa o divisiva prevista dalla struttura delle frasi del database, dall'altra con il variare dell'emozione. Infine, le analisi dei profili acustici delle emozioni del database EMOVO potrebbero servire da modello per la messa a punto di regole prosodiche per la sintesi del parlato emotivo italiano, regole che finora sono state inferite a partire da materiale relativo ad altre lingue o, per l'italiano, da analisi su *dataset* quantitativamente limitati.
- **Analisi di decoding:** sul materiale registrato possono effettuarsi numerosi test di riconoscimento, con finalità e procedure diverse rispetto al test presentato nel terzo capitolo: *in primis* test di discriminazione di emozioni con campioni più ampi e risposte con più alternative, al fine di corroborare e correggere le osservazioni fatte nello scorso paragrafo. Altri test potrebbero invece usare tutte le frasi offerte dal database, al fine di capire quali tipologie fraseologiche inibiscano o aiutino il riconoscimento di ciascuna emozione. Inoltre, tramite l'impiego di appositi programmi, come Melodyne, il materiale registrato potrebbe essere manipolato o "mascherato" in base ai più svariati parametri, dalla velocità d'eloquio al profilo intonativo, e si potrebbe studiare, tramite appositi test di riconoscimento, come queste modificazioni influiscano sulla riconoscibilità delle emozioni, e in che misura. Infine, così come le analisi sul materiale di EMOVO potrebbero fornire profili acustici per la sintesi del parlato emotivo, allo stesso modo i test di riconoscimento sul nostro materiale potrebbero fornire indispensabili modelli di riferimento per la messa a punto di sistemi di ASR (automated speech recognition) del parlato emotivo.
- **Analisi interculturali:** i dati estrapolati da esperimenti su EMOVO potrebbero essere messi a confronto con quelli relativi alle emozioni nelle altre lingue, per gettare luce su questioni cruciali circa gli aspetti culturali vs. universali della codifica e decodifica delle emozioni.
- **Comparazioni fra materiale recitato e materiale dal vivo:** come ricordato nel § 2.1, se è auspicabile lo studio delle emozioni con materiale genuino non recitato, tuttavia in fase preliminare è necessario l'impiego di materiale elicitato in laboratorio. Quest'ultimo infatti consente di eliminare eccessive variabili di ricerca capaci di soffocare i primi avanzamenti della conoscenza in questo campo. In molti casi, in una frase registrata dal vivo non ci si potrà mai concentrare in modo ottimale sulla misurazione di un parametro, perché difficilmente si potranno scervare le misurazioni del dato parametro dalle influenze e dalle covarianze che

con esso hanno i numerosissimi altri parametri che la “realtà” offre, molti dei quali difficilmente misurabili. E comunque sia, potrebbe darsi che lo scarto fra materiale recitato e materiale naturale sia veramente minimo: per stabilire ciò, sarebbero utili test progettati esclusivamente per mettere alla prova le due tipologie di materiale in modo sistematico: “this would seem to be one of the highest priorities in the field”, come afferma Klaus Scherer.<sup>17</sup>

## RINGRAZIAMENTI

Un sentito ringraziamento va al Prof. Paolo Di Giovine dell’Università di Roma La Sapienza e all’Ing. Andrea Paoloni della Fondazione Ugo Bordoni di Roma, senza il supporto dei quali la realizzazione di EMOVO non sarebbe stata possibile. Si ringraziano inoltre il Prof. Luciano Romito e il Dott. Vincenzo Galatà per la loro collaborazione alla validazione del database.

## BIBLIOGRAFIA

- AA. VV. (2008), *Comunicazione parlata e manifestazione delle emozioni, atti del I convegno GSCP, Padova, 29 novembre-1 dicembre 2004* (a cura di Emanuela Magno Caldognetto, Federica Cavicchio e Piero Così), Liguori, disponibile in e-book.
- Alpert M., Kurtzberg R. L. & Friedhoff A. J. (1963), *Transient voice changes associated with emotional stimuli*, in “Archives of general psychiatry” 8, pp. 362-365,
- Anolli L. & Ciceri R. (1992), *La voce delle emozioni*, Franco Angeli, Milano.
- Bachorowski J. A. & Owren M. J. (1995), *Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context*, in “Psychological science” 6, pp. 219-224.
- Banse R. & Scherer K.R (1996), *Acoustic profiles in vocal emotion expression*, in “Journal of personality and social psychology” 70, pp. 614-636.
- Bortz J. (1966), *Physikalisch-akustische Korrelate der vokalen Kommunikation*, in “Arbeiten aus dem Psychologischen Institut der Universität Hamburg” 9.
- Brown B. L. (1980), *The detection of emotion in voice qualities*, in Giles H., Robinson W. P. & Smith P. M. (a cura di), *Language: social perspectives*, Pergamon press, Oxford.
- Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W. & Weiss B. (2005), *A database of german emotional speech*, in Interspeech 2005, Lisbona, pp. 1517- 1520, pp. 187-200.
- Canepari L. (1985), *L’intonazione*, Liguori, Napoli.
- Cauldwell R. (2000), *Where did the anger go? The role of context in interpreting emotion in speech*, in *ISCA Workshop on speech and emotion*, settembre 2000, Belfast, pp. 127-131.
- Coleman R. F. & Williams R. (1979), *Identification of emotional states using perceptual and acoustic analyses*, in Lawrence V. & Weinberg B. (a cura di), *Transcript of the eighth symposium: care of the professional voice (part I)*, The voice foundation, New York.
- Cresti E. & Moneglia M. (2005) (a cura di), *C-ORAL-ROM Integrated reference corpora for spoken romance languages*, Benjamins, Amsterdam-Philadelphia.

---

<sup>17</sup> Scherer 2003, p. 247.

- Davitz J. R. (1964), *Auditory correlates of vocal expressions of emotional feeling*, in Davitz J. R. (a cura di), *The communication of emotional meaning*, McGraw-Hill, New York.
- Davitz J. R. & Davitz L. J. (1959), *The communication of feelings by content-free speech*, in "Journal of communication" 9, pp. 6-13.
- Dawes R. M. & Kramer E. (1966), *A proximity analysis of vocally expressed emotion*, in "Perceptual and motor skills" 22, pp. 571-574.
- De Dominicis A. (1992), *Intonazione e contesto. Uno studio su alcuni aspetti del discorso in contesto e delle sue manifestazioni intonative*, Edizioni dell'Orso, Alessandria.
- Dimitrovsky L. (1964), *The ability to identify the emotional meaning of vocal expressions at successive age levels*, in Davitz J. R. (a cura di), *The communication of emotional meaning*, McGraw-Hill, New York.
- D'Urso V., Cavicchio F. & Magno Caldognetto E. (2008), *Le etichette lessicali nelle ricerche sperimentali sulle emozioni: problemi teorici e metodologici*, In Magno Caldognetto E., Cavicchio F. & Così P. (a cura di), *Comunicazione parlata e manifestazione delle emozioni, atti del I convegno GSCP, Padova, 29 novembre-1 dicembre 2004*, Liguori, disponibile in e-book.
- Douglas-Cowie E., Campbell N., Cowie R. & Roach P. (2003), *Emotional speech: towards a new generation of databases*, in "Speech communication" 40, pp. 33-60.
- Dusenberry D. & Knowler F. H. (1939), *Experimental studies of the symbolism of action and voice II- a study of the specificity of meaning in abstract tonal symbols*, in "Quarterly journal of speech" 25, pp. 67-75.
- Ekman P. (1994), *Strong evidence for universals in facial expression: a reply to Russel's mistaken critique*, in "Psychological bulletin" 115, pp. 268-287.
- Eldred S. H. & Price D. B. (1958), *A linguistic evaluation of feeling states in psychotherapy*, in "Psychiatry" 21, pp. 115-121.
- Fairbanks G. & Hoaglin L. W. (1941), *An experimental study of the durational characteristics of the voice during the expression of emotion*, in "Science" 221, pp. 1208-1210.
- Fairbanks G. & Pronovost W. (1939), *An experimental study of the pitch characteristics of the voice during the expression of emotion*, in "Speech monographs" 8, pp. 85-90.
- Fónagy I. (1978), *A new method of investigating the perception of prosodic features*, in "Language and speech" 21, pp. 34-49.
- Fónagy I. (1983), *Emotions, voice and music*, in Sundberg J. (a cura di), *Research aspects on singing*, Payot, Parigi.
- Fónagy I. (1987), *Semantic diversity in intonation*, in Proceedings of International congress of phonetic sciences 1987, Tallinn, pp. 468-471.
- Fónagy I. (1989), *La vive voix*, Payot, Parigi.
- Fónagy I. & Magdics K. (1963), *Emotional patterns in intonation and music*, in "Zeischrift für phonetische. Sprachwissenschaft und Kommunikationforschung" 16, pp. 293-326.

- Green R. S. & Cliff N. (1975), *Multidimensional comparison of structures of vocally and facially expressed emotion*, in "Journal of perceptive psychophysiology" 17, pp. 429-438.
- Hargreaves W. A., Starkweather J. A., & Blacker K. H. (1965), *Voice quality in depression*, in "Journal of abnormal psychology" 70, pp. 218-220.
- Havrdova Z. & Moravek M. (1979), *Changes of the voice expression during suggestively influenced states of experiencing*, in "Activitas nervosa superior" 21, pp. 33-35.
- Helfrich H., Standke R. & Scherer K. R. (1984), *Vocal indicators of psychoactive drug effects*, in "Speech communication" 3, pp. 245-252.
- Johnstone T. & Scherer K. R. (2000), *Vocal communication of emotion*, in Lewis M. – Haviland J. (a cura di), *Appraisal processes in emotion: theory, methods, research*, Oxford university press, New York, pp. 271-284.
- Kienast M. & Sendlmeier W. F. (2000), *Acoustical analysis of spectral and temporal changes in emotional speech*, in Sendlmeier W. (a cura di), *Speech and signals*, Frankfurt am M., pp. 157-168.
- Kienast M., Paeschke A. & Sendlmeier W. F. (1999), *Articulatory reduction in emotional speech*, in Eurospeech 1999, Budapest, pp. 117-120.
- Klasmeyer G. (1999), *Akustische Korrelate des stimmlich emotionalen Ausdrucks in der Lautsprache*, in Wodarz H. W., Heike G., Janota P. & Mangold M. (a cura di), "Forum phoneticum" 67, pp. 1-238.
- Klasmeyer G. & Meier T. (1999), *Rhythm in Emotional Speech*, in DAGA-Tag der ASA-Tagung 99 in Berlin, CD-Rom zur ASA-EAA-DEGA Gemeinschaftstagung Berlin 99.
- Klasmeyer G. & Sendlmeier W.F. (1997), *The classification of different phonation types in emotional and neutral speech*, in "Forensic Linguistics" 4, pp. 104-124.
- Klasmeyer G. & Sendlmeier W.F. (1999), *Voice and emotional states*, in Kent R. - Ball M. (a cura di), *Voice quality measurement*, Singular publishing group, San Diego, pp. 339-359.
- Kori S. & Magno Caldognetto E. (1986), *Caratterizzazioni elettroacustiche delle emozioni in italiano*, in "Rivista italiana di acustica" 10, pp.33-39.
- Kori S. & Magno Caldognetto E. (1990), *Indici acustici e percettivi delle emozioni*, in *Scritti in onore di Lucio Croatto*, Microprint, Padova.
- Kori S. & Magno Caldognetto E. (2003), *La caratterizzazione fonetica delle emozioni: primi dati da uno studio cross-linguistico italiano-giapponese*, in Così P., Magno Caldognetto E. & Zamboni A. (a cura di), *Voce-canto-parlato. Studi in onore di Franco Ferrero*, Unipress, Padova, pp. 187-200.
- Kotlyar G.M. & Morozov V.P. (1976), *Acoustical correlates of the emotional content of vocalized speech*, in "Journal of acoustics of the academy sciences of the USSR" 22, pp. 208-211.
- Kuroda I., Fujiwara O., Okamura N. & Utsuki N. (1976), *Method for determinino pilot stress through analysis of voice communication*, in "Aviation, space, and environmental medicine" 47, pp. 528-533.

- Ladd D. R., Scherer K. R. & Silverman K. E. A. (1986), *An integrated approach to studying intonation and attitude*, in Johns-Lewis C. (a cura di), *Intonation in discourse*, College-Hill Press, San Diego, pp. 125-138.
- Ladd D. R., Silverman K. E. A., Tolkmitt F., Bergman G. & Scherer K. R. (1985), *Evidence for the independent function of intonation contour type, voice quality and f0 range in signaling speaker affect*, in "Journal of the acoustical society of America" 78, pp. 435-444.
- Levin H. & Lord W. (1975), *Speech pitch frequency as an emotional state indicator*, in "Institute of electrical and electronics engineers transactions on system, man and cybernetics" 5, pp. 259-273.
- Magno Caldognetto E. & Ferrero F. E. (1996), *Macro e micro variazioni prosodiche dipendenti dalle scelte paralinguistiche del parlante*, in Fedi F. & Paoloni A. (a cura di), *Atti delle VI giornate di studio del GFS, "Caratterizzazione del parlatore"*, Roma 23-24 novembre 1995, pp. 95-107.
- Markel N. N., Bein M. F., & Phillis J. A. (1973), *The relationship between words and tone-of-voice*, in "Language and speech" 16, pp. 15-21.
- Moneglia M. (2008), *Le unità di riferimento per l'espressione delle emozioni: dati dal corpus C-ORAL-ROM*, in Magno Caldognetto E., Cavicchio F. & Cosi P. (a cura di), *Comunicazione parlata e manifestazione delle emozioni, atti del I convegno GSCP, Padova, 29 novembre-1 dicembre 2004*, Liguori, disponibile in e-book.
- Niwa S. (1971), *Changes of voice characteristics in urgent situations*, in "Reports of the aeromedical laboratory, Japan air self-defence force" 11, pp. 246-251.
- Paeschke A., Kienast M. & Sandlmeier W. F. (1999), *f0 contours in emotional speech*, in *Proceedings of International congress of phonetic sciences 1999*, San Francisco, pp. 929-933.
- Paeschke A. & Sendlmeier W. F. (2000), *Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements*, in *SpeechEmotion 2000*, pp. 75-80.
- Plaikner D. (1970), *Die Veränderungen der menschlichen Stimme unter dem Einfluss psychischer Belastung*, tesi di dottorato, Innsbruck Universität.
- Poggi I. & Magno Caldognetto E. (2003), *Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici*, in *Atti del Convegno Il parlato italiano*, D'Auria, Napoli, cd-rom.
- Polzin T. S. & Waibel A. (2000), *Emotion-sensitive human-computer interfaces*, in *ISCA workshop on speech and emotion*, Settembre 2000, Belfast, pp. 201-206.
- Roessler R. & Lester J. W. (1976), *Voice predicts affect during psychotherapy*, in "Journal of nervous and mental disease" 163, pp. 166-176.
- Scherer K. R. (1977), *The effect of stress on the fundamental frequency of the voice*, in "Journal of acoustical society of America" 62, pp. 25-26.
- Scherer K. R. (1979), *Nonlinguistic vocal indicators of emotion and psychopathology*, in Izard C. E. (a cura di), *Emotion and personality in psychopathology*, Plenum Press, New York.



- Scherer K. R. (1984), *On the nature and function of emotion: a component process approach*, in Scherer K.R. & Ekman P. (a cura di), *Approaches to emotion*, Erlbaum, Hillsdale.
- Scherer K. R., Koivumaki J. & Rosenthal R. (1972a), *Minimal cues in the vocal communication of affect: judging emotions from content-masked speech*, in "Journal of psycholinguistic research" 1, pp. 269-285.
- Scherer K. R., Koivumaki J. & Rosenthal R. (1972b), *Mediating interpersonal expectancies via vocal cues: different speech intensity as a means of social influence*, in "European journal of social psychology" 2, pp. 163-176.
- Scherer K. R., London H. & Wolf J. (1973), *The voice of confidence: paralinguistic cues and audience evaluation*, in "Journal of research in personality" 7, pp. 31-44.
- Scherer K. R., Johnstone T. & Klasmeyer G. (2003), *Vocal expression of emotion*, in Davidson D. J., Goldsmith H. & Scherer K. R. (a cura di), *Handbook of the Affective Sciences*, Oxford university press, New York and Oxford, pp. 433-456.
- Scherer K. R., Banse R., Wallbott H. G. & Goldbeck T. (1991), *Vocal cues in emotion encoding and decoding*, in "Motivation and emotion" 15, pp. 123-148.
- Sedlacek K. & Sychra A. (1963), *Die Melodie als Faktor des emotionellen Ausdrucks*, in "Folia phoniatica" 15, pp. 89-98.
- Skinner E. R. (1935), *A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness*, in "Speech monographs" 2, pp. 81-137.
- Sobin C. & Alpert M. (1999), *Emotion in speech: the acoustic attributes of fear, anger, sadness, and joy*, in "Journal of psycholinguistic research" 28, pp. 347-365.
- Sokolov J. N. (1963), *Perception and the conditioned reflex*, Pergamon press, Oxford.
- Sulc J. (1977), *To the problem of emotional changes in the human voice*, in "Activitas nervosa superior" 19, pp. 215-216.
- Van Bezooijen R. (1984), *The characteristics and recognizability of vocal expression of emotion*, Foris, Dordrecht.
- Wallbott H. G. & Scherer K. R. (1986), *Cues and channels in emotion recognition*, in "Journal of personality and social psychology" 51, pp. 690-699.
- Williams C. E. & Stevens K. N. (1972), *Emotion and speech: some acoustical correlates*, in "Journal of the acoustical society of America" 52, pp. 1238-1250.
- Zuberbier E. (1957), *Zur Schreib- und Sprechmotorik der Depressiven*, in "Zeitschrift für Psychotherapie und medizinische Psychologie" 7, pp. 239-249.
- Zwirner E. (1930), *Beitrag zur Sprache des Depressiven*, in "Phonometrie" 3, pp. 171-187.

# LA COMPOSIZIONE FONETICA DEL PRIMO VOCABOLARIO DEL BAMBINO

°Claudio Zmarich, \*Marco Dispaldro, °Pasquale Rinaldi, °Maria Cristina Caselli

°Istituto di Scienze e Tecnologie della Cognizione, CNR, Padova-Roma

\*Università degli Studi di Padova

°[zmarich@pd.istc.cnr.it](mailto:zmarich@pd.istc.cnr.it), [marco.dispaldro@unipd.it](mailto:marco.dispaldro@unipd.it), °[cristina.caselli@istc.cnr.it](mailto:cristina.caselli@istc.cnr.it)

## 1. SOMMARIO

Chiunque si occupi del primo sviluppo linguistico sa che l'aspetto fonetico delle parole da produrre rappresenta per il bambino una sfida non meno importante di quella rappresentata dagli aspetti cognitivi. È lecito quindi supporre che la complessità nell'organizzazione del controllo motorio che è implicata nella produzione delle prime parole possa contribuire in modo decisivo sia alla selezione da parte del bambino di una certa forma lessicale piuttosto che un'altra (Schwartz & Leonard, 1982), e sia alla sua particolare realizzazione (MacNeilage *et alii*, 2000) a partire da quelle abilità già acquisite durante il babbling (Locke, 1983; Vihman, 1996). D'altra parte, ci sono prove che le capacità fonetico/fonologiche non influenzano in modo unidirezionale le conoscenze lessicali, poichè è vero anche il contrario: i bambini che possiedono un vocabolario più grande hanno un inventario fonetico più ricco e capacità fonotattiche più sviluppate.

Per approfondire le relazioni tra la frequenza dell'input lessicale e la facilità di articolazione nel primo sviluppo fonologico, in questo studio ci proponiamo di confrontare la frequenza di occorrenza dei foni consonantici iniziali delle sillabe CV e CVC nelle parole che compongono il Primo Vocabolario del Bambino (o PVB, cfr. Caselli & Casadio, 2002; Caselli *et alii*, 2007) e che sono tentate dai bambini di 18-23 mesi e dai bambini di 30-36 mesi.

Abbiamo analizzato le caratteristiche fonetiche delle parole del PVB facendo riferimento alle percentuali relative alle parole prodotte dai bambini suddivisi per fasce d'età (18-23 mesi, 30-36 mesi), e all'interno di ciascuna fascia di età abbiamo suddiviso le parole del vocabolario in quartili (1-25%, 26-50%, 51-75%, 76-100%), ed espresso il calcolo delle frequenze in funzione delle parole del terzo quartile. I risultati preliminari ci portano effettivamente a dire che le parole prodotte più tardi sono anche più complesse dal punto di vista fonetico di quelle che sono prodotte ad un'età più precoce, e che la complessità articolatoria può essere un fattore determinante per queste differenze.

## 2. INTRODUZIONE

Per acquisire la lingua nativa, il bambino deve fare essenzialmente due cose: apprendere le parole della lingua ed estrarre le caratteristiche fonologiche rilevanti di quelle parole.

A fronte della relativa abbondanza di studi che trattano gli aspetti categoriali e semantici dell'acquisizione lessicale nei bambini italiani, che ci dicono come durante la prima fase di sviluppo lessicale i bambini possiedono, attorno ai 18 mesi, un vocabolario di circa 50 parole riferite perlopiù a nomi (Caselli *et alii*, 1995; Caselli & Casadio, 2002; D'Odorico, *et alii*, 2001; Rinaldi *et alii*, 2004), mentre nella seconda fase durante la così detta "esplosione del vocabolario" avviene un incremento sia dei predicati che delle parole

funzioni (Camaioni & Longobardi, 2001), sono ancora pochissimi quelli che riguardano l'acquisizione degli aspetti fonologici.

Per i bambini di madrelingua inglese, lo sviluppo lessicale e quello fonetico sono stati studiati perlopiù indipendentemente l'uno dall'altro. Recentemente però l'interfaccia tra le abilità fonetiche e quelle lessicali nello sviluppo linguistico è stata oggetto di una serie di studi osservativi e sperimentali, specialmente nei bambini che producono meno di 50 parole (per una rassegna, cfr. Storkel & Morrisette, 2002). Questi studi hanno fornito delle evidenze preliminari a favore di una stretta interazione tra sviluppo lessicale e fonologico.

Il bambino che produce meno di 50 parole fa infatti buon uso del suo limitato inventario fonetico, e cerca di sfruttare quelle abilità già acquisite durante il *babbling*: i bersagli lessicali con fonî e strutture già presenti nel babbling del bambino saranno tentati prima ed usati di più dei bersagli lessicali lontani dal babbling (Oller *et alii*, 1976; Locke, 1983), e questi ultimi saranno invece attivamente evitati (Ferguson & Farwell, 1975; Schwartz & Leonard, 1982).

L'associazione tra lo sviluppo lessicale e fonologico è stata osservata anche in bambini con sviluppo linguistico precoce (Stoel-Gammon & Dale, 1988) e in quelli con sviluppo linguistico ritardato (Rescorla & Ratner, 1996; Paul & Jennings, 1992; Thal *et alii*, 1995). I primi conoscevano molte parole e tendevano ad usare una maggior varietà e combinazione di fonî, mentre i secondi ne conoscevano poche con una limitata varietà e combinazione di fonî.

Anche gli studi clinici permettono di verificare questa forte interazione. Per es., uno studio su un caso clinico di un bambino con un ritardo di linguaggio espressivo dimostrò che un tipo di trattamento finalizzato ad aumentare il vocabolario espressivo conduceva ad un aumento della varietà fonologica (Girolametto *et alii*, 1997), e viceversa, un altro studio evidenziò come i bambini di un gruppo sperimentale apprendevano più facilmente parole nuove che erano però costituite da fonî già nel loro repertorio (Schwartz & Leonard, 1982).

La maggior parte delle evidenze per una forte interazione nell'acquisizione lessicale e fonologica sono state limitate alla produzione di meno di 50 parole, ma un numero crescente di studi ha recentemente prolungato il periodo di interazione anche allo stadio di sviluppo successivo. Per es. Stoel-Gammon (1991) riportò che in un gruppo di 34 soggetti normali di 2 anni c'era una forte relazione tra la dimensione del vocabolario e la dimensione degli inventari fonetici. Ancora Stoel-Gammon (1998), in uno studio condotto sul MacArthur Communicative Development Inventories (CDI) (Fenson *et alii*, 1993), che è un questionario dove i genitori devono contrassegnare le parole della lista che sono presenti nel vocabolario dei loro figli, concludeva che in genere le caratteristiche fonologiche più frequenti nei target lessicali infantili (calcolate sul CDI) sono anche quelle acquisite prima (come dimostrano gli inventari fonetici sulle produzioni effettive). Inoltre i bambini con un vocabolario eccezionalmente ampio o eccezionalmente ristretto sono anche, rispettivamente, molto e poco dotati dal punto di vista fonetico-fonologico. Ci sono poi delle indicazioni che il dato sulla stretta interazione tra sviluppo lessicale e fonologico anche oltre lo stadio delle 50 parole, valga anche in prospettiva interlinguistica. Uno studio condotto sull'adattamento alla lingua mandarina del CDI (Fletcher *et alii*, 2004) indica infatti che, nei bambini cinesi di Canton, l'interazione tra abilità fonetiche e lessicali continua anche oltre lo stadio delle 50 parole.

Per quanto riguarda il panorama italiano, le ricerche sull'interazione tra sviluppo lessicale e fonologico sono solo agli inizi.

Bortolini (1993) analizzò 4 soggetti audio- e video-registrati ogni due settimane dai 10 ai 18 mesi di vita. In generale, lo sviluppo fonetico riferito agli stadi di parola (0, 4, 15, e 25

parole) mostrava una forte correlazione tra il repertorio dei suoni presenti nel *babbling* e quello dei suoni presenti nelle prime parole, e un forte progresso in termini di riduzione della variabilità intra- e inter-individuale e di aumento del numero dei foni nativi prodotti con l'aumento del vocabolario.

Lo studio di Majorano (2006) riporta le differenze nello sviluppo fonologico di quattro coppie di bambini seguiti dai 10 ai 24 mesi che alla tappa finale presentavano uno sviluppo del vocabolario disomogeneo. I due bambini che erano più avanzati dal punto di vista lessicale avevano anche dei repertori fonetici più ricchi e variati, mentre i bambini meno avanzati dal punto di vista lessicale lo erano anche nello sviluppo fonologico.

Zmarich, Stocco *et alii*, (in questo volume) riportano le frequenze di occorrenza dei segmenti fonologici, ordinati in riferimento agli stadi di parola (0, 4, 15 e 25 parole), prodotti da 5 bambini italiani, e li confrontano con i dati di quattro gruppi nazionali (francese, giapponese, inglese, svedese) composti ciascuno da 5 bambini (cfr. Vihman, 1993). I risultati consentono di affermare che nello stadio finale delle 25 parole, le differenze fonetiche fra i gruppi linguistici nazionali sono maggiori delle differenze all'interno di ogni singolo gruppo, e nella maggior parte dei casi queste diversità riflettono i *patterns* fonetico-fonologici caratteristici di ciascuna lingua.

Per finire, in Zmarich *et alii*, (in corso di pubblicazione) viene indagata la composizione fonetica delle consonanti scempie iniziali nelle parole del Primo Vocabolario del Bambino (PVB) (Caselli & Casadio, 2002; Caselli *et alii*, 2007) allo scopo di osservarne le relazioni con l'età di acquisizione lessicale e sulla frequenza con cui la parola è utilizzata nei bambini tra 18 e 36 mesi. La maggioranza delle parole è costituita da bisillabi e trisillabi, e le sillabe sono del tipo CV. All'aumentare dell'età e dell'ampiezza di vocabolario si assiste ad un aumento dei tipi consonantici più complessi, per modo e luogo di produzione, rispettivamente fricative e liquide, e labio-dentali, alveolari e palato-alveolari; inoltre si assiste ad una riduzione di occlusive e bilabiali.

Nel presente studio ci siamo proposti di approfondire lo studio precedente indagando la relazione tra il numero di parole attestate dal PVB e la loro composizione fonetica nelle fasce d'età di 18-23 mesi e 30-36 mesi, non più limitata come nello studio succitato alle consonanti iniziali di parola, ma allargata a tutte le consonanti iniziali dei tipi sillabici CV e CVC (e dunque anche interne alla parola), che da soli costituiscono oltre l'80% delle occorrenze di tutte le sillabe delle parole del PVB.

### 3. METODOLOGIA

#### 3.1 Materiali e partecipanti

Per conoscere la composizione fonetica delle parole “prodotte” dai bambini, abbiamo utilizzato il “Primo Vocabolario del Bambino” -PVB- (Caselli & Casadio, 2002; Caselli *et alii*, 2007).

Il PVB è un questionario rivolto ad utenze molteplici (tipicamente genitori, ma anche operatori sanitari di vario tipo) che permette di raccogliere dati sistematici sulla comunicazione e sul linguaggio nei primi anni di vita, e di valutarli e confrontarli con quelli di altri bambini. I dati relativi al linguaggio di bambini di età compresa tra gli 8 ed i 36 mesi di vita vengono raccolti attraverso l'uso di schede di rilevazione. Il questionario è costituito da due schede distinte: “Gesti e Parole”, utilizzata per bambini di età compresa tra gli 8 ed i 17 mesi, e “Parole e Frasi” per bambini di età compresa tra i 18 ed i 36 mesi. Precisiamo subito che le parole elencate nel PVB non rispecchiano in generale la pronuncia

che ne fa il bambino. Infatti, nella sua forma attuale, il questionario, che viene utilizzato come strumento clinico e di ricerca, viene compilato autonomamente dalle famiglie cui viene proposto: ai genitori viene richiesto di barrare la voce lessicale scritta in grafia alfabetica, come un'entrata del vocabolario. Bisogna quindi considerare, che la maggior parte delle volte i bambini alterano la pronuncia adulta della parola target; quest'ultima viene generalmente semplificata attraverso l'uso dei cosiddetti processi fonologici (Vihman, 1996). Per questo motivo è più corretto riferirsi a queste forme lessicali come le parole tentate dal bambino, o target lessicali.

Per l'obiettivo del presente lavoro facciamo riferimento ai dati normativi della sezione di Vocabolario della scheda "Parole e Frasi" raccolti su 752 bambini (Caselli *et alii*, 2007). Tale sezione comprende 670 parole distribuite all'interno di 23 categorie semantiche. Il database usato costituisce un aggiornamento di quello precedentemente costruito ed utilizzato da Zmarich & Miotti (2003) su versioni precedenti del PVB.

Il database completo delle 670 parole è stato ordinato rispetto alla percentuale di bambini che produce il target, all'interno di 3 sottogruppi di età (18-23 mesi, 24-29 mesi, 30-36 mesi). In questo modo si mette in evidenza quali parole sono tentate da quale percentuale di bambini all'interno di ciascuno dei 3 sottogruppi; questi dati saranno pertanto messi in relazione con la composizione fonetica delle parole.

### 3.2 Variabili considerate:

Nel presente studio abbiamo preso in considerazione le seguenti variabili:

- il tipo sillabico;
- i foni consonantici della tabella IPA (AAVV, 1999);
- il modo di articolazione delle consonanti (AAVV, 1999);
- il luogo di articolazione delle consonanti (AAVV, 1999).

### 3.3 Criteri di codifica e analisi dei dati

Per poter condurre i calcoli di frequenza in modo automatico abbiamo convertito sul PC i caratteri speciali dell'*International Phonetic Alphabet* (IPA) (AAVV 1989) con quelli dell'alfabeto fonetico SAMPA ([www.phon.ucl.ac.uk/home/sampa/home.htm](http://www.phon.ucl.ac.uk/home/sampa/home.htm)), basato sui caratteri ASCII.

Il database completo delle 670 parole è stato codificato, secondo i simboli fonetici, all'interno di una matrice del programma statistico Systat 8.0. Sono state eseguite le codifiche che vengono spiegate di seguito.

Abbiamo pertanto suddiviso in sillabe ciascuna delle 670 parole del PVB, sulla base della gerarchia di sonorità e dei principi di sillabificazione (Blevins, 1995; Nespor, 1993). Sono risultate, da tale divisione, un totale di 1795 sillabe, che sono state codificate secondo i seguenti criteri:

- numero progressivo delle sillabe in una data parola;
- codice di appartenenza di una data sillaba a una data parola;
- posizione della sillaba nella parola (ad es: [biscotti] = [bi] iniziale, [skot] mediana, [ti] finale; [no] = monosillabo);
- tipo sillabico, (ad esempio [bi]= CV (consonante, vocale); [skot]=CCVC).

È stato analizzato ciascun fono per modo e luogo di articolazione. I foni [j] e [w] sono stati codificati come consonanti in base alla considerazione che non costituiscono apice di sillaba. Il fono [s] in posizione iniziale di nesso consonantico intervocalico è stato attribuito alla sillaba successiva (per es.: 'bi.skot.ti' e non 'bis.kot.ti').

Per gli obiettivi del presente lavoro, le analisi sulla modalità e sul luogo di articolazione sono state effettuate solo per fonemi consonantici, in quanto il sistema vocalico italiano normalmente è già acquisito all'età di 18 mesi, età da cui parte il presente studio (Zmarich *et alii*, 2005; Zmarich & Luppari, 2006). Inoltre i fonemi analizzati sono le consonanti iniziali delle sillabe CV e CVC, che risultano essere i tipi sillabici più frequenti dell'italiano (cfr. Mancini & Voghera, 1994), questo per non complicare eccessivamente l'analisi, mantenendone nel contempo una validità di fondo. Infatti, tutti i fonemi che concorrono a formare i gruppi consonantici dell'italiano vengono prodotti anche singolarmente; inoltre risulta da analisi precedenti (Zmarich & Bonifacio, 2005) che i gruppi consonantici vengono prodotti solo a partire dai 27 mesi d'età.

Le parole sono state poi ordinate in modo decrescente in base alla percentuale di bambini che le producono all'interno di ciascuno dei due sottogruppi di età (18-23 mesi; 30-36 mesi). Utilizzando la percentuale media di bambini che producono una data parola, il database è stato poi suddiviso in quartili all'interno di ciascuna fascia di età, come di seguito:

- 1° quartile (parole prodotte da una percentuale di bambini compresa tra l'1% e il 25%);
- 2° quartile (parole prodotte da una percentuale di bambini compresa tra il 26% e il 50%);
- 3° quartile (parole prodotte da una percentuale di bambini compresa tra il 51% e il 75%);
- 4° quartile (parole prodotte da una percentuale di bambini compresa tra il 76% e il 100%).

Poiché molte delle parole comprese all'interno del 4° quartile si sovrappongono all'interno dei tre sottogruppi, le parole comprese nei 4<sup>i</sup> quartili di ciascuna delle tre fasce d'età sono state escluse dalle analisi delle parole date per acquisite (cioè attestate in oltre il 50% dei bambini). La sovrapposizione di tali dati può essere spiegata assumendo che le parole all'interno del suddetto quartile sono quelle acquisite molto precocemente e che fanno parte del cosiddetto lessico di base. Infatti la precocità di acquisizione avrebbe mascherato l'attesa distinzione sulla base della difficoltà del target.

Di conseguenza, nel condurre l'analisi evolutiva sul cambiamento delle caratteristiche fonetiche delle parole prodotte da una percentuale elevata di bambini, abbiamo considerato solo le parole prodotte da una percentuale di bambini compresa tra il 51% e il 75% (3<sup>i</sup> quartili).

#### 4. RISULTATI

Ricordiamo che in questa sede, per semplificare l'analisi e per evidenziare maggiormente le eventuali differenze, abbiamo voluto confrontare solo i dati relativi alle fasce estreme d'età, la prima (18-23 mesi) e la terza (30-36 mesi).

Le parole che costituiscono il terzo percentile della prima fascia d'età sono 45, mentre quelle che costituiscono il terzo percentile della terza fascia sono 197. La Fig. 1 presenta la frequenza dei tipi sillabici che costituiscono le parole tentate dai bambini nella prima e nella terza fascia di età, insieme con le percentuali calcolate su tutto il PVB (terza colonna di ogni sottogruppo).

Come si può vedere dalla figura 1, i due tipi sillabici CV e CVC sono rispettivamente il primo e il secondo tipo sillabico per frequenza e costituiscono da soli oltre l'80% delle sillabe che compongono le parole tentate dai bambini, e anche l'80% delle sillabe che compongono le parole dell'intero PVB (Zmarich *et alii*, in corso di pubblicazione).

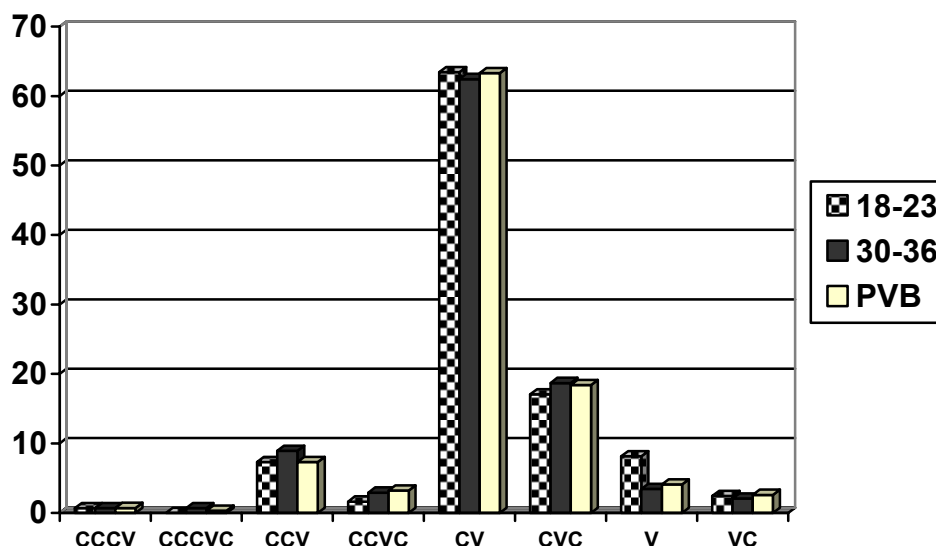


Fig.1. Frequenza % dei tipi sillabici nelle parole tentate dai bambini nella prima (18-23 mesi) e nella terza fascia di età (30-36 mesi), insieme con le percentuali calcolate su tutto il PVB. Il totale delle sillabe analizzate nella prima fascia è 123, nella terza fascia 583, mentre quelle del PVB completo assommano a 1795.

Allo scopo di esaminare come la frequenza di occorrenza delle categorie fonetiche delle parole tentate dai bambini secondo il PVB cambi con l'età, sono stati confrontati i foni consonantici (fig. 2) e le classi fonologiche naturali di modo (fig. 3) e di luogo (fig. 4) iniziali dei tipi sillabici CV e CVC, nelle parole date per acquisite nel 3° quartile dai bambini che costituivano il campione del PVB nella prima (18-23 mesi) e nella terza fascia di età (30-36 mesi). Come riferimento sono espresse anche le percentuali di occorrenza dei foni consonantici iniziali dei tipi sillabici CV e CVC per il PVB in generale (1471 occorrenze). Non sono stati riportati i foni inferiori all'1% in entrambe le fasce di età.

Come si può osservare, le percentuali del PVB generale, che meglio rispecchiano nella loro quantità le caratteristiche della lingua target, sono maggiormente approssimate dalle percentuali esibite dai target tentati dal terzo percentile dei bambini di 30-36 mesi. E' da notare soprattutto il forte aumento della vibrante /r/ passando dalla prima alla terza fascia.

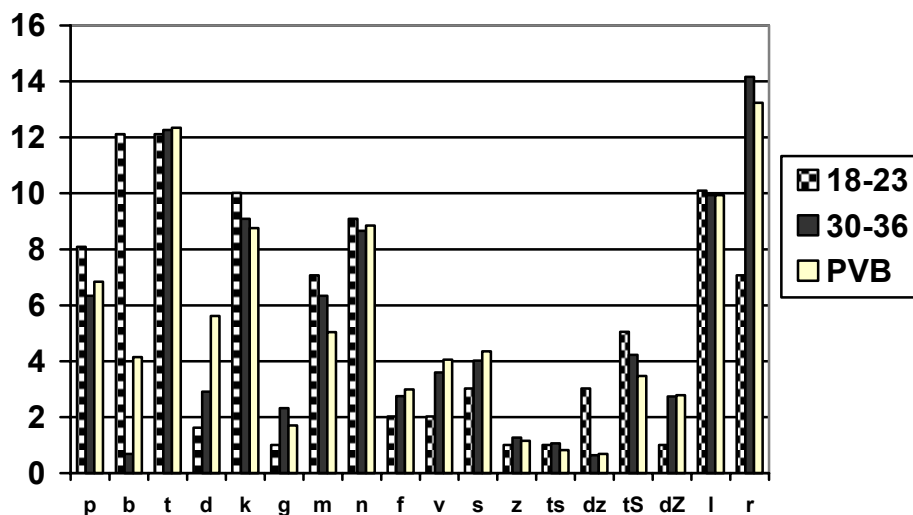


Fig. 2. Frequenza % dei foni consonantici iniziali dei tipi sillabici CV e CVC nelle parole tentate dai bambini della prima (18-23 mesi) e della terza fascia di età (30-36 mesi), insieme con le percentuali calcolate su tutto il PVB (tʃ, dʒ = /tʃ dʒ/). Il totale dei foni analizzati è di 99 per la prima fascia e di 473 per la seconda, mentre il totale per il PVB in generale è 1471. Non sono stati riportati i foni inferiori all'1% in entrambe le fasce di età.

La fig. 3 illustra il confronto tra le classi fonologiche naturali di modo iniziali dei tipi sillabici CV e CVC, nelle parole date per acquisite dal 50% al 75% dei bambini che costituivano il campione del PVB nella prima (18-23 mesi) e nella terza fascia di età (30-36). Come riferimento sono espresse anche le percentuali di occorrenza dei foni consonantici iniziali dei tipi sillabici CV e CVC per il PVB in generale (1471 occorrenze).



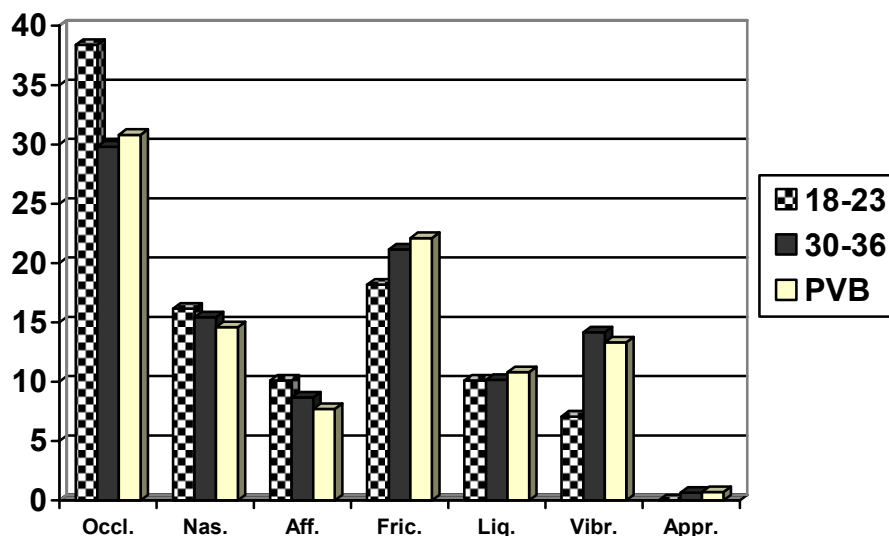


Fig. 3. Frequenza % delle classi fonologiche naturali di Modo per le consonanti iniziali dei tipi sillabici CV e CVC nelle parole tentate dai bambini della prima (18-23 mesi) e della terza fascia di età (30-36 mesi). Il totale dei foni analizzati è di 99 per la prima fascia e di 473 per la seconda, mentre il totale per il PVB in generale è 1471.

È possibile notare come anche per le categorie di modo, le percentuali relative ai bambini da 30 a 36 mesi sono più simili alle percentuali dei target adulti, calcolate su tutte le parole che compongono il PVB. Per es., dopo aver detto già della vibrante, nella fascia di età più tarda si assiste a un forte ridimensionamento delle occlusive orali, che si allineano alle percentuali valide per il PVB nel suo complesso.

La fig. 4 illustra il confronto tra le classi fonologiche naturali di luogo iniziali dei tipi sillabici CV e CVC, nelle parole date per acquisite dal 50% al 75% dei bambini che costituivano il campione del PVB nella prima (18-23 mesi) e nella terza fascia di età (30-36). Come riferimento sono espresse anche le percentuali di occorrenza dei foni consonantici iniziali dei tipi sillabici CV e CVC per il PVB in generale (1471 occorrenze).

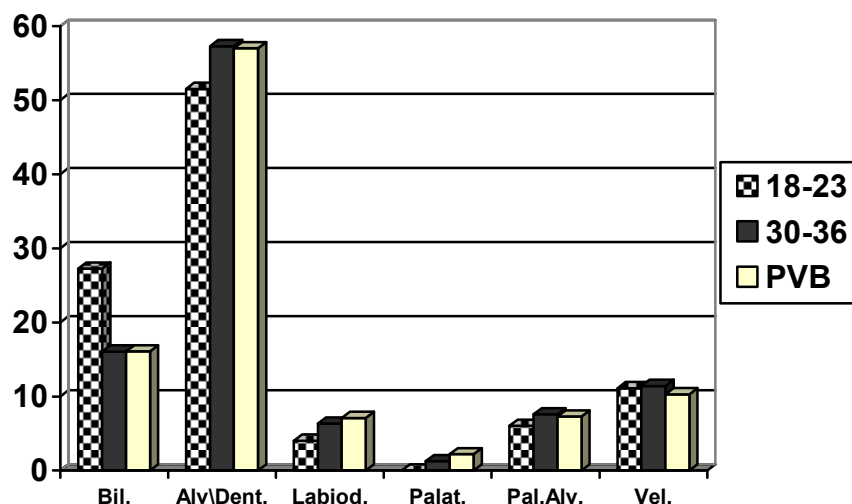


Fig. 4. Frequenza % delle classi fonologiche naturali di Luogo per le consonanti iniziali dei tipi sillabici CV e CVC nelle parole tentate dai bambini della prima (18-23 mesi) e della terza fascia di età (30-36 mesi). Il totale dei foni analizzati è di 99 per la prima fascia e di 473 per la seconda, mentre il totale per il PVB in generale è 1471.

Per quanto riguarda l'analisi relativa al luogo di articolazione, i dati più salienti riguardano la brusca riduzione delle bilabiali passando dalla prima alla terza fascia di età, e l'aumento dalla prima alla terza fascia delle alveolari/dentali, con percentuali simili a quelle che si trovano in generale nell'intero PVB.

Per finire confronteremo i dati provenienti da questo studio con quelli relativi a un'analisi osservativa dello sviluppo fonetico. In tab. 1 vengono elencate le percentuali di occorrenza nel terzo percentile del PVB per le due fasce di età qui considerate delle consonanti da considerare "precoci", "intermedie" e "tarde" nell'acquisizione in base all'analisi della produzione spontanea di un gruppo di bambini normali, valutata attraverso la compilazione degli inventari fonetici (Zmarich & Bonifacio, 2005). Le consonanti precoci sono definite come le consonanti che risultano acquisite da oltre il 50% dei bambini di Zmarich & Bonifacio (2005) a 18 mesi, le consonanti intermedie sono quelle aggiuntive che risultano acquisite da oltre il 50% dei bambini di Zmarich & Bonifacio (2005) a 27 mesi, e le consonanti tarde sono quelle che non ancora sono state acquisite da almeno il 50% dei bambini di Zmarich & Bonifacio (2005) a 27 mesi e che sarebbero state acquisite solo nei mesi successivi.

I dati esposti in tab. 1 indicano che il gruppo più giovane del presente studio ha una preferenza per i target lessicali che sono costituiti da quelle consonanti che erano state effettivamente acquisite prima dal gruppo di bambini normali dello studio osservativo (Zmarich & Bonifacio, 2005). Questa preferenza si riduce nel gruppo di bambini più grandi.

| Consonanti                         | Prima fascia (18-23) | Terza fascia (30-36) |
|------------------------------------|----------------------|----------------------|
| “Precoci” /p b t k m/              | 49,1%                | 37,2%                |
| “Intermedie” /d g n f v s tʃ l/    | 37,4 %               | 40,8%                |
| “Tarde” /ɲ, ʃ, z ts dz dʒ ʎ r j w/ | 13,5%                | 22,0%                |

Tab. 1. Proporzione delle frequenze % di occorrenza delle consonanti “precoci”, “intermedie” e “tarde”(determinate in base a studi osservativi) nei target del PVB tentati dal 3° percentile dei bambini della prima fascia (18-23 mesi) e della terza fascia (30-36 mesi)

## 5. DISCUSSIONE E CONCLUSIONE

I presenti risultati consentono di estendere anche alle consonanti iniziali dei tipi sillabici CV e CVC, indipendentemente dalla posizione iniziale o non iniziale nella parola, una delle conclusioni già riservate ai risultati dell’analisi delle consonanti iniziali di parola (Zmarich *et alii*, in corso di pubblicazione): le parole tentate ad età più precoci sono meno complesse dal punto di vista fonetico di quelle che sono prodotte ad età più tarde.

Questa conclusione è suffragata dai seguenti fatti:

- le percentuali di occorrenza delle consonanti della lingua target (quelle dell’intero PVB), sono maggiormente approssimate dalle percentuali esibite nei target tentati dal terzo percentile dei bambini di 30-36 mesi rispetto ai target tentati dai bambini di 18-23 mesi. Infatti il gruppo più giovane del presente studio ha una preferenza per i target lessicali che sono costituiti da quelle consonanti che in uno studio di tipo osservativo (Zmarich & Bonifacio, 2005) erano state acquisite prima da un gruppo di bambini normali. Questa preferenza si riduce nel gruppo di bambini più grandi;
- per le categorie relative al modo di articolazione, le percentuali relative ai bambini da 30 a 36 mesi sono più simili alle percentuali dei target adulti (per l’intero PVB). Inoltre la /r/, che è il fono più frequente della lingua italiana (cfr. Bortolini *et alii*, 1977; Batinti, 1993), ma anche uno dei più complessi da realizzare dal punto di vista articolatorio, è relativamente molto frequente solo nei target tentati dai bambini della fascia di età più tarda;
- per le categorie relative al luogo di articolazione, passando dalla prima alla terza fascia di età si ha una netta riduzione delle bilabiali. Per spiegare questo dato forse è utile richiamare la strategia “*Labial-Coronal*”, usata dai bambini anglofoni studiati da MacNeilage *et alii* (2000), e da questi autori reputata universale e non linguospecifica. Secondo gli autori, quando i bambini pronunciano le loro prime parole, che nel nostro studio sarebbero proprio quelle tentate dai bambini dalla prima fascia (18-23 mesi), usano cambiare il fonema iniziale della parola target con una consonante bilabiale in modo da facilitare l’articolazione della sillaba iniziale. Alla prevalenza della bilabiale nella fascia iniziale potrebbe però contribuire anche la salienza percettiva di tipo visivo che contraddistingue queste consonanti.

Il fatto che i nostri risultati echeggino da vicino i risultati trovati per i bambini di comunità linguistiche diverse, come l’inglese (Stoel-Gammon, 1998) e il cinese mandarino parlato a Canton (Fletcher *et alii*, 2004), potrebbe suggerire che ci sono fattori universali di tipo fonetico che sono importanti per l’acquisizione fonologica in tutte le lingue del mondo, e questi fattori potrebbero avere a che fare con la complessità articolatoria (Stokes & Surendram, 2005).

## BIBLIOGRAFIA

- AAVV (1999), *Handbook of the International Phonetic Association*, Cambridge (UK): Cambridge University Press.
- Batinti, A. (1993), *Il sistema fonologico dell'italiano*. Perugia: Guerra Edizioni.
- Blevins, J. (1995). The syllable in phonological theory. In *The Handbook of Phonological Theory* (J.A. Goldsmith, Ed.), Cambridge (MA): Basil Blackwell.
- Bortolini U. (1993), Continuità fonetica tra “babbling” e prime parole, in *Ricerche sull'acquisizione dell'italiano* (E. Cresti & M. Moneglia, editors), Bulzoni, Roma.
- Bortolini, U., Degan, C., Minnaja, C., Paccagnella & L., Zilli, G. (1977). Statistics for a stochastic model of spoken Italian. In *Proceedings of the Twelfth International Congress of Linguists*, Wien, 580-586.
- Camaioni, L. & Longobardi, E. (2001). Nature and stability of individual differences in early lexical development of Italian-speaking children. *First Language*, 15, 203-218.
- Caselli, M.C., Bates, E., Casadio, P., Fenson, J., Sanderl, L. & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10, 159-199.
- Caselli, M.C. & Casadio, P. (2002). *Il primo vocabolario del bambino* (4<sup>a</sup> ed.). Milano: Franco Angeli.
- Caselli M. C., Pasqualetti P. & Stefanini S. (2007), Parole e frasi nel “Primo vocabolario del bambino”, Milano, Franco Angeli.
- D'Odorico, L., Carubbi, S., Salerni, N. & Calvo, V. (2001). Vocabulary development in Italian children: A longitudinal evaluation of quantitative and qualitative aspects. *Journal of Child Language*, 28, 351-372.
- Fenson, L., Dale, P., Reznick, J.S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *MacArthur Communicative Development Inventory (CDI)*, San Diego (CA): Singular Publishing Group.
- Ferguson C.A. & Farwell C. (1975), Words and sounds in early language acquisition: English initial consonants in the first 50 words, *Language*, 51, 419-439.
- Fletcher, P., Chan, C. W.-Y., Wong, P. P.-T., Stokes, S., Tardif, T. & Leung, S. C.-S. (2004), The interface between phonetic and lexical abilities in early Cantonese language development. *Clinical Linguistics & Phonetics*, 18, 535-545.
- Girolametto L., Pearce P.S. & Weitzman (1997), Effects of lexical intervention on the phonology of late talkers, *Journal of Speech, Language and Hearing Research*, 40, 338-348.
- Locke, J.L. (1983). *Phonological Acquisition and Change*. New York: Academic Press.
- MacNeilage, P.F., Davis, B.L., Kinney A. & Matyear C.L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, 71, 153-163.

- Majorano M. (2006), Differenze individuali nella lallazione e nelle prime parole: studio longitudinale su quattro bambini con diversi profili di sviluppo linguistico, *Psichiatria dell'infanzia e dell'adolescenza*, 73, 265-278.
- Mancini, F., Voghera, M. (1994). Lunghezza, tipi di sillabe e accento in italiano. In *Come parlano gli italiani* (T. De Mauro, editor), Firenze: La Nuova Italia, 217-245.
- Nespor, M. (1993). *Fonologia*, Bologna: Il Mulino.
- Oller, D.K., Wieman, L.A., Doyle, W.J. & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, 2, 1-11.
- Paul R. & Jennings P. (1992), Phonological behavior in toddlers with slow expressive language development, *Journal of Speech, Language and Hearing Research*, 35, 99-107.
- Rescorla L. & Ratner N.B. (1996), Phonetic profiles of toddlers with specific expressive language development (SLI-E), *Journal of Speech, Language and Hearing Research*, 39, 153-165.
- Rinaldi, P., Barca, L. & Burani, C. (2004). Caratteristiche semantiche, grammaticali e di frequenza delle parole del <<primo vocabolario del bambino>>. *Psicologia Clinica dello Sviluppo*, 8, 119-143.
- Schwartz, R. (1988). Phonological factors in early language acquisition. In M.D. Smith, J. L. Locke (eds.), *The Emergent Lexicon: the child's development of a linguistic vocabulary*, New York: Academic press, 185-222.
- Schwartz, R. & Leonard, L. (1982), Do children pick and choose: An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, 9, 319-336.
- Stoel-Gammon, C. (1991). Normal and disordered phonology in two-years-olds, *Topics in Language Disorders*, 21-32.
- Stoel-Gammon, C. (1998). Sounds and words in early language acquisition. In P. Rhea (Ed.), *Exploring speech-language connection*. Baltimore: Paul H. Brookes Publ.
- Stoel-Gammon, C. & Dale P. (1988), *Aspects of phonological development of linguistically precocious talkers*. Paper presented at the Child Phonology Conference, University of Illinois, Champaign-Urbana.
- Stokes S. F. & Surendran D. (2005), Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children, *Journal of Speech, Language and Hearing Research*, 48, 577-591.
- Storkel H. L. & Morrisette M. L. (2002), The Lexicon and Phonology: interactions in language acquisition, *Language, Speech and Hearing Services in Schools*, 33, 24-37.
- Thal, D. J., Oroz M. & McCaw, V. (1995). Phonological and lexical development in normal and late-talking toddlers. *Applied Psycholinguistics*, 16, 407-424.
- Vihman M.M. (1996), *Phonological Development*, Blackwell Publ., Oxford.
- Zmarich C. & Bonifacio S. (2005). Phonetic inventories in Italian children aged 18-27 months: a longitudinal study. *Proceedings of INTERSPEECH'2005-EUROSPEECH, Lisboa*, 757-760.

Zmarich, C. & Miotti, R. (2003). Frequenza di occorrenza di consonanti e vocali e delle loro combinazioni nel babbling e nelle prime parole di quattro bambini italiani. In *Atti delle XIII Giornate di Studio del Gruppo di Fonetica Sperimentale* (G. Marotta, N. Nocchi, Eds.), Pisa: Edizioni ETS.

Zmarich C., Stocco D., Minozzi M. & Bonifacio S. (2005), La frequenza di occorrenza di consonanti e vocali e delle loro combinazioni nella sillabe del babbling e delle prime parole dai 10 ai 27 mesi di età. In *Misura dei parametri, Atti del 1° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV)* (P. Cosi, Ed.), Padova: EDK Editore, Brescia.

Zmarich, Dispaldro, Rinaldi & Caselli (in corso di pubblicazione), Caratteristiche fonetiche del Primo Vocabolario del Bambino.

Zmarich, Stocco, Giulivi & Vayra (in questo volume), Sviluppo fonetico e acquisizione lessicale: l'emergere della fonologia dell'italiano in una prospettiva interlinguistica.

## ETICHETTATURA FONETICA E TTS

Pier Luigi Salza \*, Claudia Barolo\*, Beata Dobrzynska^

\* Loquendo S.p.A. – Torino; ^ CELI s.r.l. – Torino

{pierluigi.salza;claudia.barolo}@loquendo.com; beata@celi.it

### 1. SOMMARIO

Nell'articolo si illustra la gestione dell'etichettatura fonetica attuata all'interno del sistema di sintesi da testo LoquendoTTS® multilingue e multilocutore basato sulla tecnica di sintesi concatenativa per selezione di unità. L'etichettatura fonetica rappresenta uno strumento delicato e al tempo stesso essenziale per lo sviluppo del sistema. Il codice per la trascrizione, infatti, deve essere “universale”, nel senso che deve permettere di rappresentare contemporaneamente, in maniera congruente e non ambigua, sia tutti i foni delle diverse lingue, quindi essere *language-independent*, sia tutte le varianti allofoniche dei singoli locutori, quindi essere *speaker-dependent*.

Un grande livello di dettaglio e di accuratezza fonetica è necessario per soddisfare la compatibilità con il segmentatore fonetico automatico dei *database* acustici registrati, mentre la non ambiguità nella rappresentazione dei foni tra lingue diverse è necessaria per attuare il *mapping* fonetico multilingue, ovvero la possibilità che qualsiasi “voce” del TTS, utilizzando il proprio sistema fonologico, pronunci parole o frasi “in prestito” da una qualsiasi delle altre lingue supportate dal sistema, e anche per consentire, infine, l'accesso a lessici fonetici esterni.

Ciò è possibile solo garantendo l'univocità del codice di trascrizione, attuata attraverso la corretta e rigorosa associazione di ciascun simbolo fonetico a tratti fonetico-articolatori (*features*) pertinenti, assimilabili a quelli dell'IPA. L'etichettatura fonetica nel sistema LoquendoTTS® si esplica a tre diversi livelli di granularità: livello di lingua o di parola: trascrizione larga o fonemica; livello di frase: trascrizione stretta; livello di voce: trascrizione allofonica. Il sistema LoquendoTTS®, attraverso questa accurata ed univoca etichettatura fonetica, è in grado di gestire una pluralità di codici: un codice interno, assimilabile a quello dell'IPA, il codice SAMPA e la sua estensione XSAMPA.

### 2. INTRODUZIONE

Il sistema LoquendoTTS® (Text-to-Speech) multilingue e multilocutore si basa sulla tecnica di sintesi concatenativa per selezione di unità non uniformi e contestuali, anche detta tecnica *corpus based*. Tale tecnica consiste nella possibilità di generare il segnale vocale a partire da un *database* di segmenti acustici registrati direttamente dalla voce naturale di uno *speaker* professionista nella sua lingua nativa. Il *database* contiene quindi i foni di una singola lingua e di un singolo parlatore. Le unità acustiche considerate, al contrario di quelle tradizionalmente utilizzate (singoli foni, coppie di foni o difoni, trifoni, demisillabe), sono di lunghezza non definita a priori (perciò non uniformi) e disponibili in differenti varianti prosodiche (dunque contestuali), distinte mediante informazioni aggiuntive di tipo sintattico-prosodico.

I sostanziali vantaggi dell'uso di tali unità sono: 1) l'**ampliamento della estensione** delle unità, che, includendo eventualmente anche parole intere o perfino parti di messaggi, riduce il numero delle giunzioni e quindi delle relative discontinuità; 2) la **contestualità**, che diminuisce la necessità di modificare drasticamente la prosodia in fase di sintesi,

evitando le conseguenti distorsioni del segnale vocale. Il risultato complessivo è che la conoscenza acustico-fonetica di una certa lingua rimane precostituita nelle unità acustiche stesse e il timbro della voce del donatore viene preservato praticamente inalterato, conservando gran parte delle sue caratteristiche naturali. La sintesi mediante questa tecnica ha una qualità acustica paragonabile a quella del parlato preregistrato.

In un sintetizzatore vocale per testi qualsiasi, la soluzione delineata comporta l'attuazione di un difficile compromesso tra estensione e contestualità delle unità acustiche (quindi il loro numero e la pluralità delle loro occorrenze) da un lato e le dimensioni del *database* acustico da registrare dall'altro, per ottenere una consistente copertura fonetica e prosodica della lingua.

Occorre definire, innanzi tutto, una fotografia statistica della lingua. Essa è realizzata mediante algoritmi in grado di analizzare una grande mole di testi trascritti foneticamente, estrarre tutte le sequenze fonetiche da 2 a "n" foni, differenziate in base alla posizione occupata nel sintagma, che è l'unità linguistica prescelta come riferimento per l'assegnazione dell'etichettatura sintattico-prosodica, ordinare infine le sequenze in base alla frequenza di occorrenza. La fotografia statistica del dominio è assunta come *target* di copertura. Si estrae quindi in modo automatico un sottoinsieme minimo di frasi da registrare, foneticamente dense e adatte a coprire il *target*. Le frasi, una volta registrate, vengono poi segmentate ed etichettate automaticamente da un segmentatore fonetico che allinea ciascuna forma d'onda con la corrispondente trascrizione fonetica. Il segmentatore necessita di un addestramento per ciascuno *speaker*, a partire da un piccolo insieme di frasi registrate e dalla loro corrispondente segmentazione manuale.

La sequenza di foni viene quindi automaticamente etichettata con i valori dei parametri acustico-prosodici intrinseci (F0, durata e intensità), estratti mediante algoritmi di analisi del segnale. Un algoritmo specifico marca ogni periodo di F0 (*pitch marker*). A ciascun fono viene infine associata l'etichetta con la categoria prosodica, che identifica il tipo di sintagma (conclusivo, sospensivo, interrogativo) e la posizione del fono nel sintagma cui appartiene. Questa classificazione ha lo scopo di permettere la selezione di varianti significativamente diverse di ciascuna possibile unità.

Il dizionario di unità acustiche contestuali non-uniformi è quindi costituito da un insieme di frasi registrate, spezzate in sintagmi, segmentate e annotate, da utilizzare in linea come serbatoio di elementi potenziali, la cui natura e dimensione non sono definite a priori, ma sono da definirsi *run-time* ad ogni selezione necessaria per generare, in fase di sintesi, un qualsiasi messaggio richiesto.

La selezione delle unità in fase di sintesi è attuata da un algoritmo che sfrutta in modo ottimale il materiale vocale immagazzinato, mirando all'individuazione delle unità migliori, cioè le sequenze più lunghe che meglio si combinano, avendo come obiettivo quello di minimizzare sia le differenze spettrali nei punti di giunzione (contestualità fonetica), sia le differenze tra i parametri acustico-prosodici intrinseci (contestualità prosodica). E' quest'ultima condizione a richiedere che almeno le unità più frequenti vengano coperte da più versioni, tra le quali poter scegliere, al momento della concatenazione, quella che, in quel determinato contesto, minimizza la necessità di modifiche dei parametri prosodici e quindi le conseguenti distorsioni.

Conoscenze linguistiche generali e specifiche di lingua sono perciò fondamentali per lo sviluppo di ogni nuova lingua che si chiede al TTS di sintetizzare. Il punto di partenza per lo sviluppo di tali conoscenze consiste nella definizione del *set* di foni occorrenti nella lingua in esame e delle regole di trascrizione grafema-fono. Si comprende quindi come la



gestione dell'etichettatura fonetica rappresenti uno strumento delicato e al tempo stesso essenziale per tutto il sistema.

Fondandosi sull'utilizzo della tecnica descritta e sul supporto di esperti consulenti madrelingua, sono state sinora sviluppate per il sistema LoquendoTTS® le lingue elencate nell'APPENDICE A. Per una descrizione maggiormente dettagliata si rimanda a (Quazza *et alii*, 2001)

In Figura 1 è riportato lo schema architetturale del sistema LoquendoTTS®, nel quale sono evidenziate alcune funzionalità del sistema in cui l'etichettatura fonetica riveste un ruolo particolarmente importante. Si tratta di: **Segmentatore automatico**, di cui si è parlato sopra; **Mapping fonetico** multilingue, ovvero la possibilità che qualsiasi "voce" del TTS possa pronunciare, utilizzando il proprio sistema fonologico, parole o frasi "in prestito" da altre lingue; **Ingresso fonetico**, ovvero la possibilità che il sistema legga una stringa scritta foneticamente (come previsto, ad es. dalla necessità di accedere ad un lessico esterno).

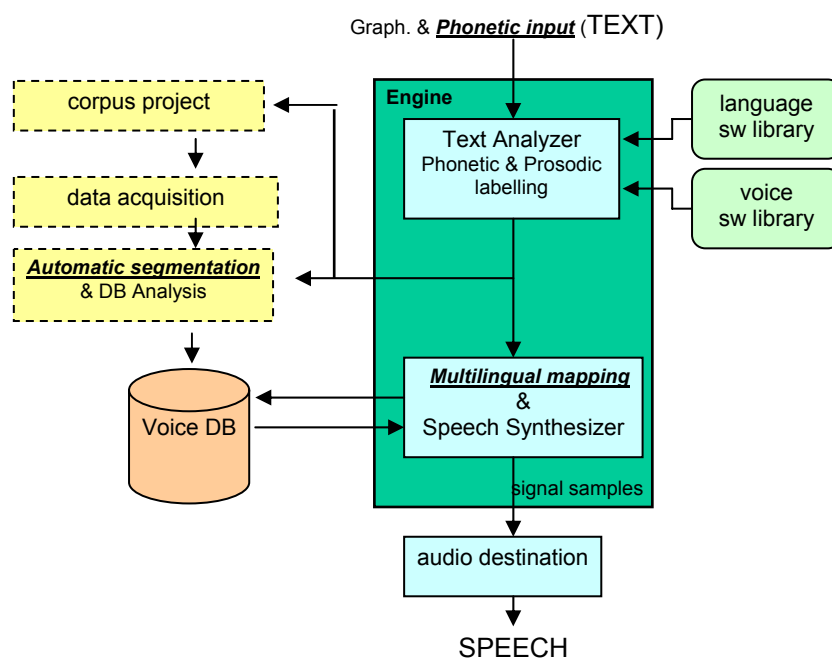


Figura 1. Schema architetturale del sistema LoquendoTTS®.

### 3. ETICHETTATURA FONETICA E TTS

Da quanto esposto risulta evidente la necessità di attuare, all'interno del sistema LoquendoTTS® multilingue e multilocutore, la gestione di un *set* di foni "potenzialmente

infinito”. Il codice per la trascrizione, infatti, deve essere “universale”, nel senso che deve permettere di rappresentare contemporaneamente, in maniera congruente e non ambigua, sia tutti i foni delle diverse lingue, quindi essere *language-independent*, sia tutte le varianti allofoniche dei singoli locutori, quindi essere *speaker-dependent*, consentendo il livello di dettaglio e di accuratezza fonetica atto a soddisfare la compatibilità con il segmentatore fonetico. L’ambiguità, nella rappresentazione dei foni, tra lingue diverse non è accettabile se si deve poter passare da una lingua all’altra (*mapping* fonetico). Ciò significa che la precisione del codice di trascrizione deve essere estremamente accurata, fondata su scelte giustificate foneticamente e pragmaticamente adattate all’applicazione.

La gestione dell’etichettatura fonetica nel sistema LoquendoTTS® si fonda sull’uso rigoroso di tratti fonetico-articolatori assimilabili a quelli dell’IPA (International Phonetic Association). Con “codice”, infatti, si intende l’insieme del simbolo e dei tratti fonetico-articolatori (*features*) associati. L’elenco dei tratti utilizzati da LoquendoTTS® per la codifica fonetica è riportato nell’APPENDICE B. Attraverso un’etichettatura attuata mediante queste *features*, LoquendoTTS® gestisce, oltre ad un codice interno assimilabile al codice dell’IPA, il codice SAMPA (Speech Assessment Methods Phonetic Alphabet) (Wells, 1997), divenuto una sorta di “standard” a livello europeo (van den Heuvel, 2003), e il codice XSAMPA (eXtended SAMPA), versione estesa del SAMPA, che possiede la stessa struttura e quindi la stessa potenzialità di specificazione dell’IPA, proposto dallo stesso autore (si veda il sito [www.phon.ucl.ac.uk/home/sampa](http://www.phon.ucl.ac.uk/home/sampa)). In Figura 2 sono riportati alcuni esempi di codifica di foni con il codice IPA, con il codice SAMPA e con quello XSAMPA.

| IPA  | SAMPA | XSAMPA | TRATTI ARTICOLATORI TTS                                |
|------|-------|--------|--------------------------------------------------------|
| [ē]  | [6~]  | [6~]   | VOWEL   RSDOPEN   CENTRAL   NASALIZED                  |
| [aū] | [aU]  | [aU]   | VOWEL   OPEN   CENTRAL   LWRDCLOSE2   BACK2   ROUNDED2 |
| [β]  | [B]   | [B\]   | CONSONANT   APPROXIMANT   BILABIAL   VOICED            |

Figura 2. Alcuni esempi di codifica di foni con il codice IPA, con il codice SAMPA e con il codice XSAMPA. RSDOPEN: Rised Open; LWRDCLOSE: Lowered Close; 2: secondo segmento vocalico del dittongo.

L’etichettatura fonetica nel sistema LoquendoTTS® si esplica a tre diversi livelli di granularità, che vanno da quello più largo, fonemico, a quello più stretto, allofonico. A ciascun livello di trascrizione corrisponde un *set* di simboli più o meno ampio, con associati tratti fonetico-articolatori, espresso in forma tabellare, come in Figura 2.

Il primo livello è quello della **trascrizione di lingua**. Si tratta di una trascrizione grafema-fonema larga, con cui il testo in ingresso viene trascritto parola per parola. Tale livello di trascrizione fa riferimento ad un *set* di foni contenuto in una Tabella di Lingua. Questo livello di trascrizione deve assicurare la compatibilità con ingressi fonetici esterni (accesso a lessici trascritti foneticamente) e consentire la costruzione di lessici fonetici da parte dell’utente.

Il secondo livello di trascrizione è quello delle regole di riscrittura fonema-fonema. Tale livello di trascrizione è detto anche livello di **trascrizione di frase**. A questo livello vengono trattate le realizzazioni fonetico-articulatorie occorrenti nel contesto di frase

mediante regole di riscrittura che agiscono sia all'interno (occorrenze allofoniche di lingua) sia ai confini tra le parole (effetto dei fenomeni di coarticolazione), incluse le deaccentazioni. Si tratta di una trascrizione stretta, o fonetica (allofonica), che fa ancora riferimento ai foni contenuti nella Tabella di Lingua. Tale trascrizione deve garantire la coerenza con le effettive realizzazioni fonetico-articulatorie dipendenti dalle caratteristiche della lingua in esame a livello di frase e deve allo stesso tempo garantire la non-ambiguità tra le diverse lingue.

Infine, vi è un terzo livello di trascrizione, detto livello di **trascrizione di voce**, che consente l'inserzione di allofoni speaker-dependent, anche diversi per le differenti voci di una certa lingua. Si tratta di una trascrizione stretta, o fonetica (allofonica) idiosincratica, che agisce ancora in base a regole di riscrittura fonema-fonema, cui corrisponde un *set* di foni più ampio contenuto in una Tabella di Voce. Tale trascrizione deve garantire l'operatività del segmenatore fonetico automatico *voice dependent* e dell'algoritmo specifico che marca ogni periodo di F0 (*pitch marker*). A tale livello di trascrizione, l'esatta attribuzione del tratto fonetico-articulatorio *voiced/voiceless* (e *devoiced*) è estremamente importante, sia per l'addestramento del segmentatore sia per il funzionamento del *pitch marker*.

Nella Figura 3 è illustrato un esempio di trascrizione su più livelli dell'enunciato in lingua Inglese britannico “*pests became*”.

Trascrizione di lingua (o di parola): Tabella di Lingua

- IPA: [ˈpɛs ts bɪ k eɪ m]
- SAMPA: [ˈp E s ts b I 'k eI m]
- XSAMPA: [ˈp E s ts b I 'k eI m]

Trascrizione di frase: Tabella di Lingua

- IPA: [ˈpʰɛs ts bɪ kʰ eɪ m]
- SAMPA: [ˈp E s ts b I 'k eI m]
- XSAMPA: [ˈp\_h E s ts b I 'k\_h eI m]

Trascrizione di voce: Tabella di Voce (voce Simon)

- IPA: [ˈpʰɛs ts ɓɪ kʰ eɪ m]
- SAMPA: [ˈp E s ts b I 'k eI m]
- XSAMPA: [ˈp\_h E s ts b\_0 I 'k\_h eI m]

Figura 3. Esempio di trascrizione multilivello dell'enunciato: “*pests became*”, in Inglese britannico. Nota: in Inglese britannico la plosiva *voiceless*, in determinati contesti, può essere aspirata oppure *unreleased*; inoltre, sono frequenti le realizzazioni *speaker dependent devoiced* delle plosive *voiced*, come [b], in contesto *voiceless* (ad es. con la voce Simon).

Dall'esempio mostrato si può osservare come il codice SAMPA non consenta il grado di accuratezza necessario per la rappresentazione degli allofoni, tanto a livello di frase quanto a livello di voce.

#### 4. IMPORTANZA DELL'ACCURATEZZA FONETICA

Come già evidenziato nello schema della Figura 1, per alcune funzionalità del sistema LoquendoTTS® l'etichettatura fonetica è cruciale. Ora mostreremo qualche maggiore dettaglio relativamente a: **Segmentatore fonetico** automatico, **Mapping fonetico** multilingue su testi "mixed language" e **mapping** fonetico con altro codice (SAMPa) per **Ingresso fonetico** al TTS.

##### 4.1 Segmentatore fonetico automatico

Il Segmentatore fonetico automatico è un software che, utilizzando il segnale del parlato registrato e la sua trascrizione fonetica, realizza automaticamente la segmentazione del *database* acustico in foni, difoni e demifoni (mezzi foni), i "mattoncini" più piccoli eventualmente utilizzabili durante la sintesi per la generazione del messaggio, quando non sono reperibili sequenze più lunghe. Il livello di trascrizione necessario è quello di voce. Il software si basa su modelli markoviani (HMM, Hidden Markov Model) e regole (Brugnara *et alii*, 1993). La trascrizione fornisce non solo l'etichettatura fonetica, ma anche l'etichettatura prosodico-posizionale della sillaba cui il fono appartiene, ad es.: prima tonica, tonica interna, atona, ultima tonica, sillaba finale di sintagma/enunciato. Il Segmentatore fonetico automatico opera sulla singola voce mediante previo addestramento. Tutte le peculiarità *speaker-dependent* devono essere rilevate dalla trascrizione e fornite al segmentatore. I foni devono essere etichettati sul *database* acustico così come effettivamente pronunciati dal parlatore perciò ogni discrepanza tra la trascrizione e la pronuncia effettiva va individuata e corretta. Un'analisi acustico fonetica manuale viene quindi effettuata mediante ascolto e visualizzazione del segnale. Si tratta, per la gran parte, di fenomeni di sonorizzazione e desonorizzazione delle consonanti o riduzione delle vocali. Un esempio è dato dalla desonorizzazione delle Occlusive e Approssimanti in Castigliano, da parte della voce Jorge: [b d̥ g β ɸ tɕ] -> [b̥ d̥̥ ɡ̥ β̥ ɸ̥ t̥ɕ̥]. In Figura 4 è riportato un esempio di trascrizione a livello di voce (voce Jorge) della parola "amaba" in Castigliano.

- IPA: [a 'm a β a]
- SAMPA: [a "m a B a]
- XSAMPA: [a "m a B\_0 a]

Figura 4. Esempio di trascrizione a livello di voce, con i codici IPA, SAMPA e XSAMPA, della parola "amaba" in Castigliano (voce Jorge).

##### 4.2 Mapping fonetico su testi "mixed-language"

La conversione testo-voce è intrinsecamente language-dependent, tuttavia le applicazioni del TTS, a partire dai testi reali di una molteplicità di applicazioni, includono sempre più spesso parole o intere frasi che sono prese in prestito da altre lingue. In questi casi i testi sono detti "mixed language". L'obiettivo è ottenere una pronuncia dei termini stranieri mantenendo la stessa voce. Trattandosi di lettura di parole o frasi in L2 presenti in un testo della lingua corrente L1, la pronuncia attesa non è esattamente quella prevista in L2, bensì una sua versione approssimata. Solitamente, infatti, il parlante di L1 tende a pronunciare questi "inserti" di L2 mantenendo il più possibile il proprio sistema fonologico.

Nel sistema LoquendoTTS® multilingue e multilocutore si è realizzato un algoritmo che abilita ciascuna voce a parlare in tutte le lingue supportate dal sistema, utilizzando il proprio sistema fonologico e basandosi su un principio detto *mapping* fonetico multilingue (Badino *et alii*, 2004). Tale principio sfrutta l'accurata e non ambigua etichettatura di fonemi e allofoni delle varie lingue/voci, più sopra descritta, effettuata mediante attribuzione di simboli associati a una stringa di tratti fonetico-articolatori. Il criterio fondamentale per la scelta del fono di L1 che più si avvicina a quello di L2 da pronunciare e mancante in L1 è che due fonemi possono essere giudicati simili quando essi sono caratterizzati da tratti fonetico-articolatori simili. In sostanza i fonemi di L2 mancanti in L1 sono realizzati, con una certa approssimazione, scegliendo i fonemi più simili nel *database* della lingua corrente L1.

La principale difficoltà del *mapping* è rappresentata dalla definizione del peso di ciascun tratto fonetico-articolatorio all'interno di una funzione di stima della similarità (o distanza) tra tratti. Particolarmente difficile è calcolare il grado di affinità tra i valori dei tratti "non binari", sia quelli che si riferiscono al modo e luogo di articolazione delle consonanti sia quelli che specificano la posizione delle vocali all'interno del quadrilatero vocalico. Un esempio molto semplice e comune di *mapping* fonetico è la pronuncia dei fonemi [p<sup>h</sup>] e [p<sup>ʰ</sup>] da parte di una voce italiana che parla Inglese: in entrambi i casi il fono scelto sarà [p]. In Figura 5 e in Figura 6 si riportano due esempi di *mapping*, rispettivamente tra Italiano e Inglese britannico e tra Italiano e Inglese americano, riguardanti la pronuncia della stessa frase: "Her party was a success".

- IPA: [h ɜ: 'p<sup>h</sup> a: t̪ i w ə z ə s ʌ k<sup>ʰ</sup> 's ɛ s]
  - SAMPA: [h 3: "p A: t i w @ z @ s V k "s E s]
  - XSAMPA: [h 3: "p\_h A: 4\_d i w @ z @ s V k\_] "s E s]
- 
- IPA: [e 'p a t̪ i w e z e s e k 's ɛ s]
  - SAMPA: [e "p a t i w e z e s e k "s E s]
  - XSAMPA: [e "p a t i w e z e s e k "s E s]

Figura 5. Esempio di *mapping* tra Italiano e Inglese britannico per la frase "Her party was a success", trascritto in IPA, in SAMPA e in XSAMPA. Prima versione: trascrizione a livello di frase in Inglese britannico; seconda versione: trascrizione del *mapping* tra l'Italiano e l'Inglese britannico (voce Paola).

- IPA: [h ɜ: 'p<sup>h</sup> a: ɹ̥ t̪ i w ə z ə s ʌ k<sup>ʰ</sup> 's ɛ s]
  - SAMPA: [h @ r: "p A: r t i w @ z @ s V k "s E s]
  - XSAMPA: [h 3': "p\_h A: r\ 4\_d i w @ z @ s V k\_] "s E s]
- 
- IPA: [e r 'p a r̥ t̪ i w e z e s e k 's ɛ s]
  - SAMPA: [e r "p a r t i w e z e s e k "s E s]
  - XSAMPA: [e r "p a r t i w e z e s e k "s E s]

Figura 6. Esempio di *mapping* tra Italiano e Inglese americano per la frase "Her party was a success", trascritto in IPA, in SAMPA e in XSAMPA. Prima versione: trascrizione a livello di frase in Inglese americano; seconda versione: trascrizione del *mapping* tra l'Italiano e l'Inglese americano (voce Paola).

#### 4.3 Mapping fonetico con altro codice: SAMPA

Al fine di conseguire una corretta pronuncia è spesso necessario ricorrere a lessici fonetici esterni. E' questo il caso, in quasi tutte le lingue, di molti nomi propri, in particolare cognomi e toponimi, che fanno eccezione alle normali regole di pronuncia. I lessici fonetici utilizzano normalmente trascrizioni con il codice SAMPA, mentre la versione estesa XSAMPA non è praticamente ancora utilizzata.

D'altro canto, è necessario rendere disponibile, oltre all'uscita acustica del TTS, anche un'uscita scritta foneticamente in un codice facilmente leggibile e immediatamente utilizzabile in determinate applicazioni.

Queste due funzioni del sistema LoquendoTTS® sono indicate come *input* e *output* fonetico del TTS. Per ciascuna lingua/voce sono state perciò create due ulteriori tabelle, una in codice SAMPA e una in codice XSAMPA.

Come già accennato in precedenza, il codice SAMPA è un codice per trascrizione larga o fonemica, la cui “grana” non è sufficientemente accurata per una trascrizione fonetica allofonica. Spesso, le trascrizioni SAMPA dei lessici esterni risultano affette da diverse criticità, in particolare: l'incongruenza e la sottospecificazione. L'**incongruenza** ha a che fare con un uso non coerente dei simboli di alcuni foni e con i diacritici. Ad esempio, nei lessici, spesso viene rispettata la non ambiguità intralingua ma non quella interlingue. Ai fini dell'intelligibilità e della fruibilità del singolo lessico (che ovviamente è monolingua) ciò può risultare sufficiente. Per il TTS multilingue ciò rappresenta invece un problema. Non è raro trovare lessici dell'Inglese britannico trascritti in SAMPA in cui l'approssimante alveolare [ɹ] è trascritta semplicemente [r], come nell'esempio “*sorry*” illustrato nella Figura 7.

- IPA: [ˈsɔːrɪ]
- SAMPA: [ˈs O: r i]
- XSAMPA: [ˈs O: r \ i]

Figura 7. Esempio di trascrizione a livello di voce, con i codici IPA, SAMPA e XSAMPA, della parola “*sorry*” in Inglese britannico (voce Simon)

Quanto alla **sottospecificazione** rispetto alle corrispondenti trascrizioni IPA, capita, ad esempio, che in un lessico Olandese si trovino utilizzati i simboli SAMPA [G], [@ n], [r], [l] in luogo dei simboli IPA [x], [ɔ̃], [ə], [t̪]. Nella Figura 8 sono illustrati i casi [@ n] e [r] nell'enunciato in lingua Olandese “*Baarzenstraat*”.

- IPA: [ˈb aə z ɔ̃ ˈs t ɛ a t]
- SAMPA: [ˈb a: r z @ n s t r a: t]
- XSAMPA: [ˈb a: @ ˘ z @ ~ n s t R a: t]

Figura 8. Esempio di trascrizione a livello di voce, con i codici IPA, SAMPA e XSAMPA, della parola “*Baarzenstraat*” in Olandese (voce Willem).

Si incontrano anche frequenti esempi di trascrizioni SAMPA in cui si evidenzia la neutralizzazione di alcune vocali con lo schwa, come mostrato nella Figura 9 per l'enunciato in lingua inglese americano “*east Los Angeles*”, e nella Figura 10 per l'enunciato in lingua tedesca “*Erle Straße*”.

- IPA: [i: s t l oʊ 's æ n dʒ ə l ɪ s]
- SAMPA: [i s t l O " s { n dZ @ l @ s]
- XSAMPA: [i: s t l oU " s { n dZ @ l I s]

Figura 9. Esempio di trascrizione a livello di voce, con i codici IPA, SAMPA e XSAMPA, dell'enunciato "east Los Angeles" in lingua inglese americano (voce Dave).

- IPA: [ɛ ʋ l e: 'ʃ t ʁ a: s ə]
- SAMPA: [E 6 l @ "S t R a: s @]
- XSAMPA: [E 6 l e: "S t R a: s @]

Figura 10. Esempio di trascrizione a livello di voce, con i codici IPA, SAMPA e XSAMPA, dell'enunciato "Erle Straße" in lingua tedesca (voce Stefan).

In tutti questi casi, il *mapping* fonetico consente di selezionare nel *set* di foni di voce quello effettivamente realizzato dallo speaker.

## 5. CONCLUSIONI

L'accurata etichettatura fonetica adottata nel sistema di sintesi multilingue e multilocutore LoquendoTTS<sup>®</sup>, fondata sull'uso rigoroso di tratti fonetico-articolatori assimilabili a quelli dell'IPA, consente una gestione qualitativamente elevata della segmentazione fonetica automatica dei *database* acustici registrati e dell'accesso a lessici fonetici esterni. Inoltre essa fornisce allo stesso tempo una soddisfacente soluzione alla possibilità che qualsiasi voce, attraverso un algoritmo di *mapping* fonetico multilingue, possa leggere parole straniere in una qualsiasi delle lingue supportate dal sistema.

## BIBLIOGRAFIA

- Brugnara, F., Falavigna, D., & Omologo, M. (1993), Automatic Segmentation and Labelling of Speech based on Hidden Markov Models, *Speech Communication*, 12, 357-370.
- Badino, L., Barolo C., Quazza S. (2004), A general approach to TTS reading of mixed-language texts, in *Proceedings of ICSLP 2004*.
- Quazza, S., Donetti, L., Moisa, L. & Salza, P. L. (2001), ACTOR<sup>®</sup>: a multilingual unit-selection speech synthesis system, in *Proceedings of 4th ISCA Tutorial and Research Workshop on Speech Synthesis 2001*, Perthshire, Scotland, August 29 - September 1.
- Van den Heuvel, H., Choukri, K., Höge, H., Maegaard, B., Odijk, J. & Mapelli V. (2003), Quality control of language resources at ELRA, in *Proceedings of EUROSPEECH 2003*, Geneva, Switzerland.
- Wells, J.C., (1997), SAMPA computer readable phonetic alphabet, in *Handbook of Standards and Resources for Spoken Language Systems* (D. Gibbon, R. Moore, & R. Winski, editors), 1997, Berlin and New York: Mouton de Gruyter, Part IV, section B.

## **APPENDICE A.**

Lingue attualmente (03-08) sviluppate per LoquendoTTS®.

|                       |                  |            |
|-----------------------|------------------|------------|
| American              | French           | Portuguese |
| Am. Span. (Columbian) | French Canadian  | Swedish    |
| Argentine             | Galician         | Turkish    |
| Brazilian             | German           | Valencian  |
| British               | Greek            |            |
| Castilian             | Italian          |            |
| Catalan               | Mandarin Chinese |            |
| Chilean               | Mexican          |            |
| Dutch                 | Polish           |            |

Lingue in fase di sviluppo:

Russian, Danish, Finnish



## APPENDICE B.

Tratti fonetico-articolatori (*features*) utilizzati da LoquendoTTS® per l'etichettatura fonetica, dove: LWRDCLOSE: Lowered Close; RSDOPEN: Raised Open; 2: secondo segmento vocalico del dittongo; LTRLTAPFLAP: Lateral TapFlap; LTRLFRICATIVE: Lateral Fricative; LTRLAPPROXIMANT: Lateral Approximant.

### COMMON FEATURES:

Class: VOWEL, CONSONANT, PAUSE, NONPHONEME  
Length: SHORT, LONG, EXTRASHORT, EXTRALONG  
Voicing: VOICED, VOICELESS

### COMMON DIACRITICS:

NASALIZED, DEVOICED, BREATHY, CREAKY

### VOWEL FEATURES:

Rounding: NONROUNDED, ROUNDED  
Height: CLOSE, LWRDCLOSE, CLOSEMID, MID, OPENMID, RSDOPEN, OPEN  
Backness: FRONT, CENTRAL, BACK  
Rounding2: NONROUNDED2, ROUNDED2  
Height2: come sopra  
Backness2: come sopra  
Stress: UNSTRESSED, STRESSED  
Tone: T1, T2, T3, T4, T5, T6, T7

### VOWEL DIACRITICS:

NONSYLLABIC, RHOTICIZED, MOREROUNDED, LESSROUNDED, ADVANCED, RETRACTED

### CONSONANT FEATURES:

Manner: PLOSIVE, NASAL, TRILL, TAPFLAP, LTRLTAPFLAP, FRICATIVE, AFFRICATE, LTRLFRICATIVE, APPROXIMANT, LTRLAPPROXIMANT, CLICK, IMPLOSIVE  
Place: BILABIAL, LABIODENTAL, DENTAL, ALVEOLAR, POSTALVEOLAR, (RETROFLEX), PALATAL, VELAR, UVULAR, PHARYNGEAL, GLOTTAL, EPIGLOTTAL

### CONSONANT DIACRITICS:

ASPIRATED, UNRELEASED, NASALRELEASE, LATERALRELEASE, SYLLABIC, WITHVOICING, LINGUOLABIAL, LABIALIZED, PALATALIZED, VELARIZED, PHARYNGEALIZED, LAMINAL, APICAL, RAISED, LOWERED

# **DIALETTOLOGIA**

## LE OCCLUSIVE PALATALI NEL DIALETTO DI SAN GIOVANNI IN FIORE (CS)

Biagio Mele, Stephan Schmid  
Università di Zurigo  
*boccarosa@gmx.ch, schmidst@pholab.uzh.ch*

### 1. SOMMARIO

Nel presente studio<sup>1</sup> viene esaminato un fenomeno particolare della parlata di San Giovanni in Fiore che si trova anche nel sistema fonetico-fonologico di altri dialetti meridionali. Si tratta delle occlusive palatali [cʲ] che appaiono in parole come [çaʃe] ‘chiave’, [βiççʰu] ‘vecchio’, [comu] ‘piombo’, oppure in [ʲiçzja] ‘chiesa’ e [ε̞ʲutu] ‘è andato’. L’occlusiva sonora ricorre sempre come geminata; oltre che in [ʲiçzja] ‘chiesa’ essa è presente in contesto raddoppiante come allofono dell’approssimante palatale /j/. L’occlusiva palatale sorda [c] ha invece statuto di fonema; diacronicamente deriva dai nessi latini (-)CL-, (-)TL- e (-)PL-.

Negli anni recenti si è assistito a un rinnovato interesse per le caratteristiche fonetiche delle occlusive palatali in area romanza, come dimostrano le ricerche sperimentali sul dialetto della Valsesia (Romano *et alii*, 2005) e sulla varietà romancia della Bassa Engadina (Schmid, in stampa). Il presente contributo si inserisce in questo filone di studi, apportando dati empirici tratti da una varietà dell’Italia meridionale.

A tale scopo sono stati registrati quattro parlanti maschi di un’età compresa tra i 43 e i 70 anni e residenti nella Svizzera tedesca, i quali hanno pronunciato 36 parole contenenti le consonanti [k g cʲ tʃ dʒ j] in vari contesti vocalici, ripetendole tre volte in frase cornice. Il corpus comprende quindi 432 ricorrenze di foni consonantici che sono stati sottoposti ad analisi acustica con l’ausilio del programma *Praat*.

Nella dimensione temporale si è tenuto conto della durata totale dei foni esaminati nonché del rapporto di durata tra le fasi di occlusione e di rilascio. In base ai dati analizzati risulta che le occlusive palatali hanno una durata totale relativamente breve, più simile a quella delle occlusive velari che non a quella delle affricate palato-alveolari; rispetto al rapporto tra tenuta e rilascio, le palatali del sangiovese si collocano in una posizione intermedia tra le velari e le palato-alveolari. Ricordiamo che le ostruenti palatali del romancio si sono invece rilevate molto più simili alle palato-alveolari che non alle velari (v. Schmid, in stampa).

Per quanto riguarda le caratteristiche spettrali, è stato calcolato il ‘Centro di Gravità’ (CdG) del rumore di rilascio (Forrest *et alii*, 1988). I risultati rivelano che le occlusive palatali del dialetto di San Giovanni in Fiore si differenziano chiaramente dalle altre due categorie di foni considerati (a differenza del retoromancio della Bassa Engadina, dove le occlusive palatali e le affricate palato-alveolari mostrano anche delle caratteristiche spettrali simili).

---

<sup>1</sup> Benché il lavoro sia frutto di una stretta collaborazione fra i due autori, il testo può essere suddiviso nel seguente modo: a BM vanno attribuiti i §§ 1, 2, 3.1-3.2, 4.1-3, a SSch i §§ 3.3, 4.4, 5.1-3, 6.

## 2. IL DIALETTO DI SAN GIOVANNI IN FIORE

Il comune di San Giovanni in Fiore è situato a settanta chilometri ad est di Cosenza ad oltre 1000 m sul livello del mare; con una superficie territoriale di 279 kmq e con 18.000 abitanti esso rappresenta il più grande ed antico centro abitato dell'altipiano silano. A San Giovanni l'uso del dialetto è in declino tra i parlanti più giovani, ma è ancora vivo nelle altre generazioni; il dialetto costituisce la varietà dominante per la comunicazione informale, vale a dire in famiglia e fra amici, anche all'interno delle comunità di emigrati sangiovanesi all'estero (particolarmente numerosi nella Svizzera tedesca).

La parlata di San Giovanni in Fiore è da annoverare tra i dialetti meridionali estremi, con cui condivide alcune proprietà fondamentali, quali ad esempio il vocalismo di tipo 'siciliano' e la presenza di consonanti retroflesse (v. Mele, in stampa). Inoltre, questo dialetto presenta delle peculiarità fonetiche che lo contraddistinguono dalla parlata del capoluogo di provincia, tra cui va menzionato innanzitutto il dittongamento di /ɪ/ tonica in /ie/ davanti alla laterale scempia /l/. Il sangiovanese ha poi in comune con altre due parlate calabresi il fono interdentale laterale [ð<sup>(l)</sup>], così come condivide con altri dialetti dell'area silana la nasalizzazione di /a/ tonica in contesto nasale (cfr. Falcone, 1976; Trumper, 1997).

Il sistema pentavocalico del dialetto in questione viene riportato nel prospetto diacronico (1) che illustra gli esiti delle vocali nel dialetto odierno rispetto alla loro genesi dal latino classico:

|     |       |   |     |  |            |   |                       |
|-----|-------|---|-----|--|------------|---|-----------------------|
| (1) | ɪfj   | > |     |  | DiŋCO      | > | [rɪkʊ]                |
|     | ĩ     | > | /ɪ/ |  | ÞIRA(M)    | > | [pɪra]                |
|     | eŋ    | > |     |  | FEŋMINA(M) | > | [fɪmɪna]              |
|     | ě     | > | /ɛ/ |  | PĚTRA(M)   | > | [pĕa]                 |
|     | Aŋ, Ǻ | > | /a/ |  | CǺPU(T)    | > | [kəpu]                |
|     | ö     | > | /ʊ/ |  | PÖNTE(M)   | > | [pʊnt <sup>h</sup> ɛ] |
|     | oŋ    | > |     |  | HoŋRA(M)   | > | [ʊra]                 |
|     | ũ     | > | /u/ |  | JŮGU(M)    | > | [jʊβu]                |
|     | uŋ    | > |     |  | CRUŋDU(M)  | > | [krʊru]               |

A differenza dello sviluppo più frequente nei dialetti meridionali estremi, il sangiovanese possiede al luogo delle vocali alte /i/ e /u/ le corrispondenti vocali semi-alte /ɪ/ e /ʊ/ (cfr. Loporcario, 1991).

La vocale semi-alta posteriore /ʊ/ rimane di solito inalterata, mentre quella anteriore /ɪ/ è interessata da un processo allofonico di dittongamento che avviene quando tale vocale è seguita dall'esito di –L– intervocalica (v. *infra*):

- (2) [apɾið<sup>(l)</sup>ɛ] 'aprile', [fiɛð<sup>(l)</sup>ʊ] 'filo', [tiɛð<sup>(l)</sup>a] 'tela', [pʊð<sup>(l)</sup>ɪtʃɛ] 'pulce'

Per quanto ci risulta, tale fenomeno non è mai stato descritto per nessun'altra varietà dialettale; in questo caso, il mutamento fonico trova la sua spiegazione nella reinterpretazione, da parte dell'ascoltatore, delle transizioni formantiche che si producono davanti a questo particolare tipo di consonante (v. Loporcario & Mele, 2004: 90-98).

Un'altra regola allofonica colpisce la /a/ tonica, che si innalza e si nasalizza nei contesti pre- e postnasale. Tale processo di assimilazione (che può essere sia regressiva che progressiva) dà origine a una vocale nasalizzata, medio-bassa, centrale, non arrotondata [ẽ], che viene articolata con la lingua piatta e abbassata mentre l'apertura labiale è più ristretta rispetto a quella dell'allofono orale [a] (v. Mele, in stampa):<sup>2</sup>

- (3) [mẽṃa] 'mamma', [mẽkṃa] 'macchina', [ẽṇima] 'anima'

Consideriamo ora brevemente anche qualche particolarità fonetica nel consonantismo di questo dialetto. Notevole risulta ad esempio la sonorizzazione di /s/ in posizione intervocalica, laddove nella grande maggioranza dei dialetti meridionali la sibilante rimane sorda. Abbiamo già accennato a un fenomeno strettamente legato al dittongamento illustrato negli esempi in (2): si tratta dell'evoluzione di (-)L- latina che viene realizzata come laterale interdentale [ð<sup>(l)</sup>], caratteristica che il sangiovanese condivide solo (a quanto sinora se ne sa) con i dialetti cosentini di Bisignano e Casole Bruzio:

- (4) [ð<sup>(l)</sup>ṃu] 'lino', [ð<sup>(l)</sup>ʊtʃɛ] 'luce', [kʊð<sup>(l)</sup>ʊrɛ] 'colore', [pʊð<sup>(l)</sup>ʊtʃɛ] 'pulce'

Per una trattazione più dettagliata di questo particolare tipo di consonante rimandiamo alle analisi acustiche in Loporcaro & Mancuso (1998), chiudendo invece questa breve rassegna delle peculiarità fonetiche del sangiovanese con la descrizione di un fenomeno di natura fonosintattica.

Il dialetto di San Giovanni in Fiore mostra infatti di aver mantenuto il raddoppiamento della consonante iniziale dopo le forme verbali di terza plurale del presente indicativo:

- (5) [mẽndʒu t<sup>h</sup>ardu] 'mangiano tardi', [sɛ ʌtsu p<sup>h</sup>riɛstu] 'si alzano presto'

Oltre che nel sangiovanese, il raddoppiamento fonosintattico si conserva in tale contesto solo nel dialetto di Conflenti, nella provincia di Catanzaro (cfr. Loporcaro, 1995).

Accingiamoci dunque ad esaminare l'aspetto del consonantismo che ci occupa in questa sede (le occlusive palatali per l'appunto), rimandando per una trattazione più estesa e dettagliata della fonetica e della fonologia del dialetto di San Giovanni in Fiore alla monografia del primo autore di prossima pubblicazione (Mele, in stampa).

### 3. LE OCCLUSIVE PALATALI

#### 3.1 *Gli esiti di occlusiva sorda con laterale in ambito italo-romanzo e nel sangiovanese*

Il nesso latino PL- ha subito sviluppi diversi all'interno dell'area italo-romanza. L'evoluzione più frequente è quella che si osserva in toscano e in gran parte dei dialetti settentrionali, dove la palatalizzazione della laterale è passata attraverso una fase intermedia [j] all'approssimante [j]: PLACE(T) > [pjatʃɛ], PLUMBU(M) > [pjombo] ecc. .

Poche sono le varietà italo-romanze che si sottraggono alla palatalizzazione del nesso PL-. Tra le aree marginali che hanno conservato la laterale, Rohlfs (1949 [1966: 252-255])

<sup>2</sup> Nelle giovani generazioni la nasalizzazione di /a/ tende a regredire in alcuni contesti fonotattici, in particolare quando tra la consonante e la vocale si trova un confine di morfema o di parola (v. Mele, 2004).

individua nel settentrione alcuni dialetti della Valtellina superiore, dell'area lombarda orientale e del Trentino, citando forme come *plómp* 'piombo' per Livigno (SO) e Bagolino (BS), oppure *plats* 'piazza' per Pejo (TN). Nell'Italia centro-meridionale tale nesso è conservato in una piccola area dell'Abruzzo: ad esempio, a Castelli (TE) si registra *plats* 'piazza' e *plómm* 'piombo'.

In alcuni dialetti dell'Abruzzo, della Calabria e della Sicilia la laterale può passare a vibrante (abr. *prand* 'pianta', cal. *praja* 'spiaggia' e sic. *prattu* 'piatto'), ed è generalizzato il rotacismo di tutti i nessi di consonante + L nel campidanese: *kramare* 'chiamare', *prus* 'più', *frore* 'fiore' ecc. (Wagner, 1951: 311). A dire il vero, anche nel sangiovese si registra qualche caso isolato di [pr] in corrispondenza di PL- etimologico, come in [prak<sup>h</sup>a] 'placca' e [pra<sup>h</sup>tanu] 'platano'; tali voci però non sono di tradizione diretta (cfr. Mele, in stampa).

Com'è noto, nella maggior parte dei dialetti dell'Italia meridionale, l'esito del nesso PL- coincide con quello del nesso CL-; infatti, in contesto iniziale alla palatalizzazione si è aggiunta la fusione delle occlusive (originariamente) labiali con quelle velari, per cui in napoletano si ha *chiano* per 'piano' così come in siciliano si ha *chiummu* per 'piombo'. Non è questa la sede per interrogarci sulla motivazione di questo mutamento fonico che potrebbe essere sia di tipo acustico-percettivo, nel senso di una reinterpretazione delle labiali che condividono con le velari il tratto 'grave' (Tuttle 1975, 1997), sia di tipo articolatorio, nel senso di un'assimilazione regressiva parziale dell'occlusiva, condizionata dall'approssimante palatale desonorizzata (Barbato 2005).<sup>3</sup>

A dire il vero, la seconda ipotesi sembrerebbe più plausibile soprattutto per quei dialetti in cui la palatalizzazione ha raggiunto uno stadio ulteriore all'esito bifonemico /kj/ (generalmente attribuito al napoletano e al siciliano; cfr. Rohlfs, 1949 [1966: 253]), quello che troviamo appunto nel dialetto di San Giovanni in Fiore e che proponiamo di rappresentare con l'esito monofonemico /c/. Tale sviluppo viene riportato non solo per tutta una serie di dialetti calabresi e siciliani, ma anche per i dialetti pugliesi e salentini e persino per il napoletano (Tuttle, 1975: 421; Barbato, 2005: 412).<sup>4</sup> Comunque, una verifica dell'esatta natura di queste consonanti nei dialetti meridionali costituisce ancora un *desideratum* della ricerca fonetica e dialettologica.

Nel presente lavoro cerchiamo di iniziare a colmare tale lacuna attraverso l'analisi dei dati di un singolo dialetto, quello di San Giovanni in Fiore; prima di presentare il metodo e i risultati dell'analisi acustica descriviamo nel prossimo paragrafo l'origine diacronica delle occlusive palatali e il loro funzionamento nel sistema fonologico del sangiovese odierno.

<sup>3</sup> Evidentemente, il discorso andrebbe esteso all'evoluzione di tutti i nessi latini di consonante + laterale (v. Repetti & Tuttle, 1987; Tuttle, 1997; Barbato, 2005). Per motivi di spazio non affrontiamo qui la vasta casistica della palatalizzazione (quasi) panromanza di C-, G- davanti a vocale anteriore (v. Celata, 2002) né quella più circoscritta della palatalizzazione di C-, G- davanti a /a/ (v. Schmid, in stampa).

<sup>4</sup> V. per esempio Valente (1975: 61) e Loporcaro (1988: 214-215) per la Puglia, Romano (1999: 1054) per il Salento, e Romito & De Rito (2002: 96-97) per il catanzarese; per la presenza delle occlusive palatali in varie regioni del Meridione si rimanda a Romano *et alii* (2005: 395-396).

### 3.2 Le occlusive palatali nel dialetto di San Giovanni in Fiore

Come già accennato, nella parlata in questione i nessi iniziali PL- e CL- hanno dato àdito a un unico esito [c]:

- (6) a. [çaβɛ] ‘chiave’, [cʊzʊ] ‘chiuso’, [cuoβʊ] ‘chiodo’  
 b. [caʦsa] ‘piazza’, [cuɲʊ] ‘piombo’, [cuβɛɾɛ] ‘piovere’, [cɛ̃ndʒɛɾɛ] ‘piangere’

Vi è un unico lessema dove in posizione iniziale assoluta appare l’occlusiva palatale sonora geminata [ʝ]: si tratta del tipo lessicale [ʝiɛzja] ‘chiesa’ < (E)CCLESIA(M)<sup>5</sup> che si presenta con una velare sonora scempia in buona parte della Romània, ad esempio nello spagnolo *iglesia*, nel francese *église* e nel friulano *glesie*. Evidentemente la parola [ʝiɛzja] non appartiene alla stessa categoria degli esempi in 6b., visto che la consonante iniziale non deriva da un nesso CL-, ma da un nesso con geminata in posizione interna -CCL-; la sonorizzazione dev’essere avvenuta prima dell’aferesi della vocale iniziale.

In posizione interna di parola, il dialetto di San Giovanni in Fiore presenta numerose parole con l’occlusiva palatale sorda geminata [çʰ]:

- (7) a. [ʝɾiçʰa] ‘orecchio’, [mɛçʰa] ‘macchia’  
 b. [kaçʰʊ] ‘cappio’, [kuçʰa] ‘coppia’  
 c. [βiçʰʊ] ‘vecchio’

Dal punto di vista sincronico notiamo innanzitutto che nel sangiovanese /p t k c/ vengono realizzate come aspirate (v. Mele in stampa).<sup>6</sup> Dal punto di vista diacronico vanno invece precisati due aspetti: da un lato si constata che tutte le parole in (7) sono originariamente dei proparossitoni, diventati parossitoni in seguito alla sincope della vocale postonica, mentre dall’altro si osserva che l’occlusiva palatale sorda non deriva solo dalla fusione dei nessi -CL- e -PL- (v. gli esempi in 7a. e b.), ma che in questo caso alla palatalizzazione partecipa anche il nesso originario -TL- (v. l’esempio in 7c.) che appare già nel latino volgare come -CL- secondario (cfr. Rohlfs 1949 [1966]: 349).

A differenza dei nessi latini di occlusiva sorda + laterale, i nessi -BL- e -GL- non hanno portato alla genesi dell’occlusiva sonora (il caso di [ʝiɛzja] rimane infatti isolato), ma la lenizione e la successiva perdita dell’occlusiva sonora danno come esito l’approssimante palatale /j/:

- (8) [jɿɲaru] ‘gomitolo’ (< \*GLOMERU(M), [jaʦʃu] ‘ghiaccio’

Tuttavia, nei contesti raddoppianti il fonema /j/ mostra come variante combinatoria l’occlusiva palatale sonora geminata [ʝ]:

- (9) [sɪçʰʊjʉtʉ] ‘sono andato’ vs. [ɛ̃çʰʉtʉ] ‘è andato’

<sup>5</sup> Ricordiamo che anche nell’italiano di Roma la parola *chiesa* appare con una geminata lessicale in posizione iniziale (Bertinetto & Loporcaro, 2005: 134).

<sup>6</sup> Le occlusive vengono aspirate anche quando ricorrono come consonanti scempie dopo vibrante o dopo nasale (v. Mele, in stampa). Per questo aspetto il sangiovanese si mostra quindi più simile al catanzarese che non al cosentino, che presenta al contrario una forte lenizione delle occlusive sorde postnasali (cfr. Marotta & Soriano, 1992).

[nu<sub>2</sub>jʊɾɛ] ‘un fiore’                      vs.        [ɪ<sub>2</sub>ʊ<sub>2</sub>ɾɪ] ‘tre fiori’  
 [ɪ<sub>2</sub>ʊ<sub>2</sub>jʊkɑɾɪ] ‘lui gioca’                      vs.        [a<sub>2</sub>ʊ<sub>2</sub>kɑtʊ] ‘ha giocato’

Ribadiamo infine che all’interno del sistema fonologico del sangiovanese l’occlusiva palatale sorda possiede lo statuto di fonema, come si evince da numerose coppie minime:

- (10) [cʊzʊ] ‘chiuso’ ~ [kʊzʊ] ‘cucio’  
 [cɛ̃nʊ] ‘piano’ ~ [sɛ̃nʊ] ‘sano’  
 [caɫ<sup>h</sup>ʊ] ‘piatto’ ~ [aɫ<sup>h</sup>ʊ] ‘gatto’  
 [kaç<sup>h</sup>ʊ] ‘cappio’ ~ [kaɫ<sup>h</sup>ʊ] ‘tolgo’

Dato l’alto rendimento funzionale dell’occlusiva palatale in posizione tanto iniziale che intervocalica, possiamo concludere che si tratta di un suono ‘vitale’ che ancora non sembra correre il rischio di fondersi con /tʃ/.

### 3.3 La descrizione fonetica delle occlusive palatali

Tuttavia, le occlusive palatali rappresentano dei suoni marcati dal punto di vista della tipologia fonologica. Infatti, la maggior parte delle lingue del mondo possiede suoni occlusivi ai tre luoghi di articolazione bilabiale, alveolare e velare, mentre sono piuttosto rare le occlusive palatali. Questo almeno è il dato che si evince consultando la banca dati degli inventari fonematici allestita all’Università della California a Los Angeles (UCLA), lo *UCLA Phonetic Segment Inventory Database* (UPSID); tale banca dati si basa su uno spoglio delle descrizioni fonologiche di 451 lingue rappresentative dal punto di vista della loro appartenenza alle grandi famiglie genetiche delle lingue del mondo (Maddieson, 1984; Maddieson & Precoda, 1991).<sup>7</sup>

| ʃ    c     | ɲ      | ʃ    tʃ    | dʒ     | ʃ    k | g      |
|------------|--------|------------|--------|--------|--------|
| (d) 11.97% | 09.53% | (e) 41.69% | 25.06% | 89.36% | 56.10% |

Tabella 1: Frequenza delle ostruenti palatali et velari nell’UPSID

In effetti, la tabella 1 mostra che nove su dieci lingue possiedono l’occlusiva velare sorda, laddove l’occlusiva palatale sorda si trova solo in una su dieci lingue. È notevole che siano quattro volte più frequenti le affricate palato-alveolari sorde, data la relativa marcatezza delle affricate rispetto alle occlusive. Invece, nella coppia delle palatali è meno accentuato lo scarto di frequenza tra la sorda e la sonora. Rarisime sono infine le occlusive palatali lunghe [ç ɲ] che sono ambedue presenti nel sangiovanese, ma appaiono solo nel 0.22% delle lingue registrate nell’UPSID. Tutto sommato possiamo desumere che la rarità delle ostruenti palatali sia da imputare a una loro difficoltà intrinseca per la produzione e/o la percezione della parola; dal punto di vista del parlante va rilevato che la distanza tra la volta del palato duro e il corpo la lingua è maggiore rispetto agli altri articolatori passivi ed attivi, così come potrebbe essere difficilmente discriminabile il rumore di rilascio delle palatali rispetto a quello di altre consonanti simili.

<sup>7</sup> V. anche il sito del laboratorio di fonetica dell’Università di Francoforte che permette una rapida consultazione dell’UPSID: <http://web.phonetik.uni-frankfurt.de>.



Ma più che per i parlanti, le occlusive palatali sembrano essere dei suoni ‘difficili’ per i linguisti. Così descrizioni diverse di una stessa lingua usano trascrizioni alternative per quello che deve essere lo stesso fonema data la sua posizione strutturale nel sistema fonologico. Nel caso dell’ungherese, lo schizzo fonetico dell’Associazione Internazionale di Fonetica (Szende, 1999: 104) caratterizza come affricate [cç ʷ] quelle che nella monografia di Siptár & Törkenczy (2000: 82-83) vengono rese come occlusive [c ʷ]. Similmente, per la varietà romancia della Bassa Engadina troviamo sia i simboli per le occlusive palatali [c ʷ] (Haiman & Benincà, 1992: 29, 33) sia una trascrizione che concepisce gli stessi suoni come affricate alveolo-palatali [tɲ dʃ] (Taggart, 1990: 11). In quest’ultimo caso è quindi soggetto a discussione sia il modo di articolazione (occlusive vs. affricate) sia il luogo di articolazione (palatale vs. alveolo-palatale; cfr. Schmid, in stampa e *infra*).

Non è questa la sede per esaminare tutta la casistica delle diverse possibilità articolatorie nella vasta gamma del tratto vocale che si estende dagli alveoli al velo duro. Alla difficoltà oggettiva posta dall’analisi di queste consonanti si aggiunge una certa proliferazione terminologica e concettuale, dato che le descrizioni sono state fornite non solo da fonetisti, ma anche da dialettologi e filologi romanzi (si pensi ad esempio alla fortuna di termini come ‘prepalatale’, ‘mediopalatale’ e ‘postpalatale’ che non vengono interpretati nella stessa maniera da tutti gli autori). Per una trattazione estesa e dettagliata della problematica rimandiamo il lettore alla meritevole *summa* in Romano *et alii* (2005, 390-407), dove si passa in rassegna la vasta letteratura sull’argomento, tenendo conto dei vari filoni e delle varie tradizioni di ricerca in ambito italiano e internazionale (cfr. anche Romano, 2007).

Limitandoci dunque a qualche sommaria osservazione di ordine espositivo, possiamo distinguere in base al modo di articolazione almeno quattro categorie: occlusive, affricate, articolazioni secondarie e nessi consonantici. Se le occlusive si possono ridurre ai tre prototipi alveolare [t d], palatale [c ʷ] e velare [k g], per le affricate è il caso di considerare le palato-alveolari (dette anche postalveolari) [tʃ dʒ], le alveolo-palatali [tɲ dʃ], le palatali *tout court* [cç ʷ] e le palato-velari [kç gʷ].<sup>8</sup> Dalle consonanti palatali vere e proprie vanno infatti distinte le articolazioni secondarie delle occlusive velari, siano esse palatalizzate [kʲ gʲ] o semplicemente anteriorizzate (o prevelari) [k̟ g̟]. Queste ultime, cioè le occlusive velari anteriorizzate, ricorrono tipicamente in francese davanti alle vocali [i e y] (per un ovvio meccanismo di coarticolazione), laddove le occlusive velari palatalizzate possono risultare dalla pronuncia *allegro* di un originario nesso con l’approssimante palatale [kj gj]; in particolare per i dialetti italiani del Meridione dobbiamo aspettarci di trovare i tre esiti diversi [kj], [kʲ] e [c] (cfr. § 3.1) che possono essere interpretate come tre fasi di una stessa evoluzione diacronica. La nostra ipotesi è che nel dialetto di San Giovanni in Fiore siamo di fronte alla terza soluzione, ovvero a delle occlusive palatali vere e proprie (che rappresenterebbero quindi la fase più innovativa), e che molti altri dialetti meridionali presentino invece esiti del secondo se non addirittura del primo tipo.

A questo punto si impone una verifica sperimentale delle descrizioni fonetiche, e il metodo più immediato consisterebbe certamente nell’ispezione dei processi fisiologici coinvolti nella produzione delle consonanti in questione. Purtroppo, nel nostro caso non è

<sup>8</sup> Dato il livello altamente semplificativo della nostra esposizione consideriamo qui solo l’articolatore passivo, ma è ovvio che bisognerebbe far riferimento anche all’articolatore attivo, a seconda della forma e della parte della lingua (lamina, predorso, dorso, ecc.) coinvolta nel gesto (v. Keating & Lahiri, 1993: 76-79).

stato possibile condurre un'analisi articolatoria (elettropalatografica, EMA, ecc.), a differenza dello studio sul dialetto della Valsesia presentato da Romano *et alii* (2005: 409-411); questi ricercatori si sono avvalsi della tecnica delle Immagini in Risonanza Magnetica (IRM), evidenziando le aree di contatto tra lingua e palato duro in varie articolazioni (palatali e non).

In mancanza di tali tecniche di analisi, il metodo più facilmente accessibile resta quello elettro-acustico. Due sono le dimensioni acustiche nelle quali le occlusive palatali dovrebbero distinguersi da altre articolazioni vicine, ovvero la dimensione temporale e quella spettrale. L'aspetto temporale può fornire delle indicazioni riguardanti il modo di articolazione: calcolando il rapporto tra la durata della fase di rilascio e la durata del segmento intero ci si aspetta di trovare dei valori più alti nel caso delle affricate che non nelle occlusive (cfr. Romano *et alii*, 2005: 414-415; Schmid, in stampa). Tuttavia, proprio nel nostro caso non si tratta di un discrimine facile, dato che nelle occlusive palatali la fase di rilascio può avere una durata sensibilmente maggiore rispetto ad esempio alle labiali o alle alveolari. Tale maggiore durata della fase di rilascio è dovuta da un lato alla distanza maggiore che separa il (pre-)dorso della lingua dal palato duro e dall'altro alla minore velocità del gesto eseguito dall'articolatore attivo.

Sul versante spettrale esistono invece vari parametri che servono come correlati acustici del luogo di articolazione delle occlusive palatali. Com'è noto, per caratterizzare le occlusive la fonetica acustica 'classica' fa riferimento alla teoria dei *loci* che si basa sulle transizioni formantiche osservabili tra il rilascio dell'occlusive e le vocali contigue. Uno sviluppo ulteriore di questa teoria consiste nel modello dell'equazione dei *loci* proposto da Sussmann *et alii* (1993) che definisce il *locus* della consonante in base a due valori della seconda formante, misurati rispettivamente all'inizio della transizione e nella fase 'stabile' della vocale. Tale modello è stato applicato con successo all'analisi delle occlusive palatali nel dialetto della Valsesia da parte di Romano *et alii* (2005: 415-417).

Nel presente contributo si è invece scelto un approccio alternativo per esaminare le proprietà spettrali delle consonanti in questione. Trattandosi di articolazioni con una fase di rilascio consistente (che anche nel caso delle occlusive velari si protrae di più rispetto alle bilabiali e alle alveolari), abbiamo deciso di focalizzare l'analisi non sulla transizione formantica nella vocale contigua, bensì sulla distribuzione dell'energia nello spettro del rumore del rilascio della consonante stessa. A tale scopo abbiamo considerato il cosiddetto 'Centro di Gravità' proposto da Forrest *et alii* (1988). Sinora questo parametro acustico è stato utilizzato soprattutto per caratterizzare il rumore delle fricative, ad esempio nello studio contrastivo su sette lingue diverse presentato da Gordon *et alii* (2002). Il Centro di Gravità (CdG) è una specie di centroide spettrale che si calcola moltiplicando tutti i valori di frequenza nello spettro con i rispettivi valori d'intensità, per dividere successivamente tutti questi prodotti per la somma di tutti i valori di frequenza dello spettro. Il CdG riflette il carattere più o meno acuto (o grave) di un rumore e può essere correlato con la grandezza della cavità anteriore della bocca durante la produzione di un suono in cui è coinvolta la lingua come articolatore attivo. I valori del CdG aumentano per le consonanti anteriori e diminuiscono per le consonanti posteriori: ad esempio, Gordon *et alii* (2002: 150) riportano nel gaelico un CdG di 4884 Hz per [s] e di 4209 Hz per [x] (valori medi su otto parlanti). Uno degli scopi della presente ricerca è quello di verificare fino in che misura il parametro del CdG può essere utilizzato per caratterizzare consonanti con un rumore di una durata relativamente breve come appunto le occlusive palatali.

## 4. DATI E METODO

### 4.1 I parlanti

Sono stati intervistati quattro parlanti maschi di un'età compresa tra i 43 e i 70 anni. Tutti e quattro risiedono da molti anni nell'area di Baden-Wettingen (situata nel Canton Argovia, nella Svizzera tedesca), ma nel loro repertorio sociolinguistico il dialetto funge da sistema primario dal punto di vista sia della biografica linguistica sia della competenza linguistica: è la loro lingua materna (ovvero la prima lingua che hanno imparato) ed è la varietà nella quale si sentono più sicuri e che costituisce il codice comunicativo in famiglia e che adempie a un'importante funzione identitaria. Presentiamo dunque i quattro parlanti nell'ordine in cui sono stati registrati (che corrisponde pure all'ordine di anzianità).

Il parlante più anziano AM è nato a San Giovanni in Fiore nel 1937 dove ha vissuto fino all'età di 19 anni. Nel 1956 si è trasferito a Wettingen (Canton Argovia, Svizzera) dove risiede tuttora. Autista in pensione, è padre di due figli con i quali ha sempre parlato in sangiovese.

Il secondo parlante AB è nato nel 1959 a San Giovanni in Fiore dove ha vissuto fino all'età di 21 anni, ad eccezione di un breve periodo trascorso in provincia di Pordenone per il servizio militare. Nel 1980 si è trasferito in Svizzera, dove dopo due anni l'ha raggiunto sua moglie, anche lei originaria di San Giovanni; due figli (un maschio e una femmina) sono nati in Svizzera. AB lavora come meccanico industriale.

Il terzo parlante GC è nato a San Giovanni in Fiore nel 1960 dove ha fatto le scuole elementari, medie e professionali. È perito meccanico e ha vissuto nel paese fino all'età di 19 anni; nel 1979 è emigrato in Svizzera, dove si è sposato nel 1984 con una donna originaria di San Giovanni; ha due figli di 23 e 13 anni.

L'ultimo parlante, GA, è nato nel 1964 a San Giovanni dove ha vissuto fino all'età di 24 anni, con l'eccezione del periodo del servizio militare; nel 1988 è emigrato in Svizzera. Lavora come elettromeccanico, è sposato e ha due figli di 12 e 8 anni.

### 4.2 Le registrazioni

Le interviste con i quattro parlanti sono state effettuate l'8 ottobre 2007 nell'abitazione privata del primo autore. Il colloquio si è svolto prevalentemente in dialetto sangiovese ed è consistito di due fasi. In un primo momento, il primo autore (dialettologo anche lui) ha rivolto agli intervistati delle domande generali sulla loro biografia e sulla loro esperienza migratoria. In una seconda fase egli ha elicitato delle parole o espressioni dialettali fornendo loro come stimolo i termini equivalenti in italiano; ottenuta l'espressione desiderata, i locutori sono stati pregati di ripeterla in frase cornice:

- (11) Aju rittu \_\_\_ a prima vota.  
Aju rittu \_\_\_ a secunna vota.  
Aju rittu \_\_\_ a terza vota.

Le interviste sono state salvate su scheda Flash mediante un registratore digitale Marantz PMD 671; è stato utilizzato un microfono Sennheiser ME66 con direttività supercardioidale/clava, gamma di frequenza di 50-20000 Hz  $\pm 2,5$  dB e coefficiente di trasmissione a vuoto di 35 mV/Pa  $\pm 2,5$  dB. I file audio sono stati registrati in formato .wav con una frequenza di campionamento di 44100 Hz e una quantizzazione di 16 Bit.

### 4.3 Il corpus

La lista delle 36 espressioni elicitate comprende parole contenenti le consonanti [k g c ɲ tʃ dʒ j] davanti alle vocali [a e i ʊ]. Nella tabella 2 sono elencati gli stimoli in italiano, mentre la tabella 3 riporta le corrispondenti voci dialettali.

|    | <b>k</b>              | <b>g</b>              | <b>c</b>               | <b>ɲ</b>   | <b>tʃ</b>           | <b>dʒ</b>  | <b>j</b>    |
|----|-----------------------|-----------------------|------------------------|------------|---------------------|------------|-------------|
| _a | la casa<br>il cavallo | sdentato<br>tre gatti | la chiave<br>la pianta | che fiato! | leggera             | giallo     | il fiato    |
| _e |                       |                       |                        | tre generi | che brutta<br>cera! | il gesso   | il genere   |
| i  | questo                |                       | pieno                  | la chiesa  | la cera             | il giro    | andare      |
| o  | la coscia             | che gonna             | piove                  | ha giocato | stupida             | la gabbia  | lui gioca   |
| u  | la culla              | tre gocce             | il piombo              | è andato   | l'asino             | il giovane | sono andato |

Tabella 2: Stimoli in italiano

|    | <b>k</b>            | <b>g</b>               | <b>c</b>              | <b>ɲ</b>       | <b>tʃ</b>            | <b>dʒ</b>  | <b>j</b>    |
|----|---------------------|------------------------|-----------------------|----------------|----------------------|------------|-------------|
| _a | a casa<br>u cavallu | sgangatu<br>tri ggatti | a chjave<br>a chjanta | chi gghjatu    | lecia                | ggiallu    | u jatu      |
| _e |                     |                        |                       | tri gghjennari | chi bbrutta<br>cera! | u ggressu  | u jennaru   |
| i  | chissu              |                        | chjinu                | a gghiesja     | a cira               | u ggiru    | jire        |
| o  | a coscia            | chi ggonna!            | chjovari              | ha gghiocatu   | ciota                | a gaggia   | illu jocari |
| u  | a culla             | tri ggucce             | u chjumu              | è gghjutu      | u ciucciu            | u ggiuvine | signu jutu  |

Tabella 3: Parole elicitate in dialetto

Nella rappresentazione ortografica delle parole dialettali usiamo per le occlusive palatali i trigrafi <chj> e <ghj> (ovvero il ‘quadrigrafo’ <gghj> nel caso della geminata sonora). Per ogni consonante da analizzare si è cercato di trovare cinque parole in cui la consonante in questione occorra in posizione iniziale davanti alle cinque vocali del sangiovanese, ma ciò non è stato possibile a causa della mancanza di lessemi dialettali. Alcune caselle strutturali rimangono quindi vuote, mentre altre sono riempite due volte.

Per l’analisi acustica si è tenuto conto di 5 parole inizianti con [k], di 3 parole realizzate con [g] geminata in fonotassi,<sup>9</sup> di 5 parole inizianti con [c], di 5 parole inizianti con [ɲ], di 5 parole contenenti [tʃ] (tutte scempie tranne l’ultima) e di 5 parole contenenti [dʒ] (intrinsecamente lunga, come di consueto nelle parlate del Meridione). Considerando che ognuno dei quattro parlanti ha ripetuto queste 28 parole tre volte, il *corpus* analizzato comprende quindi 336 foni. Si è deciso di confrontare le occlusive palatali con le consonanti ‘più vicine’, cioè le affricate palato-alveolari e le occlusive velari, al fine di evidenziare le loro caratteristiche temporali e spettrali. La diversificazione dei contesti vocalici serviva a neutralizzare, nei limiti del possibile, eventuali effetti di coarticolazione.

<sup>9</sup> Nel sangiovanese il fonema /g/ occorre prevalentemente nella sua variante combinatoria lenita [ɣ], per esempio in contesto postvocale [ʊ̃aɣʰʊ] ‘il gatto’, oppure nella variante raddoppiata [g]: [ɪ̃gaɣʰɪ] ‘tre gatti’. Per l’analisi spettrale si è tenuto conto anche della ricorrenza dell’occlusiva velare nella parola [zgaŋgaɣʰʊ] ‘sdentato’, per cui in questo caso il *corpus* ammonta a 348 foni.

#### 4.4 Parametri di analisi

Le misurazioni sono state effettuate mediante il programma *Praat* (Boersma & Weenink, 2008). Come già accennato al § 3.3, l'analisi acustica si è svolta in due dimensioni essenziali: per poter stabilire il modo di articolazione delle consonanti in questione (occlusiva vs. affricata) si è tenuto conto di vari parametri temporali (§ 5.2), laddove per la disamina del luogo di articolazione (palatale, palato-alveolare, velare) è stato applicato un solo correlato spettrale, il 'Centro di Gravità' (§ 5.3).

Nella dimensione temporale abbiamo misurato le durate delle fasi di occlusione e di rilascio (in ms), osservando le modulazioni del segnale nella forma d'onda e nello spettrogramma a banda larga; in base a queste due misurazioni abbiamo calcolato la durata totale dei segmenti nonché il rapporto tra la fase di rilascio e la durata totale (%). Quest'ultimo parametro si è rivelato particolarmente adatto per distinguere tra occlusive e affricate. Occorre precisare che per 'fase di rilascio' qui non si intende il momento dello scoppio prodotto dall'allontanamento degli articolatori (il *burst*), bensì tutto il lasso di tempo che trascorre tra l'inizio del gesto di rilascio della consonante e l'inizio del gesto vocalico successivo (che può consistere dello scoppio e di un momento più esteso di frizione oppure di un'unica fase di rumore); in questo senso, la 'fase di rilascio' corrisponde essenzialmente al VOT così come definito da Cho & Ladefoged (1997: 225).

Per quanto riguarda invece le caratteristiche spettrali, il parametro del 'Centro di Gravità' (in Hz) è stato rilevato direttamente mediante l'apposita funzione in *Praat*. Precedentemente abbiamo verificato mediante l'ispezione spettrografica la differenza tra le transizioni formantiche dovute da un lato alla coarticolazione consonantica e dall'altro alla configurazione formantica delle approssimanti palatali (§ 5.1).

### 5. CARATTERISTICHE ACUSTICHE DELLE OCCLUSIVE PALATALI NEL DIALETTO DI SAN GIOVANNI IN FIORE

#### 5.1 Ispezioni spettrografiche esemplari

Prima di presentare i risultati delle misurazioni dal punto di vista quantitativo, conviene illustrare qualche esempio spettrografico preliminare. Un confronto tra gli spettrogrammi di [c], [k] e [tʃ] offre una prima impressione delle caratteristiche temporali e spettrali di queste consonanti; inoltre, un ulteriore confronto con [j] e [ɥ] dovrebbe fornire delle indicazioni rispetto alla monofonematicità o bifonematicità degli esiti di CL- (cfr. § 3.2).

La figura 1 riporta la forma d'onda e lo spettrogramma del sintagma [a<sub>1</sub>caʃe] pronunciato dal parlante AM:

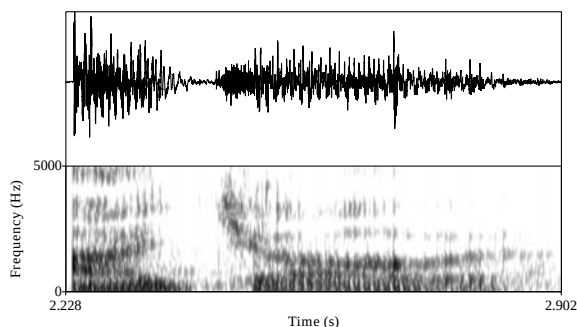


Figura 1: il sintagma [a<sub>1</sub>caʃe] pronunciato dal parlante AM

Come si evince dallo spettrogramma, la transizione della seconda formante verso il *locus* palatale è quasi più evidente nella vocale che precede l'occlusiva palatale: F2 parte da 1440 Hz nella fase stabile per raggiungere 1883 Hz all'inizio della consonante (la fase di transizione si potrae su 58 ms). Nella transizione dall'occlusiva alla vocale successiva, F2 scende da 1927 Hz fino a 1372 Hz durante un lasso di tempo di 43 ms; il precedente rumore di rilascio dura 51 ms e ha un'intensità media di 67.3 dB.

Un'immagine leggermente diversa viene offerta dalle rappresentazioni del sintagma [a<sub>ː</sub>kʌzʌ] pronunciato dallo stesso parlante:

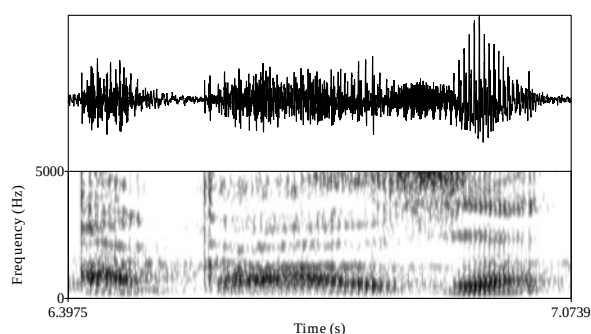


Figura 2: il sintagma [a<sub>ː</sub>kʌzʌ] pronunciato dal parlante AM

Rispetto all'occlusiva palatale, quella velare mostra con 24 ms una fase di rilascio molto più breve (e anche un'intensità media lievemente maggiore di 68.9 dB), ma anche la fase della transizione di F2 ha una durata più breve (35 ms); inoltre quest'ultima appare molto meno chiara rispetto allo spettrogramma precedente, nonostante il *locus* di F2 relativamente alto che si attribuisce in genere alle velari (cfr. Giannini & Pettorino, 1992: 195).

Vediamo dunque come si presenta l'affricata /tʃ/, pronunciata dal parlante AB nella parola *rittu ciota* 'detto stupida':

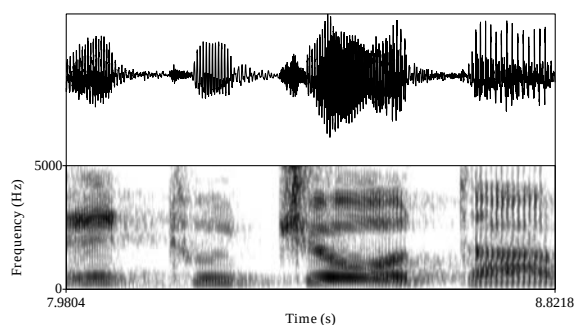


Figura 3: la sequenza [ritʰʊtʃʊta] pronunciata dal parlante AB

Nel caso dell'affricata palato-alveolare, la fase di rilascio ha una durata di 49 ms e un'intensità media di 69.5 dB; la transizione di F2 dura 49 ms e scende da 1741 Hz a 1115 Hz. Il fatto che il rilascio abbia una durata minore rispetto all'occlusiva palatale indica che gli aspetti temporali meritano di essere approfonditi da un punto di vista quantitativo (v. §

5.2). Invece, i valori dell'intensità si scostano molto poco (67.3 dB, 68.9 dB, 69.5 dB), per cui possiamo trascurare questo parametro nelle analisi successive.<sup>10</sup>

Data la notevole durata della transizione di F2 nella realizzazione sia di [c] (46 ms) che di [tʃ] (49 ms), è opportuno confrontare queste due consonanti con una sequenza di approssimante + vocale. Vediamo dunque una realizzazione di *illu jɔcari* 'lui gioca', pronunciata dallo stesso AB:

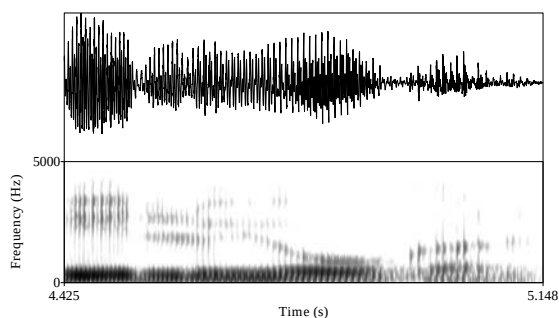


Figura 4: la sequenza [ɪlʊ.jɔkari] pronunciata dal parlante AB

Benché l'approssimante [j] sia caratterizzata da un andamento discendente di F2 paragonabile alle transizioni formantiche (da 1927 Hz a 972 Hz), essa non solo presenta una durata sensibilmente maggiore (96 ms), ma mostra anche un comportamento completamente differente dal punto di vista della coarticolazione. Data la fonematicità dell'approssimante, il punto di partenza del movimento formantico non è determinato da un segmento adiacente, ma costituisce esso stesso un *target* capace di esercitare un effetto coarticolatorio sulla vocale precedente: infatti, la /ʊ/ di *illu* viene realizzata come [y], come mostra il valore della seconda formante (1941 Hz).

Diversa è l'andamento formantico nel sintagma *ha jɔcatu*, pronunciato sempre da AB:

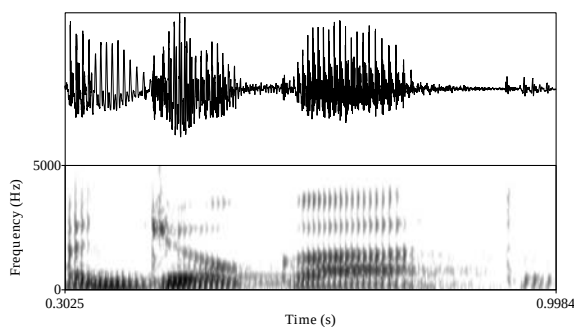


Figura 5: il sintagma [a.jɔkɪtu] pronunciata dal parlante AB

<sup>10</sup> Tale scelta è corroborata anche dall'analisi delle ostruenti palato-alveolari e palatali nel Vallader, dove un confronto quantitativo delle realizzazioni di tre parlanti non ha fornito alcuna differenza significativa riguardo all'intensità media di queste consonanti (cfr. Schmid, in stampa).

La lunghezza di [ɲ] – variante rafforzata di /j/ determinata dal lessema ‘raddoppiante’ *ha* – si manifesta piuttosto nella fase di tenuta (83 ms) che non in quella di rilascio (31 ms), il che depone a favore della natura di occlusiva (e non affricata) di questa consonante. Con una durata di 27 ms, la transizione formantica che segue il rilascio non ammonta nemmeno a un terzo della durata di [j] nello spettrogramma precedente. Alla luce di questo dato possiamo ragionevolmente dedurre che neanche nella figura 1 ci troviamo di fronte a un nesso [kj] né tanto meno di fronte a un’occlusiva velare palatalizzata [kʲ]. L’ispezione spettrografica ci invita dunque a ritenere che si tratta piuttosto di un unico gesto articolatorio, quello di un’occlusiva palatale. Tuttavia, per poter decidere sul modo di articolazione (occlusiva vs. affricata) e sul luogo di articolazione (palatale vs. palato-alveolare) delle consonanti in questione occorre esaminare varie occorrenze di più parlanti, cosa che avverrà nei prossimi due paragrafi.

## 5.2 Caratteristiche temporali delle consonanti [k c tʃ g ɲ dʒ] nel sangiovese

La figura 6 mostra il quadro complessivo delle durate nei sei foni esaminati:

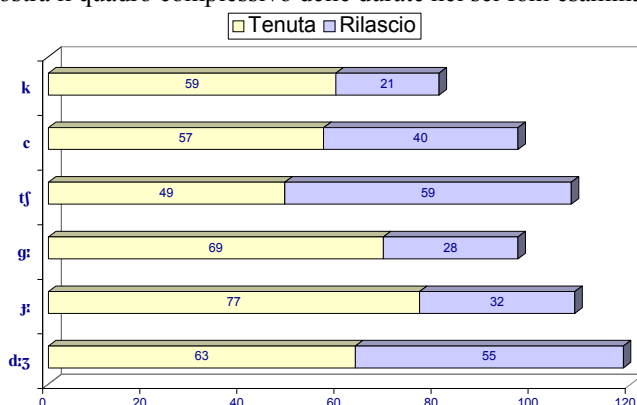


Figura 6: durate (in ms) delle fasi di tenuta e di rilascio (medie dei 4 parlanti)

Ricordiamo che i valori medi sono stati calcolati in base a 60 ripetizioni per ogni fono (4 parlanti x 3 ripetizioni x 5 parole), ad eccezione di [ɲ] di cui abbiamo misurato solo 36 ripetizioni (4 parlanti x 3 ripetizioni x 5 parole; cfr. § 4.3).

Teoricamente, l’interpretazione dei risultati viene resa più difficile per il fatto che le tre consonanti sonore [g ɲ dʒ] sono intrinsecamente lunghe, ma tale inconveniente è almeno alleviato dal fatto che tutte le sonore sono (fonologicamente) lunghe. Sorprende quindi che riguardo alla durata totale lo scarto tra sorde e sonore non sia più marcato: è vero che all’interno dei tre luoghi di articolazione le sonore sono sempre più lunghe delle sorde (97 ms di [g] rispetto a 80 ms di [k], 109 ms di [ɲ] rispetto a 97 ms di [c], 118 ms di [dʒ] rispetto a 108 ms di [tʃ]), ma la differenza tra le durate resta sempre contenuta e non corrisponde all’immagine che si ha di una consonante ‘doppia’.<sup>11</sup>

Il diagramma in figura 6 mostra invece con chiarezza una differenza di durata inerente al luogo di articolazione per cui le velari sono più brevi delle palatali, le quali a loro volta

<sup>11</sup> Ricordiamo comunque che secondo una tendenza fonetica universale le occlusive sonore hanno una durata minore rispetto alle occlusive sorde (Maddieson, 1997: 626).



sono più brevi delle palato-alveolari. Questa tendenza si manifesta sia per le sorde (valori medi della durata totale: [tʃ] = 108 ms > [c] = 97 ms > [k] = 80 ms) sia per le sonore (valori medi della durata totale: [dʒ] = 118 ms > [ʒ] = 109 ms > [g] = 97 ms). Ciò fa sì che la scempia più lunga [tʃ] abbia con 108 ms una durata media superiore alla geminata più breve [g] (97 ms). Volendo stabilire il modo di articolazione delle palatali in base alle medie della durata totale, notiamo quindi che esse si trovano in una posizione intermedia tra le (occlusive) velari e le (affricate) palato-alveolari; anzi, ad essere precisi, le palatali sembrerebbero avere caratteristiche temporali più simili alle palato-alveolari (la differenza delle durate medie è di 11 ms per le sorde e di 9 ms per le sonore) che non alle velari (dove la differenza delle durate medie è di 17 ms per le sorde e di 12 ms per le sonore).

Ma per meglio comprendere la natura temporale di occlusive e affricate non bisogna tener conto tanto della durata totale quanto delle durate delle due fasi di tenuta e di rilascio. Ad esempio, dalla figura 6 si evince che la differenza tra consonanti brevi e lunghe consiste essenzialmente nella fase di tenuta, minore nella prima categoria ([tʃ] = 49 ms < [c] = 57 ms < [k] = 59 ms) che non nella seconda ([dʒ] = 63 ms < [g] = 69 ms < [ʒ] = 77 ms). Le sorde brevi mostrano quindi per la durata dell'occlusione una gerarchia speculare a quella della durata totale; inoltre la tenuta media delle palatali è molto più simile alla tenuta delle velari che non a quella delle palato-alveolari. Nelle sonore lunghe, la fase di occlusione delle palatali supera persino quella delle velari. Ad ogni modo, le palato-alveolari hanno una tenuta più breve all'interno sia delle brevi che delle lunghe, il che costituisce un primo indizio a favore della diversa struttura temporale di queste consonanti rispetto agli altri due luoghi di articolazione.

Il quadro diventa ancora più evidente se si considera la fase di rilascio, dove le due palato-alveolari mostrano invece le durate più consistenti indipendentemente dalla sonorità (59 e 55 ms).<sup>12</sup> All'interno delle sorde, la durata del rilascio di [tʃ] (59 ms) ammonta al triplo di quella di [k] (21 ms), mentre [c] (40 ms) si colloca in una posizione equidistante; all'interno delle sonore, la durata del rilascio di [dʒ] (55 ms) ammonta al doppio di quella di [g] (28 ms), ma questa volta la consonante palatale [ʒ] (32 ms) si avvicina molto di più all'occlusiva velare che non all'affricata palato-alveolare. A differenza della durata dell'occlusione, per la fase di rilascio troviamo la stessa gerarchia secondo i luoghi di articolazione (palato-alveolari > palatali > velari) sia per le sorde che per le sonore.

Aggiungiamo infine ai valori medi delle tre durate (totale, tenuta, rilascio) una grandezza relativa, che consiste nel rapporto tra la durata del rilascio e la durata totale. Tale rapporto, espresso come percentuale (%ril), viene elencato nella tabella 4 per ognuno dei quattro parlanti (AM, AB, GC, GA); inoltre si indica la media di %ril dei quattro parlanti, nonché la deviazione standard (dev.st.) e il coefficiente di variazione (co.var.).

<sup>12</sup> In effetti, la durata della fase di rilascio (qui inteso come *burst* + frizione; cfr. §4.4) è il criterio fondamentale che sottostà alla definizione del modo di articolazione delle affricate: "Affricates are stops in which the release of the constriction is modified in such a way as to produce a more prolonged period of frication after the release" (Ladefoged & Maddieson, 1996: 90).

| %ril | AM   | AB   | GC   | GA   | media | dev.st. | co.var. |
|------|------|------|------|------|-------|---------|---------|
| k    | 0.28 | 0.26 | 0.27 | 0.25 | 0.26  | 0.01    | 0.05    |
| c    | 0.45 | 0.37 | 0.40 | 0.44 | 0.41  | 0.03    | 0.08    |
| tʃ   | 0.55 | 0.46 | 0.62 | 0.58 | 0.55  | 0.07    | 0.12    |
| g    | 0.31 | 0.21 | 0.25 | 0.42 | 0.30  | 0.09    | 0.32    |
| ɲ    | 0.35 | 0.23 | 0.24 | 0.38 | 0.33  | 0.08    | 0.26    |
| dʒ   | 0.49 | 0.49 | 0.43 | 0.52 | 0.47  | 0.04    | 0.09    |

Tabella 4: rapporto medio tra la durata del rilascio e la durata totale (4 parlanti)

La tendenza generale che emerge dalla figura 6 e dalla tabella 4 mostra che in un'affricata prototipica del sangiovese la fase di rilascio ammonta all'incirca alla metà della durata totale; nelle sorde, questo rapporto può anche superare il 50%. In un'occlusiva prototipica il rilascio può occupare un quarto o un terzo della durata totale. Riguardo alle palatali osserviamo comunque che le sonore corrispondono più fedelmente a questo prototipo, laddove le sorde hanno una struttura temporale intermedia tra quella di un'occlusiva e di un'affricata; potremmo in un certo senso considerare questo *pattern* come un caso estremo all'interno della categoria delle occlusive.

Dalla tabella 4 si evince peraltro che i quattro parlanti analizzati si comportano in modo abbastanza omogeneo, per cui possiamo ritenere questi dati tipici per il dialetto di San Giovanni in Fiore. Il coefficiente di variazione rimane piuttosto contenuto nelle sorde (0.05, 0.08, 0.12) e nel caso di [dʒ] (0.09), essendo invece più consistente per [g] (0.32) e [ɲ] (0.26). La notevole variazione nelle velari sonore è dovuta a differenze idiosincratiche tra i parlanti AB e GA e rispecchia uno stile elocutivo completamente diverso, tendente all'iperarticolazione nel primo caso e all'ipoarticolazione nel secondo caso:<sup>13</sup> nel caso di [g], una verifica delle misurazioni rivela che la tenuta media di AB (96 ms) dura più del doppio di quella di GA (45 ms).

### 5.3 Caratteristiche spettrali di [tʃ c k] e di [dʒ ɲ g] nel sangiovese

L'istogramma in figura 7 mostra il Centro di Gravità delle tre ostruenti sorde [k c tʃ] per i quattro parlanti AM, AB, GC e GA (valori medi in Hz):

<sup>13</sup> Non abbiamo calcolato la velocità di elocuzione, che potrebbe fornire un'utile misura di controllo in questo contesto.

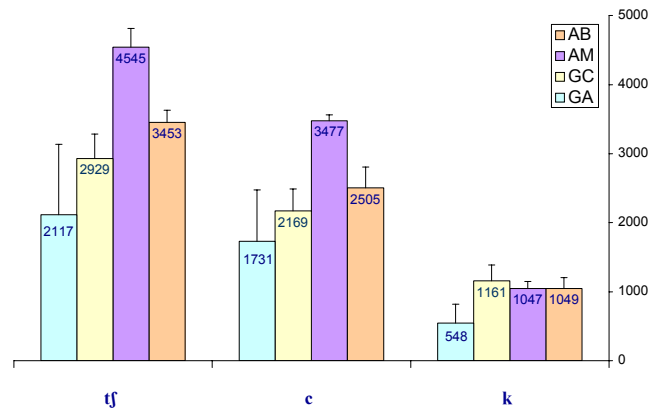


Figura 7: Centro di Gravità (in Hz) di [tʃ c k] (valori medi per ognuno dei 4 parlanti)

Il quadro non manca certo di coerenza, anche se – considerando singolarmente alcuni valori medi – si possono individuare delle sovrapposizioni. Ad esempio, il CdG di [c] rilevato presso i parlanti AB (2505 Hz) e CG (2169 Hz) è leggermente più alto di quello rilevato per [tʃ] presso il parlante GA (2117 Hz). Ancora, il parlante AM mostra un CdG di [c] (3477 Hz) più alto del CdG di [tʃ] rilevato nei tre parlanti AB, CG e GA (rispettivamente 3453 Hz, 2929 Hz e 2117 Hz). In altre parole, esiste una certa sovrapposizione dei valori misurati per l’occlusiva palatale sorda e l’affricata palato-alveolare sorda che riflette la somiglianza acustica (e quindi del rischio di confusione a livello percettivo) di queste due consonanti.

Ciononostante, vi sono altri indizi che testimoniano a favore della netta separazione delle tre consonanti in questione. Innanzitutto, va notato che il CdG di [k] è sempre inferiore a quello degli altri due foni, indipendentemente dal parlante. Osserviamo altresì che i valori medi dei quattro parlanti considerati complessivamente (ovvero la media delle quattro medie) mostrano una chiara tendenza discendente che va dall’affricata palato-alveolare (3261 Hz) all’occlusiva palatale (2471 Hz) all’occlusiva velare (951 Hz). Ma vi è di più: la stessa tendenza discendente si delinea per ciascuno dei quattro parlanti considerato singolarmente (come del resto si può vedere nell’istogramma in base al contorno ascendente-discendente che le colonne formano per ognuna delle tre consonanti), solo che pendenza di questa linea discendente immaginaria ([tʃ] > [c] > [k]) varia a seconda del parlante. La pendenza più ripida si riscontra in AM (4545 Hz > 3477 Hz > 1047 Hz); seguono AB (3453 Hz > 2505 Hz > 1049 Hz) e CG (2929 Hz > 2169 Hz > 1161 Hz) e troviamo infine la linea più ‘piatta’ in GA (2117 Hz > 1731 Hz > 548 Hz).

Da questi dati possiamo trarre due conclusioni. Da un lato viene confermata la fondamentale correlazione (negativa) tra il Centro di Gravità del rumore di frizione e la grandezza della cavità orale anteriore che si viene a creare al momento dell’articolazione consonantica; dall’altro constatiamo però che il CdG dipende in una certa misura anche da proprietà idiosincratiche dei singoli parlanti (dovute a certe abitudini articolatorie e/o alla conformazione anatomica del tratto vocale). Possiamo quindi ipotizzare che l’ascoltatore debba attuare per ogni parlante una sorta di ‘normalizzazione percettiva’ a seconda del grado di ripidità della pendenza dei valori del CdG sull’asse antero-posteriore.

E finiamo la presentazione dei risultati con uno sguardo sul Centro di Gravità nelle tre ostruenti sonore [dʒ ɲ g], indicando di nuovo i valori medi in Hz per ciascuno dei quattro parlanti analizzati :

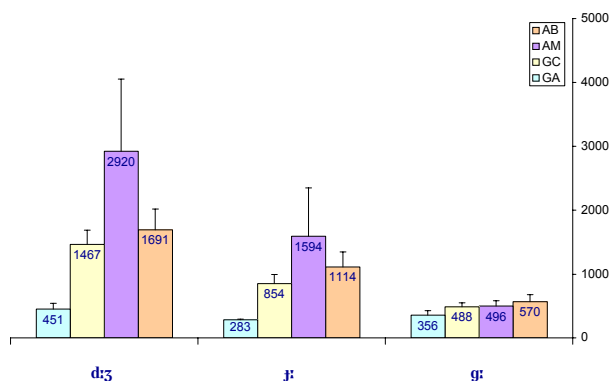


Figura 8: Centro di Gravità (in Hz) di [dʒ ɲ g] (valori medi per ognuno dei 4 parlanti)

*Mutatis mutandis*, il quadro dell'istogramma in figura 8 ripropone le stesse tendenze emerse già per le tre ostruenti sorde nell'istogramma della figura 7. La differenza più ovvia risiede nel fatto che le sonore presentano un CdG più basso rispetto alle tre ostruenti sorde, e ciò vale per ognuno dei tre luoghi di articolazione. Ricordiamo che nel caso delle sorde le medie del CdG di tutti i parlanti diminuiscono gradualmente dai 3261 Hz delle palato-alveolari ai 2471 Hz delle palatali per raggiungere i 951 Hz delle velari. Ora, nel caso delle sonore, i valori delle medie sono praticamente dimezzati, dato che la stessa linea immaginaria parte dai 1632 Hz delle palato-alveolari ai 961 Hz delle palatali per raggiungere i 478 Hz delle velari. La ragione di questa differenza è ovvia e risiede proprio nella sonorità delle consonanti in questione: la presenza di energia nelle bande di frequenza basse incide sulla media dei picchi di intensità nello spettro, abbassando notevolmente il Centro di Gravità.

In linea generale le tre ostruenti sonore mostrano però un'analogia diminuzione del CdG man mano che si procede sull'asse dell'antero-posteriorità. Anzi, in questo caso la 'linea immaginaria' del CdG ha in un certo senso una pendenza più lineare: in effetti, le palatali si collocano in una posizione equidistante (961 Hz) tra gli altri due luoghi di articolazione (1632 Hz e 478 Hz), laddove nel caso delle sorde le palatali (2471 Hz) si avvicinano di più alle palato-alveolari (3261 Hz) che non alle velari (951 Hz).

Inoltre riscontriamo nella figura 8 le stesse differenze idiosincratiche tra i parlanti emerse già nell'istogramma precedente riguardo alla pendenza della linea immaginaria del Centro di Gravità. Come abbiamo già potuto constatare nel caso delle tre ostruenti sorde, anche per le sonore in genere è AM a presentare il CdG più alto (2920 Hz > 1594 Hz > 496 Hz), benché in questo caso venga superato nelle velari da AB (1691 Hz > 1114 Hz > 570), il quale occupa normalmente il secondo posto in base alla media del CdG. Al terzo posto segue anche qui il parlante GC (1467 Hz > 854 Hz > 488 Hz), e le medie più basse si riscontrano di nuovo presso GA (451 Hz > 283 Hz > 356 Hz). Notiamo che solo in quest'ultimo parlante viene parzialmente inficiata la regola generale per cui all'interno della scala idiosincratica dei singoli parlanti il Centro di Gravità diminuisce gradualmente sull'asse dell'antero-posteriorità (dato che in questo caso il CdG delle velari è più alto di quello delle palatali).

Tutto sommato va ribadito comunque che le misurazioni del CdG nei quattro parlanti sangiovesi esaminati confermano tre risultati fondamentali: i) il Centro di Gravità correla in modo abbastanza robusto con il luogo di articolazione; ii) il Centro di Gravità va interpretato secondo una scala ‘normalizzata’ per ogni parlante; iii) nelle ostruenti sonore i valori medi del CdG sono sempre più bassi a causa dell’energia presente nelle basse bande di frequenza, pur mostrando grosso modo le tendenze i) e ii) già riscontrate per le sorde.

## 6. DISCUSSIONE E CONCLUSIONE

Accingendoci a qualche rapida considerazione conclusiva, prendiamo lo spunto proprio dall’ultimo parametro esaminato, ovvero dal Centro di gravità. I nostri risultati inducono a postulare che questo parametro – originariamente formulato per analizzare le proprietà spettrali di consonanti fricative – possa essere applicato con successo anche allo studio di altri tipi di ostruenti. Ciò vale non solo per le affricate, ma anche per quelle occlusive che vengono prodotte con una fase di rilascio consistente, com’è il caso delle velari e – soprattutto – delle palatali. Per queste costruenti il CdG offre quindi una chiave di lettura alternativa alla classica teoria dei *loci*, in quanto permette un’analisi diretta delle proprietà spettrali della consonante stessa e non un’osservazione indiretta basata su fenomeni di coarticolazione vocalica.

Ciò non toglie che il Centro di Gravità costituisce ancora una grandezza approssimativa che merita di essere approfondita e sviluppata ulteriormente.<sup>14</sup> Data la relativa scarsità degli studi condotti con questo parametro, siamo lungi dall’aver raggiunto una serie di valori di riferimento come quella canonica delle formanti vocaliche. Ad esempio, i CdG medi delle sette lingue studiate da Gordon *et alii* (2002) variano notevolmente tra di loro, e anche i parlanti delle singole lingue prese in considerazione mostrano spesso dei risultati contraddittori tra di loro; inoltre lo scarto tra il CdG dei vari luoghi di articolazione rimane in genere piuttosto ridotto.

Nel confronto interlinguistico, i nostri risultati rivestono quindi un certo interesse per quattro motivi: i) il Centro di Gravità può costituire un utile parametro spettrale per analizzare non solo fricative, ma affricate e occlusive con una fase di rilascio consistente (come lo sono le palatali e velari del sangiovese); ii) viene confermata la fondamentale correlazione (negativa) tra il Centro di Gravità del rumore di rilascio e la grandezza della cavità anteriore delle consonanti nella cui articolazione è coinvolta una parte della lingua; iii) nei dati del sangiovese esiste sì una variazione tra i singoli parlanti, ma tale variazione interindividuale è assolutamente sistematica; iv) nel sangiovese la ‘linea immaginaria’ del CdG medio (cfr. § 5.3) ha una pendenza molto più ripida rispetto a quella di altre lingue (cfr. Gordon *et alii*, 2002).

Le nostre analisi permettono di concludere che nel sangiovese le occlusive palatali possiedono un luogo di articolazione ben distinto dalle altre due categorie vicine, le affricate palato-alveolari e le occlusive velari. Soprattutto la differenziazione dalle affricate palato-alveolari depone a favore della vitalità delle occlusive palatali nel sistema del

---

<sup>14</sup> La natura approssimativa del Centro di Gravità risiede proprio nella sua natura aritmetica che indica solo il punto centrale della dispersione dell’energia nello spettro, senza tener conto della forma dell’involuppo spettrale. Per studi futuri sulle occlusive palatali potrebbero quindi rivelarsi interessanti altri parametri che sono stati applicati all’analisi delle fricative, come ad esempio la frequenza iniziale e l’ampiezza della gamma della curva spettrale, la frequenza del picco con maggiore energia, ecc. (cfr. Jones, 2005).

sangiovanese odierno. A questo proposito si rivela interessante il confronto con una parlata dove tale distinzione è invece in crisi: nel Vallader, varietà di romancio della Bassa Engadina, è stato rilevato che il Centro di Gravità di [c] (4965 Hz) e di [ɲ] (4992 Hz) supera – anche se solo di poco – quello di [tʃ] (4855 Hz) e di [dʒ] (4649 Hz); notiamo che tali valori medi, calcolati in base alle realizzazioni di tre parlanti femminili (v. Schmid, in stampa), vanno contro la consueta correlazione tra il CdG e l'asse antero-posteriore delle articolazioni linguali.

Le occlusive palatali del sangiovanese si distinguono da quelle del Vallader anche per un'altro aspetto: quello temporale. Ricordiamo che nei nostri quattro parlanti il rapporto medio tra la durata del rilascio e la durata totale (%ril) è di 0.58/0.52 per [tʃ]/[dʒ] contro 0.44/0.38 per [c]/[ɲ]; in altre parole, nelle palato-alveolari la fase di rilascio dura più della metà della durata totale, mentre nelle palatali questo rapporto è nettamente inferiore al 50%. Ora, presso le tre parlanti engadinesi il rapporto medio tra la durata del rilascio e la durata totale (%ril) è di 0.46/0.40 per [tʃ]/[dʒ] contro 0.48/0.41 per [c]/[ɲ] (Schmid, in stampa); ciò significa che nel Vallader le palatali si differenziano dalle palato-alveolari nemmeno sul piano temporale. Diversamente, nel sangiovanese possiamo distinguere in modo chiaro non soltanto tra ostruenti palatali e palato-alveolari, ma anche tra occlusive e affricate.

A questo punto si impone una riflessione di carattere più generale sulla distinzione tra affricate e occlusive. Com'è noto, riguardo alla fase di rilascio delle occlusive esistono delle tendenze universali secondo il luogo di articolazione, attribuibili in ultima analisi alla velocità del movimento eseguito dall'articolatore attivo. Dal campione di 18 lingue studiate da Cho & Ladefoged (1997: 221) si evince una chiara gerarchia, per cui in linea di massima il VOT delle alveolari è superiore a quello delle bilabiali, e soprattutto il VOT delle velari è superiore a quello delle alveolari; purtroppo, il campione di Cho & Ladefoged (1997) non contiene alcuna lingua con occlusive palatali. Considerando che il palato duro costituisce l'articolatore passivo con la maggiore distanza dal rispettivo articolatore attivo, possiamo comunque classificare dei foni con un valore %ril attorno a 0.40 come occlusive.

Ricordiamo infine che l'ispezione spettrografica delle occlusive palatali del dialetto di San Giovanni in Fiore ha permesso di individuare in questo tipo di consonante un unico gesto articolatorio (§ 5.1). Se la presente indagine strumentale ha permesso di consolidare l'analisi monofonematica già emersa in una precedente descrizione fonetica-fonologica del sangiovanese (Mele, in stampa), rimane ancora da verificare l'esatta natura fonetica degli esiti di CL- e PL- in altre varietà, ad esempio nel napoletano (cfr. § 3.1). Auspichiamo che future ricerche sperimentali possano contribuire ad approfondire le nostre conoscenze di questo interessante aspetto fonetico dei dialetti italiani meridionali.

## RINGRAZIAMENTI

Teniamo a ringraziare i quattro locutori Antonio Mele, Antonio Bilotta, Giovanni Coviello e Giovanni Ambrosio per la loro disponibilità ed il loro interesse per la nostra ricerca.

## BIBLIOGRAFIA

- Barbato, M. (2005), *Turpiter Barbarizant*. Gli esiti di cons. + l nei dialetti italiani meridionali e in napoletano antico, *Revue de linguistique romane*, 69, 405-436.
- Bertinetto, P.M. & Loporcaro, M. (2005), The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome, *Journal of the International Phonetic Association*, 35, 131-151.
- Boersma, P & Weenink, D. (2008), *Praat: doing phonetics by computer*. [www.praat.org](http://www.praat.org).
- Celata, C. (2002), Fonetica della palatalizzazione delle velari in romanzo, *Quaderni del Laboratorio di Linguistica*, 3 (nuova serie), 119-138.
- Cho, T. & Ladefoged, P. (1999), Variation and Universals in VOT: evidence from 18 languages, *Journal of Phonetics*, 27, 207-229.
- Falcone, G. (1976), *Calabria* (Profilo dei dialetti 18), Pisa: Pacini.
- Forrest, K., Weismer, G., Milenkovic, P. & R. Dougall (1988), Statistical analysis of word-initial word-final voiceless obstruents: Preliminary data, *Journal of the Acoustical Society of America*, 84, 115-123.
- Giannini, A. & Pettorino, M. (1992), *La fonetica sperimentale*, Napoli: Edizioni Scientifiche Italiane.
- Gordon, M., Barthmaier, P. & Sands, K. (2002), A cross-linguistic acoustic study of voiceless fricatives, *Journal of the International Phonetic Association*, 32, 141-174.
- Haiman, J. & Benincà, P. (1992), *The Rheto-romance languages*, London & New York: Routledge.
- Jones, M. (2005), An experimental acoustic study of dental and interdental non-sibilant fricatives in the speech of a single speaker, *Cambridge Occasional Papers in Linguistics*, 2, 109-121.
- Keating, P. & Lahiri, A. (1993), Fronted velars, palatalized velars, and palatals, *Phonetica*, 50, 73-101.
- Ladefoged, P. & Maddieson, I. (1996), *The sound of the world's languages*, Oxford: Blackwell.
- Loporcaro, M. (1988), *Grammatica storica del dialetto di Altamura*, Pisa: Giardini.
- Loporcaro, M. (1991), The natural phonological process V[+high] → [+tense] and the vowel systems of some Southern Italian dialects, *Folia linguistica*, 25, 459-489.
- Loporcaro, M. (1995), Raddoppiamento fonosintattico dopo III persone plurali del verbo nei dialetti di Conflenti (CZ) e di San Giovanni in Fiore (CS), *Rendiconti dell'Accademia dei Lincei*, s. X, 6, 543-553.
- Loporcaro, M. & Mancuso, A. (1997), Interdentale ma anche laterale: /l/ prevocalica in alcuni dialetti della (Pre)Sila, in *Unità fonetiche e fonologiche: produzione e percezione*. Atti delle 8° giornate di studio del Gruppo di Fonetica Sperimentale (P.M. Bertinetto & L. Cioni, editors), Pisa: Scuola Normale Superiore, 77-90.

- Loporcaro, M. & Mele, B. (2004), Substance (sub)segmentale et changement phonologique: diphtongaisons et monophthongaisons dans deux dialectes de l'Italie du sud, in *Nouveaux départs en phonologie. Les conceptions sub- et suprasegmentales* (T. Meisenburg & M. Selig, editors), Tübingen: Gunter Narr, 89-109.
- Maddieson, I. (1984), *Patterns of sounds*, Cambridge: Cambridge University Press.
- Maddieson, I. (1997), Phonetic Universals, in *The Handbook of Phonetics Sciences* (W. Hardcastle & J. Laver, editors), Oxford: Blackwell, 617-639.
- Maddieson, I. & Precoda, K. (1991), Updating UPSID, *UCLA Working Papers in Phonetics*, 74, 104-114.
- Marotta, G. & Soriano, P. (1992) Lenizione e sonorizzazione nell'italiano cosentino: un'analisi sperimentale, *L'Italia Dialettale*, 55, 65-122.
- Mele, B. (2004), La realizzazione di /a/ tonica in contesto nasale nel dialetto di San Giovanni in Fiore: un caso di variazione diastratica, in *Percorsi incrociati. Studi di letteratura e linguistica italiana*. Atti del Dies Romanicus Turicensis (F. Broggi, G. Nicoli, L. Pescia, T. Stein, editors), Leonforte: Insula, 107-114.
- Mele, B. (in stampa), *Fonetica e fonologia del dialetto di San Giovanni in Fiore*, Tübingen: Francke.
- Repetti, L. & Tuttle, E. (1987), The evolution of latin PL, BL, FL, and CL, GL in western Romance, *Studi Mediolatini e Volgari*, 33, 53-115.
- Rohlf, G. (1949), *Historische Grammatik der italienischen Sprache und ihrer Mundarten. I. Lautlehre*, Bern: Francke (trad.it. 1966, *Grammatica storica dell'italiano e dei suoi dialetti. Fonetica*, Torino: Einaudi).
- Romano, A. (1999), A phonetic study of a Sallentinia variety (southern Italy), in *Proceedings of the XIV International Congress of Phonetic Sciences* (J. Ohala, editor), San Francisco: University of Berkeley, 1051-1054.
- Romano, A. (2007), La fonetica sperimentale e gli atlanti linguistici: la sintesi romanza di 'pidocchio' e lo studio degli esiti palatali, in *Temas de dialectología* (J. Dorta, editor), La Laguna-Tenerife: Instituto de Estudios Canarios, 179-204.
- Romano, A., Molino, G. & Rivoira, M. (2005), Caratteristiche acustiche e articolatorie delle occlusive palatali: alcuni esempi da dialetti del Piemonte e di altre aree italo-romanze, in *Misura dei parametri. Aspetti tecnologici ed implicazioni nei modelli linguistici*. Atti del 1° Convegno Nazionale AISV (P. Cosi, editor), Brescia: EDK Editore, 389-428.
- Romito, L. & De Rito, C. (2002), Analisi elettroacustica e aerodinamica della sequenza consonantica (-)NC- nel dialetto di Catanzaro: prenasalizzazione, articolazione complessa, secondaria o nesso?, in *La fonetica acustica come strumento di analisi della variazione linguistica in Italia*. Atti delle XII Giornate di Studio del GFS (A. Regnicoli, editor), Roma: Il Calamo, 91-100.
- Schmid, S. (in stampa), Les occlusives palatales du Vallader, in *Actes du XXVe Congrès International de Linguistique et Philologie Romanes* (P. Danler, editor), Tübingen: Niemeyer.



- Sussmann, H., Hoemeke, K. & Ahmed, E. (1987), A cross-linguistic investigation of locus equation as a phonetic descriptor for place of articulation, *Journal of the Acoustical Society of America*, 94, 1256-1268.
- Szende, T. (1999), Hungarian, in *Handbook of the International Phonetic Association*, Cambridge: Cambridge University Press, 104-107.
- Sziptár, P. & Törkenczy, M. (2000), *The Phonology of Hungarian*, Oxford: Oxford University Press.
- Taggart, G. (1990), *Dictionnaire du vocabulaire fondamental: romanche vallader – français et français – romanche vallader*, Coira: Lia Rumantscha.
- Trumper, J. (1997), Calabria and southern Basilicata, in *The dialects of Italy* (M. Maiden & M. Parry, editors), London & New York: Routledge, 355-364.
- Tuttle, E. (1975), The development of PL-, BL-, and FL- in Italo-Romance: distinctive features and geolinguistic patterns, *Revue de linguistique romane*, 39, 400-431.
- Tuttle, E. (1997), Palatalization, in *The dialects of Italy* (M. Maiden & M. Parry, editors), London & New York: Routledge, 26-31.
- Valente, V. (1975), *Puglia* (Profilo dei dialetti italiani 15), Pisa: Pacini.
- Wagner, M.L. (1951), *Lingua sarda*, Basel: Francke.

# NUOVI DATI FONETICI E DIALETTOLOGICI AI MARGINI DELL'AREA LAUSBERG: LE VARIETÀ DI ALIANO E ALIANELLO<sup>1</sup>

Francesco Avolio<sup>1</sup> & Antonio Romano<sup>2</sup>

<sup>1</sup>Dip. di Storia e metodologie comparate - Università dell'Aquila;

<sup>2</sup>Dip. di Scienze del Linguaggio - Università di Torino

*francesco.avolio@cc.univaq.it, antonio.romano@unito.it*

## 1. INTRODUZIONE

La pur vasta letteratura sulle condizioni dialettali della zona calabro-lucana nota agli studiosi come “area Lausberg” non ha finora portato a conclusioni univoche (cfr. Bianchi *et alii* 2002: 761-762). Si tratta o no di un'area “arcaica” e isolata, che mostra oggi condizioni un tempo proprie di altre parti della Romania? Il suo vocalismo tonico è realmente di tipo sardo? E quali sono le aree ad essa più affini? Si trovano all'interno o all'esterno della regione amministrativa di appartenenza? Quali sono le condizioni di variazione e le dinamiche (socio)linguistiche oggi in atto?

Il contributo che qui presentiamo parte dalla convinzione che proprio un primo tentativo di risposta all'ultimo quesito consente di porre le premesse per rispondere anche agli altri. Ciò che seguita finora ad essere scarsamente disponibile è infatti la descrizione di concrete situazioni sociolinguistiche di almeno alcuni dei centri dell'area, la quale potrebbe invece fornire chiarimenti preziosi, e spiegare anche parte delle oscillazioni e delle discrepanze registrate dai vari specialisti nelle stesse località.

La nostra scelta si è rivolta su un'area ai margini settentrionali della *Mittelzone* di Lausberg, e cioè sul comune di Aliano (Matera), che, pur trovandosi in una posizione linguisticamente isolata (ormai a ridosso della vasta zona bareseggiante del Materano), mostra, però, tutti o quasi tutti i tratti caratterizzanti di quella parte della *Mittelzone* priva di dittonghi metafonetici<sup>2</sup>.

Tra le ragioni di questa scelta troviamo il fatto che, malgrado la sua notorietà letteraria (è il *Gagliano* del noto romanzo *Cristo si è fermato a Eboli* di Carlo Levi, che ha reso nota all'Italia e al mondo la condizione contadina dell'Italia meridionale), il paese resta tra le località meno descritte a livello scientifico (se si eccettuano le pochissime, ancorché preziose, informazioni fornite da Lüdtke, 1979). Inoltre, all'interno del suo territorio comunale, e in particolare tra il centro e la sua frazione di Alianello (a meno di 5 km a S, verso l'Agri, la cui parte antica è stata purtroppo evacuata per ragioni di sicurezza dopo il terremoto del 23 novembre 1980, ed è oggi un paese fantasma, del tutto inagibile come la non lontana Craco), sono state rilevate notevoli differenze dialettali, a tutti i livelli, al punto che le due comunità rappresentano, insieme, come si vedrà, una sorta di miniatura dell'intera zona<sup>3</sup>.

---

<sup>1</sup> Pur trattandosi di una ricerca programmata e svolta congiuntamente dai due autori, la lettura dei dati e la redazione finale qui presentata sono da intendersi in base alla seguente attribuzione delle distinte sezioni: §§ 1, 2 e 6 a entrambi gli autori; §3 a FA; §§ 4 e 5 (e Appendici 1 e 2) ad AR.

<sup>2</sup> È insomma quella che Martino (1991: 15) chiama *Mittelzone severior*.

<sup>3</sup> Si noti, poi, che Alianello è divenuta frazione di Aliano solo nel 1861 (cfr. Mattatelli 1999: 58) e che entrambi i centri sono sempre appartenuti alla diocesi di Tricarico (Matera),

I nostri dati si basano sulle registrazioni personalmente effettuate ad Aliano e ad Alianello nuova nell'estate 2007 e integrano le registrazioni già disponibili (per Aliano), in seguito alle indagini dell'estate 2004 (oggetto di analisi preliminare in Avolio & Romano, 2006)<sup>4</sup>.

Il contributo s'incentra sull'analisi di diversi fatti, dal vocalismo tonico alla metaforesi, dalla conservazione di -S e -T nella flessione verbale ad altri fenomeni morfologici e sintattici (come, ad Aliano, il possessivo posposto, ma non enclitico) e a qualche tratto lessicale. L'intera analisi, però, è condotta in modo da far emergere, oltre alle differenze diatopiche tra Aliano ed Alianello, le differenze diastratiche e diagenazionali interne alle due località, consentendo così di scorgere alcune delle dinamiche in atto (ad esempio, ad Aliano, la parziale perdita di -S, ma non di -T, e la progressiva espansione, in particolare nell'italiano dei giovani, dell'isocronia di sillaba tonica di tipo pugliese e in generale medio- e basso-adriatica, verso la quale Alianello mostra invece una ben maggiore resistenza) e di verificare la loro presenza nelle varietà basse del repertorio italiano. Uno spazio adeguato viene poi garantito allo studio della "percezione" e dello "status" del dialetto tra i parlanti stessi. Anche qui, il confronto con altre zone permette di articolare meglio la disamina di questo punto cruciale della questione rianalizzata oggi con nuove impostazioni teoriche (cfr. ad esempio Trumper, 1997), spesso avvalorate dalla possibilità di confronti diacronici (come proposto in Romito *et alii*, 2006) rivolti a fornire nuovi dati per riflettere sulle reali condizioni che possono avere agito in passato (e/o agiscono ancora oggi).

Dal canto nostro ci siamo proposti di procedere a descrizioni più accurate (meno normalizzanti) di situazioni attuali e di variabili fin qui trascurate (almeno per alcuni dei centri dell'area) sedimentate in quelle oscillazioni che possono essere all'origine delle discrepanze registrate dai vari specialisti nelle stesse località.

Per questo motivo abbiamo proceduto alla raccolta di una grande quantità di dati, analizzati poi, separatamente, in base a metodi complementari.

L'osservazione meticolosa, condotta con le valutazioni e le classificazioni rigorose del metodo dialettologico tradizionale, è stata infatti opportunamente integrata e completata da approfondimenti fonetici di caratteristiche segmentali e prosodiche, ottenuti con l'ausilio di metodi di analisi acustica (cfr. Rivoira & Romano, 2003, e Romano & Manco, 2004, per le prime, e Romano, 2003, per le seconde).

## 2. MATERIALI ANALIZZATI

La sezione di analisi fonetica e lessicale si basa su un corpus di produzioni ottenute col metodo dell'osservazione "partecipante"<sup>5</sup>. In molti casi si tratta di conversazioni libere condotte sulla base di un questionario specifico domanda/risposta di più di 400 *items*

---

anziché a quella di Anglona-Tursi (Matera), che invece comprende tutti i paesi della *Mittelzone* di Lausberg (annoverando fino a tempi recenti anche quelli del versante calabrese, poi aggregati alla diocesi di Cassano allo Jonio, Cosenza).

<sup>4</sup> Per la raccolta dei dati siamo riconoscenti a tutti gli informatori dei due paesi, che hanno mostrato la loro disponibilità verso di noi e nei confronti di questa ricerca; siamo debitori in particolar modo al sig. Giuseppe Salvatore D'Angelo per la sua accoglienza e per la quantità di mezzi e informazioni che ha messo a nostra completa disposizione.

<sup>5</sup> Per questa sezione, i dati raccolti nel 2007 sono stati registrati sul duplice supporto, analogico e digitale, offerto da un SONY TCS-470 (con microfono stereofonico incorporato) e da un sistema di acquisizione sonora integrato su un PC portatile TOSHIBA.

rivolto a 10 informatori di Aliano (di cui due giovani, nel 2004 FS27, PDA37, e otto anziani o di media età, nel 2004 GM89, GC79, PG78, FM71, LT70, RMDL70, VS58, nel 2007 PDA40, GC84) e 3 informatrici di Alianello di età diversa (una sessantenne, MM65, nel 2004, un'anziana, MM78, e una quarantenne, MR42, nel 2007)<sup>6</sup>.

A questa si aggiunge la sezione di analisi intonativa condotta con un questionario *AMPER* (basato su un'elicitazione per immagini e/o per ricostruzione di contesti, v. *Géolinguistique*, 2005). I locutori in questo caso sono due uomini PDA40 e GC84 (Aliano) e due donne MR42 e MM78 (Alianello). Sono stati elicitati (quasi esclusivamente dai due parlanti più giovani) più di 180 enunciati (in genere 3 ripetizioni per ogni frase), su un questionario di 60 frasi. Gli enunciati raccolti soddisfano a vari requisiti, tra i quali soprattutto quelli di avere la stessa struttura sintattica, le stesse modalità intonative e di focalizzazione, lo stesso numero di sillabe e le stesse posizioni accentuali nella struttura sillabico-lessicale della frase.

Un primo *corpus* introduttivo (mirato a far familiarizzare i locutori con il tipo di *performance* richiesto) si basa produttivamente su espansioni aggettivali (*exp.*) dei sintagmi nominali delle 18 frasi dialettali derivanti dalle frasi di partenza seguenti<sup>7</sup>.

*Il pilota vola su Tàranto./?* (senza *exp.*)

*Il pilota vola su Matera./?* (senza *exp.*)

*Il pilota vola su Paternò./?* (senza *exp.*)

*Donato si beve il caffè./?* (*exp.* in SN2 → scaldato, tórbido)

*Donato si beve il latte./?* (*exp.* in SN2 → scaldato, tórbido)

Il *corpus* sperimentale vero e proprio si basa invece su espansioni dei sintagmi verbali delle 42 frasi derivanti dalle frasi di partenza seguenti (*exp.* SN2 → càtara, curata, Paranà)<sup>8</sup>:

*La pàpera (si) mangia la fràgola./?*

*La pàpera (si) mangia la patata./?*

*La pàpera (si) mangia la maracà./?*

### 3. ANALISI DIALETTOLOGICA

#### 3.1. Principali differenze tra i due centri

**3.1.1. Vocalismo tonico.** La ricerca ha mostrato, fuor di metaforesi, condizioni ancora piuttosto chiaramente “sarde” ad Aliano, con poche eccezioni, anche se il grado di apertura delle vocali medie è diverso da quello, sicuramente aperto, di paesi vicini come Senise (cfr. Fanciullo 1988: 676 e Bianchi *et alii* 2002: 759): a volte si osserva, infatti, la confluenza di E/Ē e di O/Ō in un suono intermedio o tendenzialmente chiuso (ma cfr. §§ 3.1.2. e 4). Chiara invece la situazione delle vocali estreme, in cui confluiscono quasi sempre I/Ī, Ū/Ū̄. Da Ī si ha *illĕ* ‘lui’ ma anche ‘lei’, *linguĕ* ‘lingua’, *ricchiĕ* ‘orecchia’, *sitĕ* ‘sete’, *vivĕ* ‘bere’, ma *péscĕ* ‘pesce’; da Ū si ha *cĕpullĕ* ‘cipolla’, *muschĕ* ‘mosca’, *tussĕ* ‘tosse’, *vucchĕ* ‘bocca’ (e

<sup>6</sup> I risultati dell'analisi segmentale condotta col metodo strumentale sono fondati al momento sull'analisi acustica dei timbri di 120 vocali accentate (di cui 30 osservate dinamicamente) per la sola varietà di Aliano (loc. GC84).

<sup>7</sup> Del *corpus* introduttivo sono stati analizzati 44+22 *file* per Aliano e 39+20 *file* per Alianello.

<sup>8</sup> Del *corpus* sperimentale sono stati al momento analizzati 55 *file* per Aliano e 57 *file* per Alianello.

*surdë* ‘sorda’ oltre che ‘sordo’); da E abbiamo *sérë* ‘sera’, *stéllë* ‘stella’ (come *pérë* ‘piede’), da O si ha *fiórë* ‘fiore’, *nëpótë* ‘nipote’, *sówë* ‘sole’ (come *vóglië* ‘voglio’) ecc.

Sono state confermate anche le condizioni ‘balcaniche’ del *Vorposten* (l’‘avamposto’ di Lausberg) ad Alianello (dove Ī ha un esito diverso da I, mentre Ū ed Ŭ si fondono): *jéllë* ‘lei’, *lénchë* ‘lingua’, *névë* ‘neve’, ma *muschë* ‘mosca’, *vucchë* ‘bocca’.

**3.1.2. Metafonesi.** Il materiale raccolto consente di avanzare l’ipotesi che Aliano sia stata investita, per quanto riguarda la metafonesi delle medio-basse (*è, ò* derivanti dalle brevi latine), anche da correnti di tipo ‘sabino’<sup>9</sup>, del resto ben attestate in altri punti della Lucania, meridionale e non (ad esempio a Castelsaraceno, PZ, cfr. Lüdtké 1979: 58), ma qui, evidentemente, ‘disturbate’ dal progressivo conguaglio vocalico<sup>10</sup>: troviamo dunque vocali toniche ‘medie’ o tendenzialmente chiuse (*ajnellë* ‘agnello’, *celë* ‘cielo’, *perë* o *pérë* ‘piedi’, e poi *fóchë* ‘fuoco’, *oglië* ‘olio’, *nu prëcochë* ‘una pesca’, e, d’altro canto, in alcune forme non metafonetiche, *còrë* ‘cuore’, *còssë* ‘coscia, gamba’, con un’evidente vocale medio-bassa, cfr. §4).

Ad Alianello abbiamo invece chiari dittonghi ‘napoletani’ ascendenti (*ajnjéllë* ‘agnello’, *cjélë* ‘cielo’, *pjédë* ‘piedi’ e ancora *fwóchë* ‘fuoco’, *wóglië* ‘olio’, *pwórcë* ‘maiali’) ecc. Il nome del paese suona poi (*gh*)*alianjéllë*, rispetto ad *alianellë* del dialetto di Aliano). Netto è dunque il contrasto tra le due parlate, ben percepito dagli stessi abitanti, compresi i giovani. Esso, inoltre, ripropone il non risolto problema del rapporto, anche diacronico, tra i diversi tipi di metafonesi delle medio-basse, che però, per ragioni di spazio, non possiamo affrontare in questa sede; rinviando quindi, in prima istanza, a Castellani (1973: 82-83, che propende, non senza valide argomentazioni, per una maggiore antichità del dittongamento) e ad Avolio (1995: 37, che riassume lo *status* della situazione).

**3.1.3. Velarizzazione della laterale scempia intervocalica.** Il tratto è presente ad Aliano (*awìwë* ‘olive’ arc., *rasùwë* ‘radimadia’, *sówë* ‘sole’, *strittëwë* ‘vicoli’ ecc.), di nuovo in accordo con diversi centri della *Mittelzone*, anche sul versante calabrese (Tursi, Valsinni, Mt, Noepoli, PZ, in Calabria Oriolo, CS), ma non ad Alianello (*ulivë* ‘olive’, *rasólë* ‘radimadia’, *sólë* ‘sole’, *stréttë* ‘vicoli’ ecc.).

**3.1.4. Forme del possessivo con termini di parentela e affinità.** Ad Aliano, di nuovo in accordo con paesi posti più a Sud, oltre l’Agri, si ha il possessivo posposto, ma non enclitico, alla latina (*fratë méjë* ‘mio fratello’, *frattùwë* ‘*fratë tuë* ‘tuo fratello’, *marëtë méjë* ‘mio marito’, *sorë tuë* ‘tua sorella’), ad Alianello solo quello enclitico (*fràtëmë* ‘mio fratello’, *maritëmë* ‘mio marito’, *sòrètë* ‘tua sorella’ ecc.). Anche qui, dunque, il contrasto non potrebbe essere più netto.

**3.1.5. Conservazione di -s e -t nella flessione verbale.** Mentre ad Alianello non c’è alcuna traccia di -t, e sono stati raccolti esempi residui di -s nelle forme di seconda e quinta

<sup>9</sup> Cioè la chiusura è ‘é e ò ‘ó, nelle consuete condizioni.

<sup>10</sup> Non è forse inutile ricordare, infatti, che l’«origine monogenetica [della metafonesi], in seno alla compagine (neo)latina, è infinitamente più probabile della sua poligenesi. Tale postulato implica che la metafonesi romanza sia entrata in vigore prima del conguaglio quantitativo delle vocali latine» (Lüdtké, 1979: 57).

persona dell'indicativo imperfetto di diversi verbi<sup>11</sup> (*appëniësë* 'appendevi', *vuliësë*, *vulisë* 'volevi')<sup>12</sup>, ad Aliano la conservazione, soprattutto di -T, è invece piuttosto accurata, al punto che, a causa del suo isolamento geolinguistico per quanto riguarda tale tratto, quest'ultimo si rivela uno dei più usati dagli abitanti dei paesi vicini in funzione canzonatoria (*Luiggësë/Luicësë* 'Luigi', *patànësë* 'patate'). Più in dettaglio, le nostre inchieste con i giovani alianesi (FS27, PDA37) hanno permesso di inquadrare il meccanismo della perdita di -S nelle forme della seconda persona singolare del presente indicativo, mostrando che tale suono da un lato è oggi caduto nei verbi della cosiddetta 'classe irregolare' (che, oltre agli ausiliari, include i verbi *andare*, *dare*, *fare*, *stare*, *potere*, *sapere*, *volere*; cfr. Lüdtke, 1979: 72, Avolio 2006: 340) – in molti dei quali l'intera forma è stata ristrutturata sulla base di quelle 'contratte' sviluppatesi già nel latino volgare (e, non dimentichiamolo, con precisi riscontri nei dialetti italo-romanzi limitrofi e in italiano) –, dall'altro si mantiene nei verbi regolari che invece hanno meno sofferto la concorrenza di altre varianti. Ne deriva una sorta di opposizione, via via più evidente, tra forme attualmente monosillabiche (ma che in passato potevano essere parossitone o anche proparossitone) senza -S – in cui peraltro la sillaba tonica, anche originaria, è sempre aperta – e forme proparossitone con -S conservata tanto con la sillaba tonica aperta quanto con la chiusa. Il confronto con la terza persona singolare, in cui -T, come si è detto, mostra maggiore resistenza di -S (andando in senso contrario a quanto si è osservato ad Alianello e si osserva in altre aree di lingua romanza), aiuterà a comprendere meglio la dinamica in atto:

| Caduta di - s                                                   | Conservazione di - s                                   |
|-----------------------------------------------------------------|--------------------------------------------------------|
| <i>daj</i> ≠ <i>dasë</i> 'dài', ma <i>datë</i> 'dà'             | <i>canòscësë</i> 'conosci', <i>canòscëtë</i> 'conosce' |
| <i>faj</i> ≠ <i>fasë</i> 'fai', ma <i>fatë</i> 'fa'             | <i>càntësë</i> 'canti', <i>càntëtë</i> 'canta'         |
| <i>poj</i> ≠ <i>pojsë</i> 'puoi', ma <i>pojtë</i> 'può'         | <i>dórmësë</i> 'dormi', <i>dòrmëtë</i> 'dorme'         |
| <i>saj</i> ≠ <i>sàpësë</i> 'sai', ma <i>sàpëtë</i> 'sa'         | <i>màngësë</i> 'mangi', <i>màngëtë</i> 'mangia'        |
| <i>staj</i> ≠ <i>stasë</i> 'stai', ma <i>statë</i> 'sta'        | <i>mòrësë</i> 'muori', <i>mòrëtë</i> 'muore'           |
| <i>vaj</i> ≠ <i>và(dë)së</i> 'vai', ma <i>vàtë</i> 'va'         | <i>vénësë</i> , ma anche <i>vënë</i> , 'vieni'         |
| <i>voj</i> ≠ <i>vosë</i> 'vuoi', ma <i>votë</i> 'vuole'         | <i>vénñësë</i> 'vendi', <i>vénñëtë</i> 'vende'         |
| <i>te</i> ≠ <i>tenësë</i> 'tieni', ma <i>ténëtë</i> 'tiene, ha' | <i>vidësë</i> 'vedi', <i>vidëtë</i> 'vede'             |

<sup>11</sup> La situazione del paese si mostra dunque parzialmente diversa, e più conservativa di quella finora segnalata nella letteratura scientifica. Sulla base dei dati di Rohlf, Lausberg, Bigalke ecc., era infatti emerso che Alianello conoscesse la conservazione di -S solamente nella seconda e quinta persona dell'indicativo imperfetto di *essere* (cfr. Vårvaro 1984a: 119-121, e carta 1).

<sup>12</sup> Mattatelli (1999: 150-151) lo conferma: *zumbase* 'saltavi' *zumbiësëve* 'saltavate', *vediësë* 'vedevi' *vediësëve* 'vedevate', *viviësë* 'bevevi' *viviësëve* 'bevevate', *veniësë* 'venivi' *veniësëve* 'venivate', accanto a *iërse* *iërseve* 'eri eravate', *aviësë* *aviësëve* 'avevi avevate' ecc. (rispettiamo la grafia dell'autore).

### 3.2. Italiano regionale

Si era già a conoscenza del fatto che l'area Lausberg fosse stata raggiunta, e attraversata, dall'ondata della differenziazione vocalica per posizione (DVP), o isocronia di sillaba tonica (cfr. Fanciullo 1988: 679). Oggi possiamo constatare tanto che – a differenza di quanto accade, ad esempio, con la metaforesi – la DVP non viene bloccata né nel passaggio dalla sezione dialettale a quella italiana del repertorio (cfr. Carosella, 2005, per le complesse condizioni del Gargano), né, soprattutto (vista la zona in cui ci troviamo), in quello inverso, quanto l'ormai evidente affermarsi della DVP fra gli alianesi giovani e anche di media età (*pènnà* 'penna', ma *béne* 'bene': *mòsca* 'mosca', ma *bródo* 'brodo' ecc.); il dialetto degli anziani, però, come anche le loro varianti di italiano, ne resta ancora sostanzialmente immune<sup>13</sup>. Alianello, a parte poche occorrenze isolate (ad esempio *pëccënènnë* 'bambina', che si oppone a *piccënninnë* 'bambino', in MM82) e qualche raro caso di iniziale frangimento (*u pastò"rë* 'il pastore'), con riscontri anche negli anziani di Aliano (*u mëntó"në* 'il montone' GC84; cfr. § 4), mostra una maggiore resistenza, con il mantenimento di condizioni sostanzialmente lucane occidentali, cioè, in definitiva, campane, prive di DVP, tanto nel dialetto, quanto, anche, in italiano (con qualche cedimento nell'informatrice più giovane). Piuttosto netto è quindi l'influsso, ad Aliano, del capoluogo provinciale, Matera, e degli altri centri su cui il paese in qualche modo gravita, tutti con varianti di DVP (Stigliano, in minor misura Policoro). Si tratta forse del più cospicuo cambiamento linguistico oggi in atto in questo territorio.

### 4. SPOGLIO STRUMENTALE DEI DATI SEGMENTALI

Dall'osservazione dei risultati delle misurazioni e delle trascrizioni proposte per ciascun timbro considerato è emerso un sistema con qualità vocaliche molto variabili, ma riferibili abbastanza chiaramente a un sistema a sette vocoidi (mascherato dalla notazione a cinque vocali comunemente data nelle descrizioni locali; cfr. ad esempio, Mattatelli 1999, 2005, per altri aspetti, come si è visto, ricche di utili informazioni)<sup>14</sup>.

La classificazione dei timbri su base descrittiva si rivela tuttavia non esente da difficoltà per la ricorrente presenza di oscillazioni (nelle produzioni di uno stesso locutore!), di soluzioni intermedie e di altre caratteristiche che determinano una certa indecidibilità nelle classificazioni fonologiche. L'ausilio offerto dalle possibilità obiettive di rappresentazione acustica, condotte col metodo *Voweltrack* (già collaudato in altre occasioni, cfr. ad es. Romano & Manco, 2004), non si rivela decisivo (si veda l'Appendice 1).

<sup>13</sup> Si veda l'esempio di *pòchë* 'poco' del loc. GC84. Lo stesso parlante, come del resto la signora MM82 di Alianello, ha prodotto *u fròndë* 'la fronte' (v. §4), che però potrebbe anche derivare da una base con Ò, analoga a quella che ha generato lo spagnolo *frente* 'frente'.

<sup>14</sup> Una grammaticalizzazione di queste varietà, condotta in riferimento a un italiano scolastico, può indurre a una semplificazione in tal senso. Ovviamente di altra origine sono invece le scelte di Mancarella (1988, 1994 e 2005) per la parlata della vicina Tursi, per la quale i testi del raccoglitore locale R. Campese sono riprodotti senza distinzione di qualità per i timbri medi e sono affiancati a nuovi dati che, pur presentati come caratterizzati dalla presenza di più fini distinzioni di apertura nei timbri vocalici, ricevono una rappresentazione in un sistema a soli tre gradi di apertura a causa di una lettura condotta in una chiave filologica e dialettologica diversa.

La prima caratteristica che richiamiamo per le realizzazioni vocaliche di Aliano (come per altre località dell'area d'influenza pugliese) è la notevole instabilità dei timbri.

Per le vocali alte il fenomeno è meno evidente: le realizzazioni di /i/ e di /u/ sono limitate a una minima variazione, per quanto in un'area allargata che da [i] si estende molto frequentemente a [ɪ] (o forse meglio [j]) e da [u] si estende a [ʊ] (o forse meglio [ɯ]), soprattutto per le realizzazioni brevi.

Degne di menzione sono però anche le condizioni di realizzazione delle vocali medio-alte /e/ e /o/, le cui realizzazioni sconfinano spesso nelle aree di [ɪ] e [ʊ] (con derive di tipo [eɪ] [ou]) oppure provengono da posizioni piuttosto medie o addirittura medio-basse – in funzione del tipo sillabico (ma forse non senza eccezioni) – o talvolta centralizzate.

Per /o/ può accadere che, forse anche in funzione del condizionamento consonantico, la direzione di deriva sia diversa (come in *polmónē* [mpl] dove si registra piuttosto un tipo [ʊo] vs. *pelósē* [fsg] dove si ha invece un tipo [əʊ], v. Appendice 1). Un altro tipo di deriva particolarmente suggestivo per /o/ (o /ɔ/) è quella che si sviluppa a partire da posizioni più centrali, come in *pastórē* (o *mēndonē*, v. Appendice 1)<sup>15</sup>.

Altrettanto notevoli sono le rese delle vocali medio-alte o medio-basse anteriori con realizzazioni che si concentrano spesso in un'area media, con sviluppo da posizioni più basse per un certo numero di esempi alcuni dei quali giustificabili solo su base posizionale: *ainellē* [msg], *vètēlle* [msg], *passarēllē* [fsg], *vēcchjē* [fsg].

Anche i timbri medio-bassi posteriori, pur essendo più stabili, possono localizzarsi in una regione media (come accade per *grössē* e *voj*) oppure essere più spiccatamente medio-bassi (di tipo [ɔ] oppure [ɔ̞], come *ricòttē* o *paròwē*) con una certa centralizzazione. Il tipo sillabico influisce in un certo numero di casi in cui fa apparire una [ɔ] centralizzata invece di una delle rese di /o/ in continuazione di *Ō* (come in *fròndē*, ma v. §3), ma ad esempio sembra non incidere in alcun modo in altri (come *còrē* o *pòchē*) in cui la resa è più chiaramente medio-bassa.

Un'ultima considerazione riguarda l'attesa palatalizzazione per la vocale centrale aperta. Seppur con una certa dipendenza dal contesto consonantico (minore anteriorizzazione e tendenziale minor apertura in contesti velari o labiali), la resa di /a/ è di solito affidata a timbri stabili di tipo anteriore ([a̟], [a̠] o [æ] come in *vacchē* o *lattē*, o persino [ɛ], in condizioni di allungamento, come in *capē*) oppure soggetti a una deriva più o meno 'localizzata' (come in *cambanē* o *subbressatē*).

Date tutte queste difficoltà di classificazione, si può comprendere la ragione per la quale, dal confronto tra i numerosi studi condotti in quest'area, emergano apparenti difformità e incongruenze. Ci si può chiedere anzi come mai una discussione preliminare sulle dimensioni di variazione e sulle difficoltà di rappresentazione qui evidenziate non

---

<sup>15</sup> La nostra scelta di notazione delle vocali accentate dei due esempi è semplificatoria: mentre cerca di non tener conto di condizionamenti 'etimologici', attribuisce invece arbitrariamente maggiore peso al timbro prototipico più vicino a quello più 'duraturo' delle due vocali accentate di queste parole (rispettivamente di tipo [əo] e [əɔ]). Ci chiediamo come avrebbero trascritto questi suoni altri raccoglitori e se le scelte operate in simili situazioni dai vari autori che hanno descritto altre varietà di quest'area non condizionino troppo pesantemente i tentativi di ricostruzione diacronica.



venga affrontata, con la necessaria attenzione, almeno nelle descrizioni di taglio più scientifico<sup>16</sup>.

## 5. SPOGLIO DEI DATI SOPRASEGMENTALI

Quanto alle caratteristiche intonative analizzate per le due varietà, quelle che qui presentiamo riassumono le caratteristiche degli schemi con il maggior numero di ripetizioni coerenti (cfr. Appendice 2).

In effetti, il metodo di lavoro definito nell'ambito del progetto *AMPER* si basa sull'analisi di diverse ripetizioni della stessa struttura e sul confronto tra realizzazioni giudicate come il risultato di una stessa strategia prosodica a parità di condizioni (modalità, focus, struttura informativa generale, accenti d'insistenza e sfumature interpretative)<sup>17</sup>.

Le Figg. 1a e 1b permettono di apprezzare i profili medi stilizzati di  $f_0$  per due frasi corrispondenti nelle varietà di Aliano e Alianello (per ogni frase del corpus simili andamenti sono disponibili anche per gli altri due parametri, durata ed energia).

Dal confronto tra gli esempi illustrati possiamo osservare delle caratteristiche che ci permettono di anticipare la presenza di due diverse strategie nella realizzazione dei profili finali interrogativi.

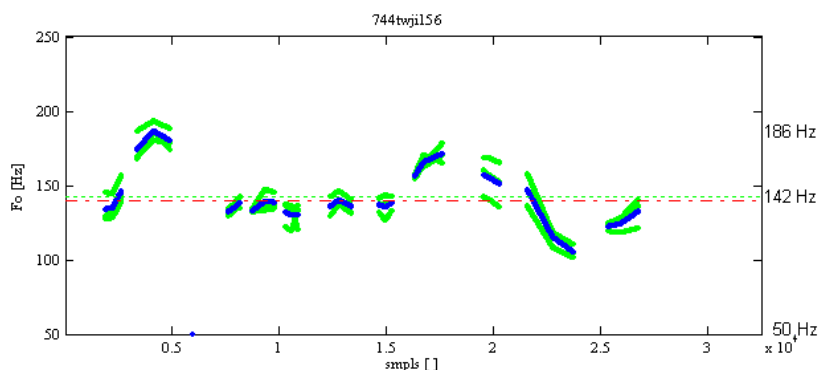


Fig. 1a. Profili stilizzati medi di tre ripetizioni dell'enunciato interrogativo *Dunàtē sē vīvētē 'u cafè mbucàtē?* [dɔ'nat(ə) sə 'vivət u ka'fɛ mbu'katə] "Donato si beve il caffè scaldato?" pronunciato dal locutore PDA40 di Aliano (si noti la caduta della terza vocale nelle rip. qui considerate).

<sup>16</sup> Una discussione sulle diverse difficoltà che impediscono di dare descrizioni tali da permettere ricostruzioni univoche è in Bianchi *et alii* (2002: 760 e segg.). La dimensione e l'ordine di grandezza dei problemi da noi sollevati sono però altri e potrebbero rivelarsi sostanziali anche in termini di descrizione sincronica.

<sup>17</sup> Le curve stilizzate o le sequenze di valori ottenute come andamenti medi sono quelle su cui si basano l'analisi e le valutazioni uditive. Il giudizio di uniformità espresso dal raccogliatore esperto o da un informatore fidato e imparziale è la prova dell'esatta allotonia degli andamenti delle ripetizioni su cui si basa l'estrazione del modello medio (a sua volta verificato con l'ascolto della riproduzione sintetica della prosodia ad esso associata, v. *Géolinguistique*, 2005) e offre il vantaggio di ridurre, nelle valutazioni dei fenomeni più sistematici, l'influenza di fattori di variazione funzionalmente ininfluenti, l'incidenza di variabili spurie o, più spesso, la presa in conto di meri errori di misura.

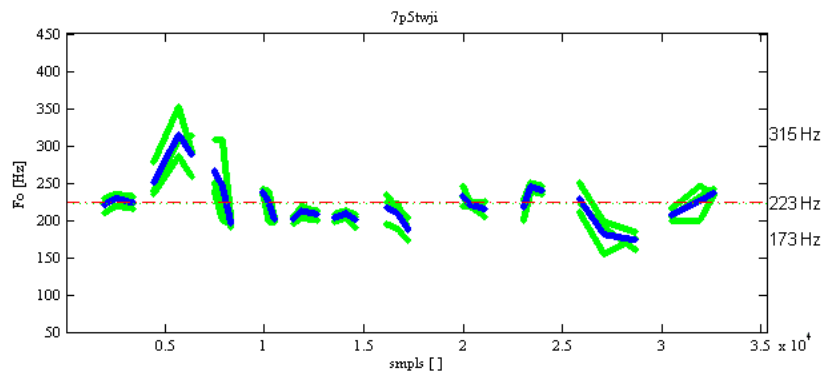


Fig. 1b. Profili stilizzati medi di tre ripetizioni dell'enunciato interrogativo *Donatè sè vèvè 'u caf(f)è mbucàtè?* [dɔ'natə sə 'vevə\_o kə'f(:)ɛ mbu'katə] "Donato si beve il caffè scaldato?" pronunciato dalla locutrice MR42 di Alianello.

Mentre ad Alianello è la vocale interessata dall'accento di frase che sembra attrarre movimenti di  $f_0$  maggiormente degni di rilievo (discendenti, con picco pretonico e, nel caso dell'esempio, risalita finale), nelle produzioni del locutore di Aliano un profilo accentuale nettamente più rilevante è presente in protonia (anche in modalità dichiarativa, v. dopo).

Un confronto interessante è quello tra gli andamenti di frasi senza espansione e di frasi con espansioni, oppure tra quelli di frasi con espansioni differenti (v. Appendice 2).

Il confronto tra modalità si realizza invece sovrapponendo gli andamenti stilizzati medi risultanti e osservando il profilo generale e le posizioni in cui le curve si accostano o si discostano maggiormente (v. Figg. 2a e 2b).

È anche da questi che emerge una diversa strategia nell'opposizione tra le due modalità che rende maggiormente notevole il rilievo melodico sulla penultima vocale accentata della frase per la varietà di Aliano che determina una divaricazione più marcata delle due curve nelle sillabe successive.

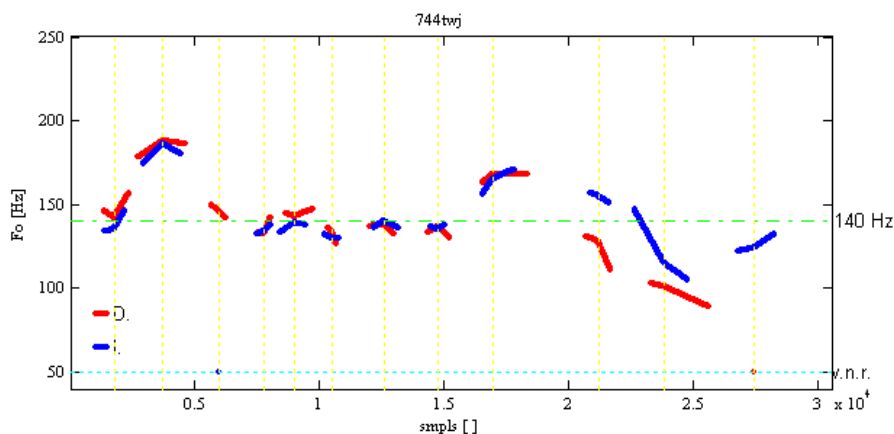


Fig. 2a. Confronto tra profili medi stilizzati delle curve di  $f_0$  alle due modalità (dichiarativa, D, e interrogativa, I) per la frase *Dunatè sè vèvè 'u cafè mbucàtè./?* (v. Fig. 1a) nelle

produzioni del locutore PDA40 di Aliano (si noti la caduta della terza vocale nelle rip. qui considerate).

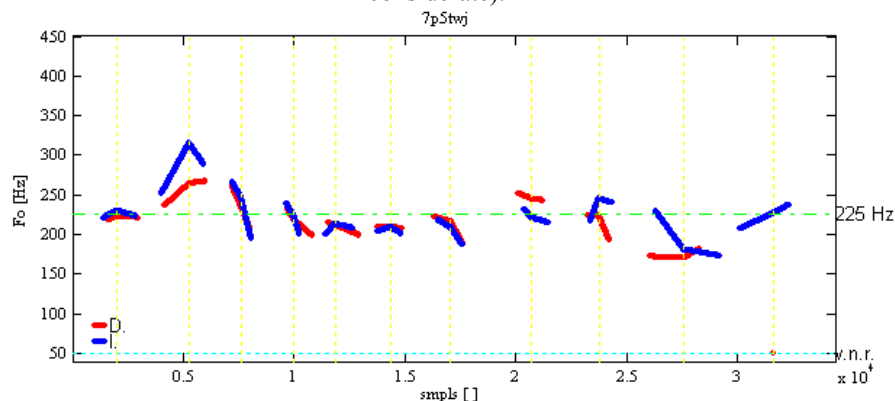


Fig. 2b. Confronto tra profili medi stilizzati delle curve di  $f_0$  alle due modalità (dichiarativa, D, e interrogativa, I) per la frase *Donàtë sè vévë 'u caf(f)è mbucàtë?* (v. Fig. 2a) nelle produzioni della locutrice MR42 di Alianello.

In seguito alla stilizzazione e alla prototipizzazione (derivante dai vari confronti tra le curve stilizzate delle strutture con gli elementi prominenti nelle diverse posizioni), i profili intonativi caratteristici delle due modalità per le frasi senza espansione nella varietà di Aliano possono essere schematizzati come in Fig. 3a. La Fig. 3b mostra invece le conseguenze sui due profili schematici nel momento in cui si presenta una seconda struttura accentuale come espansione del sintagma verbale.

Per Aliano possiamo confermare le osservazioni preliminari in Avolio & Romano (2006: 106), secondo cui le configurazioni intonative globali presentano schemi interrogativi poco discosti dai corrispondenti dichiarativi, con andamenti declinanti e rilievi melodici in corrispondenza dei due macrosintagmi intonativi presenti nelle frasi. La distinzione di modalità sembra affidata prevalentemente alla velocità locale dei movimenti melodici (pendenze più accentuate, profili con maggiore velocità su vocali più brevi) e a una diversa organizzazione temporale che porta alla riduzione e alla tendenza alla cancellazione di vocali inaccentate (ma non quella dell'articolo postverbale, v. dopo).

Il primo picco accentuale risulta di solito allineato piuttosto a metà della prima vocale accentata (con valori in genere superiori di 6 o 7 semitoni (sT) rispetto alla frequenza d'attacco). Il *Contorno Terminale di Modalità* (CTM) delle domande totali qui considerate si presenta sempre con un rilievo melodico in prossimità dell'ultima sillaba accentata (a lungo predisposto sulle sillabe precedenti) sulla quale si svolge un profilo discendente più alto e più veloce che nella dichiarativa (in cui si manifesta come declinante)<sup>18</sup>. Nel caso di ossitoni finali la risalita finale (di 4 o 5 sT) è talvolta compressa all'interno della durata della vocale accentata (in tal caso la vocale è quasi doppia) oppure è solo accennata oppure ancora è totalmente omessa (v. dopo). Essendo associata a un profilo di energia nettamente declinante, può presentare anche un breve innesco di discesa terminale.

<sup>18</sup> In media lo scarto di  $f_0$  sull'ultima vocale accentata è pari a  $97 \pm 13 \text{ Hz/s}$  alla modalità dichiarativa ( $\Delta/D_{\sigma ND}$ ) e a  $313 \pm 42 \text{ Hz/s}$  alla modalità interrogativa ( $\Delta/D_{\sigma NI}$ ).

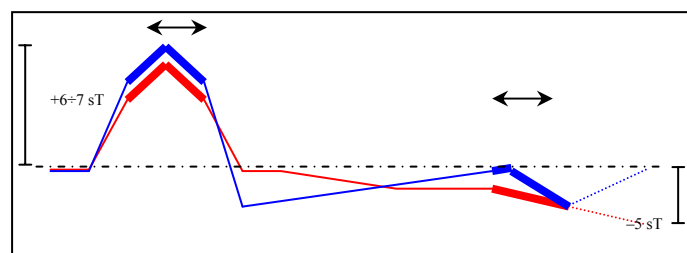


Fig. 3a. Confronto tra gli schemi intonativi prototipici delle due modalità (dichiarativa, D, e interrogativa, I) nelle frasi senza espansione prodotte dal locutore di Aliano.

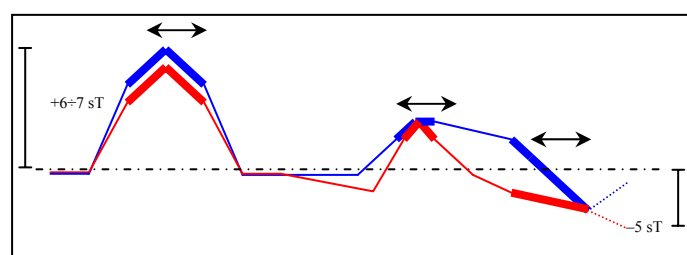


Fig. 3b. Confronto tra gli schemi intonativi prototipici delle due modalità (D e I, v. Fig. 3a) nelle frasi con un'espansione nel sintagma verbale prodotte dal locutore di Aliano.

Allo stesso modo, per Alianello, i confronti tra le curve stilizzate delle diverse strutture permettono di definire i profili intonativi caratteristici schematizzati come in Fig. 4a (per le frasi senza espansioni) e in Fig. 4b (per le frasi con un'espansione del sintagma verbale).

Per Alianello, le osservazioni preliminari ci portano a registrare configurazioni intonative globali abbastanza distinte per le due modalità: gli schemi dichiarativi si presentano con andamenti declinanti e, in assenza di espansioni, con un solo rilievo melodico significativo; gli schemi interrogativi presentano invece due rilievi, il secondo dei quali declinante, con un rallentamento progressivo, lungo la durata dell'ultima vocale accentata.

Rispetto al *CTM* delle interrogative di Aliano, quello di Alianello, si caratterizza per un maggior rilievo melodico (1 o 2 sT sopra la media) subito prima dell'ultima sillaba accentata (repentinamente dopo l'ultima sillaba non accentata precedente) da cui poi il profilo discende veloce (più che nella dichiarativa)<sup>19</sup>. La risalita finale (di massimo 3 sT) è seguita da un accenno di discesa che viene troncato se la vocale accentata è in posizione finale assoluta (pur essendo in questi casi la vocale raddoppiata)<sup>20</sup>.

<sup>19</sup> Anche in queste produzioni (come in quelle del locutore di Aliano)  $\Delta/D_{\sigma ND}$  risulta in media pari a  $63 \pm 8 \text{ Hz/s}$  in opposizione ai  $311 \pm 56 \text{ Hz/s}$  di  $\Delta/D_{\sigma NI}$ .

<sup>20</sup> Per assicurare lo sviluppo melodico tipico del *CTM* sugli ultimi elementi non accentati, anche in assenza di questi (nel caso di finale ossitona), i parlanti fanno ricorso a un altro tipo di 'propagginazione', una sorta di paragoge di vocale identica all'ultima vocale. Ad Alianello, nel segmento paragogico si manifesta la risalita finale (ma non l'accenno di discesa che, pertanto, consideriamo troncato). Anche se potrebbe trattarsi di caratteristiche stilistiche individuali (entrambi i parlanti della stessa località presentano però grosso modo le stesse tendenze, distinte da quelle degli altri dell'altra località), osserviamo che una

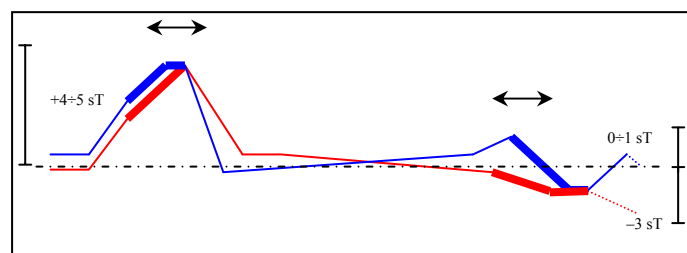


Fig. 4a. Confronto tra gli schemi intonativi prototipici delle due modalità (D e I, v. Fig. 3a) nelle frasi senza espansione prodotte dalla locutrice di Alianello.

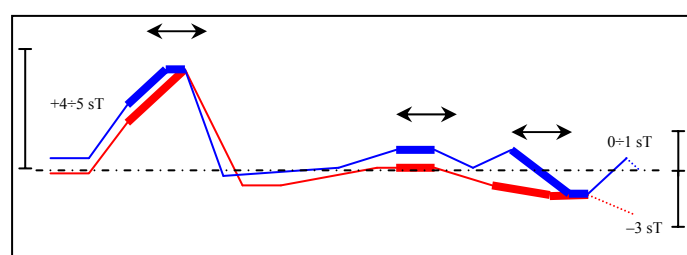


Fig. 4b. Confronto tra gli schemi intonativi prototipici delle due modalità (D e I, v. Fig. 3a) nelle frasi con un'espansione nel sintagma verbale prodotte dalla locutrice di Alianello.

Il contrasto così individuato, unitamente all'ascesa finale preservata e alla tendenza al troncamento (dell'accenno di discesa), richiama per Alianello alcune caratteristiche del napoletano da noi già ritrovate per la varietà di Ravello (Avolio & Romano, 2006: 107).

Quanto al primo picco accentuale (che, a parità di condizioni, presenta di solito massimi contenuti, 4 o 5 sT), notiamo invece un certo ritardo rispetto ai dati di Aliano: la vocale accentata risulta infatti totalmente in ascesa per entrambe le modalità con un leggero anticipo nel caso della domanda.

Un'ultima osservazione riguarda la tendenza alla preservazione della vocale dell'articolo postverbale e al mantenimento di una maggiore lunghezza di questa rispetto alle vocali precedente e seguente (v. vocale n° 7 negli istogrammi di durata in Appendice 2). Della maggiore significatività dell'allungamento di questa vocale nei dati di Aliano (vs. Alianello) testimoniano anche le barre d'errore assolute ( $\pm$  dev. standard) visibili sugli istogrammi.

## 6. CONCLUSIONI

Che dire di fronte a questi dati? Forse «continuare a discettare se il vocalismo della *Mittelzone* sia più siciliano o più sardo può isterilire la questione» (Fanciullo, 1988: 678). Se da un lato, sulla scia di Fanciullo, i dati di Aliano e Alianello confermano, per il vocalismo tonico, che «tipologicamente, il confronto più immediato è tra il vocalismo della *Mittelzone* e il vocalismo napoletano» (*ivi*), dall'altro, come vedremo, è l'intera

---

paragoge con troncamento si presenta nell'88% dei profili interrogativi finali di Alianello, mentre una paragoge (per un *CTM* finale discendente-ascendente) è presente nel 28% di quelli di Aliano. Nelle produzioni del parlante giovane di questa località, il troncamento presente in tutti gli altri casi presenta solo il tratto discendente del *CTM*.

problematica linguistica dell'area Lausberg che andrebbe sottratta alla rigida contrapposizione *sardo/non sardo*, o *arcaico/non arcaico*, per essere meglio inserita nel quadro delle dinamiche storiche che hanno portato alla formazione dei diversi tipi dialettali dell'Italia centro-meridionale. La sorprendente somiglianza, a ogni livello, fra la parlata di Alianello e il napoletano, ad esempio, riscontrabile un po' in tutta la media Val d'Agri, ma sensibilmente di meno in zone lucane e campane più prossime alla città partenopea (come la stessa alta Val d'Agri ed il Vallo di Diano), attende ancora una compiuta spiegazione: non possiamo più accontentarci, infatti, di parlare genericamente di 'influsso napoletano', più o meno recente e superficiale. Esistono relazioni più profonde e antiche, che forse oggi riusciamo finalmente a intravedere.

Sulla base di questi dati, quindi, le conclusioni, sia pur provvisorie, portano a inquadrare quest'appendice dell'area Lausberg come la parte che meglio mostra, in sincronia, gli sviluppi romanzi che, in diacronia, hanno portato alla formazione del tipo linguistico napoletano. Tali sviluppi appaiono però oggi non *in vitro*, ma intrecciati a tendenze di diversa origine (soprattutto pugliese) e, ovviamente, all'interferenza con la sezione italiana del repertorio.

Sempre a proposito del vocalismo, una questione rilevante che abbiamo sollevato riguarda la difficoltà di classificazione dei timbri registrati per le vocali accentate che, soprattutto ad Aliano, si sono per lo più sottratte a una chiara logica distribuzionale su base esclusivamente diacronica, mostrando in molti casi una certa tendenza a situarsi in posizioni intermedie. Non si tratta quindi di sforzarsi di rapportare questi timbri a un sistema a cinque o a sette (con dittongamento o senza, metafonetico o no), quanto, piuttosto, di discutere delle dimensioni di variazione e delle difficoltà di rappresentazione che timbri di tal genere, spesso soggetti al frangimento, pongono sia al raccoglitore che al linguista storico.

## BIBLIOGRAFIA

- AMPER, v. *Géolinguistique* (2005). Projet AMPER - Atlas Multimédia Prosodique de l'Espace Roman, *Géolinguistique*, hors série 3 (éd. J.P. Lai).
- Avolio F. (1995). *Bommèsprə. Profilo linguistico dell'Italia centro-meridionale*. San Severo: Gerni.
- Avolio F. (2003). "Gli « indicatori geografici » come fonte per gli studi dialettali: esempi lucani". In M.T. Greco (a cura di), *Gli "indicatori geografici" della Basilicata nord-occidentale nei territori delle Comunità montane del Marmo e del Meandro*, Napoli: Università degli Studi "L'Orientale", 27-44 [*Quaderni degli Annali dell'Istituto Orientale di Napoli* (nuova serie - 6)].
- Avolio F. (2006). "Dialetofonia giovanile, questionari, 'nuove' e 'vecchie' attività: appunti di ricerca". In G. Marcato (a cura di), *Giovani, lingue e dialetti, Atti del convegno internazionale di studi* (Sappada, luglio 2005), Padova: Unipress, 331-344.
- Avolio F. & Romano A. (2006). "La prosodia degli enunciati dichiarativi e interrogativi in tre aree dialettali dell'Italia centro-meridionale (Abruzzo, Basilicata e Campania)". In R. Savy & C. Crocco (a cura di), *Atti del II congresso AISV* (Salerno, 30 nov.-2 dic. 2005), Padova: ISTC/EDK ed., 99-110.

- Bianchi P., De Blasi N., Fanciullo F. (2002). "La Basilicata". In M. Cortelazzo, C. Marcato, N. De Blasi & G.P. Clivio (a cura di), *I Dialetti Italiani. Storia Struttura Uso*, Torino: UTET, 755-792.
- Bigalke R. (1980). *Dizionario dialettale della Basilicata*. Heidelberg: Winter.
- Bigalke R. (1994). *Basilicatese*. München / Newcastle: Lincom Europa.
- Bigalke R. (1995). "Il ferrandinese nell'ambito dei dialetti lucani". In *Bollettino della Biblioteca Provinciale di Matera*, XVI, 25-26, 130-138.
- Canepari L. (2002). "Fonetica naturale". In A. Regnicoli (a cura di), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia (Atti delle XII Giornate di Studio del "Gruppo di Fonetica Sperimentale" dell'Associazione Italiana di Acustica, Macerata, 13-15 dicembre 2001)*, Roma: Il Calamo, 269-276.
- Caratù P. (1988). "Italienisch: Areallinguistik X. b) Südlukanien". In G. Holtus, M. Metzeltin, Schmitt Christian (a cura di): *Lexikon der Romanistischen Linguistik (LRL)*, vol. IV, *Italiano, Corso, Sardo*. Tübingen: Niemeyer, 688-694.
- Carosella M. (2005). *Sistemi vocalici tonici nell'area garganica settentrionale fra tensioni diatopiche e dinamiche variazionali*. Roma: Edizioni Nuova Cultura.
- Castellani A. (1973). *I più antichi testi italiani. Edizione e commento*. Bologna: Pàtron.
- Cerrato L. & Cutugno F. (1993). "Il problema della rappresentazione tempo/frequenza dei fenomeni vocalici dinamici". In E. Magno Caldognetto & F. Ferrero (a cura di), *Le vocali: dati sperimentali, problemi linguistici, applicazioni tecnologiche (Atti delle III Giornate di Studio del "Gruppo di Fonetica Sperimentale" dell'Associazione Italiana di Acustica, Padova 19-20 novembre 1992)*, Roma: Esagrafica, 61-71.
- De Blasi N., Di Giovine P. & Fanciullo F. (a cura di) (1991). *Le parlate lucane e la dialettologia italiana (Studi in memoria di Gerhard Rohlf)*. Galatina: Congedo.
- Fanciullo F. (1988). "Italienisch: Areallinguistik X. a) Lukanien". In G. Holtus, M. Metzeltin, Ch. Schmitt (a cura di), *Lexikon der Romanistischen Linguistik (LRL)*, vol. IV, *Italiano, Corso, Sardo*, Tübingen: Niemeyer, 669-688.
- Lausberg H. (1939). *Die Mundarten Südlukaniens*. Halle (Saale).
- Lüdtke H. (1979). *Lucania*. Pisa: Pacini ["Profilo dei dialetti italiani", 17].
- Mancarella G.B. (1988 o 1989). "Ricerche linguistiche a Tursi. Per una interpretazione del vocalismo tonico della Lucania meridionale". *Studi linguistici salentini*, 16, 3-114.
- Mancarella G.B. (1994). *Lessico dialettale di Tursi*. Lecce: Del Grifo.
- Mancarella G.B. (2005). *Il dialetto di Tursi*. Lecce: Del Grifo.
- Martino P. (1991). *L' "area Lausberg". Isolamento e arcaicità*. Roma: Il Calamo [Quaderni del Dipartimento di studi glottoantropologici dell'Università "La Sapienza"].
- Mattatelli F.P. (1999). *Alianello tra fatti e parole*. Matera: Grafiche Paternoster.
- Mattatelli F.P. (2005). *Aliano. Storia sociale e linguistica*. Aliano: Circolo culturale "N. Panevino".

- Parlangèli O. (1971). "Note di storia linguistica italiana (a proposito dell'area Lausberg)". In E. Coseriu e W. D. Stempel (a cura di), *Sprache und Geschichte. Festschrift für Harri Meier zum 65. Geburtstag*, München: Fink, 353-372.
- Rivoira M. & Romano A. (2003). "Analisi acustica del sistema vocalico del dialetto di Rorà (Val Pellice)". In G. Marcato (a cura di), *I dialetti e la montagna* (Atti del Convegno *Int. di studio*, Sappada/Plodn-Sauris, 2-6 luglio 2003), Padova: Unipress, 159-166.
- Rohlf G. (1966-69). *Grammatica storica della lingua italiana e dei suoi dialetti* (3 voll.). Torino: Einaudi.
- Romano A. (1999). "Analyse des structures prosodiques des dialectes et de l'italien régional parlés dans le Salento: approche linguistique et instrumentale". *Tesi di Dottorato Université de Grenoble III* (2 voll.) [Lille: Presses Univ. du Septentrion (1 vol.), 2001].
- Romano A. (2003). "Accento e intonazione in un'area di transizione del Salento centro-meridionale". In P. Radici Colace, G. Falcone, A. Zumbo (a cura di), *Storia politica e storia linguistica dell'Italia meridionale. Atti del convegno internazionale di studi parlangeliani* (Messina, 22-23 maggio 2000), Messina-Napoli: Ed. Scientifiche Italiane, 169-181.
- Romano A. & Manco F. (2004). "Incidenza di Fenomeni di Riduzione Vocalica nel Parlato Spontaneo a Bari e a Lecce". In F. Albano-Leoni, F. Cutugno, M. Pettorino & R. Savy (a cura di), *Il Parlato Italiano* (Atti del Convegno Naz. Napoli, 13-15 Febbraio 2003), Napoli: D'Auria (CD-ROM).
- Romito L., Galatà V., Lio R., Stillo F. (2006). "La metafonía nei dialetti dell'area Lausberg: un'introspezione sulla natura della sillaba". In R. Savy & C. Crocco (a cura di), *Atti del II congresso AISV* (Salerno, 30 nov.-2 dic. 2005), Padova, ISTC/EDK ed., 538-565.
- Savoia L.M. (1989). "Teoria generativa, modelli fonologici e dialettologia. La propagazione di u in una varietà lucana". In *Rivista Italiana di Dialettologia*, 11, 1987 [ma 1989], 185-263.
- Savoia L.M. (1997). "Il vocalismo a tre gradi dell'area calabro-lucana". In *Studi offerti a Gabriella Giacomelli dagli amici e dagli allievi* (Quaderni del Dipartimento di Linguistica dell'Università degli studi di Firenze). Padova: Unipress, 363-375.
- Trumper J. (1980). "La zona Lausberg ed il problema della frammentazione linguistica". In F. Albano Leoni (a cura di), *I dialetti e le lingue delle minoranze di fronte all'italiano, Atti dell'XI Congresso internazionale di studi della SLI* (Cagliari, 27-30 maggio 1977), Roma: Bulzoni, vol. 2, 267-303.
- Trumper J. (1997). "Calabria and southern Basilicata". In M. Maiden & M. Parry (a cura di), *The dialects of Italy*, Londra- New York: Routledge.
- Vàrvaro A. (1984a). "Sulla nozione di area isolata: il caso della Lucania". In Id., *La parola nel tempo. Lingua, società e storia*. Bologna: Il Mulino, 127-144.
- Vàrvaro A. (1984b). "Sui modelli della storia linguistica: caduta o conservazione di -S in Lucania". In Id., *La parola nel tempo. Lingua, società e storia*. Bologna: Il Mulino, 117-125.



## APPENDICE 1

In quest'appendice sono riprodotti alcuni diagrammi che rappresentano l'evoluzione del timbro della porzione centrale dei nuclei vocalici accentati presenti in 30 parole selezionate tra le produzioni del locutore GC84 (di Aliano, MT) estratte da produzioni ottenute nell'interazione relativa alla prima porzione del questionario dialettologico (relativa al lessico degli animali d'allevamento).

L'evoluzione timbrica (deriva) è raffigurata sulla base di misure formantiche effettuate in diversi punti del timbro e riportate sequenzialmente sul piano F1-F2 col metodo *Voweltrack*<sup>21</sup>.

Le realizzazioni osservate sono estratte da porzioni terminali (prepausali) delle produzioni (di solito piuttosto curate anche se in un registro vocale piuttosto grave e poco energetico): solo le traiettorie più stabili sono state selezionate; quelle qui proposte sono state accuratamente verificate anche con misure manuali puntuali.

I diagrammi offrono due sistemi di riferimento, presentati sottoforma di centroidi delle aree di dispersione ed ellissi di equiprobabilità (ottenute considerando in media l'85% dei punti). Entrambi i sistemi di riferimento si presentano con sette timbri. Il primo dei due è relativo a 15 realizzazioni di ogni vocale misurata puntualmente (nella porzione di maggiore stabilità) in produzioni (talvolta in italiano) di più certa classificazione dello stesso locutore GC84 e si presenta inscritta in un triangolo vocale complessivo ristretto, con centroidi ravvicinati tra loro e verso il centro e con ellissi più larghe e spesso visibilmente sovrapposte. Il secondo sistema di riferimento è relativo allo stesso numero di realizzazioni (complessivamente sorvegliate, con pronuncia forte scandita) da parte di un parlante standard (un giovane attore teatrale attivo nel torinese) e si presenta con un triangolo vocalico più esteso, con centroidi distanziati e con ellissi nettamente più concentrate. Rispetto a questo il parlante alianese presenta soprattutto vocali alte abbassate (ma raramente centralizzate), una /a/ più alta e avanzata, frequenti realizzazioni centralizzate soprattutto di /ε/.

---

<sup>21</sup> Ciascun nucleo vocalico è analizzato strumentalmente facendo ricorso a una rappresentazione sperimentale espressamente concepita per descrivere nuclei vocalici instabili. Tale tecnica, originariamente sviluppata in ambiente Matlab<sup>TM</sup>, presso il *Centre de Dialectologie de Grenoble*, è stata messa a punto e battezzata *Voweltrack* (Romano, 1999). La tecnica consiste nel riportare automaticamente su un diagramma F<sub>1</sub>-F<sub>2</sub> l'evoluzione temporale delle formanti del nucleo vocalico. Le formanti sono misurate sulla base di analisi automatiche LPC a 20 coeff. su una finestra da 256 campioni (62,5 ms) precedente, all'interno di un'estensione temporale individuata manualmente, con un passo di 40 campioni (2,5 ms). Il tracciamento delle posizioni successive del timbro misurato sul piano F<sub>1</sub>-F<sub>2</sub> avviene sulla base di due stime: una prima stima tracciata in verde, e una seconda stima (che esclude alcune misure devianti rispetto alla traiettoria media) rappresentata in rosso. Sulla necessità di questo tipo di rappresentazioni si veda Canepari (2002). Una discussione sui metodi di rappresentazione delle variabili acustiche legate all'evoluzione dinamica delle vocali è già in Cerrato & Cutugno (1993).

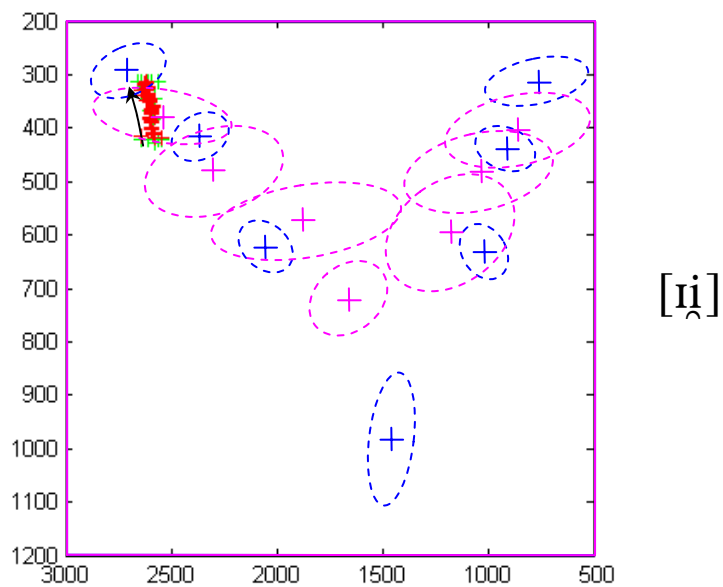


Figura A1. Diagramma *Voweltrack* del timbro della vocale accentata di *gallinē* ‘gallina’ (GC84, Aliano, MT).

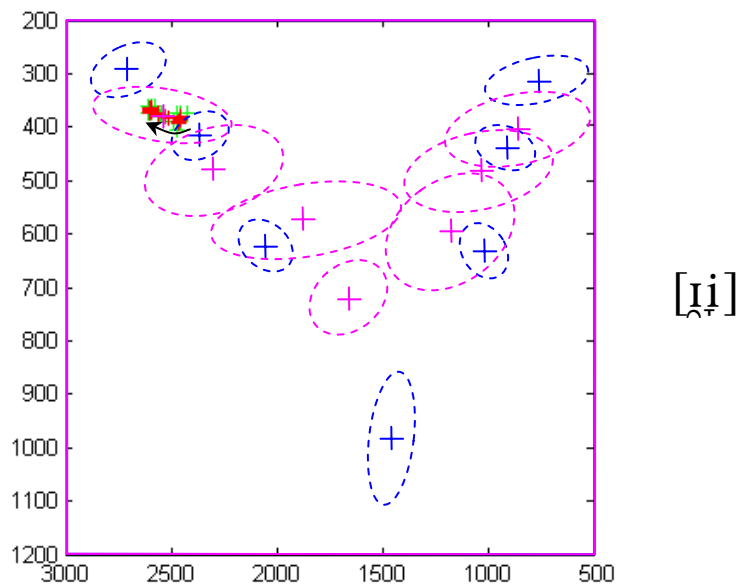


Figura A2. Diagramma *Voweltrack* del timbro della vocale accentata di *nunn esistinē* ‘non esistevano’ (GC84, Aliano, MT).

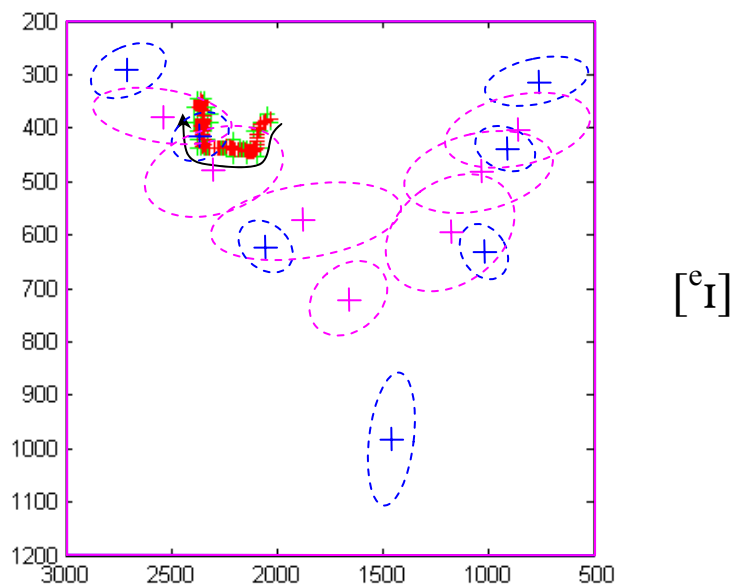


Figura A3. Diagramma *Voweltrack* del timbro della vocale accentata di 'a sauzizzë 'salsiccia' (GC84, Aliano, MT).

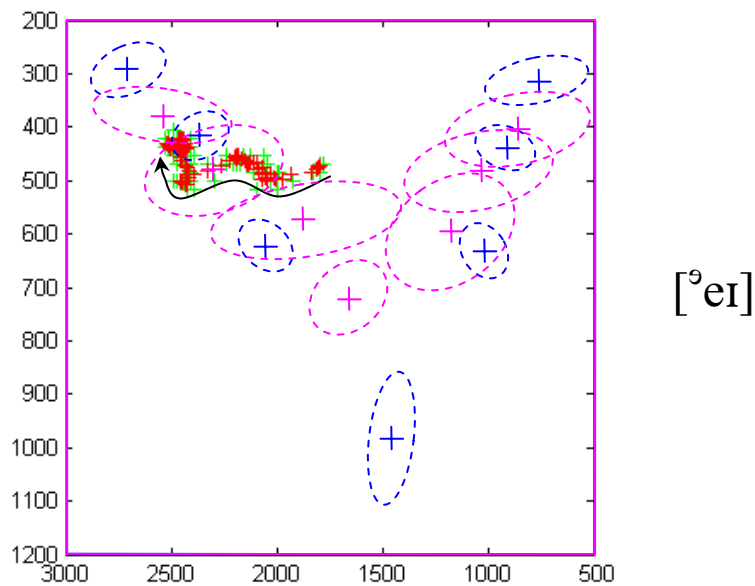


Figura A4. Diagramma *Voweltrack* del timbro della vocale accentata di *trapetë* 'gancio per il maiale (tirapiedi)' (GC84, Aliano, MT).

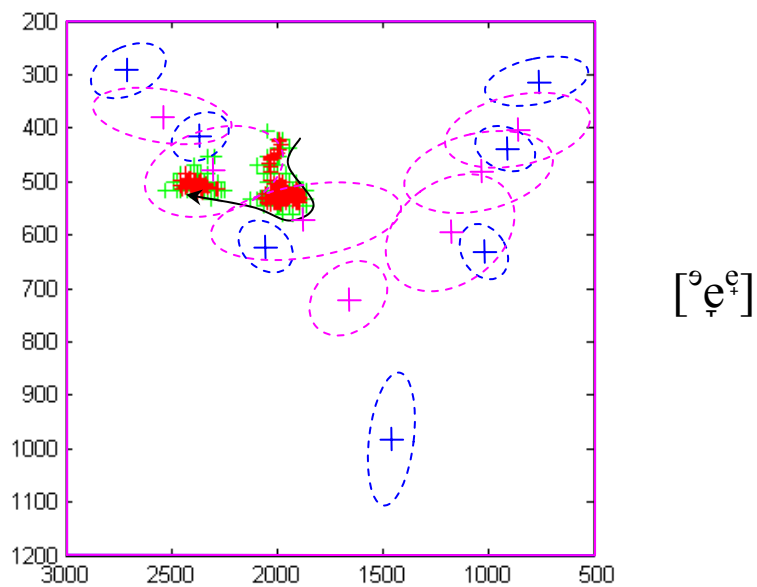


Figura A5. Diagramma *Voweltrack* del timbro della vocale accentata di 'u *passërellë* 'ː' (GC84, Aliano, MT).

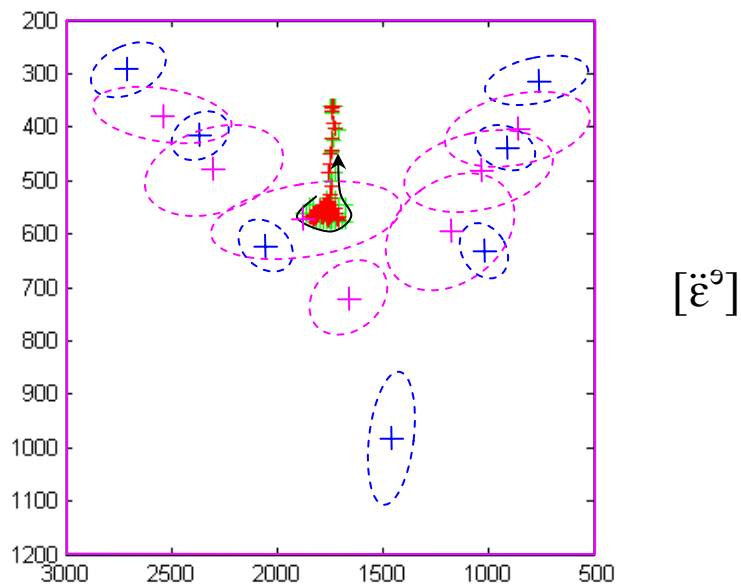


Figura A6. Diagramma *Voweltrack* del timbro della vocale accentata di 'a *passërellë* 'ː' (GC84, Aliano, MT).

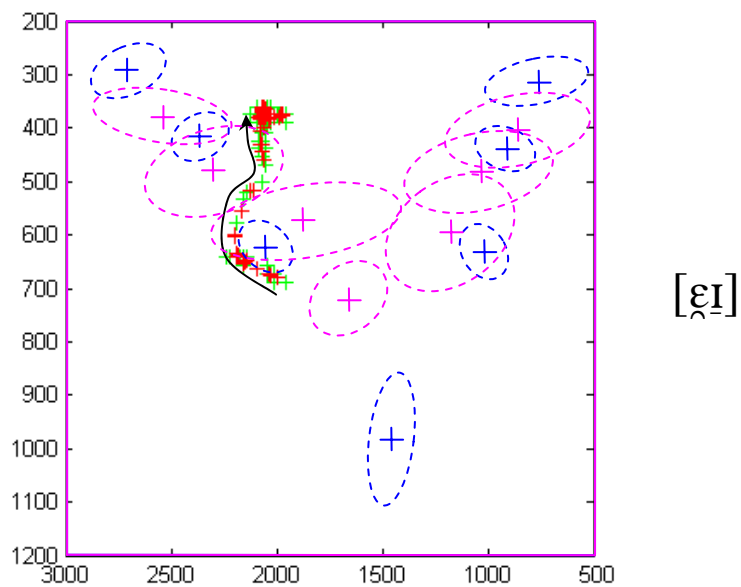


Figura A7. Diagramma *Voweltrack* del timbro della vocale accentata di *l'ainellè* 'l'agnello' (GC84, Aliano, MT).

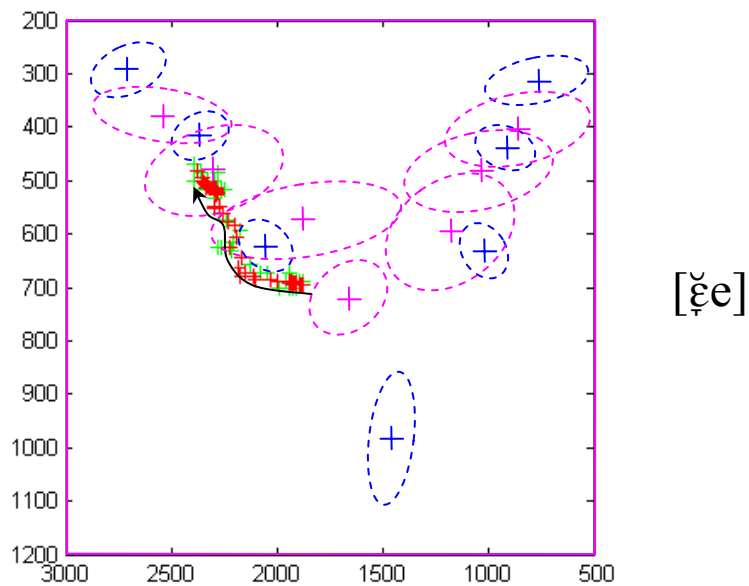


Figura A8. Diagramma *Voweltrack* del timbro della vocale accentata di un'altra realizzazione di *l'ainellè* 'l'agnello' (GC84, Aliano, MT).

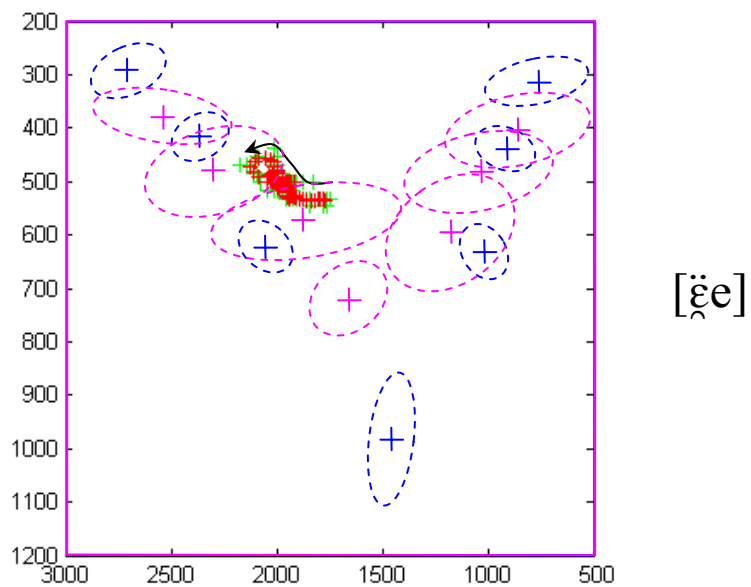


Figura A9. Diagramma *Voweltrack* del timbro della vocale accentata di *pécurë* 'pecore' (GC84, Aliano, MT).

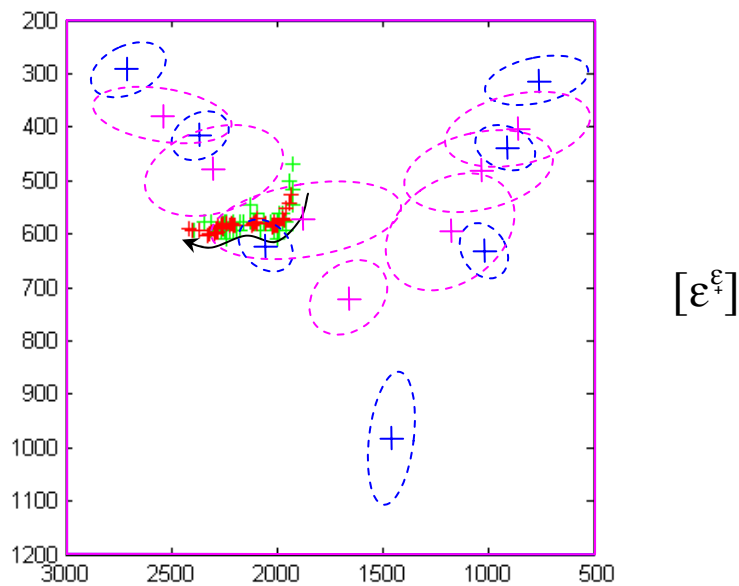


Figura A10. Diagramma *Voweltrack* del timbro della vocale accentata di *vetellë* 'vitello' (GC84, Aliano, MT).

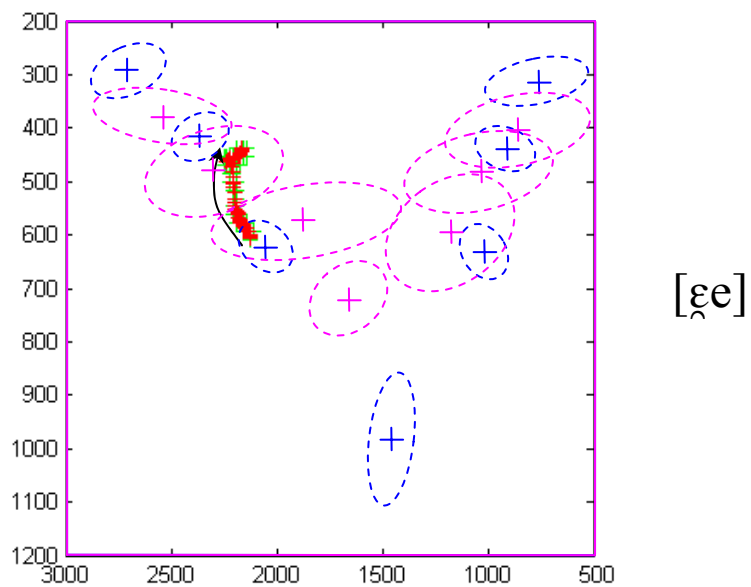


Figura A11. Diagramma *Voweltrack* del timbro della vocale accentata di 'nu *jencarellë* (*jengarellë*) 'un manzo' (GC84, Aliano, MT).

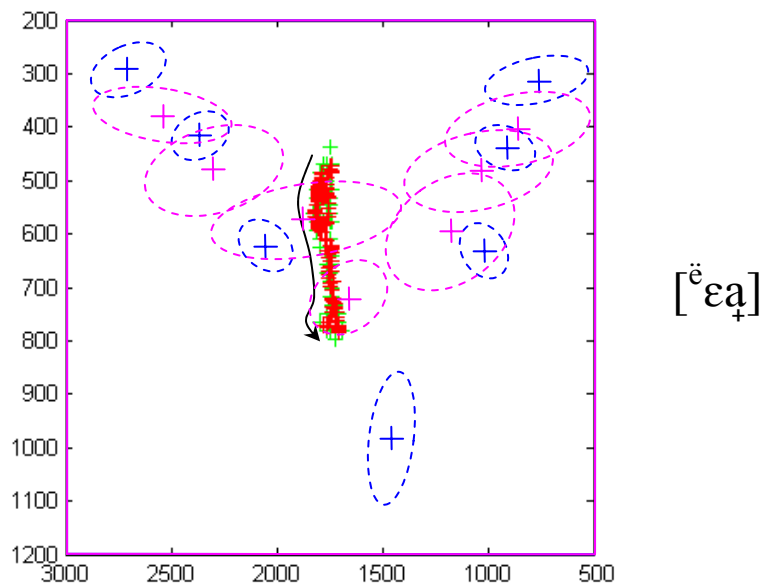


Figura A12. Diagramma *Voweltrack* del timbro della vocale accentata di 'a *subbrë(s)atë* 'la soppressa' (GC84, Aliano, MT).

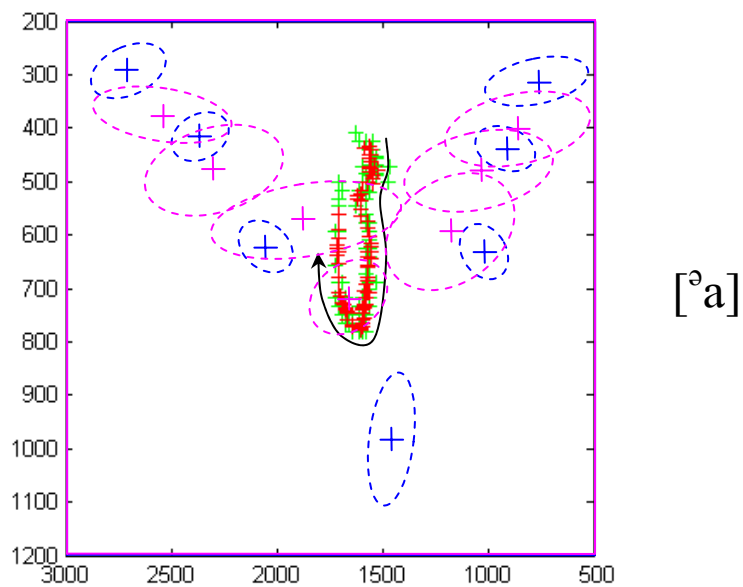


Figura A13. Diagramma *Voweltrack* del timbro della vocale accentata di 'u lardë 'il lardo' (GC84, Aliano, MT).

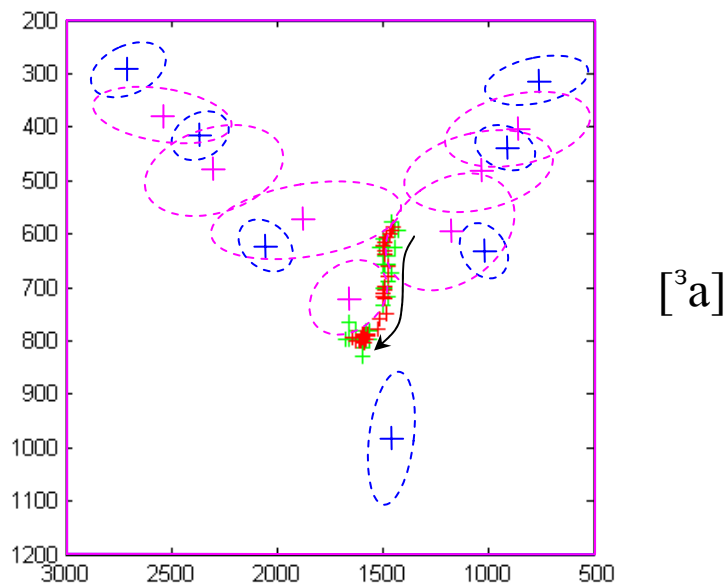


Figura A14. Diagramma *Voweltrack* del timbro della vocale accentata di *ttre vvacchë* 'tre vacche' (GC84, Aliano, MT).



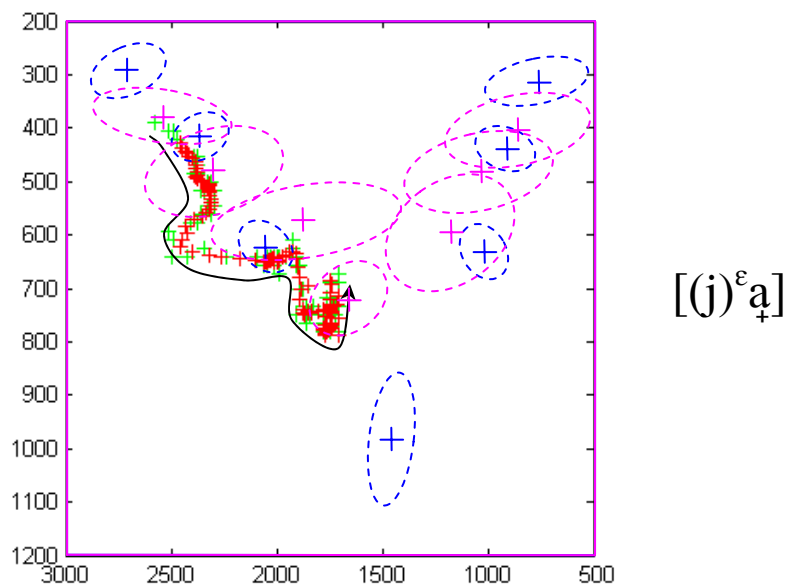


Figura A15. Diagramma *Voweltrack* del timbro della vocale accentata di 'nu (v)*jazzë* 'un ricovero per le pecore' (GC84, Aliano, MT).

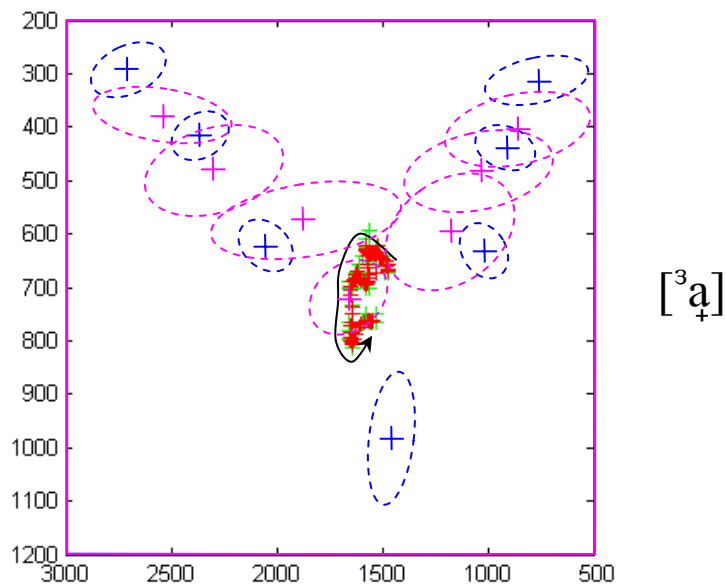


Figura A16. Diagramma *Voweltrack* del timbro della vocale accentata di 'na *sëpalë* 'una siepe' (GC84, Aliano, MT).

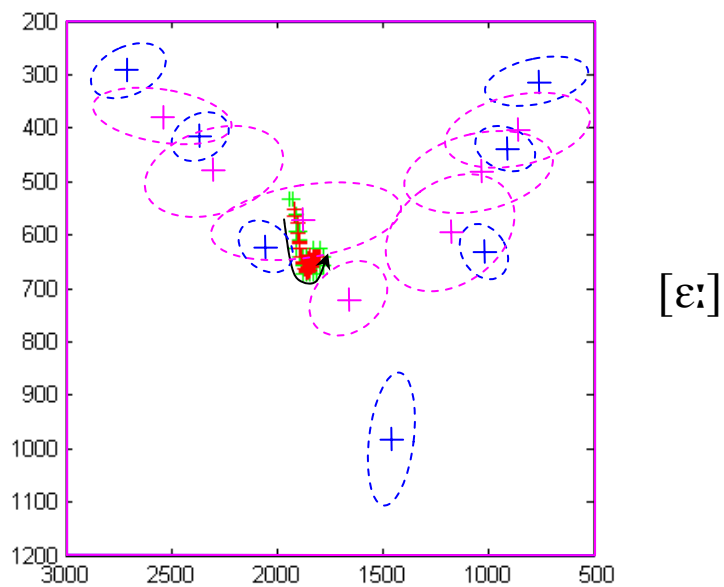


Figura A17. Diagramma *Voweltrack* del timbro della vocale accentata di 'a *capĕ* 'la testa, il capo' (GC84, Aliano, MT).

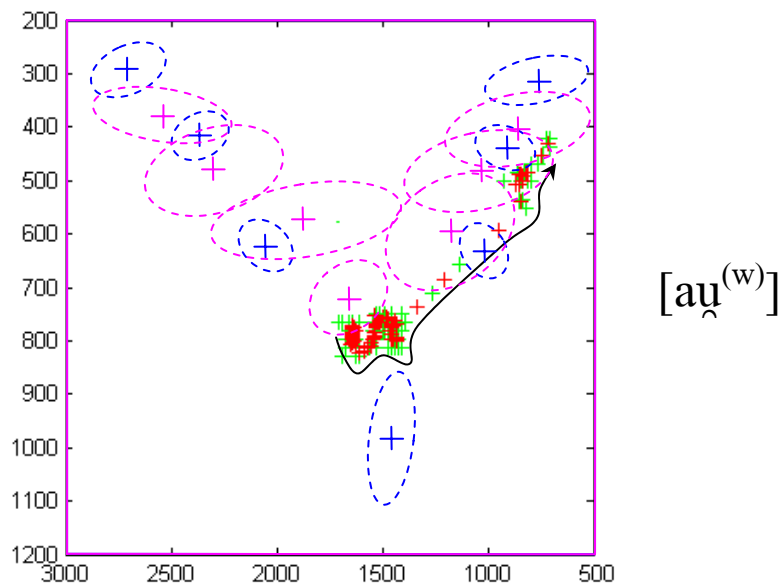


Figura A18. Diagramma *Voweltrack* del timbro della vocale accentata di 'u *tàwrĕ* 'il toro' (GC84, Aliano, MT).

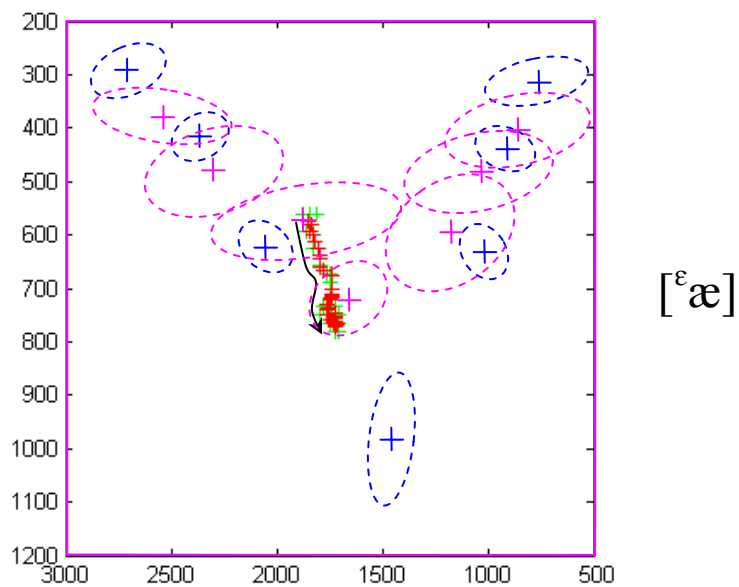


Figura A19. Diagramma *Voweltrack* del timbro della vocale accentata di *i canē* 'i cani' (GC84, Aliano, MT).

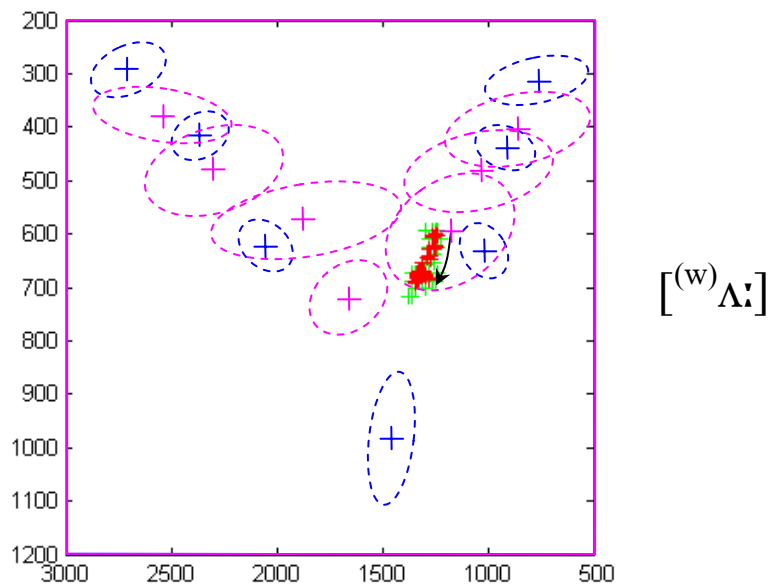


Figura A20. Diagramma *Voweltrack* del timbro della vocale accentata di *'u quanē* 'il cane' (GC84, Aliano, MT).

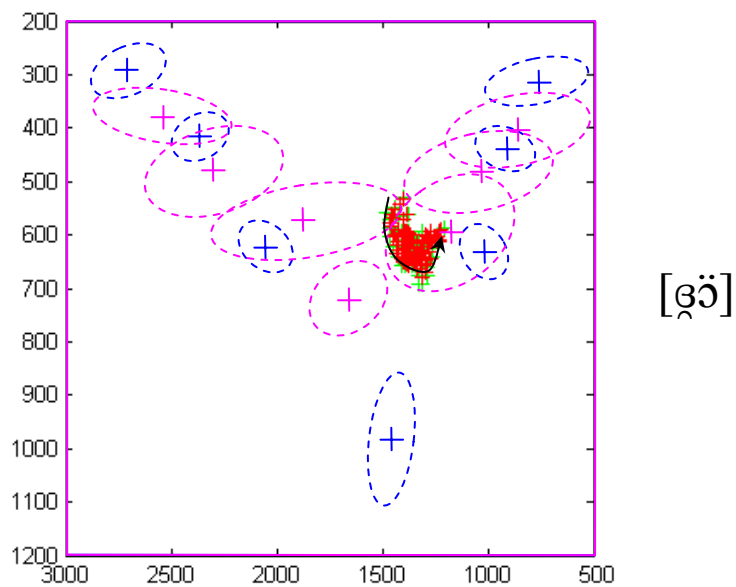


Figura A21. Diagramma *Voweltrack* del timbro della vocale accentata di 'u *frontē* (*frondē*) 'la fronte' (GC84, Aliano, MT).

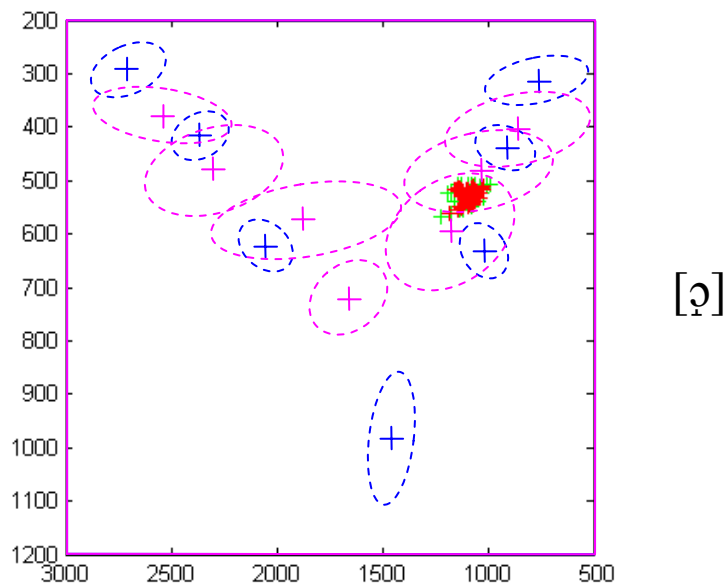


Figura A22. Diagramma *Voweltrack* del timbro della vocale accentata di *pochē* ‘poco’ (GC84, Aliano, MT).

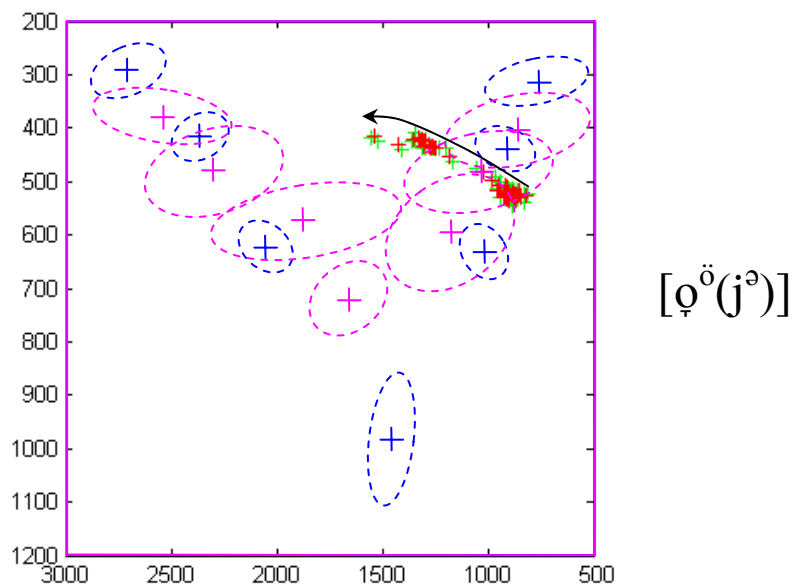


Figura A23. Diagramma *Voweltrack* del timbro della vocale accentata di 'u voj(*ē*) 'il bue' (GC84, Aliano, MT).

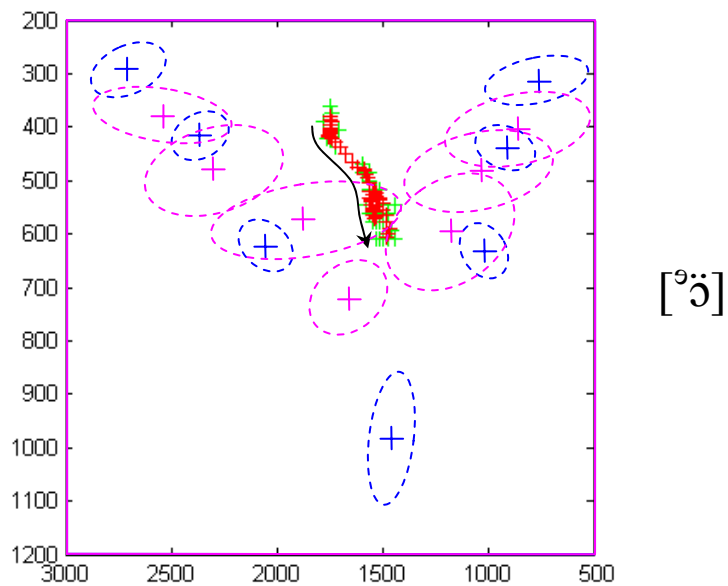


Figura A24. Diagramma *Voweltrack* del timbro della vocale accentata di 'u mendo(*nē*) 'il montone' (GC84, Aliano, MT).

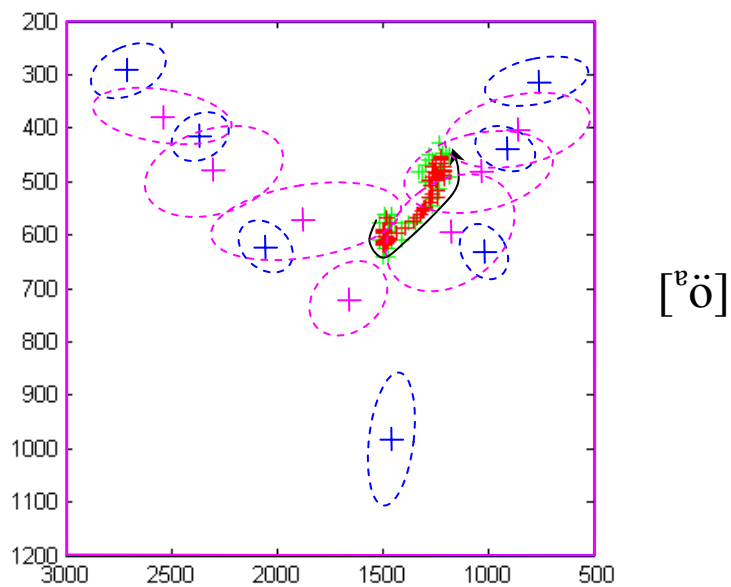


Figura A25. Diagramma *Voweltrack* del timbro della vocale accentata di *pellë pëlosë* ‘pelle pelosa’ (GC84, Aliano, MT).

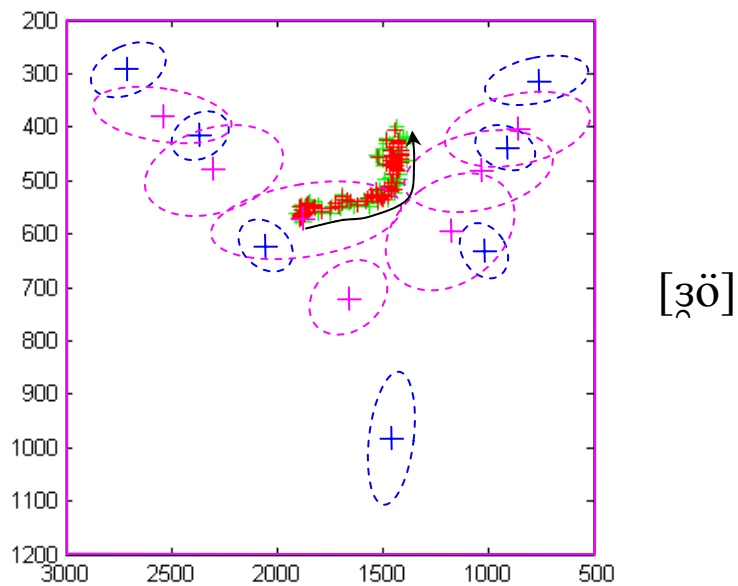


Figura A26. Diagramma *Voweltrack* del timbro della vocale accentata di *’u pastorë* ‘il pastore’ (GC84, Aliano, MT).

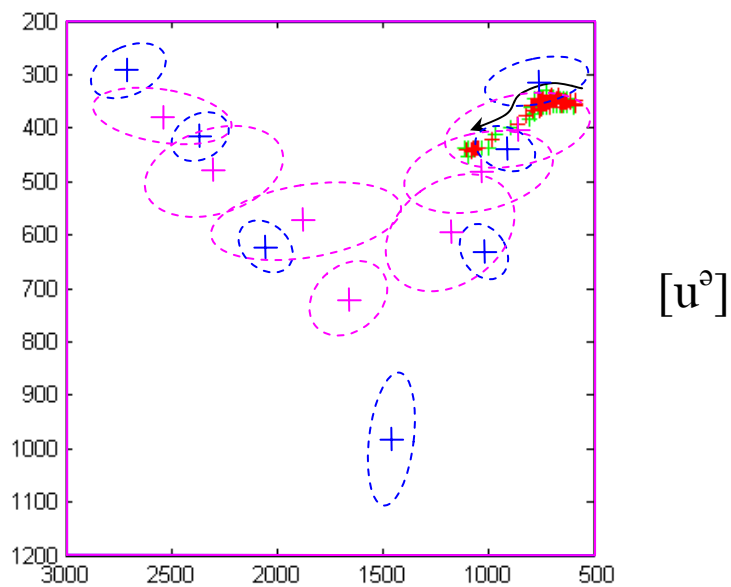


Figura A27. Diagramma *Voweltrack* del timbro della vocale accentata di 'a *murrë* 'il branco (moltitudine di animali)' (GC84, Aliano, MT).

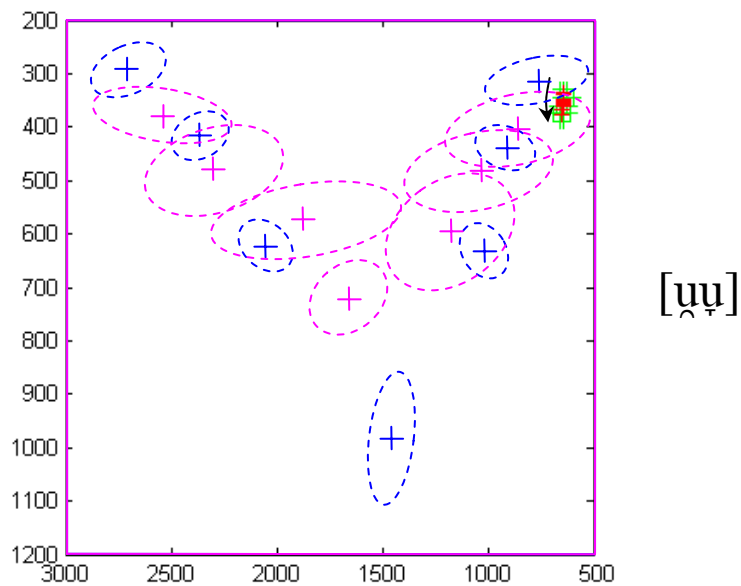


Figura A28. Diagramma *Voweltrack* del timbro della vocale accentata di *paludë* 'palude' (GC84, Aliano, MT).

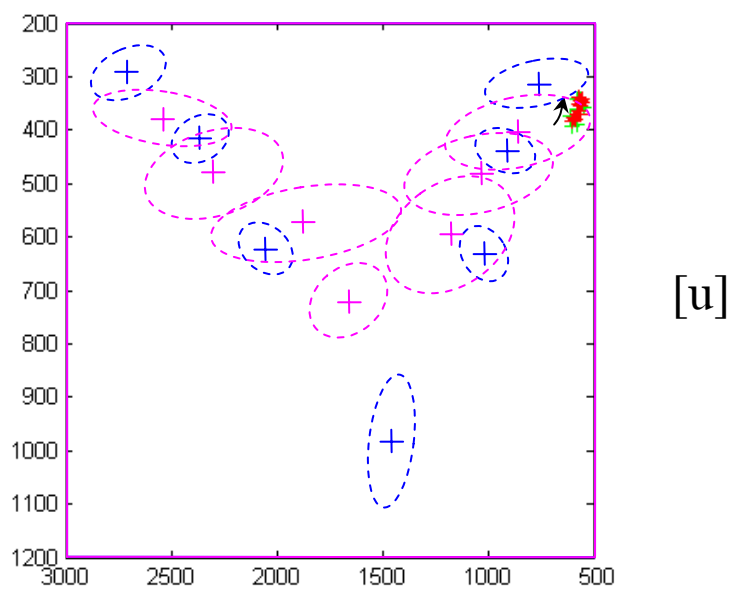


Figura A29. Diagramma *Voweltrack* del timbro della vocale accentata di 'a *vulpë* 'la volpe' (GC84, Aliano, MT).

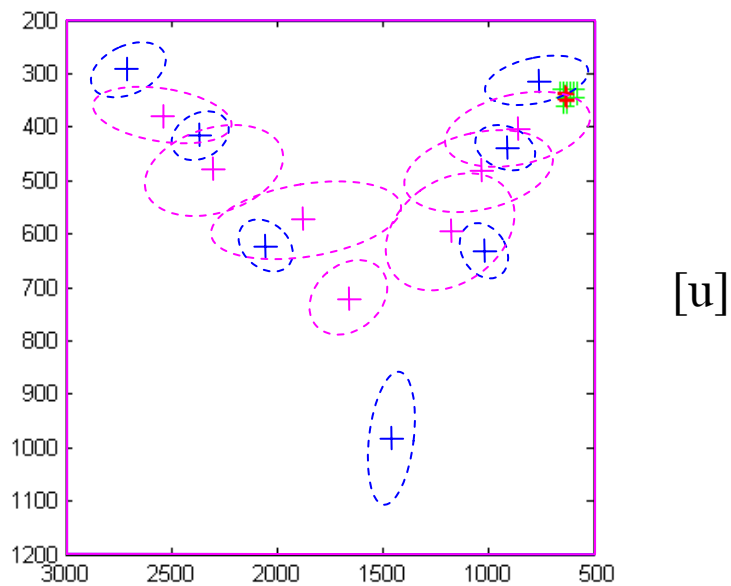


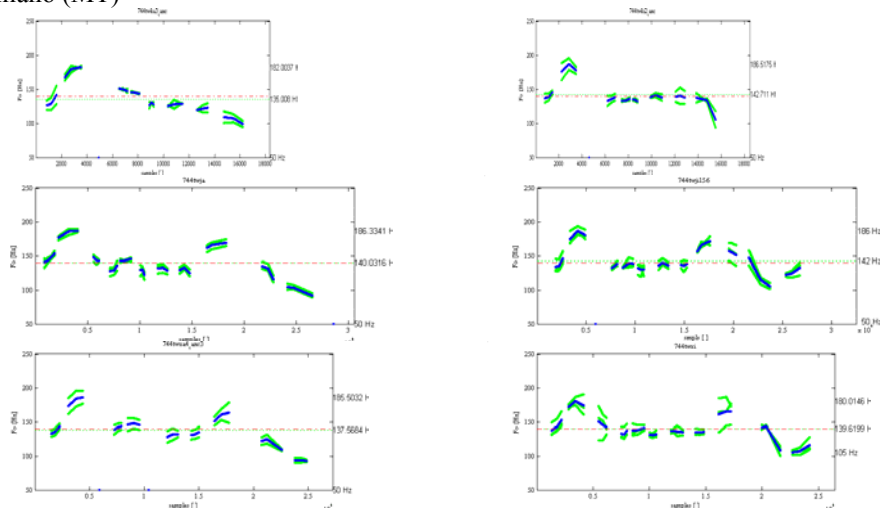
Figura A30. Diagramma *Voweltrack* del timbro della vocale accentata di *tutt'u mundë* 'tutto il mondo' (GC84, Aliano, MT).



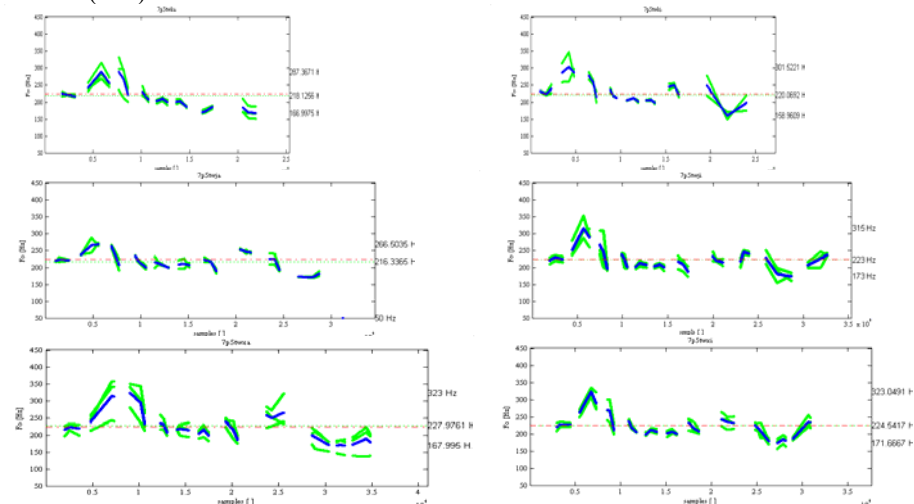
## APPENDICE 2

Quest'appendice esemplifica la stilizzazione delle curve di  $f_0$  per diverse ripetizioni delle frasi del *corpus* introduttivo (v. testo) per il locutore PDA40 (di Aliano, MT) e la locutrice MR42 (di Alianello, MT). I grafici alla sinistra si riferiscono a enunciati dichiarativi mentre quelli a destra a enunciati interrogativi<sup>22</sup>.

Aliano (MT)



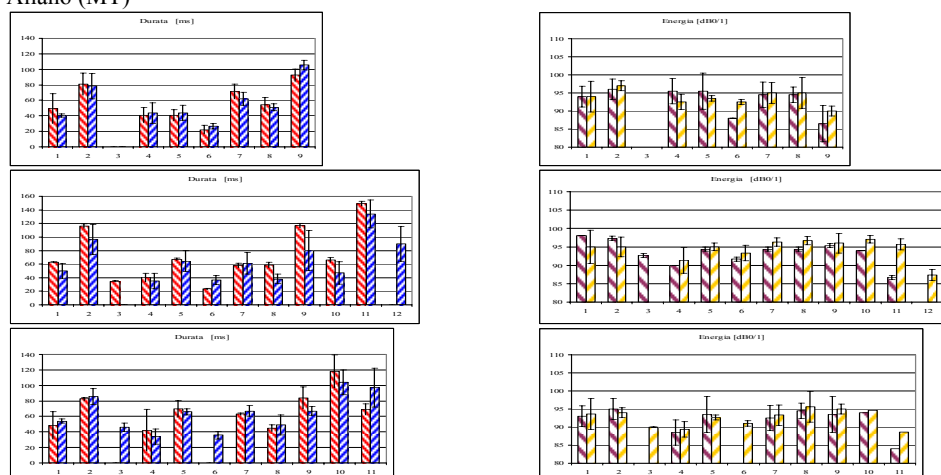
Alianello (MT)



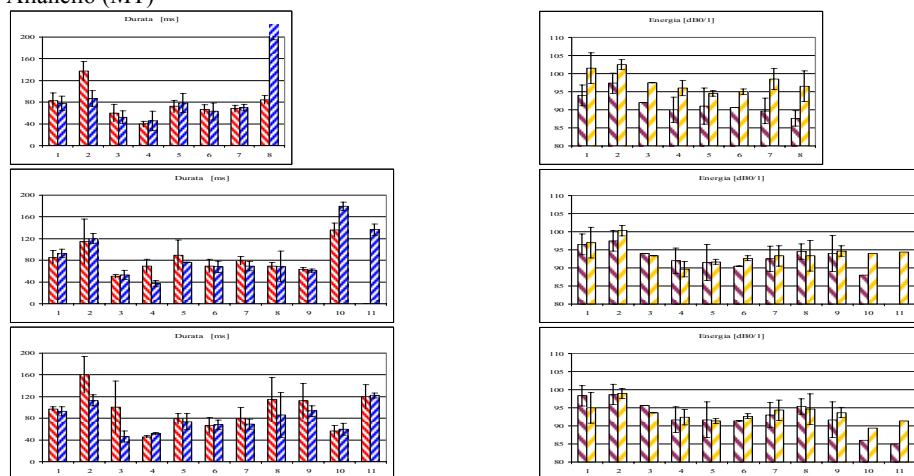
<sup>22</sup> Per ciascuna varietà i grafici più in alto sono relativi agli enunciati senza espansioni (rispettivamente *Dunatë sê vivëtë 'u cafë./?* e *Donatë sê vevë 'u caf(f)ë./?*), quelli nella riga successiva presentano un'espansione nel sintagma verbale con un aggettivo parossitono (*Dunatë sê vivëtë 'u cafë mbucatë./?* e *Donatë sê vevë 'u caf(f)ë mbucatë./?*), quelli nella terza riga presentano invece un'espansione con un aggettivo proparossitono (*Dunatë sê vivëtë 'u cafë truóv(ël)ë./?* e *Donatë sê vevë 'u caf(f)ë tróvëlë./?*).

Gli istogrammi seguenti esemplificano il tipo di rappresentazione grafica usata per l'analisi delle sequenze di valori di durata (a sinistra) ed energia (a destra) per le vocali degli enunciati dei grafici precedenti (le barre con tratteggio discendente sono relative alla modalità dichiarativa, quelle col tratteggio ascendente alla modalità interrogativa)<sup>23</sup>.

#### Aliano (MT)



#### Alianello (MT)



<sup>23</sup> La numerazione delle barre corrisponde, con la sola eccezione dei casi di sinalefe ( ), all'ordine sequenziale delle vocali realizzate almeno in un caso. Nei primi istogrammi di Aliano si ha quindi *Dunatë sě vīvētē 'u cafē./?*; in quelli successivi *Dunatë sě vīvētē 'u cafē mbucatē./?* e negli ultimi due *Dunatë sě vīvētē 'u cafē truón(ēl)ē./?*. Negli istogrammi di Alianello si ha nell'ordine *Donatë sě vėvė 'u caf(f)ē./?*; *Donatë sě vėvė 'u caf(f)ē mbucatē./?*; *Donatë sě vėvė 'u caf(f)ē trónvėlē./?*.

## IL PROGETTO BOMARZO

Amedeo De Dominicis e Pamela Mattana<sup>1</sup>  
Laboratorio di Fonetica, Università degli Studi della Toscana  
[dedomini@unitus.it](mailto:dedomini@unitus.it), [pamela.mattana@gmail.com](mailto:pamela.mattana@gmail.com)

### 1. SOMMARIO

Nel corso degli ultimi 50 anni sono stati raccolti, da vari centri di ricerca (soprattutto orientati allo studio dei dialetti), numerosi materiali che documentano l'uso della lingua da parte di parlanti diversi per provenienza geografica, età, livello di cultura. In non rari casi, le varietà dialettali documentate costituiscono la testimonianza viva di un uso ormai scomparso.

Tuttavia, i supporti materiali su cui tali raccolte sono conservate sono spesso obsoleti e ciò, oltre a renderne difficile la consultazione, mette a rischio la loro stessa conservazione. Sarebbe possibile e soprattutto auspicabile riversare su nuovi supporti, di tipo digitale, questo prezioso giacimento, creando una base di dati facilmente consultabile secondo procedure unitarie dalla comunità degli studiosi.

D'altro canto, sarebbe possibile e altrettanto auspicabile realizzare un archivio centralizzato, organizzato in una piattaforma di facile uso, in cui i punti di raccolta siano indicati sulla mappa, con la possibilità di *clickare* su un punto qualsiasi, vedere l'elenco dei materiali immagazzinati (ovunque possibile, anche sotto forma di trascrizione), e sentire la corrispondente realizzazione sonora.

In tal modo, questi materiali oggi virtualmente inesistenti per il mondo della ricerca, sarebbero nuovamente utilizzabili per l'intera comunità scientifica.

Dalla consapevolezza dell'importanza linguistica e storica di questo patrimonio a rischio, nel novembre 2004 nasce il progetto Calliope (le "soffitte" della voce), uno sviluppo del precedente progetto COVAID (aprile 2002)<sup>2</sup> (De Dominicis, 2002).

### 2. DEFINIZIONE DEI MATERIALI DI INTERESSE

I documenti vocali di interesse per i due progetti sono tutte le registrazioni analogiche di rilevanza linguistica, con particolare riguardo a dialetti, comprese le parlate alloglotte d'Italia, le varietà locali dell'italiano (incluse quelle di comunità estere), le varietà di italiano L2, il parlato registrato in manifestazioni folkloriche. In tale lista si comprendono non solo documenti puramente acustici, ma anche di tipo audiovisivo. Nel caso di materiali già digitalizzati, occorre valutare la compatibilità dei criteri adottati, per eventualmente predisporre la ridigitalizzazione a partire dall'originale.

### 3. PARTECIPANTI

Inizialmente, i partecipanti al progetto COVAID erano il Laboratorio di Fonetica dell'Università della Toscana (Amedeo De Dominicis), il Laboratorio di Linguistica della Scuola Normale Superiore di Pisa (Pier Marco Bertinetto), l'Università di Padova (Alberto Zamboni), la sezione di "Fonetica e Dialettologia" di Padova dell'Istituto di Scienze

---

<sup>1</sup> I paragrafi 1-4 sono di Amedeo De Dominicis, i paragrafi 5-8 di Pamela Mattana.

<sup>2</sup> **COVAID** è l'acronimo di **C**onservazione e **V**alorizzazione degli archivi vocali dell'**I**taliano e dei suoi **D**ialetti.

Tecnologie della Cognizione del CNR (I.S.T.C.) (Emanuela Magno Caldognetto) e la Fondazione “Ugo Bordoni” di Roma (Andrea Paoloni). In seguito, si è aggiunto il Laboratorio di Fonetica dell’Università della Calabria (Luciano Romito).

#### **4. OBIETTIVO**

L’obiettivo dei due progetti è la creazione di un archivio storico dell’italiano parlato e delle sue varietà, con particolare attenzione alle differenze diacroniche, diastratiche e diafasiche, che sia di facile consultazione da parte degli studiosi. Tuttavia, la mancata attuazione del progetto su scala nazionale ha rallentato la preliminare definizione degli *standard* di lavoro. In primo luogo, si sarebbe dovuto preliminarmente precisare la struttura della base di dati, le modalità di consultazione della stessa e come dinamicamente ampliarla ed aggiornarla. Allo stesso modo, avrebbero dovuto essere definite le modalità di individuazione, sul territorio nazionale, dei materiali fonici a vario titolo raccolti da organismi o persone. Per quanto riguarda il formato di archiviazione audio, lo standard di diffusione dei documenti audio su internet non è affidabile (il formato MP3 a 32 Kbyte sec., ad es., non consente un’analisi spettrografica attendibile) (Paoloni, 2004). Inoltre, sarebbe necessaria una scheda di trascrizione, altrimenti il materiale sarebbe simile a quello oggi udibile, ad es. sul sito delle Teche RAI (Parola, Santoli, 2004).

Inoltre, resta da decidere quali siano le modalità atte a motivare la collaborazione degli istituti e degli studiosi alla collezione dei dati. Infine, la costituzione di un archivio centralizzato creato grazie alle donazioni di un insieme di contributori pone alcune questioni.

Innanzitutto occorre stabilire dove collocare materialmente l’archivio. Probabilmente la soluzione migliore è di affidarlo ad un Ente tecnico *super partes*, come il Ministero delle Comunicazioni o la Discoteca di Stato.

Inoltre, occorrerebbe motivare i detentori (privati o pubblici) di giacimenti vocali a divenire contributori del progetto. A tal fine sarebbe utile che la proprietà dei documenti contenuti nell’archivio rimanesse ai contributori (con tutela del nome del proprietario, dell’istituto di appartenenza e del raccoglitore). A fronte della disponibilità a riversare i documenti sonori nell’archivio centralizzato, il progetto dovrebbe fornire ai contributori i seguenti vantaggi: riversamento su supporto non deperibile del proprio materiale, che resterebbe ovviamente a disposizione dell’ente contributore; documentazione e catalogazione (e, in prospettiva, trascrizione) dell’immenso patrimonio sommerso; in prospettiva, accesso gratuito all’intera banca dati da parte dei contributori, previo consenso - in forme da concordare - dei proprietari dei singoli fondi documentali (si può ipotizzare per es. un ragionevole arco di tempo, trascorso il quale il documento prima riservato all’uso esclusivo del proprietario possa divenire consultabile per tutti i contributori. Ciò permetterebbe di completare eventuali ricerche scientifiche in corso); possibilità di figurare all’interno del catalogo, con evidente guadagno di immagine.

#### **5. COME NASCE IL PROGETTO BOMARZO**

Il territorio della Toscana è soggetto - seppur in forma poco pubblicizzata - ad indagini e raccolte a carattere etnografico. Questo materiale è eterogeneo: in parte composto da manufatti del vivere quotidiano in ambiente montano o rurale (confluiti poi nel museo di Canepina (VT), in parte da foto e video di rappresentazioni sacre o feste patronali, ma anche - e in parte rilevante - da registrazioni su supporto magnetico, come bobine o audiocassette. Molti dei materiali sopra elencati sono oggi nel fondo audio della Provincia

È nata, quindi, una collaborazione fondata sulla convinzione che la tutela dell'identità storica del territorio e la sua valorizzazione siano il fondamento del suo stesso sviluppo culturale.

Il fondo ha dei registri di inventario e schede documentarie nelle quali figurano non solo l'anagrafico del documento, ma anche le trascrizioni di documentazione (cfr. Figura 1).

Lo stato di conservazione e documentazione del materiale, all'epoca del progetto, non rispettava gli *standard* di conservazione dei materiali sia cartacei che non cartacei (Canazza, 2003; Carucci, 1993). Il fondo presenta lacune sia dal punto di vista

407

documentaristico (descrizione interna alle schede), che archivistico (codici delle schede, numerazioni, divisioni in faldoni e inventario).

## **7. PROGETTO BOMARZO**

Il periodo di elicitazione dei dati relativi a Bomarzo è ristretto in un arco di tempo che va dal 31-12-1978 al 31-12-1979. Gli intervistatori erano principalmente due donne e un uomo originari della zona della Tuscia; lo scopo del loro lavoro è stato quello di raccogliere materiale per una descrizione delle tradizioni popolari nel detto territorio, al fine di tutelare un patrimonio culturale che ha come suo veicolo primario l'oralità.

Per quanto concerne gli informatori censiti, possiamo rilevare che sono un numero piuttosto elevato, ben 61 divisi tra 38 uomini e 23 donne. Per quel che riguarda la loro età, è rilevante il dato che 11 fossero nati negli ultimi anni dell'800 e 50 nei primi decenni dello scorso secolo (Arduini, Leuzzi, Palmisciano 1983).

Il lavoro consta di un totale registrato pari a 1934.769.15 minuti, cioè circa 32 ore, suddiviso in 24 cassette audio di tipo ferro-cromo da 90 minuti, registrate su ambedue i lati, con larghezza di banda di 8000 Hz.

Per quanto concerne le schede anagrafiche corredate di trascrizioni, il Laboratorio di Fonetica dispone di una copia sia cartacea che su formato PDF, mentre gli originali sono rimasti in possesso della Provincia.

Il nostro lavoro si è articolato in cinque fasi.

### *7.1 Fase prima*

La prima fase ha richiesto competenze di tipo archivistico. Infatti, abbiamo rilevato l'intero fondo audio della provincia di Viterbo, effettuando il sondaggio del fondo e la consultazione degli inventari cartacei ad esso relativi. Attraverso la loro messa a confronto siamo giunti alla stima specifica: quantità, stato di conservazione, luogo di conservazione, lacune o materiale disperso, ecc.. È stata la fase più delicata del lavoro, senza la quale non sarebbe possibile giungere alla selezione del materiale che è stato successivamente sottoposto al trattamento di conservazione e documentazione. La scelta della porzione di fondo sulla quale operare è stata effettuata sulla base di parametri di opportunità; con questo termine ci riferiamo allo stato di conservazione del materiale, al suo prestigio e alla sua quantità (Carucci, 1993; Carucci, 1998; Londei, 2004; Zanni Rosiello, 1996).

### *7.2 Fase seconda*

La seconda fase ha richiesto competenze di carattere acustico e di restauro dell'audio. Il materiale in nostro possesso era conservato in audio-cassette del tipo ferro-cromo della durata di 90 minuti. In primo luogo, è stata verificata la qualità del segnale analogico, che in alcune parti risulta saturato, mentre in altre presenta una dinamica tanto insufficiente da comprometterne l'udibilità. Di conseguenza, è stato effettuato il riversamento da analogico a digitale (formato WAV, frequenza di campionamento 44.1 kHz e quantizzazione 16 bit) e successivamente abbiamo sottoposto il segnale così digitalizzato a filtraggio ed a normalizzazione in ampiezza. Facciamo presente che esistono quindi due copie digitali: la copia conforme all'originale e la copia che ha subito il filtraggio e la normalizzazione.

Il materiale raccolto è affetto da una serie di imperfezioni e disturbi acustici, dovuti alla impreparazione dei raccoglitori originari e alle finalità amatoriali che essi si prefiggevano. I segnali sono spesso saturati e i rumori ambientali contenuti nelle registrazioni (passaggio di bus, mezzi agricoli o altri macchinari) li rendono poco intelligibili. Le registrazioni sono state condotte in luoghi conviviali e in occasioni particolari come feste e celebrazioni, quindi il parlato è spesso sovrapposto e in alcuni casi più parlanti si alternano nel corso

della stessa sessione di registrazione. Si tratta di disturbi per i quali non è facile attuare un protocollo di restauro (Cosi, 2001; Canazza, 2004; De Dominicis, 2006; De Mezzo, 2004).

### 7.3 Fase terza

La terza fase ha carattere specificamente documentale. Le schede originali di documentazione, redatte dai raccoglitori su supporto cartaceo, sono state digitalizzate in formato PDF.

Successivamente, i dati contenuti nelle schede sono stati inseriti in un foglio *Excel*. I campi già presenti sono stati integrati con dati relativi alla durata, alla qualità del segnale, ecc. (cfr. Figura 2). Infine, le schede in formato *Excel* sono state convertite in formato *Access*, al fine di consentirne una più agevole consultazione attraverso *query*.

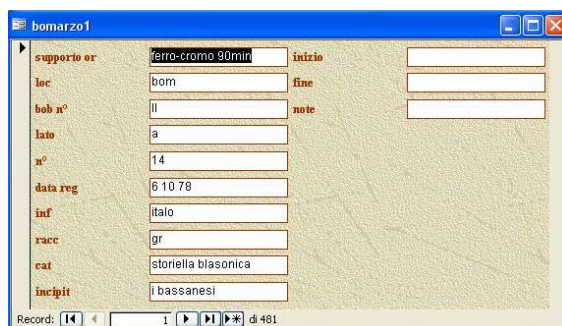


Figura 2: Scheda *Access*.

### 7.4 Fase quarta

Nella quarta fase si è sviluppato un prototipo – assai primordiale – di trascrizione linguistica del *corpus*.

Le schede cartacee di documentazione non costituiscono un valido punto di partenza in tal senso, in quanto redatte da mani diverse, con competenze amatoriali e con criteri non scientifici. La trascrizione è estremamente disomogenea: in genere, è ispirata all'ortografia dell'italiano, talvolta con tentativi di rappresentare le forme dialettali con segni diacritici.

Inoltre le schede cartacee forniscono una trascrizione del contenuto parlato della registrazione spesso lacunosa: rappresentano solo il turno di parola dell'intervistato e non dell'intervistatore, descrivono sommariamente l'argomento elencando solo le parole ritenute "interessanti", trascrivono solo il *cammeo* (barzelletta, canzone) e non l'intero svolgersi della conversazione in cui è inserito.

Perciò, il lavoro di trascrizione è stato svolto *ex novo*. Il segnale è stato etichettato attraverso le risorse offerte dal *software* di elaborazione audio *Praat*. Ai segnali audio sono stati associati file *TextGrid* dotati di due *Tiers* di annotazione: uno per la trascrizione ortografica e l'altro per annotazioni generali, ove vengono indicati vuoti di segnale (0), sovrapposizioni di più persone (sov.), risate (ridono), ma anche rumori (R). Ovviamente, è possibile aggiungere altri *Tiers* per segnalare, ad es., il cambio di locutore o semplicemente per distinguere un rumore ambientale da un colpo di tosse o da una sovrapposizione. La scelta di *Praat* è stata dettata proprio per questa possibilità di poter aggiungere livelli di annotazione (cfr. Figura 3).

La trascrizione ortografica è stata effettuata su tutto il materiale ad eccezione di quello cantato.

### 7.5 Fase quinta

Il Laboratorio di Fonetica ha intrapreso la trascrizione e l'annotazione fonetica del materiale di Bomarzo al fine di compiere un lavoro di descrizione della varietà locale. Ad oggi questo materiale non è mai stato oggetto di studio linguistico, ma da una prima analisi del *corpus* emergerebbero i tratti caratteristici dell'area centro-meridionale (Grassi *et al.*, 2001: 71-160), come il raddoppiamento fonosintattico, la sonorizzazione delle occlusive intervocaliche e il rotacismo della laterale preconsonantica. Certamente queste prime rilevazioni attendono la validazione dai dati del *corpus*.

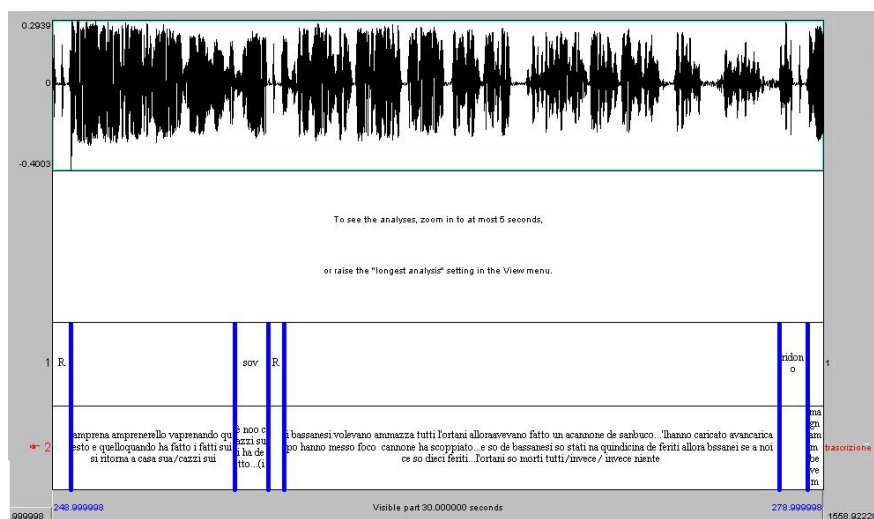


Figura 3: Schermata Praat con trascrizione e annotazioni.

## 8. COSA RIMANE DA FARE? IL FUTURO

Il progetto Bomarzo, in quanto parte del più ampio progetto Calliope, per svilupparsi ha bisogno di potersi avvalere di *standard* non solo relativi ai criteri di trascrizione fonetica, ma anche alla struttura del data-base e delle gerarchie classificatorie sottostanti, ai livelli di analisi; come pure sarebbe necessario disporre di un *thesaurus* classificatorio omogeneo e stabile che permetta una facile interrogazione dei dati (Weston, 2002).

Per quanto riguarda specificamente l'annotazione relativa alla qualità del segnale e alla sua intelligibilità, sarebbe auspicabile la possibilità di poter etichettare i due fattori con criteri graduali, potendo inoltre segnalare per chi è fruibile un determinato *frame*. Ad esempio, sarebbe utile poter indicare che una data porzione di segnale è una fonte valida per la ricostruzione lessicale oppure per la rilevazione della F0.

Ovviamente, quanto detto finora è attuabile, nelle migliori prospettive, solo nella lunga durata e con largo investimento di risorse; infatti, si richiede la digitalizzazione di tutto il patrimonio che allo stato attuale non è stato ancora interamente censito, la sua schedatura, trascrizione, inserimento in una banca dati ed etichettatura in senso documentario-multimediale (*tag e link*) (Calvo, Roncaglia, 2000).

Si sta lavorando inoltre al fine di una messa in rete del materiale.



## BIBLIOGRAFIA

- Arduini M., Leuzzi M. D., Palmisciano M.G. (1983), *Tradizioni orali a Bomarzo alcuni repertori di una ricerca*, Viterbo: Amministrazione Provinciale di Viterbo.
- Calvo M., Roncaglia G. (2000), *Internet 2000: manuale per l'uso della rete*, Bari: Laterza.
- Canazza S. (2003), <http://www.dei.unipd.it/~musica/Dispense/Restauro.pdf>
- Canazza S. (2004), Il restauro del segnale vocale mediante proiezione locale, in: De Dominicis A., Mori L., Stefani M. (editors) (2004), 39-52.
- Carucci P. (1993), *Le fonti archivistiche: ordinamento e conservazione*, Roma: La Nuova Italia Scientifica.
- Carucci P. (1998), *Manuale di archivistica per l'impresa*, Roma: Carocci.
- Cosi P. (2001), Il restauro digitale del segnale audio, in: De Dominicis A. (ed.) (2001), 189-199.
- De Dominicis A. (2001), La voce come bene culturale, la linguistica come strumento di catalogazione e restauro, in: De Dominicis A. (ed.) (2001), 7-26.
- De Dominicis A. (2002), Co.Va.I.D. (COnservazione e VAlorizzazione degli archivi vocali dell'Italiano e dei suoi Dialetti), *La comunicazione*, numero unico speciale a cura di Giuseppe Rinaldo e Roberto Piraino contenente gli Atti della Conferenza *TIPI (Tecnologie Informatiche nella Promozione della lingua Italiana)*, LI, 97-98.
- De Dominicis A. (2006), Il restauro "perfetto". In Canazza, Sergio & Casadei Turronei Monti, Mauro (eds.) *Ri-mediazione dei documenti sonori*, Udine: Forum, 603-608.
- De Dominicis A. (ed.) (2001), *La voce come bene culturale*, Roma: Carocci.
- De Dominicis A., Mori L., Stefani M. (editors) (2004), Costituzione, gestione e restauro di corpora vocali, Atti delle XIV giornate di studio del GFS (A.I.A.), Roma: Esagrafica.
- De Mezzo G. (2004), Conservazione dei documenti sonori: il riversamento conservativo di registrazioni su nastro magnetico, in: De Dominicis A., Mori L., Stefani M. (editors) (2004), 27-38.
- Grassi C., Sobrero A.A., Telmon T. (2001), *Fondamenti di dialettologia italiana*, Roma-Bari: Laterza.
- Londei L. (2004), *Elementi di archivistica*, Roma: Jouvence.
- Paoloni A. (2004), Restauro e memorizzazione del segnale vocale, in: De Dominicis A., Mori L., Stefani M. (editors) (2004), 53-66.
- Parola L., Santoli P. (2001), Il catalogo multimediale della RAI. Le Teche radiofoniche, in: De Dominicis A. (ed.) (2001), 201-208.
- Weston P. G. (2002), *Il catalogo elettronico: dalla biblioteca cartacea alla biblioteca digitale*, Roma: Carocci.
- Zanni Rosiello I. (1996), *Andare in archivio*, Bologna: il Mulino.

## LA RETROFLESSIONE IN ALCUNI CENTRI DELLA PROVINCIA DI CATANZARO: VERSO UNA MAPPA REGIONALE

Luciano Romito & Paola Antonietta Scuticchio

Laboratorio di Fonetica - Dipartimento di Linguistica – Università della Calabria

[luciano.romito@unical.it](mailto:luciano.romito@unical.it) [paolas2@virgilio.it](mailto:paolas2@virgilio.it)

### SOMMARIO

Lo scopo di questo lavoro è fornire una mappatura delle realizzazioni fonetiche dei suoni retroflessi nella provincia di Catanzaro.

I risultati di questa indagine andranno a confluire in una ricerca già in atto presso il Laboratorio di Fonetica dell'Università della Calabria dal titolo "I corpora vocali come patrimonio culturale: archiviazione, conservazione e restauro di materiale sonoro calabrese" (cfr. Romito, Lio 2006). Buona parte del materiale utilizzato per le analisi del presente lavoro è stata reperita all'interno dell'Archivio Sonoro Calabrese<sup>1</sup>

In questa sede ci proponiamo di delineare, attraverso una comparazione di dati empirici, la storia dell'evoluzione del fenomeno fonologico della retroflessione nei dialetti della provincia catanzarese. Riteniamo che tale studio potrà chiarire non solo i vari passaggi storici dello sviluppo della laterale –LL–, ma ancora più importante, di prevedere le sue future trasformazioni.

Il lavoro si sviluppa attraverso quattro fasi principali:

1. raccolta di tutto il materiale di ricerca relativo alla evoluzione dell'articolazione retroflessa come esito di –LL– sia dal punto di vista dialettologico che fonetico acustico ed elettropalatografico;
2. analisi percettiva ed elettroacustica dei suddetti dialetti;
3. comparazione dei dati ottenuti con quelli precedentemente analizzati;
4. stesura di una completa mappatura del materiale sonoro analizzato.

Per quanto riguarda la prima fase, il fine è quello di ripercorrere le principali teorie che nel corso degli anni sono subentrate in ambito dialettologico (ci si riferisce in particolare ad Alessio 1954, Rohlfs 1966, Falcone 1976, Bianco 1981, Trumper e al 1988 e seguenti), ed in ambito fonetico-acustico (cfr. Romito et al. 1996, 1999a, 1999b, 2000, 2001 e 2002).

I suoni retroflessi in Calabria hanno suscitato un notevole interesse nei dialettologi che si sono occupati della ricerca linguistica nell'area dell'Italia Meridionale e nei fonetisti per la particolarità articolatoria di tali suoni. Questi sono distribuiti eterogeneamente su tutto il territorio e diacronicamente sono esiti della laterale di origine latina –LL– o –L– o dei nessi consonantici con una vibrante che precede o che segue.

Riguardo gli esiti di –LL–, questi vanno considerati come la conseguenza della perdita dell'opposizione di lunghezza –L– vs –LL–, ciò causò lo *slittamento* articolatorio di –L– in [d, r], nei dialetti dove –LL– si era conservata.

---

<sup>1</sup> Nell'Archivio Sonoro Calabrese i punti di raccolta e di localizzazione di tutti i materiali sonori sono indicati su una mappa virtuale che permette di interrogare un punto, di vedere l'elenco dei materiali immagazzinati e di ascoltare la corrispondente realizzazione sonora. In alternativa è possibile utilizzare un motore di ricerca comune e per campi multipli su tutto il materiale sonoro presente al suo interno (cfr. Romito&Lio, 2006).

Secondo Hála (1964) “*Sont rétroflexes uniquement les consonnes articulées avec la pointe recourbée en arrière*“, mentre sono cacuminali, “*celle formées par l’extrémité de la pointe appliquée verticalement contre le palais*”. Dalle più recenti ricerche svolte, gli esiti di -LL- nella provincia di Catanzaro, salvo pochissime eccezioni (come ad esempio i comuni di Andali e Squillace), sarebbero da considerarsi cacuminali piuttosto che retroflessi (cfr. gli studi elettropalatografici effettuati da Romito & Soriano nel 1999).

E’ soprattutto sulla base di questi recenti studi che prende le mosse la seconda fase del nostro progetto di ricerca. Infatti attraverso l’analisi elettroacustica ci proponiamo di verificare l’eventuale evoluzione del fenomeno ‘retroflessione’.

I risultati ottenuti proseguono gli studi di ricerca precedentemente effettuati (cfr. Romito, Milelli 1999, Romito, Soriano 1999, Romito et al 2000), e si inseriscono in un archivio di materiale sonoro precedentemente restaurato e catalogato nel database sopra citato.

La terza fase del progetto prevede la comparazione dei dati dialettologici, fonetico-acustici e quelli presenti in letteratura, mentre la quarta ed ultima fase sarà incentrata sull’analisi di tutti i dati in nostro possesso e sulla costruzione di una mappatura tematica sul fenomeno fonologico della retroflessione nel territorio della provincia di Catanzaro, di una etichettatura fonetica del materiale sonoro analizzato ed archiviato e su una possibile evoluzione futura di tale processo.

L’osservazione diacronica sull’evoluzione della -LL- latina nei dialetti calabresi, dimostra una doppia tendenza. Da una parte si mira al massimo rafforzamento con l’esito fonetico della retroflessione e dalla parte opposta alla massima riduzione con l’esito fonetico di una vocale o addirittura di una cancellazione. Tra le due tendenze sono presenti innumerevoli realizzazioni fonetiche nei diversi dialetti che potrebbero essere considerati come differenti stadi dello stesso processo fonologico o processi differenti e del tutto scollegati tra loro.

## 1. LA RETROFLESSIONE IN CALABRIA

Le consonanti retroflesse contribuiscono a caratterizzare buona parte dei dialetti centro-meridionali, essendo presenti non solo in Sicilia, Calabria e Salento, ma anche in Sardegna, nella Corsica meridionale e in alcuni centri della Toscana nord-occidentale. In queste zone la retroflessione coinvolge, con una grande diversificazione di esiti, la laterale geminata. In Calabria, le pronunce retroflesse sono state importate con ogni probabilità dalla Sicilia, secondo un’espansione da ovest verso est del processo.

In particolare si è considerato come punto di partenza per la nostra analisi lo studio effettuato da Bianco (1981)<sup>2</sup> che illustra come la distribuzione delle forme retroflesse sia geograficamente molto irregolare e gli esiti da lui individuati, accanto al mantenimento di /l:/, sono cinque: un’occlusiva geminata retroflessa sonora [ɖ:], una variante indebolita “di colorito palatale”; una forma rotacizzata, [r], l’approssimante [j], una laterale retroflessa geminata [ɭ:].

Le varianti registrate da Rohlf e Falcone (1976)<sup>3</sup> coincidono con quelle riscontrate da Bianco, ma a queste deve essere aggiunta una realizzazione affricata apicale sublaminale [ɖ:ʒ], documentata per le parlate di Cosenza e Catanzaro da Soriano & Mancuso (1998) e Romito & Soriano (1998) nonché una occlusiva sia scempia [d] che doppia [d:].

<sup>2</sup> Bianco E. (1981), “Gli sviluppi di LL in Calabria”, in M. Cortelazzo (cur.), *La ricerca dialettale*, 12, Pisa: Pacini.

<sup>3</sup> Falcone G. (1976), *Calabria*, Pisa: Pacini.

Secondo quanto riportano Bianco e altri autori, la Calabria meridionale costituisce una delle pochissime zone romanze che attestano la presenza di uno stadio retroflesso non ancora delaterizzato (insieme all'Abruzzo e alla Sardegna).

In Calabria è presente una diffusione molto ampia degli esiti cacuminali dei gruppi -tr- e -str-, ma si può riscontrare una diffusione ancora più diversificata e irregolare degli esiti cacuminali di -LL-.

Romito & Belluscio (1996)<sup>4</sup> attraverso un'analisi elettropalatografica hanno evidenziato la differenza tra l'articolazione di /dd/ e quella di /d̠d̠/ nel dialetto della città di Catanzaro. Mentre per la prima si può parlare chiaramente di geminata e di occlusiva, vista la sua durata e la sua totale chiusura, lo stesso non può essere affermato per la realizzazione di /d̠d̠/.

Romito, Trumper, Mendicino<sup>5</sup> basano le loro misurazioni sia sul livello elettropalatografico (con lo studio dei coefficienti di anteriorità<sup>6</sup>) che su quello elettroacustico. Gli stessi autori dimostrano che per il dialetto di Catanzaro, il fonema /d/ è articolato in posizione dentale, mentre /l:/ e /d̠:/ sono decisamente più arretrati e inoltre che, mentre /d̠:/ perde contatto, /d:/ ne guadagna.

## 2. ANALISI PERCETTIVA ED ELETTROACUSTICA

La seconda fase del nostro lavoro si è concentrata sull'analisi percettiva ed elettroacustica di alcuni dialetti della provincia di Catanzaro esattamente: Catanzaro, Amato, Badolato, Borgia, Botricello, Falerna, Girifalco, Lamezia Terme, Maida, Pentone, Pianopoli, Sellia, Taverna e Zagarise.



Figura 1: Comuni nella Provincia di Catanzaro presi in analisi

<sup>4</sup> Studio elettropalatografico dell'opposizione fonematica [ll], [dd], [d̠d̠] nel dialetto di Catanzaro e [ɬ], [ʎ], [d], [ð] nella parlata albanese di San Basile, AIA XXIV Convegno Nazionale Trento, 12-14 giugno 1996.

<sup>5</sup> Luciano Romito et al (2000).

<sup>6</sup> Fontdevila J., M.D. Pallarès, D. Recasens (1994), The Contact Index Method of Electropalatographic Data Reduction, JP 22.

La scelta di questi comuni è dettata dalla volontà di analizzare il fenomeno partendo dalla città di Catanzaro, nelle sue diramazioni verso il nord (Pre-Sila e Sila), verso la costa ionica (settentrionale e meridionale), ed infine ad ovest verso la costa tirrenica (passando dalla zona collinare catanzarese verso la Piana di Lamezia ).

La nostra indagine è partita dal presupposto di considerare Catanzaro come punto di partenza e “centro guida” su cui poter confrontare i dati successivamente raccolti.

### 3. MATERIALI E METODI

Sulla base degli studi e delle ricerche precedentemente citate è stato strutturato il materiale linguistico utile alla nostra analisi predisponendo, in un primo momento, una lista di parole “*non-sense*” ed in seguito una lista di circa sessanta parole di *uso comune* in cui, il nesso consonantico in esame è stato posto in diversi contesti intervocalici ed accentuali al fine di analizzare la durata temporale della consonante retroflessa e della vocale conseguente, oltre che i valori frequenziali delle prime quattro formanti della vocale in questione (analisi della porzione stazionaria e della transizione della vocale precedente e conseguente il nesso).

Ci si è avvalsi infatti della teoria dei loci al fine di individuare i luoghi dei segmenti presi in esame, osservando l’andamento delle frequenze formantiche delle vocali adiacenti.

La lista, inoltre, è stata inserita in frasi cornice al fine di ricreare una condizione di fonazione il più simile possibile alla normale coarticolazione comprendendo anche parole contenenti la realizzazione dei foni scempi e geminati [d, dd] e [l, ll], sempre in contesto intervocalico, con lo scopo di effettuare un controllo e una comparazione con il fono oggetto di indagine.

Le registrazioni, effettuate nei quattordici comuni presi in analisi, sono state rivolte a otto parlanti per paese, di sesso maschile, suddivisi in due fasce di età al fine di individuarne una possibile evoluzione ed effettuate in ambienti informali al fine di garantire una maggiore spontaneità delle produzioni.

Tutte le registrazioni effettuate sono state archiviate e catalogate nel già citato Archivio Sonoro Calabrese.

### 4. OSSERVAZIONE ED ANALISI DEI DATI

I dialetti in cui sono stati riscontrati esiti occlusivi, sicuramente subapicali, comprendono otto comuni dei quattordici analizzati, ossia Amato, Borgia, Botricello, Catanzaro, Falerna, Girifalco, Sellia e Taverna.

Successivamente ad una prima analisi percettiva, si è effettuato un confronto dei valori frequenziali di transizione della seconda formante delle vocali adiacenti all’esito nei dialetti presi in esame, con i valori di transizione della seconda formante dei modelli di riferimento costituiti dalle vocali poste in contesti occlusivi.

Per verificare, invece, sia la natura geminata che occlusiva di tali segmenti si è confrontata in un primo luogo la durata dell’intero segmento con i foni di controllo sicuramente occlusivi e geminati e, in un secondo momento, si è confrontata la durata della fase di rilascio dei due suddetti segmenti.

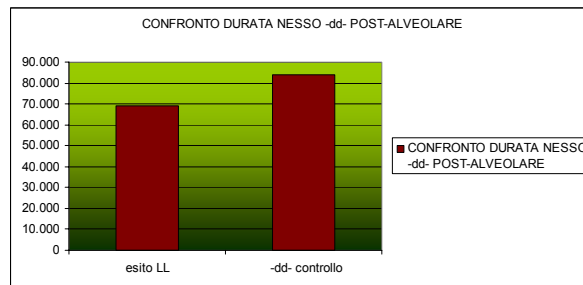


Figura 2: Confronto durata del nesso -dd- postalveolare

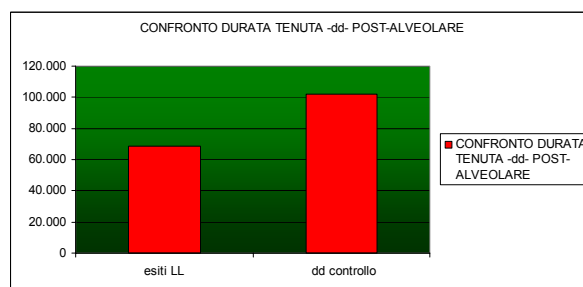


Figura 3: Confronto durata tenuta del nesso -dd- postalveolare

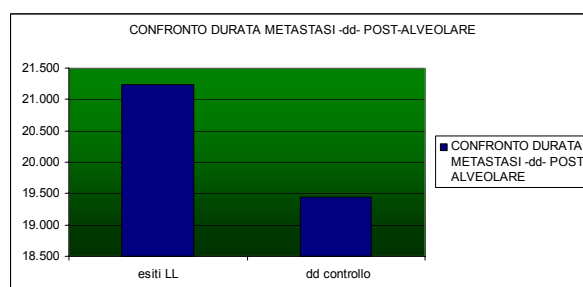


Figura 4: Confronto durata metastasi del nesso -dd- postalveolare

A conferma delle nostre ipotesi, si è notato, che il nesso in esame è sicuramente geminato, ma che è visibilmente più breve nella sua durata totale, rispetto alla produzione del fono di controllo. Per quanto riguarda il confronto fra la fase di metastasi dei segmenti tendenzialmente post-alveolari, si è riscontrata una durata maggiore che conferma gli studi precedentemente effettuati su tali tipi di dialetto (Romito-Belluscio, 1995; Romito-Sorianello, 1999) che finiscono per considerare l'esito latino come affricato sonoro lungo e cacuminale più che retroflesso.

Infine, a conferma di quanto detto, sulla base dell'osservazione del sonogramma, si è infatti notata la presenza di frizione durante la durata del 'burst' che ha confermato la presenza di fruscio ad alte frequenze tipica di un suono lievemente affricato e la presenza costante della barra vocale nelle basse frequenze rivela che il fono in questione è sicuramente sonoro.

#### 4.1 Esiti laterali di tipo -ll-

I dialetti in cui l'esito -LL- sembra essersi conservato sono inferiori in numero rispetto a quelli occlusivi e precisamente riguardano quattro paesi dei quattordici analizzati, ossia Lamezia Terme, Pentone, Pianopoli e Zagarise.

Da una prima analisi attraverso l'osservazione del sonogramma e dell'oscillogramma, si è riscontrata la struttura formantica tipica di una laterale e quasi la stessa ampiezza d'onda rispetto alle vocali adiacenti.

Dal confronto dei valori frequenziali di transizione delle vocali adiacenti con i modelli di riferimento, si è osservato che il luogo di articolazione dei suddetti dialetti è tendenzialmente post-alveolare.

Infine, si è confrontata la durata dei diversi esiti per confermare la lunghezza del segmento, considerandolo dai risultati riscontrati come sicuramente geminato ma visibilmente più breve nella sua durata rispetto alla produzione dei fono di controllo.

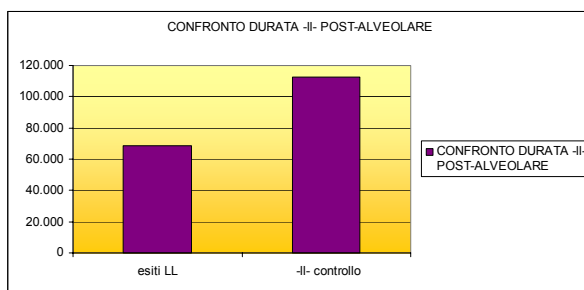


Figura 5: Confronto durata -ll- postalveolare

#### 4.2 Esito vibrante di tipo /r/

Dalle analisi acustiche effettuate, solo uno dei dialetti analizzati sembra avere tali caratteristiche, vale a dire il dialetto del comune di Badolato.

L'esito -LL- nel dialetto di Badolato presenta delle caratteristiche acustiche del tutto differenti rispetto alla maggior parte dei dialetti sul territorio della provincia di Catanzaro. Tali caratteristiche acustiche sono state facilmente percepibili anche solo attraverso una prima analisi uditiva e successivamente confermati attraverso un'analisi specifica dei dati acustici rilevati.

In tale dialetto l'esito -LL- assume caratteristiche acustiche simile ad un fono vibrante /r/. Si è notato che le formanti sul sonogramma presentano quell'alternanza di silenzi relativi alla chiusura del tratto e di bande formantiche relativi alla sua apertura tipici di un suono vibrante.

Infine, per confermare ulteriormente le caratteristiche acustiche di tale esito simile a vibrante, si è notato che il valore della prima formante tende sempre ad essere intorno ai 400 Hz, così come evidenzia Fant (1960) nella descrizione di tali foni.

| Filename        | Fonema  | F1         | F2   | F3   |
|-----------------|---------|------------|------|------|
| <b>Badolato</b> | ca-rr-u | <b>469</b> | 1371 | 2056 |
| <b>Badolato</b> | ca-rr-u | <b>397</b> | 1168 | 2381 |
| <b>Badolato</b> | co-r-u  | <b>383</b> | 1149 | 1896 |
| <b>Badolato</b> | i-r-a   | <b>410</b> | 1754 | 2526 |
| <b>Badolato</b> | i-r-u   | <b>367</b> | 1303 | 1940 |
| <b>Badolato</b> | ca-rr-u | <b>342</b> | 1122 | 2089 |
| <b>Badolato</b> | co-rr-u | <b>426</b> | 1072 | 1552 |
| <b>Badolato</b> | i-r-a   | <b>447</b> | 1781 | 2475 |
| <b>Badolato</b> | i-r-a   | <b>322</b> | 1821 | 2695 |
| <b>MEDIA</b>    |         | <b>395</b> |      |      |

Tabella 1: Valori formantici dell'esito -ll- nel Comune di Badolato

#### 4.3 Esiti semivocalici di tipo /j/

Dalle analisi acustiche effettuate, solo uno dei dialetti analizzati sembra avere tali caratteristiche, vale a dire il dialetto del comune di Maida.

Tali caratteristiche acustiche sono state facilmente percepibili anche solo attraverso una prima analisi uditiva e successivamente confermati attraverso un'analisi specifica dei dati acustici rilevati.

In tale dialetto l'esito -LL- assume caratteristiche acustiche simile ad una semivocale anteriore tipica del dialetto meridionale /j/.

| Filename         | Fonema   | F1         | F2          | F3   |
|------------------|----------|------------|-------------|------|
| <b>maida2050</b> | ca-j-u   | <b>297</b> | <b>1937</b> | 2535 |
| <b>maida2050</b> | ca-j-u   | <b>363</b> | <b>1852</b> | 2398 |
| <b>maida2050</b> | capi-j-i | <b>250</b> | <b>2045</b> | 3176 |
| <b>maida2050</b> | cuo-j-u  | <b>323</b> | <b>1831</b> | 2398 |
| <b>maida2050</b> | ga-j-na  | <b>350</b> | <b>1999</b> | 3039 |
| <b>maida5080</b> | capi-j-i | <b>448</b> | <b>1925</b> | 2689 |
| <b>maida5080</b> | co-j-u   | <b>354</b> | <b>1802</b> | 2323 |
| <b>maida5080</b> | i-j-a    | <b>341</b> | <b>1932</b> | 2748 |
| <b>maida5080</b> | mo-j-ica | <b>334</b> | <b>1858</b> | 2528 |
| <b>maida5080</b> | sta-j-a  | <b>400</b> | <b>1891</b> | 2849 |
| <b>MEDIA</b>     |          | <b>346</b> |             |      |

Tabella 2: Valori formantici dell'esito -ll- nel Comune di Maida



## 5. NESSI CON VIBRANTE

### 5.1 *La retroflessione nel contesto vibrante*

Uno dei contesti principali per lo sviluppo delle consonanti retroflesse è la presenza di una vibrante precedente. Bhat (1973:43) osserva che una mono- o polivibrante alveolare ha la proprietà di indurre la retroflessione di una consonante seguente, anche senza possedere essa stessa una pronuncia retroflessa. Accanto alla retroflessione indotta da una vibrante precedente, è possibile che una consonante diventi retroflessa anche per effetto di una vibrante seguente. Questo processo è considerato minoritario da parte di Hamann (2003:86), ma in realtà esempi di questo tipo di processo non sembrano mancare nelle lingue del mondo.

Bhat (1973) cita quattro esempi di retroflessione indotta da una /r/ seguente, uno dei quali si riferisce al lugamba, lingua parlata in Uganda in cui /t/ e /d/ diventano retroflesse quando si trovano davanti ad una fricativa retroflessa originatasi da /r/; della lingua cham del Vietnam, dove /tr/ > /ʈ/ e di alcune lingue sinotibetane dove tutti i nessi iniziali di occlusiva seguita da /r/ hanno dato origine ad una retroflessa. Hamann aggiunge inoltre i casi della lingua pakistana yidgha (/str/ e /sr/ > /ʂ/) del pashto e altre lingue iraniche orientali moderne (/sr/ > /ʂ/), in cui /n/ ha generalmente una pronuncia postalveolare nel gruppo /nr/ che viene realizzato come /ŋd/ (Balasubramanian 1982a).

La retroflessione dei nessi con occlusiva coronale seguita da /r/ è presente anche nei dialetti romanzi, in particolare nel siciliano, calabrese meridionale, salentino meridionale e còrso settentrionale. In queste parlate, la realizzazione retroflessa corrisponde generalmente ad un'affricata [ʈ(:)ʂ], come è stato dimostrato sperimentalmente da Sorianello & Mancuso (1998).

La retroflessione del gruppo /tr/ è stata ampiamente documentata per i dialetti italiani meridionali e studi recenti hanno mostrato che l'esito del processo corrisponde molto spesso alla realizzazione monofonematica di un'affricata [ʈ(:)r]<sup>7</sup>.

Un solo correlato acustico è stato individuato come caratterizzante, cioè l'abbassamento della terza formante. Questo fattore è correlato principalmente con il gesto di ritrazione della radice della lingua ed è presente ogni qualvolta l'apice della lingua venga lasciato libero di muoversi e di vibrare.

L'abbassamento della terza formante, com'è noto, costituisce un tratto che le vibranti condividono con i suoni retroflessi. In alcuni casi, una flessione particolarmente marcata di F3 nelle transizioni permette di differenziare una vibrante postalveolare retroflessa da una alveolare non retroflessa.

Tuttavia, nel parlato spontaneo, quando la velocità di eloquio è sostenuta, è difficile poter stabilire l'effettivo grado di retroflessione di un suono così breve come una monovibrante o un'approssimante apicale.

### 5.2 *Analisi dei dati*

Analizzando percettivamente e acusticamente tali nessi non si è riscontrata alcuna differenza preponderante nei vari dialetti. Di seguito i valori formantici successivi al fruscio intorno ai 400 Hz, caratteristici della presenza di una vibrante:

---

<sup>7</sup> Cf. Sorianello & Mancuso (1998), Loporcaro (2001).

| NESSO | F1  | F2   |
|-------|-----|------|
| -str- | 464 | 2181 |
| -tr-  | 408 | 1673 |
| -ntr- | 469 | 1642 |
| MEDIA | 447 |      |

Tabella 3: Valori formantici dei nessi con vibrante

In conclusione, sulla base dei dati acustici rilevati dall'analisi dei nessi con vibrante, -str-, -tr-, ed -ntr-, si è potuto osservare che tali nessi possono essere definiti come costituiti dall'insieme di due suoni, ossia da un nesso retroflesso-tr (tr- > [tɾ(r)]) preceduti da una nasale nel caso di -ntr-, e da una fricativa più vibrante nel caso di str- (str- > [ʃr]).

## 6. CONCLUSIONI

Dalla mappatura elaborata si è riscontrata una chiara preponderanza di esiti occlusivi di tipo /dd/, a partire da Catanzaro, nei comuni pre-silani fino ad arrivare sulla costa ionica.

Ciò che risulta dalle nostre analisi acustiche è la rilevanza che riveste l'esito retroflesso, non tanto per le sue peculiari caratteristiche riscontrabili dal punto di vista fonetico, ma per il carattere distintivo che dal punto di vista fonologico permette i parlanti dei dialetti, in cui questo fenomeno si verifica, di poter utilizzare il tratto distintivo [± arretrato] per opporre parole dal significato differente: ['kaɖɖu] 'callo' vs ['kaddu] 'caldo'.

Si è rilevato, tuttavia, anche un esito occlusivo nel comune di Falerna, circondato in realtà da dialetti con esiti /ll/, tipico della Piana di Lamezia. Del tutto differenti invece, gli esiti riscontrati in due soli dialetti, quello di Maida e di Badolato in cui si sono rilevati esiti quali /r/ e /j/.



Figura 6: Mappatura degli esiti -ll- nei comuni analizzati nella Provincia di Catanzaro

Confrontando i dati da noi rilevati con gli esiti riscontrati nelle analisi effettuate nel corso degli anni (vedi tabella), sia dialettologiche che elettroacustiche nell'area di nostro interesse, abbiamo riscontrato una concordanza nei risultati negli esiti di Catanzaro,

Lamezia e Badolato (Rholfs 1966, Falcone 1976, Bianco 1981, Romito-Sorianello 1998, Trumper-Galdini 1999).

Per quanto riguarda invece il dialetto del comune di Maida l'analisi acustica da noi effettuata contrasta con i dati dialettologici riscontrati in precedenza (Rholfs 1966, Falcone 1976, Bianco 1981), i quali evidenziano un esito occlusivo cacuminale del tutto simile a quello da essi riscontrato a Catanzaro.

Gli stessi studi dialettologici hanno evidenziato nel dialetto di Falerna un esito occlusivo cacuminale scempio che discorda con l'esito lungo cacuminale lievemente affricato da noi rilevato.

Per questi esiti discordanti, come per i restanti esiti presenti nella mappatura privi di altri riferimenti di ricerca con cui poter effettuare un confronto, ci proponiamo di effettuare in futuro ulteriori analisi elettroacustiche in modo da riuscire a fornire un quadro diacronico sempre più preciso dell'esito di ll in tali dialetti.

| COMUNI     | Rholfs<br>1966 | Falcone<br>1976 | Bianco<br>1981 | Romito-Sorianello<br>1998 | Trumper-Galdini<br>1999 | Romito-Scuticchio<br>2008 |
|------------|----------------|-----------------|----------------|---------------------------|-------------------------|---------------------------|
| Amato      |                |                 |                |                           |                         | ɖɖʒ                       |
| Badolato   | ɾ              | ɾ               | ɾ              |                           |                         | r                         |
| Borgia     |                |                 |                |                           |                         | ɖɖʒ                       |
| Botricello |                |                 |                |                           |                         | ɖɖʒ                       |
| Catanzaro  | dd             | dd              | dd             | ɖɖʒ                       | ɖ:                      | ɖɖʒ                       |
| Falerna    | dd             | dd              | dd             |                           |                         | ɖɖʒ                       |
| Girifalco  |                |                 |                |                           |                         | ɖɖʒ                       |
| Lamezia    | ll             | ll              | ll             |                           |                         | l:                        |
| Maida      | dd             | dd              | dd             |                           |                         | j                         |
| Pentone    |                |                 |                |                           |                         | l:                        |
| Pianopoli  |                |                 |                |                           |                         | l:                        |
| Sellia     |                |                 |                |                           |                         | ɖɖʒ                       |
| Taverna    | d              | d               | d              |                           |                         | ɖɖʒ                       |
| Zagarise   |                |                 |                |                           |                         | l:                        |

Tabella 4: Confronto esiti –ll– con studi dialettologici ed elettroacustici effettuati in precedenza

## BIBLIOGRAFIA

- Bhat D.N.S. (1974b), "Retroflexion and retraction", *Journal of Phonetics* 2, 233-237;
- Bianco E. (1981), "Gli sviluppi di LL in Calabria", in M. Cortelazzo (cur.), *La ricerca dialettale*, 12, Pisa: Pacini;
- Celata C. (2002), "Fonetica della palatalizzazione delle velari in romanzo". *Quaderni del Laboratorio di Linguistica*, Scuola Normale Superiore 3, 119-138;
- Falcone, G., (1991) *Appunti per un saggio storico-linguistico sui dialetti della Calabria*, in *Vivarium Scyllacense, Bollettino dell'Istituto di Studi su Cassiodoro e sul Medioevo in Calabria*, Davoli Marina (CZ): Sud Grafica;
- Hála P.B. (1964), "Apical, cacuminal, rétroflexe, coronal, dorsal", *Phonetica* 11, 186-195.
- Hamann S. (2002a), "Retroflexion and Retraction revised". In T.A. Hall, B. Pompino-Marschall & M. Rochon (eds.), *ZAS Working Papers* 28, 13-25.
- Ladefoged P. & I. Maddieson (1996), *The sounds of the world's languages*, Oxford: Blackwell;
- Rohlf G. (1932), *Dizionario dialettale delle Tre Calabrie*, I, Halle / Milano: M. Niemeyer / Hoepli.
- Rohlf G. (1966[=1949]), *Grammatica storica dell'italiano e dei suoi dialetti, Fonetica*, Torino Einaudi;
- Rohlf G. (1985), *Latinità ed ellenismo nel mezzogiorno d'Italia, Studi e ricerche dalla Magna Grecia alla Grecia Italica*, Frama Sud ;
- Romito L. & G.M.G. Belluscio (1996), "Studio elettropalatografico dell'opposizione fonematica [ll], [dd], [lll] nel dialetto di Catanzaro e [...], [ʎ], [dd] nella parlata albanese di San Basile". In *Atti del XXIV Convegno Nazionale dell'A.I.A.*, Trento 12-14 giugno 1996, 141-144;
- Romito L. & P. Sorianello (1998), "Ridefinizione delle consonanti retroflesse nei dialetti calabresi". Comunicazione presentata al V Convegno della Società Internazionale di Linguistica e Filologia Italiana (S.I.L.F.I.), Catania, ottobre 1998;
- Sorianello P. & A. Mancuso (1998), "Le consonanti retroflesse nel cosentino: analisi preliminare". In *Atti delle VIII Giornate di Studio del GFS*, Pisa 18-19 dicembre 1997, 142-154;
- Trumper J. & M. Maddalon (1988), "Converging divergence and diverging convergence: the dialect language conflict and contrasting evolutionary trends in modern Italy". In J.C. Auer & A. Di Luzio (eds.), *Variation and Convergence*, Berlin: Mouton de Gruyter, 217-259.

# LA METAFONIA IN ALCUNI CENTRI DEL NORD CALABRIA: VERSO UNA MAPPA REGIONALE

Luciano Romito & Daniela Gagliardi  
Laboratorio di Fonetica-Università della Calabria  
*luciano.romito@unical.it; fradalu2002libero.it*

## 1. INTRODUZIONE

Il fatto che la Calabria<sup>1</sup>, pur così nettamente definita e rinchiusa dai suoi confini naturali, sia linguisticamente una regione tutt'altro che unitaria, è noto e riportato in lavori scientifici fin dell'inizio del '900. Le cause sono sicuramente da ricercare nella sua stessa dislocazione geografica e nell'isolamento dovuto alle vicende storiche della regione nel contesto italiano e meridionale. In base ai dati tradizionali e al vocalismo tonico, si possono distinguere tre aree linguistiche: a) area meridionale, fino al limite di una linea che va da Vibo Valentia a Punta Stilo; b) area centrale, che comprende i paesi della provincia di Catanzaro; c) area settentrionale, rappresentata dai dialetti dei paesi della provincia di Cosenza.

A questa suddivisione si aggiunge, nella parte più settentrionale, la fascia di territorio linguistico comune alla Lucania, la cosiddetta zona Lausberg, a vocalismo arcaico e conservazione delle vocali finali, il cui limite meridionale calabrese si estende da Cirella sul Tirreno a Villapiana sullo Ionio, girando al disopra del Monte Pollino (per una completa trattazione della suddivisione della Calabria si veda Romito L. et al. 2005).

La Calabria, come la Basilicata, è territorio di conservazione più che di innovazione linguistica. La situazione odierna tuttavia, è tutta da verificare, alla luce non soltanto dell'evoluzione che si è prodotta negli ultimi decenni, ma anche degli sconvolgimenti sociali di questi ultimi anni (emigrazione, mutamenti socio-economici, ecc.), che hanno alterato profondamente il tessuto culturale tradizionale della regione.

In relazione a quanto detto, l'obiettivo di questo lavoro è contribuire alla definizione di una mappatura completa della distribuzione e caratterizzazione del fenomeno in Calabria.

Il lavoro prevede la consultazione di una ricca "letteratura" comprensiva di atlanti dialettali e di studi di dialettologia<sup>2</sup>.

Il presente lavoro non può, dunque, evitare di tenere in considerazione che nucleo della fonetica di un dialetto è il sistema vocalico, ossia lo schema secondo il quale le vocali *toniche* del latino evolsero verso le corrispondenti del dialetto.

La metaforia colpisce le vocali toniche *e/ε* ed *o/ɔ* quando la vocale finale della parola originaria latina è *i* oppure *u*. In particolare, ciò avviene per i sostantivi e gli aggettivi maschili singolari (terminazioni latine *-us*, *-um*) e plurali (terminazione latina *-i*), rispetto ai

---

<sup>1</sup> Riguardo la situazione dialettale presente in Calabria e la sua ineludibile frammentarietà si rimanda ai lavori di Trumper et al. 1981, 1988, 1995, 1997, Falcone 1976, Maiden 1997, Rohlf 1966, Bianco 1981 e Martino 1991.

<sup>2</sup> Es.: Romito et al. (1996) e Loparco et al. (1998); Rensch (1964); Falcone (1991); Martino (1991); Trumper (1997).

corrispondenti femminili singolari e plurali (terminazioni *-a*, *-ae*), ma anche nella coniugazione di alcuni verbi<sup>3</sup>.

La metafonìa può avvenire quindi:

- per **innalzamento**: la *e* e la *o* passano normalmente a *i* e, rispettivamente a *u*;
- per **dittongazione**: la *e* e la *o* dittongano generalmente in *ie* e in *uo* (ma anche *ua*) ascendenti o discendenti;
- per **monottongamento**: molto spesso, il dittongo è ritratto sul primo componente, e così l'esito metafonetico diventa [ie] > [i:], [uo] > [u:].

Nelle varie indagini sul fenomeno linguistico della metafonìa in Calabria, è emersa tale situazione sicuramente degna di attenzione. Il sud della zona arcaica conosciuta come Mittelzone<sup>4</sup> è caratterizzata dalla presenza di molti dialetti in cui i dittonghi metafonici si sono successivamente monottongati dando come esito fonetico, *in sillaba chiusa*, a delle vocali lunghe [i:] ed [u:].<sup>5</sup> L'evoluzione si sarebbe definita in questi termini: \*['lettʉ] > ['liettʉ] > ['li:ttʉ], \*['koktʉ] > ['kuottʉ] > ['ku:ttʉ]. Il secondo elemento del dittongo originario cade, ma la vocale conserva lo stesso la durata originaria. In netta contraddizione con la teoria fonologica, ci troveremmo di fronte ad una sillaba chiusa tonica con vocale lunga, definita come "extralunga" o "superpesante".

## 2. MATERIALI E METODI

Le nostre ricerche sul campo si sono concentrate nella Calabria Nord Occidentale, è più precisamente in quell'area del cosentino sud orientale che accoglie i paesi di Rogliano, Mangone, Figline Vegliaturo, Pianecrati, Aprigliano, Pietrafitta, Pedace, Serra Pedace e spostandoci più ad est San Giovanni in Fiore (nel lavoro precedente di Romito et al. 2005 era stata investigata l'Area Lausberg). I dialetti di questi paesi rispecchiano un tipo di vocalismo tipicamente 'siciliano', a tre gradi di apertura e fanno parte del secondo 'corridoio' individuato da Trumper (1988). Quest'area come ben sappiamo, è fortemente caratterizzata dalla presenza della metafonìa, e sebbene l'esito della dittongazione sia quello preponderante, (Aprigliano, Pietrafitta, Pedace, Serra Pedace, San Giovanni in Fiore presentano infatti metafonìa per dittongazione) sono tanti i paesi i cui dialetti presentano esiti metafonici per innalzamento. Infatti nei dialetti di Pianecrati, Figline Vegliaturo, Mangone e Rogliano è possibile verificare un chiusura delle vocali toniche medio basse /ɛ/ ed /ɔ/ in /e/ ed /o/ in presenza di vocali atone finali come /i/ ed /u/ facendo sì che questi paesi, normalmente caratterizzati da un sistema vocalico a tre gradi di apertura, ne sviluppino uno a cinque nel caso di esiti metafonici.

<sup>3</sup> Nella stessa regione, in alcuni paesi della provincia di Cosenza come Belvedere, Cassano ecc è presente anche una dittongazione spontanea (da lavori in corso presso il Laboratorio di Fonetica dell'UNICAL).

<sup>4</sup> Martino P., *L'area Lausberg*, editrice Il Calamo, Roma, 1991.

<sup>5</sup> Cfr. Trumper B. J. (1997), *Calabria and South Basilicata*, in Maiden M. e Parry M. (eds.), *The dialects of Italy*, Routledge, Londra; New York, pp. 355-64; Abete G. (2005), Sulla questione della sillaba superpesante: I dittonghi discendenti in sillaba chiusa nel dialetto di Pozzuoli, in AISV 2005 II Convegno Nazionale, Analisi prosodica - Teorie, modelli e sistemi di annotazione; Romito L., Galatà V., Lio L., Stillo F. (2005), La metafonìa nell'area di Lausberg: una introspezione sulla natura della sillaba, in AISV 2005 II Convegno Nazionale, Analisi prosodica - Teorie, modelli e sistemi di annotazione.

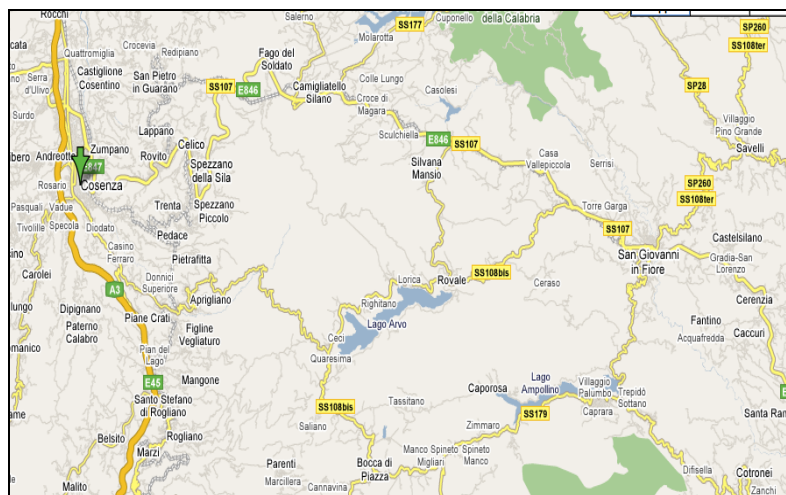


Figura 1: Area oggetto delle nostre indagini

Il campione scelto è femminile e questo perché meno esposto alle contaminazioni linguistiche derivate da esperienze lavorative in altre regioni o in altre nazioni, o più semplicemente perché per varie contingenze si ritrovano a lavorare in realtà che non sono sempre legate al paese d'origine.

Il campione è suddiviso in base alla variabile età 30 e 40 anni e Over 50 tutte con un grado di scolarizzazione medio. Tale differenziazione del campione è motivata dalla necessità di verificare una eventuale trasformazione ed evoluzione del fenomeno nel tempo. Tutte le registrazioni sono state archiviate e catalogate nell'Archivio Sonoro Calabrese<sup>6</sup>.

Il materiale linguistico utilizzato è stato definito attraverso una lista di parole programmata in funzione delle ipotesi fonologiche di partenza. Di conseguenza tale lista doveva includere parole che presentassero delle alternanze metafoniche al plurale o al singolare dei sostantivi e nei verbi (le prime tre persone del presente indicativo). Le parole includono per lo più bisillabi e trisillabi in cui le vocali toniche (nodo focale delle nostre ricerche) si trovano sia in sillaba aperta che in sillaba chiusa.

### 3. ANALISI DEI DATI

Le nostre analisi si sono definite in funzione della verifica di una serie di ipotesi di partenza:

- accertare innanzitutto la presenza o meno della metaforia nell'area posta ad indagine;

<sup>6</sup> Archivio di materiali sonoro catalogati e conservati per scopi di ricerca, progettato e prodotto all'interno di un progetto di ricerca in atto presso il Laboratorio di Fonetica dell'Università della Calabria dal titolo "I corpora vocali come patrimonio culturale. Archiviazione, conservazione e restauro di materiale sonoro calabrese" (Romito&Lio, 2006)

- determinare la tipologia di metafonìa, vale a dire se i paesi siano interessati da metafonìa per innalzamento o per dittongazione (in quest'ultimo caso stabilire se il dittongo è ascendente o discendente);
- verificare la presenza o meno di trasformazioni ed evoluzioni del fenomeno in funzione della variabile dell'età.

I metodi di indagine ai quali si è fatto riferimento riguardano l'analisi formantica delle vocali toniche semplici in sillaba aperta e sillaba chiusa e l'analisi formantica delle vocali toniche metafoniche. Sono stati, inoltre, elaborati i grafici delle *aree di esistenza vocalica* comprendenti le vocali toniche semplici e le vocali metafoniche nei paesi caratterizzati da metafonìa per innalzamento. Per i paesi interessati da metafonìa per dittongazione sono stati elaborati i grafici delle aree di esistenza vocalica all'interno dei quali sono state inserite le medie delle vocali dei dittonghi esito metafonico, sia per quanto riguarda l'elemento tonico che quello atono. In quest'ultimo caso è stato anche analizzato l'andamento del pitch per stabilire se il dittongo è ascendente o discendente.

### 3.1 Metafonìa per dittongazione

Le nostre analisi acustiche, relative ai paesi di Aprigliano, Pietrafitta, Pedace, Serra Pedace, S. Giovanni in Fiore, accertano e confermano la presenza di metafonìa per dittongazione.

L'andamento della F0, la misura della durata e della struttura formantica confermano che in entrambe le fasce di età prese in considerazione il dittongo metafonico è di tipo discendente.

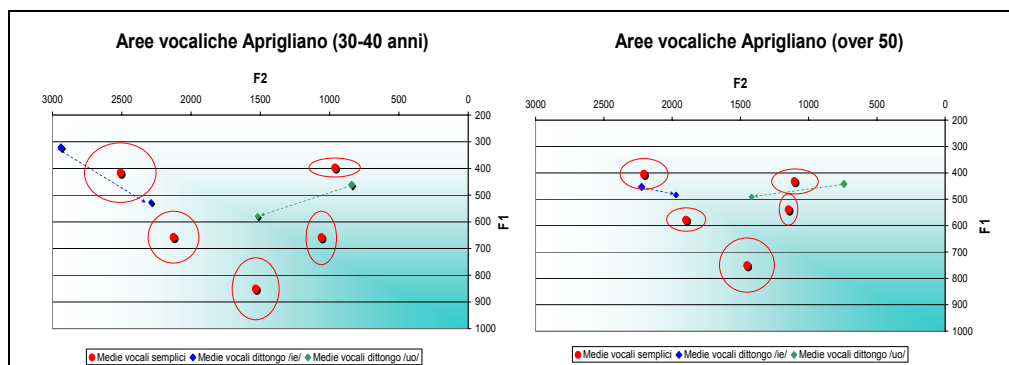


Figura 2: Aprigliano



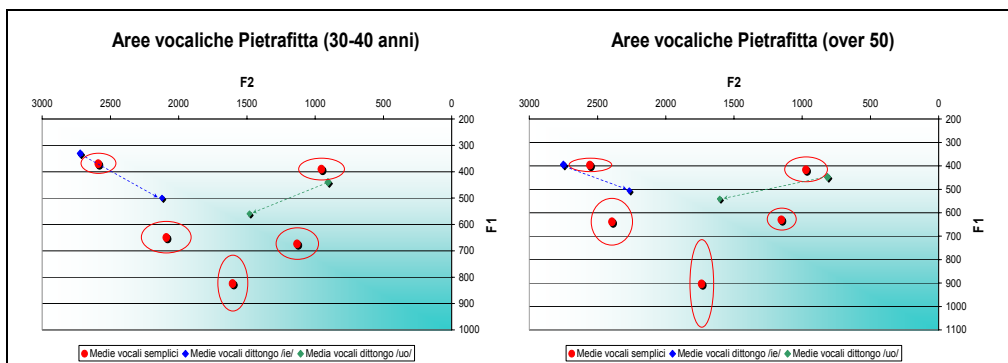


Figura 3: Pietrafitta

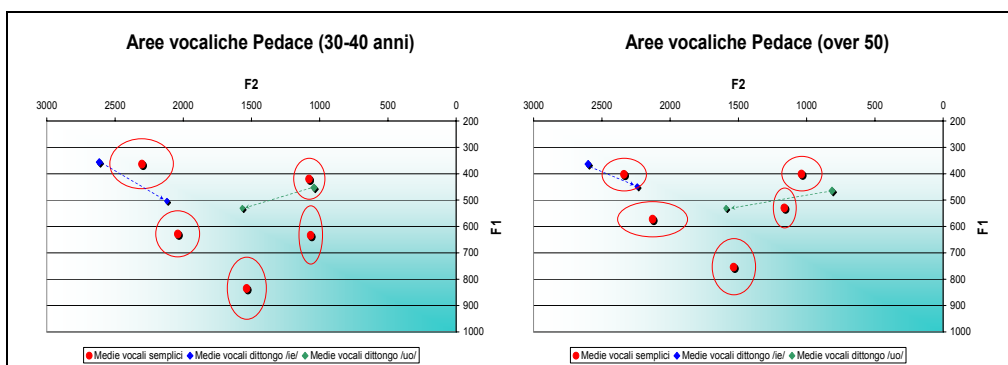


Figura 4: Pedace

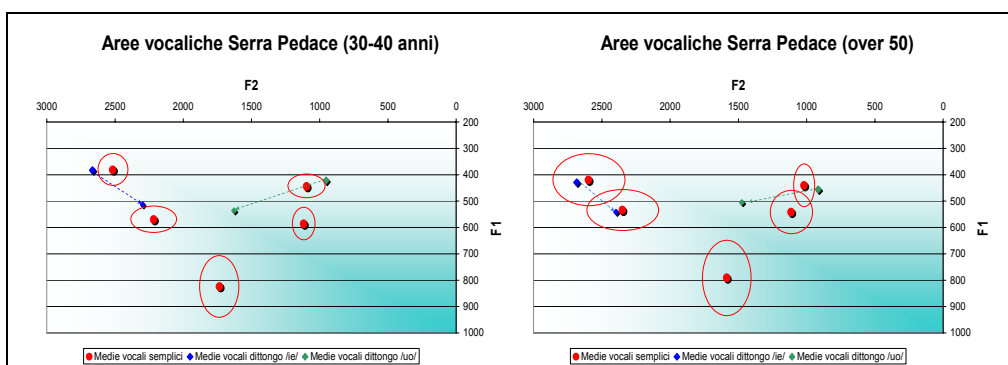


Figura 5: Serra Pedace

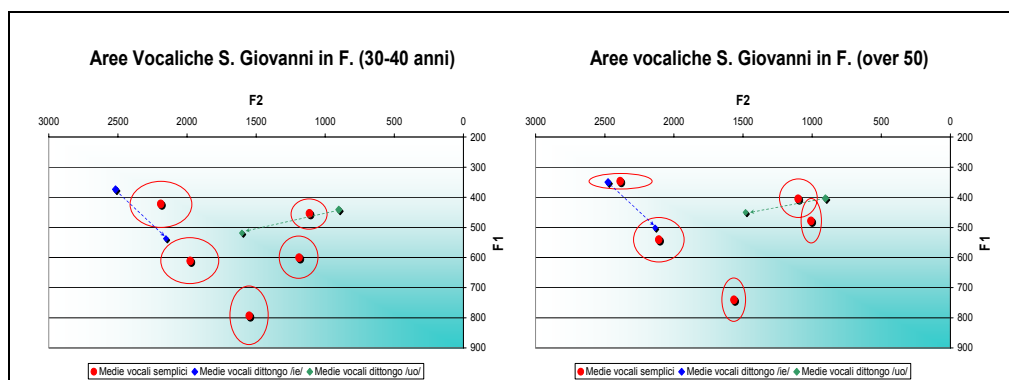


Figura 6: San Giovanni in Fiore

### 3.2 Metafonia per innalzamento

Pianecrati rientra in quei paesi del cosentino sud orientale in cui il fenomeno della metaforia determina trasformazioni particolari per quanto riguarda l'inventario fonologico dei parlanti. Questo paese (così come tutti quelli da noi studiati) si colloca in quell'area caratterizzata da vocalismo siciliano a tre gradi di apertura e cinque vocali, in cui l'esito metafonico dittongante è quello preponderante. Ma a Pianecrati (così come a Figline Vegliaturo, Mangone e Rogliano, anch'essi paesi oggetto del nostro studio) la metaforia si realizza non per dittongazione ma per innalzamento, con una particolarità: l'evoluzione non è da vocali toniche medie a vocali toniche alte (/e/ > /i/ ed /o/ > /u/), ma da medio basse a medio alte (/ɛ/ > /e/ ed /ɔ/ > /o/), facendo sì che un normale sistema a tre gradi e cinque vocali si evolva (in presenza di metaforia) in un sistema a cinque gradi e sette vocali. Ciò trova conferma nei grafici che rappresentano le aree di esistenza vocalica dei parlanti del paese, analizzati nelle due fasce di età 30-40 anni e over 50 dalla analisi statistiche.

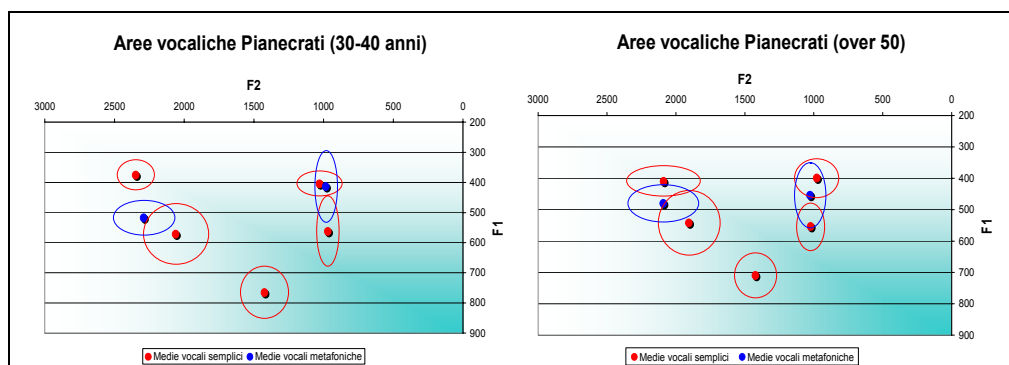


Figura 7: Aree di esistenza vocalica-Pianecrati

L'analisi T-student, utilizzata per la comparazione dei dati, dimostra che la differenza tra i valori formantici della /e/ e della /ɛ/ oscillano tra un 79 ed un 99%. Un dato interessante che non conferma quanto affermato riguarda il confronto tra la vocale /o/ metafonica e la vocale /u/ nei giovani. In questo caso la dittongazione avviene proprio per innalzamento come se il processo fosse ancora in atto e soprattutto nei giovani venisse reinterpretato volgendo ad una semplificazione del sistema.

| TABELLA CONFRONTO T-Test (Valori F1) |             |             |             |             |
|--------------------------------------|-------------|-------------|-------------|-------------|
| 30-40 anni                           |             |             |             |             |
|                                      | /ɔ/ vs /o/M | /ɛ/ vs /e/M | /u/ vs /o/M | /i/ vs /e/M |
| Valore T-Test                        | 0,015219    | 0,21053     | 0,792212    | 0,000129    |
| Valore di differ. in %               | 98,48       | 78,95       | 20,78       | 99,99       |
| Over 50                              |             |             |             |             |
| Valore T-Test                        | 0,031943    | 0,120683    | 0,14884     | 0,016068    |
| Valore di differ. in %               | 96,81       | 87,93       | 85,12       | 98,39       |

Tabella 1: analisi statistica-Pianecrati

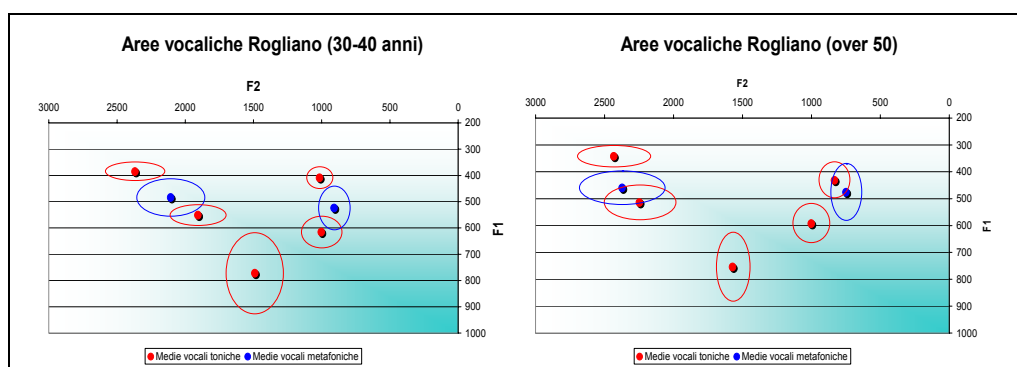


Figura 8: Aree di esistenza vocalica-Rogliano

| TABELLA CONFRONTO T-Test (Valori F1) 30-40 anni |             |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|-------------|
|                                                 | /o/ vs /o/M | /e/ vs /e/M | /u/ vs /o/M | /i/ vs /e/M |
| Valore T-Test                                   | 0,020891    | 0,026316    | 0,003787    | 0,006808    |
| Valore di differ. in %                          | 97,91       | 97,37       | 99,62       | 99,32       |
| Over 50                                         |             |             |             |             |
| Valore T-Test                                   | 0,005989    | 0,073975    | 0,41105     | 0,000119    |
| Valore di differ. in %                          | 99,40       | 92,60       | 58,90       | 99,99       |

Tabella 2: Analisi statistica-Rogliano

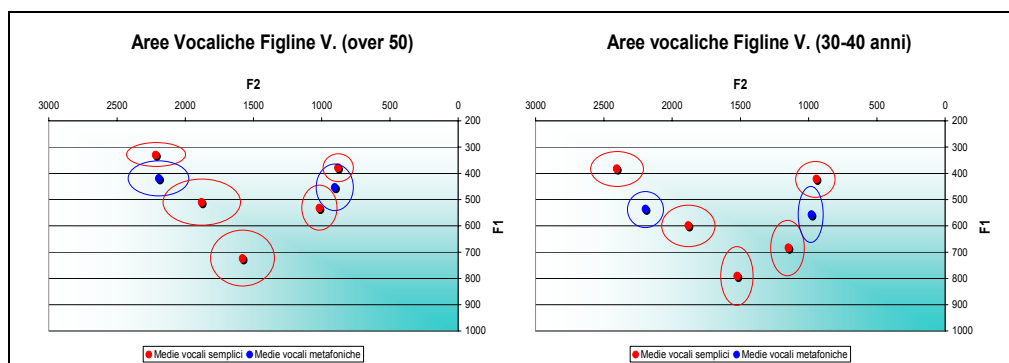


Figura 9: Aree di esistenza vocalica-Figline Vegliaturo

| TABELLA CONFRONTO T-Test (Valori F1) |             |             |             |             |
|--------------------------------------|-------------|-------------|-------------|-------------|
| 30-40 anni                           |             |             |             |             |
|                                      | /ɔ/ vs /o/M | /ɛ/ vs /e/M | /u/ vs /o/M | /i/ vs /e/M |
| Valore T-Test                        | 3,20        | 6.04        | 1,04        | 0,01        |
| Valore di differ. in %               | 99,97       | 99,94       | 99,99       | 100,00      |
| Over 50                              |             |             |             |             |
| Valore T-Test                        | 0,001382    | 0,05484     | 0,002104    | 0,001034    |
| Valore di differ. in %               | 100,00      | 99,95       | 100,00      | 100,00      |

Tabella 3: Analisi statistica - Figline Vegliaturo

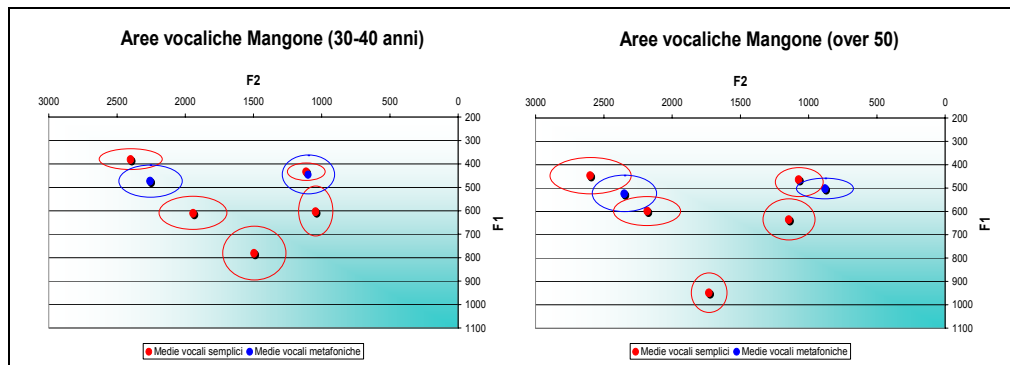


Figura 10: Aree di esistenza vocalica - Mangone

| TABELLA CONFRONTO T-Test (Valori F1) |             |             |             |             |
|--------------------------------------|-------------|-------------|-------------|-------------|
| 30-40 anni                           |             |             |             |             |
|                                      | /ɔ/ vs /o/M | /ɛ/ vs /e/M | /u/ vs /o/M | /i/ vs /e/M |
| Valore T-Test                        | 0,008912    | 0,001484    | 0,666213    | 0,01701     |
| Valore di differ. in %               | 99,11       | 99,85       | 33,38       | 98,30       |
| Over 50                              |             |             |             |             |
| Valore T-Test                        | 0,005256    | 0,059337    | 0,12559     | 0,05129     |
| Valore di differ. in %               | 99,47       | 94,07       | 87,44       | 94,87       |

Tabella 4: Analisi statistica - Mangone

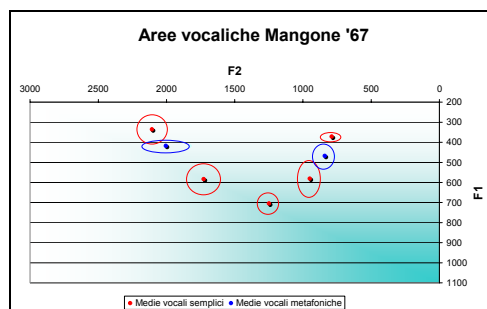


Figura 11: Area di esistenza vocalica – Mangone '67

### 3.3 Mappatura della metaforia nell'area sottoposta a studio

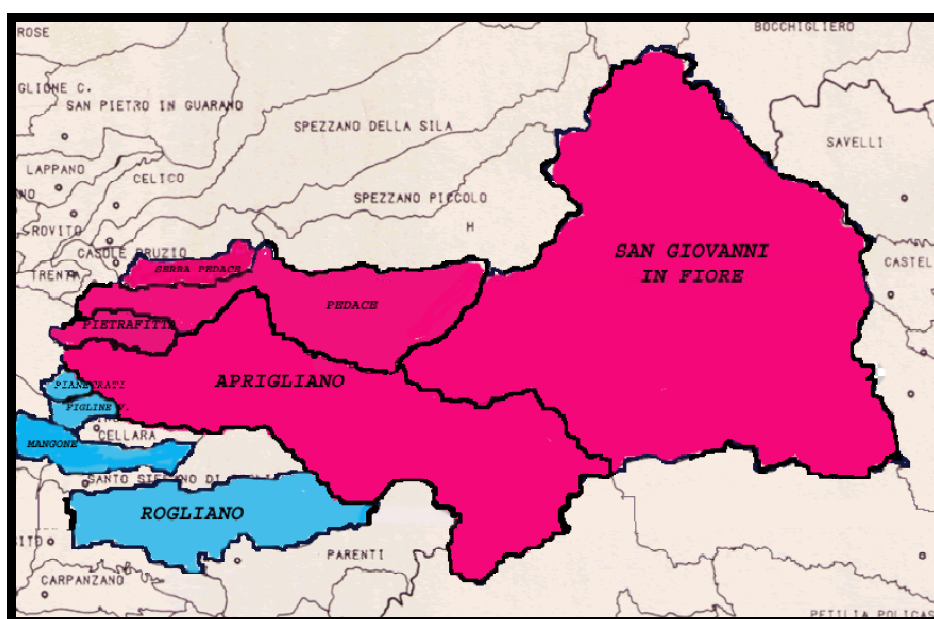
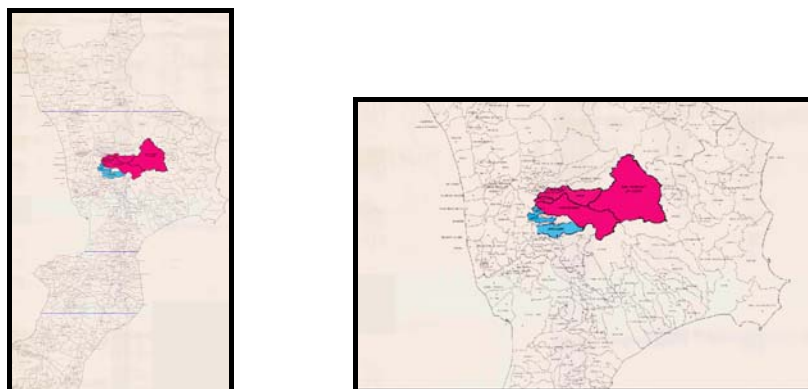


Figura 12: ● Metaforia per dittongazione: /e/ > /ie/ ed /o/ > /uo/; ● Metaforia per innalzamento: /e/ > /e/ ed /o/ > /o/

|                                                                                                                                                                                               | Aprigliano  | Pietrafitta | Pedace      | Serra<br>Pedace | S.<br>Giovanni<br>in F. | Pianecrati | Figline V. | Mangone    | Rogliano   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|-------------|-------------|-----------------|-------------------------|------------|------------|------------|------------|
| <b>Rohlf G. (1962:1966:1969:1972:1977)</b>                                                                                                                                                    | +metafonia  | +metafonia  | +metafonia  | +metafonia      | +metafonia              | +metafonia | +metafonia | +metafonia | +metafonia |
| <b>Falcone G. (1974:1976:1991)</b>                                                                                                                                                            | +metafonia  | +metafonia  | +metafonia  | +metafonia      | +metafonia              | +metafonia | +metafonia | +metafonia | +metafonia |
| <b>Trumper J. B.<br/>(1980:1982:1987:1988:1995:1997)</b>                                                                                                                                      | e>ie - o>uo | e>ie - o>uo | e>ie - o>uo | e>ie - o>uo     | e>ie - o>uo             | ɛ>e - ɔ>o  | ɛ>e - ɔ>o  | ɛ>e - ɔ>o  | ɛ>e - ɔ>o  |
| <b>Romito L., Gagliardi D. (2008)</b>                                                                                                                                                         | e>ie - o>uo | e>ie - o>uo | e>ie - o>uo | e>ie - o>uo     | e>ie - o>uo             | ɛ>e - ɔ>o  | ɛ>e - ɔ>o  | ɛ>e - ɔ>o  | ɛ>e - ɔ>o  |
| <b>+ <u>metafonia</u>: gli studi di Rohlf e Falcone individuano macroaree con presenza o assenza della metafonia, senza definire poi nello specifico le caratteristiche di ciascun paese.</b> |             |             |             |                 |                         |            |            |            |            |

Tabella 5: tabella riassuntiva degli esiti

#### 4. CONCLUSIONI

Le analisi statistiche evidenziano come l'area oggetto di studio (nove paesi del cosentino sud occidentale) siano chiaramente interessati dal fenomeno della metafonìa. Cinque dei nove paesi presentano metafonìa per dittongazione (Aprigliano, Pietrafitta, Pedace, Serra Pedace, S. Giovanni in Fiore: o>uo, e>ie), mentre i rimanenti quattro per innalzamento (Pianecrati, Figline V., Mangone, Rogliano: ε>e, ɔ>o). I dati elettroacustici raccolti trovano ulteriori conferme negli studi dialettologici precedenti e, soprattutto in quelli effettuati nel '97 da J. B. Trumper che nei medesimi paesi oggetto della nostra indagine riuscì a verificare non solo la presenza del fenomeno, ma anche a riscontrare i medesimi esiti. Dunque approfonditi studi dialettologici prima ed elettroacustici poi, ci permettono di arrivare alle medesime conclusioni.

#### BIBLIOGRAFIA

Abete G., "Sulla questione della sillaba superpesante: I dittonghi discendenti in sillaba chiusa nel dialetto di Pozzuoli" in AISV 2005 II Convegno Nazionale: *"Analisi prosodica"*, Teorie, modelli e sistemi di annotazione.

Alessio G., I dialetti della Calabria, in AA. VV., in *Almanacco Calabrese*: rassegna annuale di vita e problemi regionali, Istituto Grafico Tiberino, Roma 1964.

Alessio G., *La stratificazione linguistica del Brutio*, in Atti del I Congresso Storico Calabrese, Collezione Meridionale Editrice, Roma, 1954.

Calabrese A., Metafonia in Salentino, in *Rivista di grammatica generativa*, Unipress, Padova, 1985.

Calabrese A., Metaphony revisited. *Rivista di linguistica*, X, 1, 1998.

Cole J., Deconstructing Metaphony, in *Rivista di linguistica*, vol. X, n° 1, 1997.

Cortellazzo M., *Profilo dei dialetti Italiani*, Pisa, Pacini, 1974.

Cutugno F., Falcone M., Sulla distribuzione statistica dei valori formativi delle vocali, in *Atti del XXIV Convegno Nazionale dell'Associazione italiana di Acustica*, pp. 163-168, Trento, 12-14 giugno 1996.

Falcone G., Appunti per un saggio Storico-linguistico sui dialetti della Calabria, in *Bollettino dell'Istituto di studi di Cassidoro e sul Medioevo in Calabria*, n° II. 1991.

Falcone G., Calabria, in *Profilo dei dialetti italiani*, M. Cortelazzo (a cura di), vol. XVIII, Pacini, Pisa, 1976.

Falcone G., Innovazione e conservazione nei dialetti calabresi, in *Dal dialetto alla lingua*, Pacini, Pisa, 1974.

Fanciullo F., Morfo-metafonìa, in Cipriano P. e Mancini M (a cura di) *Miscellanea di studi linguistici in onore di Walter Belardi II*, il Calamo, Roma, 1994.



- Fava E., Magno Caldognetto E., Studio sperimentale delle caratteristiche elettroacustiche delle vocali toniche ed atone in bisillabi italiani, SLI, Studi di fonetica e fonologia, Bulzoni, Roma, 1976.
- Ferrero F., Problemi spettroacustici di misurazione delle vocali: un contributo, in *Atti delle VII Giornate di studio del Gruppo di Fonetica Sperimentale* (A.I.A.), pp. 235-264, vol. XXIV, 1997.
- Gentili, A., *Fonetica del dialetto cosentino*, Milano, Bernardoni di C., Rebeschini e C., 1897 (ristampa Cosenza, MIT, 1980).
- Giannini A., Accuratezza delle misure spettrografiche: il rilevamento delle formanti, in *Atti delle II Giornate di Studio del Gruppo di Fonetica Sperimentale* (A.I.A.), pp. 37-47, vol. XX, 1994.
- Hjemslev L., The syllable as structural unit, in *Proceedings of the Third International Congress of Phonetic Sciences* (1938).
- Hyman L. M., A theory of phonological weight, Forbis publications, Dordrecht, Holland/Cinnamison, USA, 1985.
- Jeffrey W. Kaze, Metaphony and two models for the description of vowel systems, in *Phonology*, vol. 8, 1991.
- Maiden M., Dispaced metaphony and the morphologization of metaphony, *Rom.Phil.*, XXXIX, 1985.
- Maiden M., *Interactive morphology. Metaphony in Italy*, Routledge, Londra, New York, 1991.
- Maiden M., Metaphony, in Maiden M e Parry M. (eds.), *The dialects of Italy*, Routledge, Londra, New York, 1997.
- Martino P., *L'area Lausberg*, editrice Il Calamo, Roma, 1991.
- Pulgram E., *Sillabe, Word, Nexus, Cursus*, The Hague, Mouton, 1970.
- Rohlf's G., *Dizionario dialettale delle tre Calabrie*, II voll., Cosenza, editrice Casa del Libro, 1966-69.
- Rohlf's G., Grammatica storica della lingua italiana e dei suoi dialetti I, Fonetica, Einaudi, Torino, 1966.
- Rohlf's G., Le due Calabrie, in AA: VV., *Almanacco Calabrese: rassegna annuale di vita e problemi regionali*, Istituto Grafico Tiberino, Roma 1962.
- Rohlf's G., Nuovo dizionario dialettale della Calabria, Ravenna, 1977.
- Rohlf's G., Studi e ricerche su lingua e dialetti d'Italia, Sansoni, Firenze, 1972.
- Romito L., Cenni elettroacustici dell'accento in alcune varietà dell'italiano, in *Atti delle IV Giornate di Studio del Gruppo di Fonetica Sperimentale* (A. I. A.), a cura di P. L. Salsa, pp. 107-119, vol. XXI, Torino, 11-12 Novembre 1993, 1994.
- Romito L., Galatà V., Lio L., Stillo F., *La metafonía nell'area di Lausberg: una introspezione sulla natura della sillaba*. Dipartimento di Linguistica. Laboratorio di fonetica. Università degli Studi della Calabria, 2005.

Trumper B. J. E Maddalon M., Converging divergence and diverging convergence: the dialect language conflict and contrasting evolutionary trends in modern Italy, edited by Auer P. and Di Luzio A., de Gruyter, Variation and Convergence, study in social dialectology, Berlino-New York, 1987.

Trumper B. J. E Maddalon M., L'italiano regionale tra lingua e dialetto. Presupposti ed analisi, Brenner, Cosenza, 1982.

Trumper B. J., Calabria and South Basilicata, in Maiden M. e Parry M. (eds.), *The dialects of Italy*, Routledge, Londra; New York, 1997. pagg.355-64.

Trumper J. B., La zona Lausberg e il problema della frammentazione linguistica, Atti "SLI", 16/1, Roma, 1980.

Trumper J. B., Maddalon M., Converging Divergence and Diverging Convergence: the dialect: Language Conflict and Contrasting Evolutionary trends in modern Italy, in Auer P., Di Luzio (a cura di), Variation and convergence Studies Dialectology, Berlin- New York, De Gruyter, 1988, pp.217-29.9.

Trumper J. B., Maddalon M., Chiodo G., L'influenza di eventi macrosismici su alcune discontinuità linguistiche (Calabria), in Pellegrini, G. B. (ed.) *Saggi dialettologici di area italo-romanza* (nuova Raccolta), pag. 91, CNR, Centro di dialettologia o. Parlangeli, Padova, 1995, pp. 89-106.

Venneman T., Preference Law for Syllable Structure and explanation of sound change, Mouton de Gruyter, Berlin, New York- Amsterdam, 1988.

Vogel I., *La sillaba come unità fonologica*, Zanichelli, Bologna, 1982.

# **FONOLOGIA**

# **“GENERATIVO”: DESIGNAZIONI E CONTENUTI DI UN TERMINE NEI MODELLI FONOLOGICI DAL SECONDO NOVECENTO A OGGI**

Manuela Lo Prejato

Università di Roma “La Sapienza” – Dipartimento di Studi filologici, linguistici e letterari  
*manuela\_lo\_prejato@hotmail.com*

## **1. SOMMARIO**

Con questo contributo propongo una riflessione di taglio storico-linguistico attorno al termine “generativo” e al senso che esso ha assunto dalla seconda metà del Novecento a oggi. In particolare, effettuo quest’osservazione dal punto di vista della fonologia, analizzando quali modelli – e in base a quali caratteristiche – si possano designare appunto come “generativi” o “non-generativi” o anche “post-generativi”. Indagando il senso di “generativo” nei vari modelli, a partire da *The Sound Pattern of English* di Chomsky&Halle (1968) fino ad arrivare alla Teoria dell’Ottimalità, cerco anche un possibile principio ordinatore di tali modelli, numerosi e difficili da sistematizzare.

Da un punto di vista storico, non è semplice individuare la nascita e la fine di una “fonologia generativa”. Si possono comunque avanzare alcuni interrogativi: “generativo” è un modello derivazionale o un modello formale? E come si evolve il generativismo originario? Con quali ripercussioni sui sostenitori o sugli oppositori del fondatore Chomsky? Chi si appropria o rifiuta l’etichetta di “generativo”? Perché?

Il percorso di indagine prende le mosse dall’articolo del 1993 di Ney, *On generativity. The history of a notion that never was*, in cui l’autore si riferisce a un passo di *Knowledge of Language: Its Nature origin and use* (1986) dove Chomsky dichiara che “generativo” non significa altro che “esplicito” (e non anche “produttivo”).

Rispetto a tale precisazione chomskiana, si vede poi come i vari modelli fonologici possano essere collocati.

## **2. IL SENSO DEL TERMINE “GENERATIVO”: LA RICERCA DI UN PRINCIPIO ORGANIZZATORE NEL QUADRO DELLA FONOLOGIA CONTEMPORANEA**

Dall’ultimo trentennio del Novecento a oggi, gli studi fonologici hanno prodotto una moltitudine di modelli dai confini non sempre definiti, con diverse sovrapposizioni concettuali e differenze di soli piccoli particolari. Gli autori stessi oscillano dall’uno all’altro modello. Per avere un’idea di tale affastellamento, basta scorrere l’elenco riportato da Durand&Laks:

«A very incomplete and partly arbitrary list (in alphabetical order) of the types of phonology would include: articulatory phonology, autosegmental phonology, CV phonology, computational phonology, declarative phonology, dependency phonology, government phonology, harmonic phonology, laboratory phonology, lexical phonology, metrical phonology, natural phonology, natural generative phonology, non-linear phonology, radical CV phonology, tridimensional phonology. In parallel with this, we also find approaches such as moraic theory, optimality theory, theory of constraints and repair strategies, underspecification theory, and so on» (Durand&Laks, 2002: 27).

Rispetto a tali approcci, può essere utile e interessante porsi la domanda circa cosa si possa o non si possa chiamare “generativo”, sia per trovare un primo principio organizzatore dei vari modelli – generativi, non generativi, post-generativi? –, sia per capire quali conseguenze porti in concreto tale definizione.

### 3. LA “FONOLOGIA GENERATIVA”: LA DIFFICILE INDIVIDUAZIONE DELLA SUA NASCITA

Da un punto di vista storico, non è facile individuare la nascita e la fine di tale categoria di “fonologia generativa”:

«[...] it is not an easy task even to decide where this story should begin, nor where it ends. The simple answer to the question as to where it begins is with Noam Chomsky’s efforts to describe the morphophonemics of modern Hebrew (Chomsky 1951), but the historical roots of generative phonology lie deeper and earlier [...]. Where generative phonology ends is a much harder question to answer [...]» (Goldsmith&Laks, 2005: 1).

Rispetto al termine *post quem*, è d’altro canto invalso farlo coincidere con la pubblicazione di *The Sound Pattern of English* da parte di Chomsky&Halle (1968; *SPE*): si tratta senza dubbio di una scelta convenzionale, ma che, come ogni convenzione, ha dei fondamenti che possono essere argomentati.

Anderson (1985: 228) non esita a dichiarare che: «There are surely few years that are so clearly marked as watersheds in the history of phonology as 1968». E anche Encrevé è deciso nell’affermare l’importanza di *SPE*, considerando il lavoro di Chomsky&Halle appunto come uno spartiacque:

«[...] le livre clé de la phonologie au XX<sup>e</sup> siècle, duquel seuls les *Grundzüge der Phonologie* rédigés trente ans plus tôt par Troubetzkoy peuvent, pour l’importance théorique, être rapprochés. L’un et l’autre ouvrages d’ailleurs sont dédiés à Roman Jakobson, lien évident entre les auteurs respectifs de l’un et de l’autre. Aujourd’hui, chacun reconnaît une phonologie pré-*SPE*, une phonologie de type *SPE* et une phonologie post-*SPE* : *SPE* s’est imposé comme repère absolu dans l’histoire de la phonologie» (Encrevé, 1997: 100-101).

Per quanto riguarda la distinzione tra una fonologia pre-, post- e *SPE* stessa, è necessario fare alcune precisazioni. Innanzitutto, il materiale pubblicato nel 1968 in forma di libro circolava, benché non strutturato, già dall’inizio degli anni Sessanta<sup>1</sup>. Inoltre, alcune intuizioni sviluppate da Chomsky&Halle si potevano trovare almeno già in Harris (1951) e in Hockett (1955), oltre che nell’analogo *Sound Pattern of Russian* del solo Halle (1959).

Ciò detto, resta indubbio che il libro di Chomsky&Halle costituisca il *magnum opus* del movimento fonologico generativo (Goldsmith&Laks, 2005) e possa essere considerato l’iniziatore di una temperie:

---

<sup>1</sup> Cfr. Anderson, 1985: 328. Anderson nota comunque che *SPE* assume in pieno il proprio valore soltanto nel 1968, appunto come opera di pubblico dominio, non più appannaggio dei soli “iniziati”.

«Many of the frameworks which have emerged are characterizable in relation to *SPE* and, indeed, some textbooks published in the 1990s have simply called themselves ‘generative phonology’ or ‘phonology in the generative grammar’ to emphasize the continuity of a tradition within which they saw various developments» (Durand&Laks, 2002: 27-28).

#### 4. LA “RITRATTAZIONE” DI CHOMSKY

Per rintracciare e caratterizzare il senso originario del termine “generativo”, conviene comunque risalire e riferirsi alle affermazioni del suo fondatore.

In un articolo dal titolo provocatorio, *On generativity. The history of a notion that never was*, Ney (1993) compie appunto un percorso simile, partendo da un passo di *Knowledge of Language* (1986) in cui Chomsky dichiara che “generativo” non significa altro che “esplicito”. In una recensione al libro chomskiano, McCawley (1988) esprime sorpresa al riguardo, avendo l’impressione che qualcosa sia cambiato rispetto agli anni Sessanta, quando gli appariva che “generativo” fosse il dispositivo astratto che specificasse gli elementi della lingua. Chomsky (1991), a sua volta, si stupisce di tali affermazioni:

«I have always understood a generative grammar to be nothing more than an explicit grammar. Some apparently have a different concept in mind. For example [...], McCawley (1988) notes that I interpret the concept here as meaning nothing more than explicit, as I have always done (see, for instance, Chomsky 1965: 4), and concludes erroneously that this is a ‘sharp change’ in my usage that gives the enterprise an entirely different cast from that of the 1960s, when the task, as he perceives it, was taken to be ‘specifying the membership of a set of sentences that is identified with a language’ (pp. 355-356; McCawley takes the set of sentences to be what I have called that ‘structure’ of the language, that is, the set of structural descriptions.) But the characterization that he gives does not imply that ‘generative’ means anything more than ‘explicit’; there is, furthermore, no change in usage or conception, at least for me, in this regard» (Chomsky 1991: 449 n.1).

Tale citazione, per intero, rispecchia un modello chomskiano di puntualizzazioni o, se si vuole, ritrattazioni e ripensamenti, che ricorrerà spesso negli sviluppi della sua teoria. Ed è così che Ney commenta, rimarcando, non senza sarcasmo, le contraddizioni:

«[...] the conventional concept of generativity never existed in the mind of the man who is usually thought of as the creator or progenitor of generative linguistics. Yet, strangely enough, in the very article where Chomsky berates McCawley for misunderstanding the Chomskyan position on the meaning of ‘generate,’ Chomsky (1991: 425) refers to “... the words of the sentence generated.” Should this read “... the sentence produced...” or “... the sentence made explicit ...”?» (Ney, 1993: 441-442).

In effetti, il punto della questione – e ciò che ha ripercussioni anche sulle fonologie – è proprio questo confine tra il produrre e l’esplicitare, tra il predire e il rendere manifesto. In questa distinzione si può rintracciare, anche, il passaggio da un generativismo trasformatore a un generativismo *tout court* o un post-generativismo, tentando di interpretare il significato e le conseguenze di tale evoluzione; vanno intese non solo le caratteristiche oggettive dei modelli, ma anche le intenzioni, manifestate o alluse, dei propugnatori.

Circa l'identificazione di "generativo" con "esplicito", non solo diversi commentatori hanno espresso la propria costante consapevolezza al riguardo – da Newmeyer, a Hymes a Fought a Botha (Ney, 1993) – ma lo stesso Chomsky già dagli anni Sessanta ha proposto una definizione che, nel 1986, apparentemente non è variata: «If the grammar is, furthermore, perfectly explicit [...] we may (somewhat redundantly) call it a *generative grammar*» (Chomsky, 1965: 4).

D'altro canto, non si può negare che viga un'interpretazione consueta e comune di "generare", nel senso di un processo deduzionista secondo cui, tramite un sistema finito di regole, si assegna una descrizione strutturale a un numero infinito di frasi. Bisogna capire dove si annidi l'errore, se un fraintendimento effettivamente c'è stato:

«At this point, the question might be asked: How could McCawley have come to such a misunderstanding? [...] What is even more surprising is that some of the transformational generative grammarians share the same misunderstanding that McCawley has, that is, they hold to the conventional definition of 'generate' as 'specify the members of an infinite set'. [...]

The question then might be raised: Where do the conventional interpretations of the notions 'generate' come from if they do not in fact come from the writings of Chomsky himself? [...] Harris (1954:259) and Hockett (1954:232) used the term 'generate' in the conventional sense before Chomsky ever popularized the term and [...] Chomsky himself (1975a:40-41) acknowledges his debt to Harris at least on this matter. However, perhaps the concepts which later became associated with the term 'generate' antedate the works of Chomsky by some time» (Ney, 1993: 442-443).

Si deve immaginare, allora, che soltanto *gli echi* di tali concetti, presenti in Chomsky, possano essere stati alla base delle interpretazioni di McCawley e di tutta un'altra vastissima schiera di studiosi? Forse. Ma non si può ignorare che, nella medesima opera del 1965 e in numerosi ulteriori luoghi, lo stesso Chomsky ha fatto altre, diverse dichiarazioni, difficilmente equivocabili sul senso comune di "generare".

«The same work that contains the 'generative-is-explicit' statement (Chomsky 1965), also contains the following:

(6) Given a descriptive theory of language structure, we can distinguish its *weak generative capacity* from its *strong generative capacity* in the following way. Let us say that a grammar *weakly generates* a set of sentences and that it *strongly generates* a set of structural descriptions [...].

Other works of Chomsky written during the 1960s and 1970s have the conventional definition for 'generate'.

(7) ... a grammar that generates an infinite set of utterances with structural descriptions (Chomsky 1961:236).

(8) The grammar, then, is a device that (in particular) specifies the infinite set of well-formed sentences and assigns each of these one or more structural descriptions. Perhaps we should call such a device a *generative grammar* (Chomsky 1964:9).

(9) ... an internal representation of his language (i.e. a generative grammar) ... (Chomsky 1964:35).

(10) ... the ability of a transformational grammar to *weakly generate*—that is, to generate the sentences of—a natural language ... (Chomsky 1966:48)

In the context of all of the examples above [...], it is difficult to see how ‘explicit’, an adjective, can serve as a synonym for ‘generate’, a verb. Only in (9) above, where the adjective ‘generative’ is used does the adjective, ‘explicit’, substitute well. Furthermore, in (8), ‘specify’ is used in the same context and with the same sense as ‘generate’» (Ney, 1993: 444).

Gli esempi si potrebbero moltiplicare, così come si potrebbero moltiplicare i nomi dei linguisti che intendono “generare” nel senso di “produrre” o “specificare”. Riporto qui una sola, significativa affermazione di Hewson: «[...] Guillaume modifies the *langue/parole* dichotomy, by making it generative (in the natural sense, not in the artificial, quasi mathematical sense)» (Hewson, 1976: 332). Il senso convenzionale, deduttivo di “generare”, assurge, in questo caso, a termine di paragone paradigmatico nei confronti di una processualità diversa, cinetica e naturale.

Si può comunque immaginare, delle due accezioni, che l’una richiami immediatamente l’altra, come emerge dalle parole di Goldsmith&Laks:

«First, the goal of the working phonologist was to develop fully *explicit*, algorithmic phonologies which *generate* the surface forms of a language and only those. [...] Second, the *explicit* phonologies developed should employ derivational means (that is, sequential, processual analyses) to *generate* the forms of a language» (Goldsmith&Laks, 2005: 6; corsivo mio).

Rispetto a Chomsky e alla sua definizione di “generare”, ci si può allora chiedere perché egli abbia rigettato con tanta forza il senso convenzionale di “produrre”, “specificare”, salvando solo quello meno intuitivo di “esplicitare”.

«This question is legitimate since the conventional meaning of generate, ‘specify’ or ‘characterize’ has existed in concert with the less-known, secondary meaning of ‘generate’ — ‘explicit’ — at least since 1965 (Chomsky 1965:4 and 60-61). Perhaps, part of that answer might be found in the fact that building a device which ‘recursively enumerates’ all and only the grammatical sentences of a language has proved to be an impossibility. Moreover, the difficulty in specifying ‘all and only the grammatical sentences of a language’ stems from the virtual impossibility of drawing a distinction between grammatical and non- or semi-grammatical sentences [...]» (Ney, 1993: 447-448).

Chomsky non sarebbe nuovo a un meccanismo epistemologico del genere, in base al quale riprendere e riformulare le affermazioni del passato, secondo le nuove esigenze del presente. Ciò che disorienta sempre, però, è una sorta di occultamento rispetto a tale evoluzione di pensiero, in sé non estranea a un naturale percorso scientifico:

«So the final question is: Why is it so important to Chomsky on this issue to insist that he has not recently changed his position? [...] This would seem to indicate that Chomsky is willing to change and acknowledge change on some issues, but that certain issues are so important to him that he feels a responsibility to maintain consistency on them. The



definition requiring generative to means explicit might be one such issue. Another such issue might be the formal properties of grammar» (Ney, 1993: 449).

## 5. “PRODURRE” VS “ESPLICITARE”: LE DOMANDE PER LA FONOLOGIA

La riflessione sul produrre e l’esplicitare, si diceva, risulta fervida di conseguenze per la definizione delle fonologie. Attorno a queste ci si può porre infatti diverse domande: “generativo” è dunque un modello derivazionale o un modello formale? Come si evolve il generativismo originario? Con quali ripercussioni sui sostenitori o sugli oppositori del fondatore Chomsky? Chi si appropria o rifiuta l’etichetta di “generativo”? Perché? Nei vari modelli fonologici, che peso hanno le regole e quale le rappresentazioni? Che ruolo hanno le regole trasformazionali e quale quelle di costituenza?

In particolare la distinzione tra regole e rappresentazioni è stata fatta propria da Anderson (1985), il quale se n’è servito per suddividere e classificare i vari paradigmi fonologici venuti alla ribalta nel corso del ventesimo secolo. Da parte sua, Coleman (1998), dopo aver passato in rassegna i diversi approcci contemporanei alla fonologia, tenta di ridurre la distanza tra le due diverse scelte (quella derivazionale o quella rappresentazionale) o, più precisamente, cerca di favorire un metodo esclusivamente rappresentazionale:

«In generative grammar, a distinction is made between linguistic representation, which are formal or mental objects of linguistic theory, and rules, which are (usually derivational) relations between representations of different kinds. [...]

Canonical transformational rules are quite rare in phonological theory (Chomsky and Halle 1968: 427, Kenstowicz and Kisseberth 1979), but context-sensitive rewrite rules are commonplace. If it can be shown that they are unnecessary, however, they should be dropped, so that rule component of phonological theory is constrained as much as possible. [...] I shall [...] produce a new theory which has no transformational or context-sensitive rewriting rules (or any rules at all, in the traditional sense), only representations» (Coleman, 1998: 1-2).

Coleman si riferisce a una distinzione che va chiarita e che informa di sé tutti i recenti sviluppi fonologici. Egli sottolinea la differenza tra cosiddette regole trasformazionali e regole, invece, di costituenza, arrivando a voler negare l’utilità di entrambe. Ora, questa è una posizione estrema, in un contesto generale della fonologia contemporanea che ha rifiutato esclusivamente il versante trasformazionale, salvando quello di costituenza e configurazionale, rappresentativo. Tutto ciò trova espressione ordinata nelle parole di Durand&Laks:

«In recent years, descriptions have again focused on rules. This assertion may seem surprising in view of the fact that many current models claim to dispense with rules altogether (and in fact one famous book edited by Goldsmith [...] is entitled *The Last Phonological Rule*). The use of the term ‘rule’ in this context must however be clarified. Until comparatively recently, successive Chomskyan generative grammars have included two types of mechanisms, both referred to as rules historically. The first of these have been ‘constituency rules’ (‘phrase structure rules’ in syntax) and the second, transformational rules. For a very long time now, constituency rules [...] have not been construed in procedural rewriting terms (e.g. replace the symbol Syllable by the symbols Onset and Rhyme). Rather, constituency rules are regarded as well-formedness conditions or, in a

more fashionable terminology, 'constraints'. By and large, current models have not questioned the need for such descriptive devices. What, on the other hand, has been rejected by many phonologists is the use of transformational rules, which were at the core of the *SPE* approach [...]» (Durand&Laks, 2002: 30).

Ecco, dunque, che riflettendo sulla differenza tra regole e rappresentazioni si affaccia il concetto di “vincolo”, ampiamente diffuso nella fonologia contemporanea. Questo, sviluppatosi grazie anche al modello per “principi e parametri” chomskiano, favorisce oggi le analisi configurazionali.

## 6. ALCUNE POSSIBILI RISPOSTE

### 6.1 Cosa sicuramente non è generativo

La questione può essere innanzitutto circoscritta definendo i suoi termini negativi, cosa cioè possiamo chiamare con una certa sicurezza “post-generativo” o “non generativo”. Vorrei riferirmi, da questo punto di vista, agli approcci alla fonologia cosiddetti “eterodossi” o “eteroclitici”: almeno alla fonologia articolatoria, alla fonologia di laboratorio, ai modelli di matrice cognitiva (quello di Coleman del 2002, per esempio), che per la verità non si possono definire nemmeno fonologie in senso stretto, quanto programmi di ricerca sperimentali e multidisciplinari.

La fonologia articolatoria nasce infatti da una decisa rivalutazione della fonetica e pone l'accento sull'articolazione, dal momento che l'aspetto multidimensionale di questa riesce a spiegare una serie di fenomeni fonologici, in particolare quelli relativi alla sovrapposizione dei gesti produttivi lungo la linea temporale.

La fonologia di laboratorio, scegliendo un'autodefinizione che potrebbe suonare come un ossimoro, punta, a sua volta, a uno studio concreto del linguaggio, accogliendo ogni apporto scientifico in attinenza con esso e configurandosi come l'intersezione di varie diverse discipline.

Coleman, dal suo canto, con la svolta del nuovo millennio abbandona la visione agnostica – dal punto di vista cognitivo – della fonologia dichiarativa e vede la forma delle parole archiviate nella mente come memorie fonetiche poggianti anche su conoscenze statistiche e semantiche.

### 6.2 Cosa sicuramente è generativo

Rispetto alle fonologie propriamente dette, partiamo dalla constatazione che il modello del 1968 di Chomsky&Halle, basato su regole e derivazioni, corrisponde senza dubbio a un paradigma che è sia formale sia esplicito sia astratto sia trasformativo. Come detto, esso rappresenta, inoltre, il termine *post quem* convenzionale della fonologia generativa:

«En 1968, la phonologie se réveille radicalement épurée. Inventaire rapide : ont été froidement abandonnés, explicitement ou implicitement, pêle-mêle, le phonème et le niveau phonémique, le rapport son-sens, la notion même de système phonologique et, par suite, les concepts d'opposition (primitives, équipollentes...), de corrélation, de neutralisation (et d'archiphonème), mais aussi, de façon plus surprenante, le syllabe, tout le domaine prosodique et l'idée même d'unité suprasegmentale, comme d'ailleurs le concept de « long component » qui semblait si prometteur peu avant (et auquel Chomsky (1957) se référerait encore avantageusement...), et le cœur même du travail de ces *Grundzüge* qu'on rêve tant de détrôner, l'étude, universelle, des «systèmes phonologiques des oppositions

distinctives»: systèmes vocaliques, systèmes consonantiques, systèmes prosodiques» (Encrevé, 1997: 101-102).

### 6.3 La zona di incertezza

Per le altre fonologie, la scelta più comoda e sicura sarebbe dunque di definirle secondo un criterio cronologico obiettivo, appunto come “post-*SPE*”. Non rifiutando tale “scorciatoia” metodologica, ripercorriamo, però, anche le evoluzioni del modello standard e le posizioni dei vari protagonisti nell’uso del termine “generativo”.

Prendendo Chomsky come punto di riferimento teorico, già abbiamo ricordato la sua definizione perentoria e polemica, a un certo punto del proprio percorso scientifico, di “generativo” come soltanto di “esplicito”, con l’esclusione, dunque, dell’altro senso diffuso di “generazione”, quello di un processo deduttivo in base al quale, tramite un sistema finito di regole, si assegna una descrizione strutturale a un numero infinito di frasi. Nella percezione comune a vari studiosi, mettere da parte questo senso sedimentato, vicino al “produttivo” bloomfieldiano, non è facile. Ciononostante, esaminando le fonologie successive a *SPE*, pare che almeno in alcuni casi la definizione di Chomsky sia stata accolta, se per esempio la Teoria dell’Ottimalità, che decisamente dichiara di rifiutare le regole trasformazionali, comunque non rinuncia a chiamare se stessa “generativa”. Ma prima di arrivare all’Ottimalità, che costituisce il caso estremo e, rispetto a questo discorso, il più ambiguo e interessante, vediamo come si schierano le altre fonologie successive al lavoro di Chomsky&Halle.

La fonologia generativa naturale, prima di tutte, pur annunciando la rottura nei confronti di *SPE*, addirittura nel nome continua a mantenere l’etichetta di “generativa”. E la stessa rottura non è in realtà così effettiva e rivoluzionaria. Ugualmente non sovvertitrice nei fatti è la fonologia naturale, la quale comunque intende presentarsi come una versione radicale del modello della Hooper, abdicando già nella propria definizione al termine di “generativo”.

Le altre fonologie (in particolare i modelli multilivello) sull’argomento in questione non si esprimono e non si autodefiniscono esplicitamente. Ma già la fonologia lessicale, per esempio, pur non incorporando nella sua denominazione l’etichetta di “generativa”, comunque si presenta come una semplice *correzione* del modello di Chomsky&Halle. Il generativismo, d’altro canto, si è configurato per lungo tempo come un riferimento teorico del tutto pervasivo: nella cornice da esso tracciata, si può immaginare che, per una sorta di “silenzio assenso”, i modelli che non si sono espressi chiaramente contro tale temperie alla temperie stessa hanno in qualche modo aderito. A questo proposito, Goldsmith&Laks inducono all’idea di un *mood* generativo, piuttosto suggestiva:

«[...] should the theories of phonological representations studied in the 1980s be considered generative phonology? — most phonologists would agree that they should: certainly from a methodological, epistemological or theoretical point of view, they share the flavor, the spirit, the mood of generative phonology; but many of the core questions asked in the 1980s had stronger historical roots in pre-generative than in generative phonology» (Goldsmith&Laks, 2005: 1).

Sia la fonologia autosegmentale sia la fonologia metrica, allora, pur presentandosi come modelli configurazionali in cui le regole derivazionali non trovano spazio, in ogni caso si muovono in un ambito prettamente generativo: «[...] autosegmental phonology and metrical phonology were based on arguments that made, and still make, perfect sense within the very theoretical heart of generative phonology» (Goldsmith, 1990: 2).

Ma vediamo allora cos'è per Goldsmith, tra i principali protagonisti di queste fonologie, il generativismo: «If I had to summarize in just one statement the basic goal of the enterprise, it might be this: we attempt to formulate general models and principles of phonological analysis which can be successfully applied to a wide range of languages» (Goldsmith, 1990: 2-3). Il senso di “generativo”, lontano già da quello di “trasformazionale”, slitta così verso quello più neutro di “generale”. È interessante, però, che proprio dall'interno di questo inquadramento teorico si sviluppino le critiche, alla fine distruttive, per la fonologia generativa standard:

«Comme le souligne Goldsmith (1990, p. 2), c'est sur la base d'arguments internes, au nom de principes théoriques génératifs et en vertu des objectifs scientifiques réaffirmés de ce courant, que *SPE* a été progressivement déconstruit et finalement abandonné. Cela explique que son programme scientifique originel et son cadre épistémologique restent quant à eux actifs et continuent à orienter les recherches en apparence les plus éloignées de la logique initiale. [...] Cela explique le paradoxe apparent de la phonologie actuelle qui, totalement affranchie du cadre chomskyen classique, reste pourtant remarquablement unifiée et homogénéisée par le programme scientifique et la méthodologie qui étaient à son fondement, par le style argumentaire et la conception du domaine popularisé par les phonologues générativistes des années soixante» (Laks, 1997: 9).

#### 6.4 Un caso particolare: la Teoria dell'Ottimalità

Per la Teoria dell'Ottimalità (OT) a rigore si deve parlare di “generativismo”, perché sono i suoi stessi esponenti ad ascrivere a tale temperie. Nell'etichetta di “generativa” assegnata all'Ottimalità si devono però sottendere diverse specificazioni: il termine va letto, infatti, come “generativa, ma non standard, non originaria, non trasformazionale”, sottolineando dunque non tanto la re-interpretazione di “generativo” data da Chomsky nel senso di “esplicito”, quanto l'interpretazione di Foley e di Coleman:

«Many linguists view phonology and phonetics as forming a continuum, with phonological descriptions presented in the same formal language as phonetic descriptions, for instance in the assemblage of supposed universally distinctive phonetic properties. Though this is the most orthodox position, reflected in phonology textbooks, it is not a view which is shared by everyone. Foley (1977) calls this hypothesis 'Transformational Phonetics', rather than 'generative phonology', a terminological distinction which I shall adopt, since it is possible to support the generative approach to phonology without extending that approach to phonetic interpretation. The non-derivational approach to phonology demands a rigorously maintained distinction between phonology, on the one hand, and phonetics, on the other» (Coleman, 1998: 168-169).

OT, infatti, com'è noto distingue la grammatica universale (che comprende un alfabeto linguistico, un insieme di vincoli e due funzioni, GEN-erate ed EVAL-uate) dalla grammatica di ogni singola lingua (che comprende invece la forma base dei morfemi e la gerarchia per i vincoli). OT si trova però in una posizione che può sembrare ambigua, come insinuato per esempio da LaCharité&Paradis (2000) fin dal titolo del proprio intervento: *Derivational Residue: Hidden Rules in Optimality Theory*. Secondo i due autori, infatti, benché la teoria dell'Ottimalità dichiarò di aver abbandonato l'uso delle regole (LaCharité&Paradis, 2000: 229: «OT is a highly articulated effort to account for phonological regularities in a language in terms of constraints, rather than rules. Indeed, OT

is billed as a rule-free alternative to derivational constraint-based theories [...]»), nei fatti si riferisce a operazioni di trasformazione: «Indeed, we see this as a fundamental paradox in OT: it repudiates the notions of being procedural or derivational, but it nonetheless relies on the operations that defined rules to transform an input form into an output surface form» (LaCharité&Paradis, 2000: 213).

Sostengono LaCharité&Paradis, difatti, che la funzione GEN-erate dell'Ottimalità altro non sarebbe che un mascheramento delle vecchie regole generative:

«[...] OT researchers have thus far been preoccupied with the selection of one candidate from a field of potential output representations, not with the operations behind the provision of candidates. Because of that, OT has been able to maintain—speciously, we think—that it has done away with rules. However, if in order to generate that field of candidates, or even one optimal candidate, GEN, even in its most restricted conception, must have the power to insert content as well as structure, and to delete structure, then we must conclude that, contrary to its claims, OT does indeed rely on rules, as well as constraints. GEN might then be seen to be a cover-term for the most recent incarnation of the phonology's rule component» (LaCharité&Paradis, 2000: 230).

Andando allora al noto paragone di Archangeli&Langendoen (1997) tra la grammatica e la rete di un pescatore che voglia pescare solo un certo tipo di pesci, LaCharité&Paradis mettono in evidenza il potere generativo eccessivo dell'Ottimalità:

«While the analogy is enlightening, there is a crucial difference that we feel cannot be overlooked: the fisherman's net gathers what is already in the environment, whereas GEN actually creates the undesirables which the separator must subsequently deselect. Surely the idea that GEN is able to insert or delete any amount of content or structure casts doubt on the ideas that GEN constitutes a *single operation* [...]» (LaCharité&Paradis, 2000: 230).

## 7. CONCLUSIONI

In conclusione, si può osservare che, esclusi alcuni approcci eterodossi alla fonologia (fonologia articolatoria, fonologia di laboratorio, modelli di matrice cognitiva), l'etichetta e il riferimento teorico al generativismo non sono venuti mai realmente meno per i modelli post-*SPE*, sino alla Teoria dell'Ottimalità. Quello che è cambiato è il modello operativo, da derivazionale a configurazionale, e in certi casi nemmeno in termini del tutto nuovi e convincenti, se LaCharité&Paradis (2000) possono appunto parlare di un "residuo derivazionale" e di "regole nascoste" persino nella Teoria dell'Ottimalità stessa.

A proposito poi dell'opzione configurazionale, in anni recenti predominante, alcuni temi ci riportano anche ad altri scenari, di stampo pre-generativo. Il concetto di "vincolo", infatti, sviluppatosi sul modello dei "principi e parametri" chomskiani e largamente presente nella fonologia contemporanea, trova le sue radici già in lavori classici, in quelli almeno fonotattici di Grammont, Martinet e Saussure; la novità, rispetto al passato, è che il discorso si muove adesso su un piano essenzialmente sincronico, mentre precedentemente era stato ancorato a dibattiti di linguistica storica.

## BIBLIOGRAFIA

- Anderson, S.R. (1985), *Phonology in the Twentieth Century*, Chicago University Press.
- Archangeli, D. & Langendoen, K. (editors), (1997), *Optimality Theory. An Overview*, Oxford: Blackwell.
- Chomsky, N. (1951), *Morphophonemics of Modern Hebrew*, unpublished Master's thesis, Univ. of Pennsylvania.
- Chomsky, N. (1961), Some Methodological Remarks on Generative Grammar, *Word*, 17, 219-239.
- Chomsky, N. (1964), *Current Issues in Linguistics Theory*, The Hague: Mouton.
- Chomsky N. (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge, MA (trad. it. Aspetti della teoria della sintassi, in Id. (1969-70), *Saggi linguistici*, 3 voll., Torino: Boringhieri; vol. 2:39-258).
- Chomsky, N. (1966), *Topics in the Theory of Generative Grammar*, The Hague: Mouton.
- Chomsky, N. (1975), *The Logical Structure of Linguistic Theory*, New York: Plenum Press.
- Chomsky, N. (1986), *Knowledge of Language: Its Nature origin and use*, New York: Praeger Scientific.
- Chomsky, N. (1991), Some notes on economy of derivation and representations, in *Principles and parameters in comparative grammar* (R. Freidin, editor), MIT Press, Cambridge, MA, 417-454.
- Chomsky, N. & Halle, M. (1968), *The Sound Pattern of English*, New York: Harper and Row.
- Coleman, J. (1998), *Phonological Representations. Their names, forms and powers*, Cambridge University Press.
- Coleman, J. (2002), Phonetic Representations in the Mental Lexicon, in *Phonetics, Phonology and Cognition* (J. Durand & B. Laks, editors), Oxford University press, 96-130.
- Durand, J. & Laks, B. (2002), Phonology, from phonetics to cognition, in *Phonetics, Phonology and Cognition* (Eid., editors), Oxford University press, 10-49.
- Encrevé, A. (1997), L'ancien et le nouveau. Quelques remarques sur la phonologie et son histoire, *Langages*, 125, 100-123.
- Foley, J. (1977), *Foundations of Theoretical Phonology*, Cambridge University Press.
- Goldsmith, J. (1990), *Autosegmental and Metrical Phonology*, Oxford: Blackwell.
- Goldsmith, J. & Laks, B. (2005), Generative phonology: its origins, its principles, and its successors, in *The Cambridge History of Linguistics* (L.R. Waugh & J.J. Joseph, editors), Cambridge University Press  
[<http://hum.uchicago.edu/~jagoldsm/Papers/GenerativePhonology.pdf>, 1-19, da cui cito].
- Halle, M. (1959), *The Sound Pattern of Russian. A Linguistic and Acoustical Investigation*, The Hague: Mouton.

- Halle, M. (1964), Phonology in Generative Grammar, in *The Structure of Language: Readings in the Philosophy of Language* (J. Fodor & J. Katz, editors), Englewood Cliffs, New Jersey: Prentice Hall Inc., 334-352.
- Harris, Z.S. (1951), *Methods in Structural Linguistics*, Chicago University Press.
- Harris, Z.S. (1954), Transfer Grammar, *IJAL*, 20, 259-270.
- Hewson, J. (1976), Langue and parole since Saussure, *Historiographia Linguistica*, III:3, 315-348.
- Hockett, C.F. (1954), Two Models of Grammatical Description, *Word*, 10, 210-231.
- Hockett, C.F. (1955), *A Manual of Phonology*, Indiana University Publications in Anthropology and Linguistics (Memoir 11 dell'*International Journal of American Linguistics*, 21: 4, part 1), Baltimore: Waverly Press.
- Kenstowicz, M. & Kisseberth, C. (1979), *Generative Phonology : Description and Theory*, New York Academic Press.
- LaCharité D. & Paradis C. (2000), Derivational Residue: Hidden Rules in Optimality Theory, in *Optimality Theory. Phonology, Syntax, and Acquisition*, Oxford University Press (J. Dekkers, F. van der Leeuw & J. van de Weijer, editors), 211-233.
- Laks, B. (1997), *Phonologie accentuelle. Métrique, autosegmentalité et constituance*, Paris: CNRS.
- McCawley, J.D. (1988), Review of Chomsky (1986), *Language*, 64, 355-365.
- Ney, J.W. (1993), On generativity. The history of a notion that never was, *Historiographia Linguistica XX :2/3*, 441-454.

# **SIGNAL PROCESSING & PATTERN RECOGNITION**



## UN SISTEMA MULTIMODALE PER LA SEGMENTAZIONE DI TELEGIORNALI BASATO SULL'INDIVIDUAZIONE AUTOMATICA DEGLI SPEAKER

Leandro D'Anna<sup>1</sup>, Gennaro Percannella<sup>2</sup>, Carlo Sansone<sup>3</sup>, Mario Vento<sup>2</sup>

<sup>1</sup>Dipartimento di Studi Linguistici e Letterari, Università di Salerno  
Via Ponte don Melillo, 1 – Fisciano (SA), Italy

<sup>2</sup>Dipartimento di Ingegneria dell'Informazione e Ingegneria Elettrica, Università di Salerno  
Via Ponte don Melillo, 1 – Fisciano (SA), Italy

<sup>3</sup>Dipartimento di Informatica e Sistemistica, Università di Napoli "Federico II"  
Via Claudio, 21 – Napoli, Italy  
*ldanna, pergen, mvento@unisa.it; carlosan@unina.it;*

### ABSTRACT

Story segmentation is a basic step towards effective news video indexing.

All the solutions to story segmentation may be basically ascribed to two approaches. According to the first, segmentation is accomplished by directly finding story boundaries. Such boundaries are typically obtained by looking for the occurrences of some specific event (a sequence of black frames, the co-occurrence of a silence in the audio track and a shot boundary in the video track, etc.), or an abrupt change of some features at a high semantic level, as a topic switch.

The other approach performs story segmentation according to the following news program model assumption: given that each shot of the news video can be classified as an anchor shot or a news report shot, then a story is obtained by linking each anchor shot with all successive shots until another anchor shot, or the end of the news video, occurs.

According to the above news story model, automatic anchor shot detection (ASD) becomes the most challenging problem to partition a news video into stories.

In the scientific literature there are many papers that propose ASD algorithms. The majority exploits only video information, but in the last years the use of audio as an additional source of information for video segmentation has been rapidly raised up. Several proposals use audio features for directly individuating news boundaries, by means of a silence or a speaker change detector, in order to strengthen or to weaken the boundaries provided by the analysis based on ASD video techniques. Common drawbacks of such approaches are the use of supervised model-based techniques, that are not general enough, as they require an a priori definition and construction of an anchor shot model, and the unsystematic use of audio information. This is not effective due to its incoherence with video and then yields to a misleading shot classification.

In order to overcome these limitations, in this paper we present an algorithm for anchor shot detection with two audio/video stages.

The proposed algorithm first creates a set of templates in an unsupervised way. Each template represents a different anchor shot model within a video. In the second stage, a video similarity metric is used in order to retrieve a set of candidate anchor shots, which might have been missed by the first stage, and classify them by evaluating the audio similarity with respect to the templates. Finally the shots are classified by comparing them, from an audio point of view, with all the templates, when the news video is presented by only one anchorperson and with the closest template when two anchorpersons are present.

In order to choose the most suitable comparison without any a priori information about the actual number of anchorpersons, we also propose an automatic selector, based only on the audio track that is able to classify a news video as presented by one or two anchorpersons.

This automatic selector is based on the observation that each speaker is typically characterized by a specific distribution of the fundamental frequency ( $f_0$ ): the idea is then to verify if the  $f_0$  distribution calculated on all the audio samples belonging to a single speaker can be approximated by a given distribution (Log-logistic).

In the affirmative case, there is only one speaker; otherwise the number of speakers is at least two.

In particular, the audio similarity is calculated by comparing a candidate shot with all the templates extracted by the first stage. This is the best choice if the news video is presented by only one anchorperson.

On the contrary, a different criterion for defining audio similarity should be used when a news video is presented by two anchorpersons. In this case, in fact, only the audio track of the closest template (from an audio point of view), among all the extracted ones, should be used within the comparison with a candidate anchor shot. This is due to the fact that different templates can now refer to different anchorpersons and then their audio tracks no longer belong to the same speaker.

If the number of anchorpersons is known, the correct criterion to be applied can be a priori chosen. It happens if a broadcaster would employ the proposed method for analyzing all the editions of its news videos. In the general case, however, archiving companies work with large quantities of videos from different sources; moreover, we are interested in developing a completely unsupervised method.

The overall method proposed here has been tested on a database composed by several videos from two of the main Italian broadcasters. The performance of our method was evaluated in terms of Precision, Recall and F-measure. Moreover, we also compared our algorithm with respect to other state-of-the art anchor shot detection algorithm, achieving a significant performance improvement and so demonstrating its effectiveness.

## SOMMARIO

La segmentazione in story è il passo fondamentale per una efficace indicizzazione dei telegiornali.

Tutte le soluzioni presenti in letteratura per la segmentazione possono essere suddivise in due tipologie di approccio al problema.

Secondo il primo approccio, la segmentazione viene realizzata contestualmente alla determinazione dei confini tra una story ed un'altra. Tali confini sono ottenuti, tipicamente, cercando le occorrenze o di alcuni specifici eventi (ad esempio una sequenza di frame neri, la presenza contemporanea di un silenzio nella parte audio e di un confine di shot nella traccia video, etc..) o il brusco cambiamento di features di livello più alto quali quelle di tipo semantico legate ad un cambiamento di topic.

L'altro approccio, invece, realizza la segmentazione in story sulla base di un modello di telegiornale assunto a priori: supposto che ogni shot di un telegiornale sia classificato come anchor shot o news report, allora una story è ottenuta collegando ogni anchor shot a tutti i successivi shot finché o un nuovo anchor shot viene individuato o termina il telegiornale.

Secondo questo modello di story, quindi, l'individuazione automatica degli anchor shot (ASD) diviene il problema fondamentale per la suddivisione di un telegiornale in story.

Nella letteratura scientifica ci sono molti lavori in cui vengono proposti algoritmi per l'ASD. La maggioranza di essi sfrutta solo le informazioni video, ma negli ultimi anni si è sempre più affermata anche l'importanza dell'audio come una sorgente suppletiva di informazioni.

In alcuni lavori le features audio vengono usate direttamente per l'individuazione dei confini di una notizia attraverso l'uso dei silenzi o di opportune funzioni per l'individuazione del cambio di speaker allo scopo, poi, di rafforzare o indebolire le indicazioni sulla presenza di confini ottenute mediante un'analisi in video.

Tuttavia questi approcci hanno due grossi inconvenienti: il primo è dato dal fatto che usano tecniche model based supervisionate e quindi non abbastanza generiche, il secondo è legato al fatto che le informazioni audio non sono utilizzate in maniera sistematica e quindi in maniera alquanto arbitraria.

Tutto questo causa una errata classificazione degli shot in tutti quei casi in cui c'è una incoerenza tra le due fonti informative.

Allo scopo di superare queste limitazioni, in questo articolo presenteremo un algoritmo per l'individuazione degli anchor shot a due stadi integrati audio/video.

L'algoritmo proposto inizialmente crea un insieme di template di anchor in maniera non supervisionata. Ogni template rappresenta un differente modello di anchor shot all'interno del video. Successivamente, attraverso una metrica di similitudine sulla base del video, si costruisce un insieme di shot possibili candidati ad essere degli anchor shot. Allo scopo di individuare possibili anchor shot mancanti, ogni elemento di questo insieme viene, poi, confrontato con tutti i template di cui sopra in cui è presente un solo anchor e con il template più vicino nel caso in cui ci siano più anchor sulla base di un criterio di confronto in audio.

Allo scopo di scegliere il più opportuno metodo di confronto senza usare informazioni a priori, abbiamo sviluppato un modulo per la selezione automatica del numero di speaker sulla base della sola traccia audio.

Questo modulo si basa sulla osservazione che ogni speaker è tipicamente caratterizzato da una specifica distribuzione della frequenza fondamentale ( $f_0$ ): l'idea è che per verificare che una certa distribuzione della  $f_0$ , calcolata a partire dall'insieme dei campioni audio di un certo shot, appartenga ad un solo speaker è che essa debba essere approssimata da una certa distribuzione (Log-Logistica). In caso affermativo tutti i campioni audio apparterranno ad un solo speaker altrimenti ad almeno due differenti.

In particolare la similarità in audio è calcolata confrontando lo shot candidato con tutti i templates estratti durante il primo stadio. Questa è la scelta migliore nel caso in cui nel video ci sia solo un anchor. Al contrario un criterio differente per definire la similarità audio deve essere usato quando nel video sono presenti più di un anchor. In questo caso, infatti, solo la traccia audio del template più vicino (dal punto di vista audio) potrà essere utilizzato durante il confronto poiché template differenti possono ora riferirsi a differenti anchor e quindi anche le loro tracce audio non saranno più di un solo speaker. Occorre notare che se il numero di anchor è conosciuto, il criterio corretto può essere applicato a priori senza usare il modulo proposto. Ciò avviene quando l'emittente televisiva vuole impiegare il metodo proposto per analizzare tutte le edizioni delle sue news. Nel caso più in generale, però, le aziende lavorano con grandi insiemi di telegiornali di differenti emittenti: in questi casi un metodo non supervisionato quale quello proposto risulta essere di gran lunga preferibile.

Il sistema proposto è stato sperimentato su di un database costituito da numerosi video di emittenti italiane. Le prestazioni dello stesso sono state valutate in termini di Precision,

Recall e F-measure. Infine abbiamo confrontato il nostro algoritmo con altri algoritmi allo stato dell'arte ottenendo significative performance e dimostrando così l'efficienza dell'algoritmo proposto.

## **1. PREMESSA**

Una sequenza video è una ricca sorgente di informazione contenente parlato, musica, immagini, testo, ecc., quindi, per la natura stessa dei dati multimediali è difficile realizzare ricerche approfondite in vasti archivi basandosi sulle classiche interrogazioni testuali. Tutto questo significa che le risorse multimediali dovrebbero essere invece indicizzate, memorizzate e recuperate in un modo molto simile all'attività di processamento di tali informazioni da parte del cervello umano. In quest'ottica, una prima fase prevede la costruzione di indici di materiale audio/video prevalentemente realizzata manualmente mediante associazione di un limitato numero di parole chiave (keyword) all'oggetto di interesse. Una delle problematiche più interessanti da affrontare in questo settore sia da parte dell'industria dell'ICT e sia della comunità scientifica internazionale diventa allora quella dell'indicizzazione automatica dei contenuti multimediali messi a disposizione. Infatti, nonostante l'indicizzazione manuale sia correntemente il metodo più accurato, il processo automatico risulta molto appetibile in quanto fornirebbe uno strumento specialistico affidabile e a costo minimo. Tuttavia i processi finora sviluppati sono molto dispendiosi sia per l'elevato sforzo computazionale richiesto sia per lo spreco della risorsa tempo.

La fase finale di questo processo prevede l'uso di un sistema di elaborazione automatico in grado di selezionare in maniera diretta, attraverso un opportuno motore di ricerca applicato ad un database di informazioni "metadate", le informazioni audio-video di particolare interesse in modo semplice ed intuitivo. Per la creazione automatica di tali database, è necessario, quindi, un utilizzo di opportuni sistemi automatici, basati su metodi innovativi, per analizzare e classificare i dati multimediali sia dal punto di vista dell'audio che dal punto di vista del video, con tecniche di analisi multimodale.

In particolare il nostro interesse è in particolar modo focalizzato al "news video processing", ovvero all'analisi, identificazione e catalogazione dei filmati di telegiornali. Infatti le "broadcast news" sono molto preziose per gli analisti governativi, per i fornitori di informazioni (es. le agenzie ANSA) e per gli utenti finali e, poiché nuovi eventi accadono tutti i giorni nel mondo intero, è umanamente "utopico" che una persona possa tener traccia contemporaneamente tutte le notizie di tutte le emittenti televisive. Per realizzare tutto ciò è quindi necessario predisporre di un sistema di "News Video Segmentation" capace cioè di segmentare un filmato di telegiornale in unità semantiche omogenee: le notizie.

## **2. DESCRIZIONE DELLA RICERCA SVOLTA**

Il presente lavoro si occupa dell'analisi automatica dei filmati di telegiornale allo scopo di suddividere un TG in news stories, ossia in singole notizie coerenti da un punto di vista semantico. Nel caso particolare dei filmati di telegiornale considerando la sola parte video, una prima semplice caratterizzazione è data dall'alternarsi di scene relative alla presenza del solo cronista del giornale e di scene relative ai servizi.

Una prima separazione delle scene del filmato potrebbe dunque essere fatta riconoscendo la sola presenza del cronista (vedi figura 1); in tal modo, indirettamente, si ritroverebbero tutte le parti che appartengono ai singoli servizi, ottenendo così una prima

separazione del filmato in due distinte classi, che potremmo definire classe Cronista (Anchor shot) e classe News (Non anchor shot).

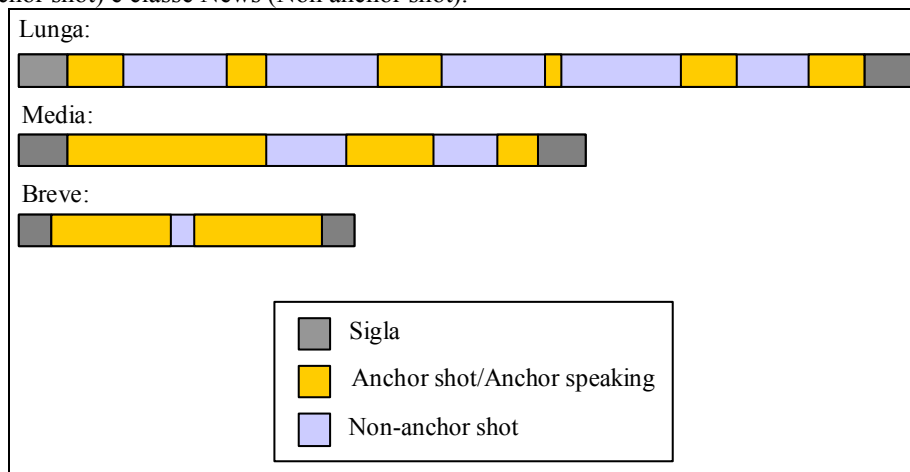


Figura 1: Esempi di suddivisione in anchor e news in un tg di diversa durata

Da un punto di vista dell'audio, un filmato di un telegiornale può essere strutturato come nella figura 2.

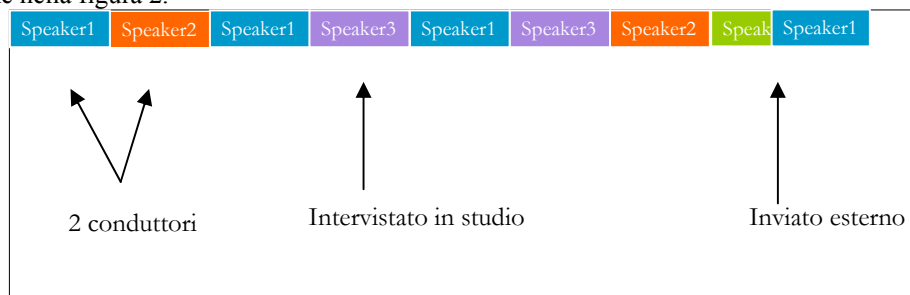


Figura 2: Esempio di suddivisione in turni dell'audio di un telegiornale

In questo caso si ha l'alternanza di diversi speaker che hanno una presenza nel filmato in percentuale molto diversa a seconda che siano speaker da studio o inviati esterni o intervistati in studio ad esempio. Tale alternanza suddivide la traccia audio in porzioni omogenee, dette anche turni, ognuna caratterizzata dalla presenza di un solo speaker.

In realtà, la singola segmentazione in video ed in audio è soggetta ad errori che impediscono di segmentare e classificare ogni segmento in maniera non ambigua. Inoltre anche per i segmenti univocamente identificati sia in video che in audio non sempre avviene una corrispondenza tra i confini temporali individuati dalla due fonti informative. Tutto questo rende impreciso il processo di costruzione di archivi di dati multimediali con la conseguenza che l'indicizzazione degli stessi non è sempre soddisfacente.

Dallo studio dello stato dell'arte, che è stato di fatto il punto di partenza del presente lavoro, possiamo suddividere i vari approcci per la segmentazione automatica di filmati di telegiornali in tre grosse categorie distinte in base a quale fonte informativa viene utilizzata (video e/o audio). Da ora in poi, escludiamo dalla nostra analisi tutti quei sistemi che utilizzano anche il testo (titoli e sottotitoli) essendo legati direttamente alla lingua utilizzata

nel telegiornale. All'interno dell'ultima categoria, ossia quelle che integra il video con l'audio, vi possono essere due approcci all'integrazione. Il primo prevede l'integrazione dell'informazione a priori ossia effettuando la stessa prima di effettuare la vera e propria segmentazione mentre il secondo approccio si basa sull'integrazione a posteriori delle singole segmentazioni indipendenti effettuate in video ed in audio.

### **3. DESCRIZIONE DELL'APPROCCIO**

L'approccio integrato, utilizzato dal nostro sistema, nasce dall'esigenza di superare alcune delle grosse limitazioni presenti nell'approccio video e nell'approccio audio. In particolare nel primo caso risulta difficile associare allo stesso speaker inquadrature diverse. In questi casi un sistema basato sul video potrebbe considerare la porzione di filmato con un'inquadratura diversa, se statisticamente poco occorrente, addirittura come quella appartenente ad un diverso speaker o anche ad un inviato da esterno. D'altro canto uno speaker da studio che commenta un filmato senza essere inquadrato sarebbe completamente omesso dal sistema basato sul video. Per quanto riguarda i sistemi basati solo sull'audio, poi, i maggiori problemi dipendono dalla estrema complessità delle caratteristiche acustiche. Questo significa che se da una parte si riesce abbastanza facilmente a distinguere il parlato dalla musica o dal rumore, diviene molto complesso identificare parlatori dello stesso sesso utilizzando clip di lunghezza non molto elevata e senza modelli addestrati a priori ed indipendentemente dalla lingua del parlante.

Queste considerazioni hanno quindi spinto molti gruppi di ricerca ad utilizzare una tecnica di segmentazione multimodale ossia che fa uso di entrambi i canali informativi servendosi così dei punti di forza di ognuno di essi. L'approccio utilizzato dal nostro sistema è dunque del tipo integrato a priori e si basa su un utilizzo integrato delle informazioni sulla segmentazione provenienti dai due canali. Una prima grossolana segmentazione in video viene successivamente raffinata da una successiva segmentazione in audio che a sua volta viene effettuata sulla base della precedente segmentazione in video. Tutto questo può essere ripetuto più volte fino ad ottenere la miglior segmentazione possibile senza alcuna scelta arbitraria.

In realtà, molto spesso è la componente audio che riesce a risolvere un'ambiguità presente in video. Consideriamo, ad esempio, il caso in cui in video compare un'esplosione caratterizzata da una rapida variazione della luminosità. La stessa scena può presentarsi durante un filmato lanciato durante un telegiornale o durante una pubblicità o durante un film di guerra. La sola analisi in video è quindi incapace di attribuire a quali tra queste categorie appartiene la scena in questione. Invece se consideriamo la stessa scena in audio potremmo notare che nel primo caso ci sarà sicuramente del parlato prima della scena in questione e questo parlato avrà una musica di sottofondo nel secondo caso. Nel terzo caso, poi avremmo una porzione abbastanza lunga in video caratterizzata da un volume molto più elevato rispetto alla media per evidenziare una scena d'azione. Quindi spesso l'audio è una fonte informativa da cui non si può prescindere per l'analisi automatica dei filmati.

In letteratura vi è utilizzata un'ampia e consolidata serie di features per risolvere il problema dell'anchor shot detection in audio. Sotto questo punto di vista, esiste una prima differenziazione tra features che possono essere di tipo più generale e meno precise nel descrivere uno speaker e features più specifiche ma meno generali che descrivono bene un singolo parlante. Le prime nascono con l'obiettivo di caratterizzare maggiormente una certa classe di suoni (ad esempio silenzio, parlato, rumore, musica,...) e sono legate più alle caratteristiche fisiche del segnale audio visto come un insieme di onde acustiche caratterizzate da una energia, una durata e uno spettro armonico ad esempio. Nel caso

specifico del parlato, oggetto della nostra ricerca, esse tendono a classificare più un gruppo di speaker che uno speaker in sé. Per quanto riguarda invece le features più specifiche, esse tendono a descrivere dettagliatamente un singolo parlante e la sua articolazione dei suoni fondamentali (foni). Tale features modellano la sorgente acustica, ossia le corde vocali e la conformazione degli organi fonatori, mediante una serie di coefficienti più o meno legati alle caratteristiche fisiche della stessa (quali spessore delle corde vocali, frequenze di risonanza, lunghezza del tratto glottideo, etc.). In tal modo si riesce a rappresentare il singolo parlante in maniera molto accurata. Quando, però, il numero degli speaker aumenta tali features risultano sempre più insoddisfacenti per effettuare una corretta discriminazione degli stessi.

Un ulteriore differenziazione delle features nasce da due considerazioni empiriche: in media un uomo utilizzando solo l'udito è capace di discriminare con una elevata precisione tutte le porzioni in cui compare uno stesso speaker all'interno di edizioni differenti di un telegiornale. Questo presuppone che questa capacità umana possa integrare informazioni provenienti da vari livelli temporali di analisi estendendo sia le informazioni specifiche a livello di suoni elementari (foni) e sia quelle legate alla semantica di quanto detto. Questo impone la necessità di individuare una varietà nella scelta delle features che vanno da quelle locali (20ms), calcolate su finestre di analisi a livello dei singoli foni, a features globali calcolate su finestre di analisi di dimensioni molto maggiori (>200ms).

#### **4. DESCRIZIONE DEL SISTEMA**

Il sistema sviluppato nasce con lo scopo di migliorare le prestazioni di un sistema già realizzato dal Mivia Lab dell'Università di Salerno basato anche esso sull'analisi del video e dell'audio con sole features di tipo locale. Tale sistema, quindi, rappresenta il sistema di base rispetto alla quale il sistema sviluppato si muove. La struttura generale del sistema in questione è riportato in figura 3 (per una descrizione dettagliata vedi De Santo et alii, 2005).

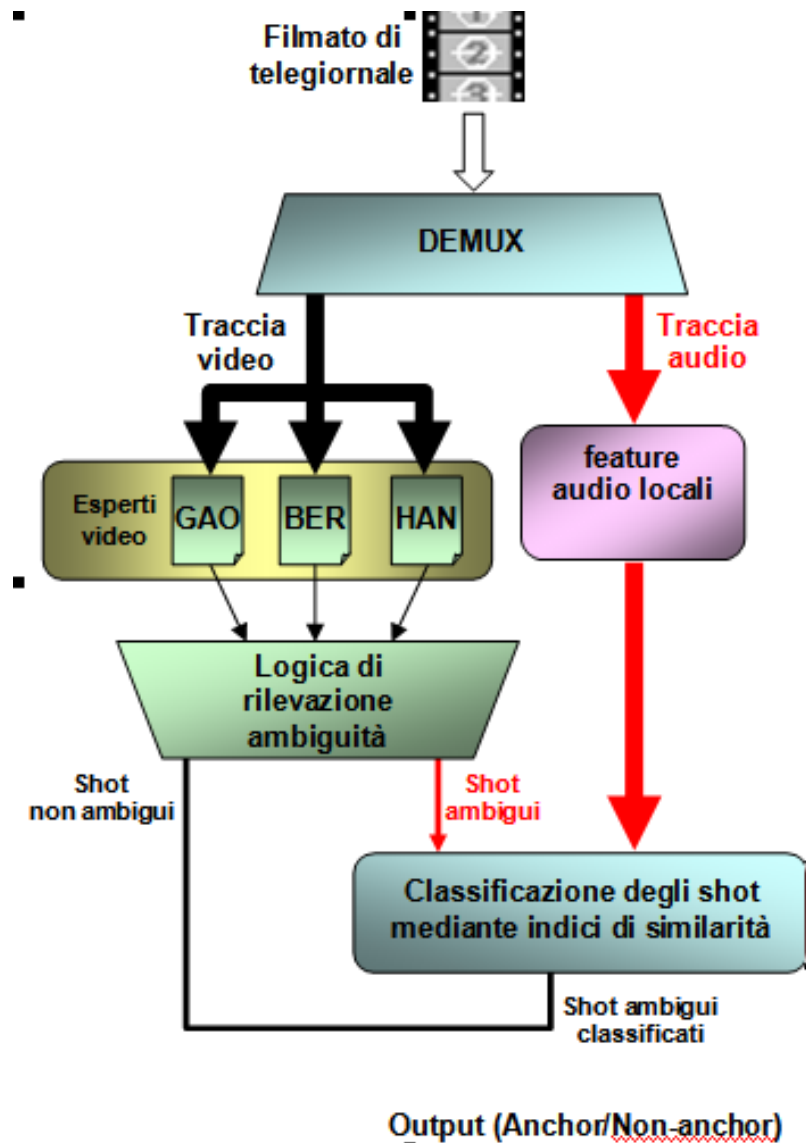


Figura 3: Architettura del sistema di base

Come si può osservare esso propone una modalità di analisi integrata a priori degli shot ambigui in video. Quindi, a partire da analisi video basata su multiesperti, si suddivide l'insieme degli shot di un telegiornale in tre classi: anchor shot certi (accordo totale multiesperti), news report certi (accordo totale multiesperti) e anchor shot dubbi.

A questo proposito l'approccio per l'analisi dell'audio utilizzato dal sistema di base si basa sull'ipotesi forte che le variazioni delle features locali siano specifiche del singolo speaker nel senso che ogni parlante sia caratterizzato da specifiche variazioni, diverse da quelle di un altro speaker, e che tali variazioni, quindi, permetterebbero di distinguerli. In realtà le features su base locale, proprio per il loro carattere, tendono a essere troppo



sensibili alle variazioni legate al contenuto sia sul piano fonetico che di parola. Tutto questo comporta una riduzione della capacità di discriminazione delle stesse.

Per superare tale limite nel sistema sviluppato è stato necessario introdurre sia una finestra di analisi di dimensioni temporali medie (100ms) o grandi ( $>1s$ ) rispetto a quelle adoperate per le features locali e sia delle features globali più legate alle caratteristiche fisiologiche dello speaker e meno dipendenti dal contenuto linguistico. Dunque nel sistema sviluppato è stato sostituito il modulo per la classificazione dei frame ambigui basato sulle sole features locali con un altro che utilizza sia feature di tipo locale che feature di tipo globale per migliorare i risultati ottenuti dal sistema base. Inoltre, allo scopo di ridurre i tempi di computazione, è prevista un'analisi delle features più mirata, nel senso che tale analisi sarà limitata di volta in volta alle sole porzioni dell'audio interessanti per una certa feature.

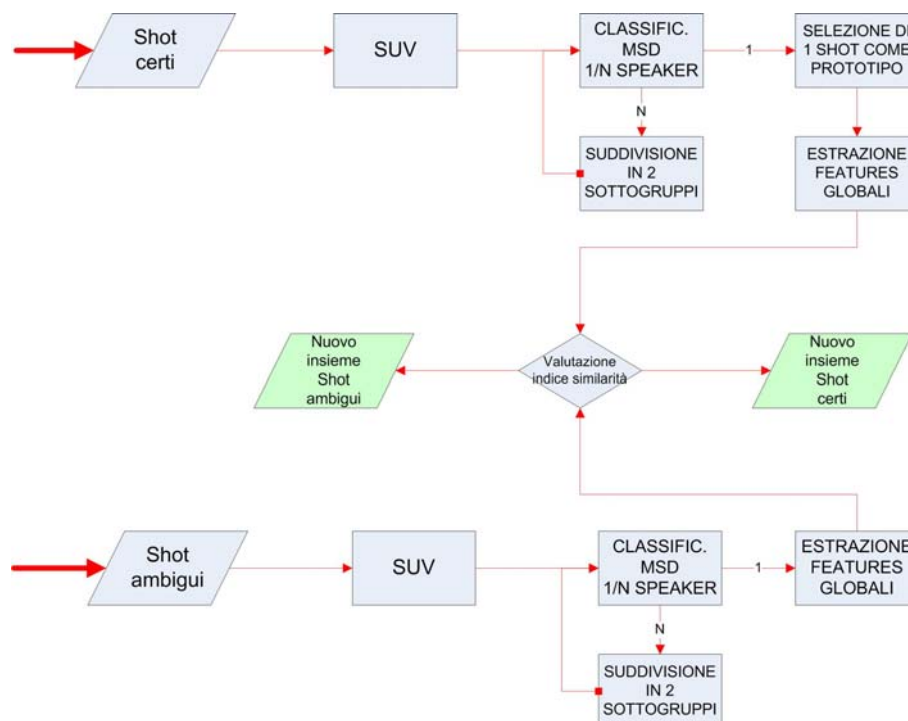


Figura 4: Architettura del sistema sviluppato

Nella figura 4, vi è uno schema che illustra l'architettura del sistema sviluppato (per una descrizione dettagliata dei moduli SUV e MSD vedi D'Anna et alii 2006). In particolare, esso agisce a partire da due set di anchor shot, uno certo l'altro ambiguo, individuati dal classificatore video ed il suo obiettivo è quello di raffinare i due insiemi di partenza fornendo così una segmentazione più accurata. Per far questo, si applica ai due insiemi prima un algoritmo (vedi figura 5) per l'individuazione delle porzioni salienti in audio (SUV)

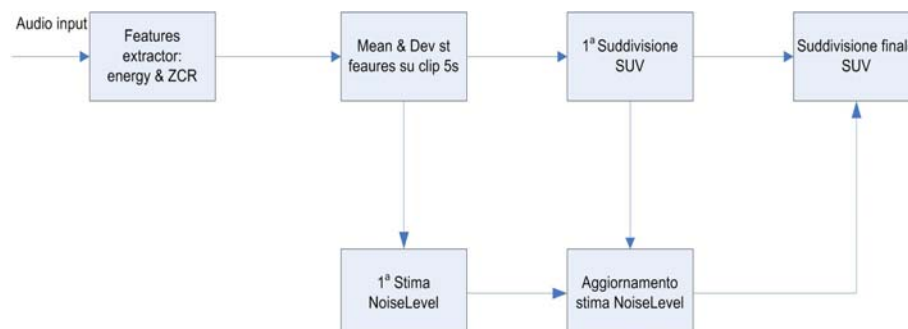


Figura 5: Schema di funzionamento dell'algoritmo SUV

e poi, sulla base della feature locale  $f_0$ , si applica ad ogni shot un classificatore (MSD) che permette di discriminare se in quello shot c'è un solo speaker o più di uno (vedi figura 6).

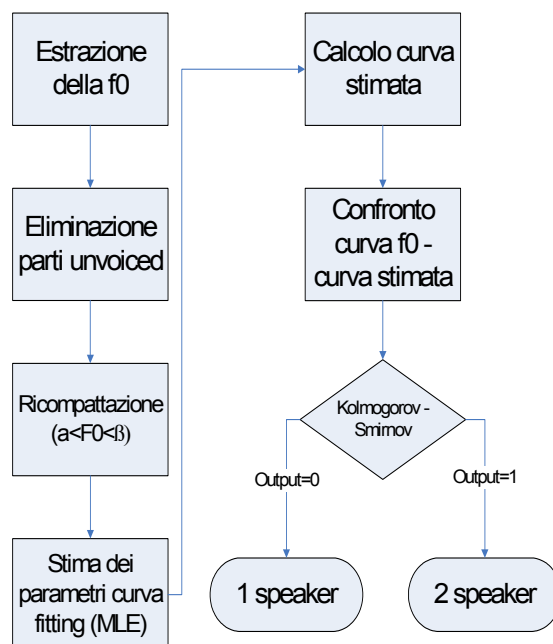


Figura 6: Schema a blocchi dell'algoritmo MSD.

Sulla base di questi risultati, per gli anchor shot certi, si costruiscono tutta una serie di prototipi relativamente a queglii shot in cui è stato identificata la presenza di un solo speaker. Per tutti gli altri shot, si suddivide ognuno in 2 sottogruppi di frame riapplicando l'algoritmo MSD. Il medesimo algoritmo si applica all'insieme degli anchor shot dubbi con l'obiettivo di separare in questi ultimi quelli in cui c'è un solo speaker in audio da quelli in cui c'è più di uno speaker. A questo punto, dopo aver estratto il set di features globali, per ognuno dei prototipi individuato sia negli anchor shot certi che in quelli dubbi, si effettua una valutazione dell'indice di similarità fissando un prototipo nell'insieme degli anchor

shot certi e confrontando quest'ultimo con tutti quelli individuati nell'insieme degli anchor shot dubbi. Nel caso in cui questo confronto dia esito positivo, il prototipo estratto dall'insieme degli anchor shot dubbi verrà associato all'insieme degli anchor shot certi ingrandendo questo insieme e riducendo l'insieme degli anchor shot dubbi.

Il processo verrà ripetuto per tutti i prototipi individuati nell'insieme degli anchor shot certi.

Il nostro sistema, dunque, permette sia di migliorare le prestazioni del sistema di base fornendo un insieme di anchor shot certi più ricco rispetto a quello del sistema di base e sia, grazie alla possibilità di individuare le porzioni salienti per effettuare il confronto di speaker, di ridurre in maniera considerevole (40%) i tempi di computazione (vedi D'Anna et alii, 2006).

## 5. CORPUS UTILIZZATO

Il dataset utilizzato durante tutta la sperimentazione è composto da filmati (per un totale di circa 16 ore) relativi a varie edizioni dei telegiornali delle emittenti nazionali italiane RAI 1 (26 edizioni del TG1) e CANALE 5 (16 edizioni del TG5) con una copertura in termini di ore di trasmissione approssimativamente uguale per ciascuna delle due emittenti. Le sessioni di registrazione sono avvenute nell'arco di tempo che va da giugno 2003 ad agosto 2004.

Riportiamo due tabella riassuntive per entrambe le reti, con l'elenco delle edizioni acquisite e le relative durate.

| Orario | Data       | Durata  | Durata in frame | Numero di shot | Anchor shot | Anchor                |
|--------|------------|---------|-----------------|----------------|-------------|-----------------------|
| 11.30  | 25/06/2003 | 0.06.27 | 9675            | 66             | 7           | Cristina Guerra       |
| 11.30  | 26/06/2003 | 0.05.05 | 7625            | 46             | 7           | Cristina Guerra       |
| 11.30  | 27/06/2003 | 0.05.23 | 8075            | 62             | 8           | Cristina Guerra       |
| 13.30  | 23/06/2003 | 0.25.45 | 38625           | 320            | 18          | Tiziana Ferrario      |
| 0.00   | 10/07/2003 | 0.09.59 | 14975           | 83             | 12          | Puccio Corona         |
| 0.55   | 14/07/2003 | 0.22.03 | 33075           | 200            | 15          | Puccio Corona         |
| 11.30  | 14/07/2003 | 0.05.20 | 8000            | 61             | 7           | Stefano Ziantoni      |
| 11.30  | 15/07/2003 | 0.05.31 | 8275            | 59             | 7           | Stefano Ziantoni      |
| 20.00  | 10/07/2003 | 0.32.54 | 49350           | 379            | 22          | Paolo di Giannantonio |
| 20.00  | 11/07/2003 | 0.32.02 | 48050           | 345            | 21          | Paolo di Giannantonio |
| 20.00  | 14/07/2003 | 0.32.17 | 48425           | 334            | 19          | Lilli Gruber          |
| 0.15   | 25/07/2003 | 0.24.42 | 37050           | 263            | 21          | Paolo Giani           |
| 11.30  | 17/07/2003 | 0.05.30 | 8250            | 66             | 9           | Stefano Ziantoni      |
| 11.30  | 23/07/2003 | 0.05.23 | 8075            | 81             | 9           | Cristina Guerra       |
| 13.30  | 22/07/2003 | 0.24.37 | 36925           | 281            | 18          | Tiziana Ferrario      |
| 13.30  | 23/07/2003 | 0.26.17 | 39425           | 291            | 18          | Tiziana Ferrario      |

|       |            |                |               |             |            |                    |
|-------|------------|----------------|---------------|-------------|------------|--------------------|
| 13.30 | 26/07/2003 | 0.27.27        | 41175         | 344         | 17         | Tiziana Ferrario   |
| 13.30 | 29/07/2003 | 0.26.38        | 39950         | 277         | 17         | Francesco Giorgino |
| 17.00 | 25/07/2003 | 0.09.31        | 14275         | 83          | 13         | Manuela de Luca    |
| 17.00 | 26/07/2003 | 0.09.57        | 14925         | 127         | 11         | Manuela de Luca    |
| 22.50 | 11/07/2003 | 0.07.31        | 11275         | 66          | 10         | Manuela Lucchini   |
| 22.50 | 14/07/2003 | 0.07.51        | 11775         | 86          | 10         | Manuela Lucchini   |
| 22.50 | 25/07/2003 | 0.10.57        | 16425         | 105         | 13         | Raffaele Genah     |
| 17.00 | 02/09/2003 | 0.10.27        | 15675         | 128         | 12         | Manuela de Luca    |
| 17.00 | 04/09/2003 | 0.11.27        | 17175         | 122         | 11         | Manuela de Luca    |
| 22.50 | 02/09/2003 | 0.05.34        | 8350          | 69          | 8          | Raffaele Genah     |
|       |            | <b>6.36.35</b> | <b>594875</b> | <b>4004</b> | <b>340</b> |                    |

Tabella 1: Dataset RAI1

| <b>Orario</b> | <b>Data</b> | <b>Durata</b> | <b>Durata in frame</b> | <b>Numero di shot</b> | <b>Anchor shot</b> | <b>Anchor</b>                   |
|---------------|-------------|---------------|------------------------|-----------------------|--------------------|---------------------------------|
| 1.00          | 01/07/2003  | 0.24.45       | 37125                  | 197                   | 13                 | Giuseppe Brindisi               |
| 20.00         | 01/07/2003  | 0.34.27       | 51675                  | 354                   | 16                 | Lamberto Sposini                |
| 20.00         | 07/07/2003  | 0.37.02       | 55550                  | 317                   | 15                 | Didi Leoni & Fabrizio Summonte  |
| 1.00          | 28/12/2003  | 0.19.40       | 29500                  | 163                   | 9                  | Gianluigi Gualtieri             |
| 20.00         | 27/11/2003  | 0.35.36       | 53400                  | 341                   | 15                 | Lamberto Sposini                |
| 1.00          | 12/02/2004  | 0.24.54       | 37350                  | 180                   | 9                  | Paolo di Mizio                  |
| 1.00          | 13/02/2004  | 0.24.46       | 37150                  | 171                   | 10                 | Paolo di Mizio                  |
| 20.00         | 12/02/2004  | 0.30.20       | 45500                  | 269                   | 14                 | Cesara Buonamici                |
| 13.00         | 17/07/2004  | 0.34.34       | 51850                  | 304                   | 22                 | Didi Leoni & Fabrizio Summonte  |
| 13.00         | 19/07/2004  | 0.35.12       | 52800                  | 242                   | 21                 | Barbara Petri&Fabrizio Summonte |
| 20.00         | 10/07/2004  | 0.31.08       | 46700                  | 235                   | 13                 | Enrico Mentana                  |
| 20.00         | 03/07/2004  | 0.30.20       | 45500                  | 279                   | 11                 | Cesara Buonamici                |
| 20.00         | 11/07/2004  | 0.30.39       | 45975                  | 282                   | 14                 | Enrico Mentana                  |
| 20.00         | 12/07/2004  | 0.32.55       | 49375                  | 298                   | 13                 | Lamberto Sposini                |

|       |            |                |               |             |            |                                   |
|-------|------------|----------------|---------------|-------------|------------|-----------------------------------|
| 20.00 | 20/02/2004 | 0.37.31        | 56275         | 308         | 16         | Lamberto Sposini                  |
| 13.00 | 05/08/2004 | 0.35.04        | 52600         | 268         | 18         | Didi Leoni &<br>Fabrizio Summonte |
|       |            | <b>8.45.53</b> | <b>788825</b> | <b>3979</b> | <b>229</b> |                                   |

Tabella 2: Dataset Canale 5

Il dataset di Canale 5 ha la particolarità di contenere cinque edizioni del TG interessate dalla presenza di due anchorperson in studio, tipicamente un uomo e una donna.

Il dataset di RAI1 ha invece la particolarità di essere composto da filmati di durata eterogenea. Infatti, distinguiamo in esso le cosiddette edizioni “ordinarie” e quelle “ridotte”, queste ultime caratterizzate da situazioni in cui l’anchor-person commenta “a voce” le immagini dei servizi News.

Una prima distinzione è stata effettuata, quindi, in termini di durata:

- Edizioni brevi (durata < 15 min): 16 edizioni RAI
- Edizioni medio-lunghe (durata > 15 min): tutte le edizioni di C5 e il rimanente delle edizioni RAI

Per mostrare la variabilità degli anchor shot al variare sia dell’emittente televisiva sia della particolare edizione del telegiornale, si riportano nella tabella 3 degli esempi di shot in cui è presente l’anchor, classificandoli sulla base dell’inquadratura di quest’ultimo nella scena.

|                                                                                     |                                                                                     |                                                                                       |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|  |  |  |
| Anchor in<br>primo piano                                                            | Anchor in<br>primo piano                                                            | Anchor in<br>primo piano                                                              |

|                                                                                                                                                  |                                                                                                                                                  |                                                                                                                                                    |
|--------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
|  <p>Anchor dietro alla scrivania con piccolo monitor a lato</p> |  <p>Anchor dietro alla scrivania con piccolo monitor a lato</p> |  <p>Anchor dietro alla scrivania con piccolo monitor a lato</p> |
|  <p>Anchor con ampio monitor</p>                               |  <p>Anchor con video e inviato</p>                             |  <p>Anchor con video e inviato</p>                             |
|  <p>Due anchor</p>                                            |  <p>Anchor in piedi sullo sfondo</p>                          |  <p>Panoramica studio</p>                                     |

Tabella 3: Diverse inquadrature dell'anchor nei vari shot per Tg della Rai e Canale 5

I filmati relativi al dataset sono stati registrati via satellite tramite un decoder satellitare con hard disk interno (a 6000 Kbit/sec in formato MPEG-2) e trasferiti su PC tramite scheda di acquisizione video (Pinnacle DC10+). La qualità finale dei filmati è stata ridotta a 2250 Kbit/sec, 25 frame/sec, ad una risoluzione di 720\*576 pixel, in formato MPEG1, con audio stereo a 44.100 Hz (384 Kbit/sec).

La traccia audio del Tg viene estratta, dal filmato MPEG, mediante l'applicazione VirtualDub: essa è nel formato WAVE a 44.1kHz e 16 bit stereo. Poiché in effetti i due canali della traccia audio si equivalgono, per ridurre i tempi di calcolo nell'estrazione delle feature (che vedremo più avanti) è stato reputato necessario trasformare in mono la traccia stessa mediante l'uso del software Cool Edit Pro 2.0.

## 6. RISULTATI

### *Modulo SUV*

Allo scopo di verificare l'ipotesi che le parti più salienti e caratteristiche di uno speaker sono quelle voiced, abbiamo fatto un confronto delle prestazioni tra il sistema di base e il sistema di base con il mascheramento utilizzando sia l'intero dataset di Rai1 che quello di Canale 5. Nelle tabelle 4 e 5 sono mostrati i risultati ottenuti.

| <i>RAI1</i>      | <b>Video senza masch.</b> | <b>Video con masch.</b> |
|------------------|---------------------------|-------------------------|
| <b>Precision</b> | 95,0%                     | 96,1%                   |
| <b>Recall</b>    | 88,5%                     | 87,2%                   |
| <b>F</b>         | 91,6%                     | 91,4%                   |

Tabella 4: Confronto prestazioni del sistema di base senza e con mascheramento sul dataset di RAI 1

| <i>Canale5</i>   | <b>Video senza masch.</b> | <b>Video con masch.</b> |
|------------------|---------------------------|-------------------------|
| <b>Precision</b> | 87,3%                     | 86,8%                   |
| <b>Recall</b>    | 84,3%                     | 85,8%                   |
| <b>F</b>         | 85,8%                     | 86,3%                   |

Tabella 5: Confronto prestazioni del sistema di base senza e con mascheramento sul dataset di Canale 5

Possiamo osservare che il sistema di base funziona meglio sui telegiornali di RAI1 caratterizzati dalla presenza di un solo anchor in studio rispetto a quelli di Canale 5 in cui sono presenti più anchor. Questo è un comportamento del tutto generale perché il primo dataset ha un carattere più omogeneo rispetto al secondo. Per quanto riguarda, poi, gli indici di prestazione possiamo notare che essi rimangono sostanzialmente identici. Tutto questo è chiaramente indicativo del fatto che il processo di mascheramento non fa perdere informazioni a riguardo del parlante e che quindi tale processo effettivamente individua le parti salienti per discriminare speaker differenti. Da notare che i tempi di computazione si sono ridotti del 40% pur fornendo risultati sostanzialmente identici da un punto di vista dei risultati.

### *Modulo MSD*

Allo scopo di valutare le prestazioni del modulo MSD (multiple speaker detector) e la sua capacità di discriminare uno speaker da più speaker differenti abbiamo utilizzato il dataset dei 16 telegiornali di Canale5 essendo l'unico tra i due disponibili con la caratteristica di avere casi in cui c'è uno speaker o a volte due speaker come conduttori. Nella tabella 6 sono mostrati nel caso migliore, ossia fissata la distribuzione e l'intervallo di ricompattazione, gli errori medi ottenuti dal sistema distinguendo il sesso nei casi in cui c'è uno speaker.

| % Errore medio | M (1sp) | F (1sp) | M+F (2sp) |
|----------------|---------|---------|-----------|
| AUTOCOR        | 15%     | 28%     | 38%       |
| CEPSTRUM       | 5%      | 29%     | 53%       |
| AUTOCOR OTTIM. | 15%     | 21%     | 38%       |

Tabella 6: Errore medio percentuale nel caso migliore.

Come si può osservare la tecnica del cepstrum fornisce i migliori risultati nel caso di speaker maschili con il 5% di errore in media, mentre fornisce le prestazioni peggiori nei telegiornali con due speaker con un errore del 53%. Invece la tecnica della autocorrelazione in entrambi le versioni ottiene l'errore medio più basso, pari al 38%, con due speaker. Nel caso di speaker femminile i miglior risultati sono ottenuti dall'autocorrelazione ottimizzata con un errore medio pari al 21%.

#### *Sistema complessivo*

Concludiamo questo paragrafo con un'analisi dei risultati ottenuti dal sistema complessivo confrontando gli stessi con quelli ottenuti dal sistema di base inteso come quel sistema che utilizza il video e le feature audio di livello locale. Nella tabella 7 sono indicati i migliori risultati ottenuti dal sistema di base e dal nuovo sistema proposto sia nel caso del dataset di Rai 1 sia nel caso del dataset di Canale 5.

| Sistema di base Rai 1<br>(340 Anchor shot)    |     |           |        | Sistema sviluppato Rai1<br>(340 Anchor shot)     |     |           |        | Diff.<br>Relat. |
|-----------------------------------------------|-----|-----------|--------|--------------------------------------------------|-----|-----------|--------|-----------------|
| correct                                       | 301 | Precision | 0,9495 | correct                                          | 333 | Precision | 0,9708 | 2,1%            |
| false                                         | 16  | Recall    | 0,8853 | false                                            | 10  | Recall    | 0,9794 | 10,4%           |
| miss                                          | 39  | F         | 0,9163 | miss                                             | 7   | F         | 0,9751 | 5,9%            |
| Sistema di base Canale 5<br>(229 Anchor shot) |     |           |        | Sistema sviluppato Canale 5<br>(229 Anchor shot) |     |           |        | Diff.<br>Relat. |
| correct                                       | 193 | Precision | 0,9495 | correct                                          | 216 | Precision | 0,9391 | 6,6%            |
| false                                         | 28  | Recall    | 0,8853 | false                                            | 14  | Recall    | 0,9432 | 10,0%           |
| miss                                          | 36  | F         | 0,9163 | miss                                             | 13  | F         | 0,9412 | 8,3%            |

Tabella 7: Risultati del sistema di base e del sistema sviluppato sui due dataset

I risultati tengono conto del numero di anchor shot correttamente identificati (correct), del numero di anchor shot erroneamente identificati (false) e del numero di anchor shot non individuati (miss) e sono espressi secondo i tre classici parametri di valutazione: precision, recall e figura di merito F.

Per quanto riguarda il numero di anchor shot correttamente individuati, le prestazioni del sistema proposto migliorano quelle del sistema di base diminuendo anche il numero di anchor shot omissi. Inoltre si osserva una riduzione del numero di falsi positivi introdotti dal sistema. Nell'ultima colonna della tabella 7 vi è confronto tra i due sistemi per i due dataset utilizzati.

Nel caso del dataset di RAI 1, osserviamo un moderato aumento della precision pari al 2,1% a fronte di un cospicuo aumento del 10,4 % della recall. Questo indica che il sistema proposto tende a diminuire in maniera maggiore il numero dei anchor shot missed rispetto



al numero di falsi positivi introdotti. Per quanto riguarda invece il dataset di canale 5, si nota un miglioramento sia della precision del 6,6% che della recall del 10%. Questo indica che il sistema proposto, nel caso di un dataset più difficoltoso, ottiene prestazioni migliori aumentando sia il numero degli anchor shot correttamente individuati sia diminuendo i falsi positivi che il numero di missed.

|                                          | DATASET | PRECISION | RECALL | F     |
|------------------------------------------|---------|-----------|--------|-------|
| (Chen et alii, 2002)                     | 2h      | 89%       | 92%    | 90,4% |
| (Perez-Freire et alii, 2004)             | 13h     | 90,5%     | 90,5%  | 90,5% |
| (Lan et alii, 2004)                      | 5h      | 92,6%     | 90,5%  | 91,5% |
| Sistema di base (De Santo et alii, 2005) | 6h      | 95%       | 88,5%  | 91,6% |
| (Iyengar et alii, 2000)                  | 1,5h    | 89,2%     | 95%    | 92%   |
| Sistema proposto                         | 6h      | 97,1%     | 97,9%  | 97,5% |
| (Kim et alii, 2005)                      | 6h      | 97,6%     | 99,7%  | 98,6% |

Tabella 8: Confronto con vari sistemi sviluppati in letteratura

Da un confronto del valore della figura di merito F dei migliori sistemi presenti in letteratura che impiegano un approccio integrato audio-video a priori (vedi tabella 8), osserviamo che il sistema preposto si comporta in maniera molto soddisfacente collocandosi al secondo posto assoluto su un dataset omogeneo, per quanto riguarda la durata, a quello utilizzato dal miglior sistema attualmente in letteratura (Kim et alii, 2005).

Riepilogando, da quanto detto, emerge chiaramente che sia l'utilizzo dell'analisi audio su scala globale e sia l'utilizzo delle tecniche di mascheramento e di individuazione di speaker multipli, permette di migliorare i risultati del sistema di base sia in termini di velocità di computazione che di indici di performance. Tali miglioramenti sono stati più cospicui nel caso del dataset di Canale 5 che è caratterizzato da un maggiore difficoltà intrinseca rispetto a quelli ottenuti con il dataset di RA11.

## 7. CONCLUSIONI

Il sistema realizzato ha avuto l'obiettivo di risolvere alcuni dei problemi del sistema di base per la segmentazione in anchor shot. Innanzitutto è stato affrontato il problema dei tempi di computazione. A tale scopo è stato sviluppato e testato un modulo per la selezione delle porzioni salienti in audio, intese come quelle porzioni che effettivamente caratterizzano un anchor rispetto ad un altro. I risultati raggiunti dal sistema originale con l'utilizzo di questo modulo sono praticamente identici a quelli ottenuti dal sistema di base ma riducendo i tempi di computazione del 40% circa. Successivamente, allo scopo di migliorare i risultati del sistema originale, è stato individuato un nuovo set di features audio globali. Tali features sono caratterizzate da un livello di granularità più grande rispetto alle features locali e catturano caratteristiche dell'anchor più generiche ed indipendenti da ciò che viene pronunciato dallo stesso.

Infine è stato sviluppato un modulo per individuare la presenza di un anchor o di più anchor sulla sola base audio che usa il profilo della frequenza fondamentale sull'intero shot e che permette di eliminare alcune ambiguità introdotte dal sistema originale. Dai risultati ottenuti, quindi, il sistema proposto si dimostra più robusto ed affidabile in generale del sistema di base ottenendo dei risultati di tutto rispetto nel panorama della letteratura di settore.

## BIBLIOGRAFIA

De Santo M., Percannella G., Sansone C., Vento M. (2005), "Combining experts for Anchorperson Shot Detection in News Videos", in *Pattern Analysis and Applications*, Springer-Verlag, Berlin. Vol. 7 - 4, pp. 447- 460.

D'Anna L., Percannella G., Sansone C., Vento M. (2006), "Un sistema automatico per la caratterizzazione degli speaker in flussi multimediali" in Atti del 2° convegno AISV, 30 novembre -2 dicembre 2005 Salerno, EDK Editore, Padova, pag.718-730.

Chen S.-C., Shyu M.-L., Liao W., Zhang C. (2002), "Scene change detection by audio and video clues", in Proceedings of ICME 2002, pp.365-368.

Perez-Freire L., Garcia-Mateo C. (2004), "A multimedia approach for audio segmentation in TV broadcast news", in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 369-372.

Iyengar G., Neti C. (2000), "Speaker change detection using joint audio-visual statistics", RIAO (Recherche d'Information Assistée par Ordinateur), Paris.

Lan D.-J., Ma Y.-F., Zhang H.-J. (2004), "Multi-level anchorperson detection using multimodal association", in Proceedings of 17th International Conference on Pattern Recognition, vol. 3, pp. 890-893.

Kim S.-K., Hwang D. S., Kim J-Y. and Seo Y-S. (2005), "An Effective News Anchorperson Shot Detection Method Based on Adaptive Audio/Visual Model Generation", in *Lecture Notes in Computer Science*, vol.3568, pp. 276-285, Springer Verlag, Berlin.

# EFFETTO DEL RUMORE AMBIENTALE SULLA STIMA DELL'SNR NELLE SOTTOBANDE DI INTERESSE DEGLI INDICI BIOMETRICI NELLE COMPARAZIONI FONICHE

Francesco Beritelli

Università di Catania – Dipartimento di Ingegneria Informatica e delle Telecomunicazioni  
*beritelli@diit.unict.it*

## 1. SOMMARIO

Il rumore ambientale rappresenta una delle principali cause di degradazione delle prestazioni in diverse aree applicative dell'elaborazione digitale della voce ed, in particolare, nel campo della codifica, riconoscimento, e della biometria vocale.

In generale, infatti, il comportamento di un sistema di elaborazione della voce dipende dal livello di potenza del rumore presente nell'ambiente, ma anche dal tipo di rumore che può essere di diversa natura (*car, bus, office, restaurant, street, stadium*, ecc) presentando, caso per caso, caratteristiche statistiche e spettrali diverse.

Nell'ambito del riconoscimento automatico o semi-automatico del parlatore, in particolare, il rumore ambientale rappresenta una delle principali cause di alterazione degli indici acustici utilizzati nella fase di identificazione/verifica biometrica.

L'attendibilità di un sistema di comparazione fonica, come è noto, dipende dalla quantità del materiale fonico disponibile, in particolare dalla numerosità delle vocali presenti nella sequenza oggetto della perizia, e dalla qualità del segnale fonico. Il primo aspetto incide sul potere risolutivo del sistema, il secondo ha un impatto sulla corretta stima degli indici biometrici. La qualità delle registrazioni è, quindi, fondamentale per avere una buona accuratezza nella identificazione del parlatore.

E' importante comprendere bene il diverso impatto che il rumore può avere sull'attendibilità di un sistema di comparazione fonica in funzione delle varie caratteristiche tempo-frequenza che esso presenta.

In questo lavoro, in particolare, si analizza il diverso effetto del rumore ambientale sulle prime tre formanti, ovvero gli indici biometrici tipicamente utilizzati nell'ambito delle comparazioni foniche. L'analisi viene effettuata attraverso il calcolo del rapporto segnale-rumore (SNR) nelle sottobande di interesse relative alle prime tre formanti al variare della vocale e del tipo di rumore.

## 2. CARATTERISTICHE SPETTRALI DEI RUMORI AMBIENTALI

La prima fase di questo lavoro è stata rivolta alla caratterizzazione dei rumori ambientali con lo scopo di analizzare e mettere in evidenza le diverse caratteristiche tempo-frequenza dei rumori ambientali.

Il database dei rumori raccoglie un insieme di registrazioni relative a diversi tipi di rumore ambientale, ciascuna della durata di 3 minuti, campionata a frequenza  $F_s = 8 \text{ kHz}$  ed quantizzata linearmente a *16 bit*.

I rumori contenuti nel database sono riconducibili alle seguenti categorie:

- **Bus**, registrazione effettuata all'interno di un autobus durante il suo percorso;
- **Construction**, registrazione del rumore prodotto dalle apparecchiature presenti in un cantiere edile;

- **Dump**, registrazioni del rumore del tubo di scarico di un'automobile in varie condizioni di marcia;
- **Factory**, registrazione dei suoni tipici di un fabbrica;
- **Car**, registrazioni effettuate all'interno dell'abitacolo di una autovettura;
- **Office**, registrazioni effettuate durante le ore di lavoro in un ufficio;
- **Pool**, registrazione dei suoni tipici di una piscina all'aperto;
- **Restaurant**, registrazioni dei rumori tipici di un ristorante;
- **Shopping**, registrazione dei suoni avvertibili in un centro commerciale;
- **Station**, registrazione dei suoni tipici degli ambienti di cui si compone una stazione ferroviaria;
- **Street**, registrazione dei suoni che si avvertono durante una passeggiata per le strade di una città;
- **Stadium**, registrazione dei suoni che si avvertono durante lo svolgimento di una manifestazione sportiva in uno stadio;
- **Train**, registrazione dei suoni che si avvertono all'interno di uno scompartimento di un treno durante la sua marcia.

Una prima analisi riguarda la distribuzione spettrale dei vari tipi di rumore per capire la loro differente incidenza sulle tre sottobande relative ai range di interesse delle prime tre formanti.

La Figura 1, mette in evidenza la densità spettrale di potenza di 4 differenti tipi di rumore presenti nel database. La prima osservazione riguarda la non uniformità dello spettro: ad esclusione del caso di rumore "construction" che rimane pressoché costante per tutta la banda fonica (0÷4 kHz), in generale la densità spettrale decresce con la frequenza e presenta profili diversi da rumore a rumore.

Un'altra considerazione riguarda la stazionarietà/non stazionarietà del rumore: nel primo caso (es. *car* e *bus* noise) la densità spettrale a lungo termine (Fig. 1) con buona approssimazione coincide con quella a breve termine, mentre nel secondo caso (es. *stadium*, *office*, *restaurant* noise) lo spettro a lungo termine è solo una rappresentazione che deriva da una media degli spettri a breve termine.

In generale, quindi, attraverso un'analisi per sottobande, si può osservare una differente incidenza del rumore ambientale e, quindi, a parità di potenza del segnale utile, un diverso SNR al variare delle sottobande.

Sulla base di queste considerazioni sarebbe opportuno valutare la qualità del singolo dato biometrico calcolando l'SNR nella relativa banda di interesse e, più in generale, adattare dinamicamente un sistema di elaborazione digitale della voce in funzione del livello e del tipo di rumore.

Per esempio, potrebbe essere opportuno adattare il set di parametri tipicamente estratti da un segnale in un sistema di pattern matching, i filtri di pre/post filtraggio, le soglie nei blocchi di detection, ecc.

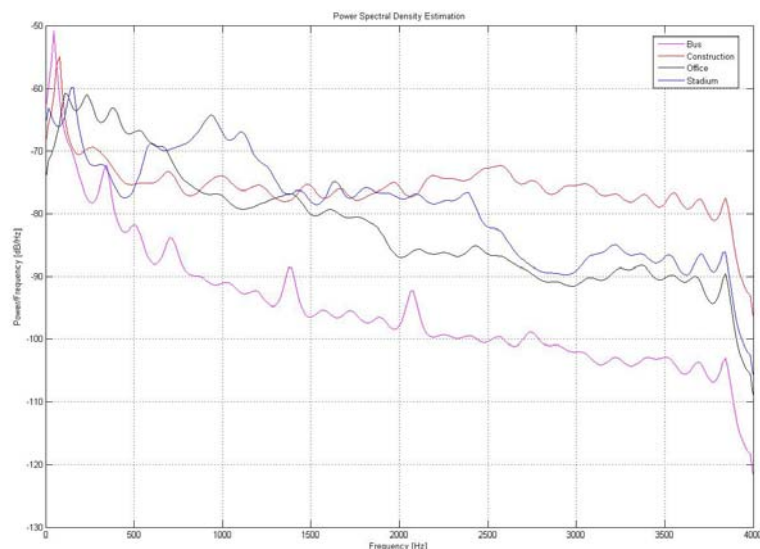


Fig. 1: Densità spettrale di potenza di alcuni tipi di rumori ambientali: bus (linea viola), construction (linea rossa), office (linea nera), stadium (linea blu).

### 3. INDICI BIOMETRICI E SOTTOBANDE DI INTERESSE

Gli indici biometrici maggiormente utilizzati nei metodi parametrici di identificazione automatica del parlatore sono la frequenza fondamentale o pitch ( $F_0$ ) e le prime 3 formanti ( $F_1$ ,  $F_2$ ,  $F_3$ ) calcolate per le quattro vocali dell'italiano: a, e, i ed o. In generale, uno dei parametri utilizzati per la valutazione della qualità di un segnale fonico (principalmente qui si fa riferimento a quello relativo alle intercettazioni in quanto il saggio fonico viene tipicamente registrato in ambienti non rumorosi) è rappresentato dal rapporto Segnale/Rumore (S/N o SNR) ovvero dal rapporto tra la potenza del segnale utile (la voce) e quella del disturbo (che in questo lavoro supponiamo essere solo il rumore ambientale). In genere, nelle tecniche parametriche, durante la fase di segmentazione si scartano le vocali che presentano un SNR calcolato nell'intera banda fonica inferiore a 6 dB in quanto esiste un legame tra SNR e l'alterazione dei valori degli indici biometrici rispetto ai valori nominali che essi assumono quando il segnale fonico è *clean* (ovvero quello registrato in ambienti non rumorosi).

Una importante considerazione riguarda il fatto che una misura dell'SNR su tutta la banda fonica (0÷4 kHz) non tiene conto del diverso effetto del rumore sulle bande di frequenza di interesse dei singoli indici biometrici con una possibile duplice conseguenza: lo scarto di misure "buone" o l'inserimento di dati biometrici "distorti", ovvero caratterizzati da valori che si discostano da quelli effettivi. La prima comporterebbe una riduzione del potere risolutivo, e, la seconda, l'introduzione di dati alterati. In entrambi i casi il risultato del test di identità potrebbe essere compromesso, soprattutto in quei casi in cui il test di comparazione è relativo a due voci appartenenti allo stesso parlatore (falsa esclusione).

Il presente lavoro propone di estendere l'analisi della qualità del segnale fonico ad una analisi più dettagliata rivolta ai singoli dati biometrici. In particolare, per tener conto della diversa distribuzione della densità spettrale di potenza tra varie tipologie di rumori, si calcola l'SNR nelle singole sottobande di interesse relative alle 3 prime formanti: F1, F2 e F3. La Tabella 1 mostra gli intervalli considerati per le sottobande di interesse in cui in genere cadono i valori delle 3 formanti. I valori sono quelli adottati all'interno di IDEM, il tool sviluppato dalla Fondazione Ugo Bordoni per le analisi foniche comparative.

Per semplicità in questo lavoro non si distingue tra una vocale aperta o chiusa in quanto si suppone che il range di frequenza della singola sottobanda non dipende dal tipo di coarticolazione nella pronuncia della vocale. Naturalmente per tenere conto di un sistema vocalico più ampio si tratta semplicemente di considerare nello studio ulteriori range di variazione delle formanti.

I passi più significativi per lo svolgimento di tale lavoro possono essere riassunti come segue:

- Creazione dei database nelle varie condizioni di SNR e tipologie di rumore ambientale attraverso una somma digitale del rumore con il segnale *clean*;
- Segmentazione delle vocali presenti nelle sequenze *clean*;
- Stima dell'SNR nelle 3 sottobande di interesse degli indici biometrici;
- Analisi dei risultati ottenuti.

In particolare il database vocale di partenza è costituito da conversazioni "*clean*" in lingua italiana della durata media di 3 minuti con un tempo di OFF (i tratti di pausa in cui è presente solo il rumore ambientale) medio di circa 5 secondi ed un tempo medio di ON (i tratti di parlato) di circa 3 secondi (la distribuzione dei tempi di ON e OFF è di tipo esponenziale). Il segnale vocale è campionato ad una frequenza di 8kHz e quantizzato linearmente a 16 bit per campione. Il database è stato marcato individuando i tratti legati al set delle vocali *a, e, i, o* presenti nelle conversazioni.

A partire dal database descritto, aggiungendo digitalmente le sequenze di rumore, sono stati creati un totale di 12 altri database, utilizzando tre tipi diversi di rumore (Car, Office, Restaurant), e quattro differenti rapporti segnale-rumore (0dB, 10dB, 20dB, 30dB).

Il rumore di tipo *Car* appartiene alla categoria dei rumori *stazionari*, mentre i rumori *Restaurant* e *Office* appartengono alla categoria dei rumori *non stazionari*.

Per quanto riguarda il calcolo dell'SNR questo è stato calcolato all'uscita di 3 filtri passa banda i cui range di applicazione dipendono dalla vocale in esame in base alla Tabella 1.

| Vocale   | Formante   |            |            |
|----------|------------|------------|------------|
|          | $F_1$ (Hz) | $F_2$ (Hz) | $F_3$ (Hz) |
| <b>A</b> | 478÷752    | 1154÷2700  | 2141÷3700  |
| <b>E</b> | 343÷700    | 1480÷2170  | 2191÷2750  |
| <b>I</b> | 220÷434    | 1746÷2450  | 2451÷3100  |
| <b>O</b> | 379÷715    | 720÷1202   | 2000÷2800  |

Tabella 1: Sottobande di interesse delle formanti al variare delle vocali

Partendo dalla marcatura delle singole vocali, si calcola l'SNR come rapporto tra la potenza del segnale utile e la potenza di rumore, centrando, con una finestra di 512

campioni, la vocale e il corrispondente rumore e applicando separatamente i tre filtri relativi ai range di interesse delle tre formanti.

#### 4. RISULTATI

Le figure 2, 3 e 4 rappresentano l'andamento dell'SNR misurato per un set di 25 esempi delle vocali a, e, i, o, estratte da una conversazione caratterizzata da un SNR medio di 10 dB in presenza di rumore *car*, *office* e *restaurant*. I grafici riportano l'andamento dell'SNR relativo alla sottobanda di interesse della prima formante (spezzata in blu), alla sottobanda della seconda formante (spezzata in verde), alla sottobanda di interesse della terza formante (spezzata in rosso) e, infine, l'SNR totale (spezzata in nero) ovvero l'SNR calcolato sull'intera banda fonica 0÷4 kHz. Una prima semplice considerazione riguarda la variabilità dell'SNR nel caso di rumore *car*, che, essendo di tipo stazionario, è naturalmente causata dalla non stazionarietà del segnale vocale. In generale, gli SNR calcolati nelle sottobande di interesse relative alle tre formanti superano il valore dell'SNR totale, ma si verificano anche, in alcuni casi, delle situazioni inverse.

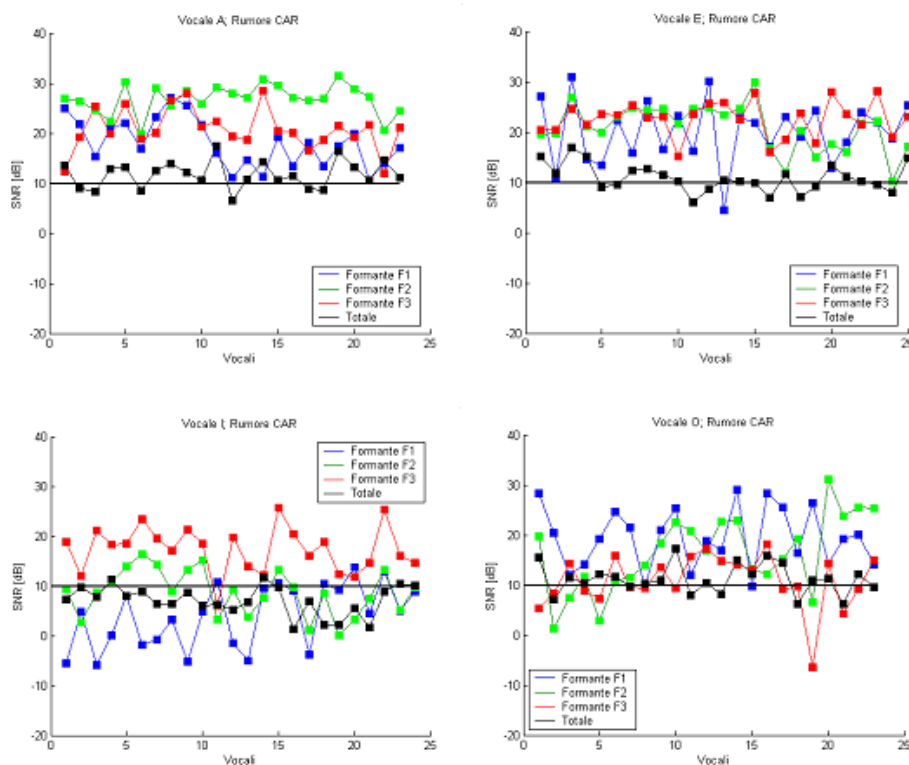


Fig. 2: SNR per un set di 25 esempi di vocali (rumore CAR, 10 dB)

Si veda, in particolare, l'SNR della prima formante per la vocale “i” e l'SNR della seconda e terza formante per la vocale “o” nel caso di rumore *restaurant*. In generale, quindi, si possono osservare una serie di casi in cui sebbene l'SNR totale superi la soglia

critica dei 6dB, l'SNR relativo ad alcuni indici biometrici è inferiore alla soglia critica e una serie di casi in cui si verifica una situazione opposta in cui l'SNR totale è inferiore alla soglia critica e mentre quello calcolato nelle sottobande relative ad alcune formanti risulta sopra soglia. Nel primo caso l'approccio proposto porterebbe ad uno scarto dei dati biometrici con SNR inferiore alla soglia critica mentre nel secondo caso potrebbero recuperare alcuni indici biometrici che da una analisi dell'SNR totale verrebbero scartati. Si analizzi ad esempio il grafico relativo alla vocale "i" nel caso di rumore *car*: il primo dato della serie presenta un SNR totale che supera i 6 dB e quindi, secondo l'approccio tradizionale, si utilizzerebbero tutti i tre dati delle formanti nelle successive fasi di matching, quando la misura dell'SNR relativo alla prima formante presenta addirittura un SNR negativo e, quindi, questo dato andrebbe sicuramente scartato.

Se si analizza il sedicesimo dato della stessa serie si osserva invece che, sebbene l'SNR totale sia inferiore alla soglia critica dei 6 dB, e quindi si scarterebbe l'intera vocale, i valori dell'SNR nelle tre sottobande di interesse delle formanti sono superiori alla soglia critica e, quindi, questa metodologia salverebbe i relativi dati aumentando il potere risolutivo, ovvero l'affidabilità del test. Analoghi ragionamenti possono essere fatti analizzando i grafici relativi alle altre vocali e agli altri tipi di rumore.

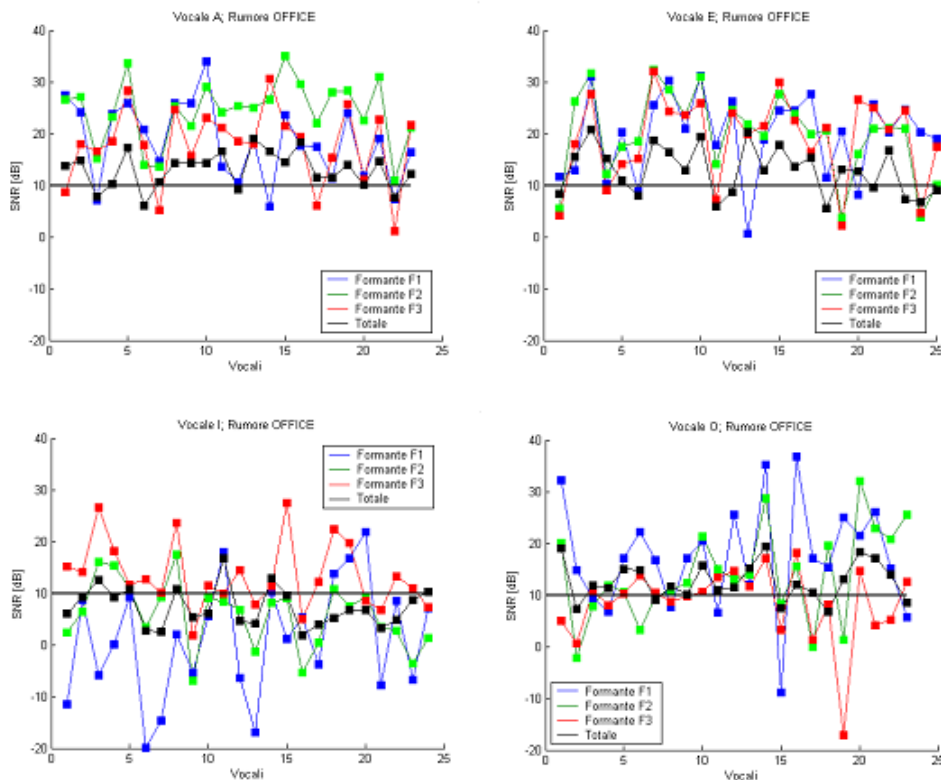


Fig. 3: SNR per un set di 25 esempi di vocali (rumore OFFICE, 10 dB)



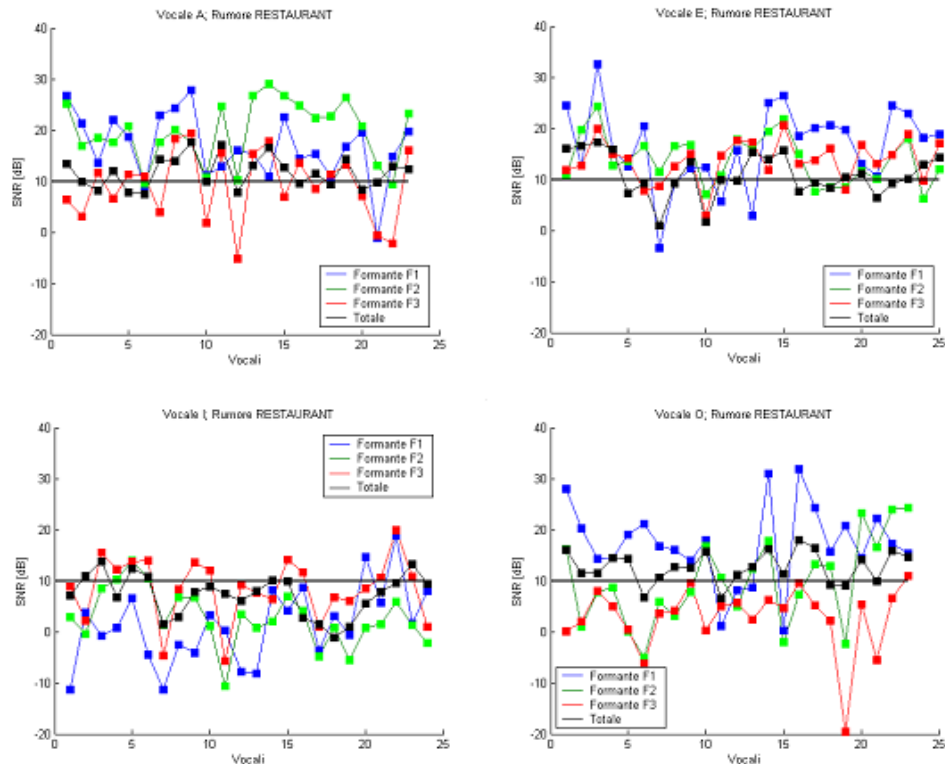
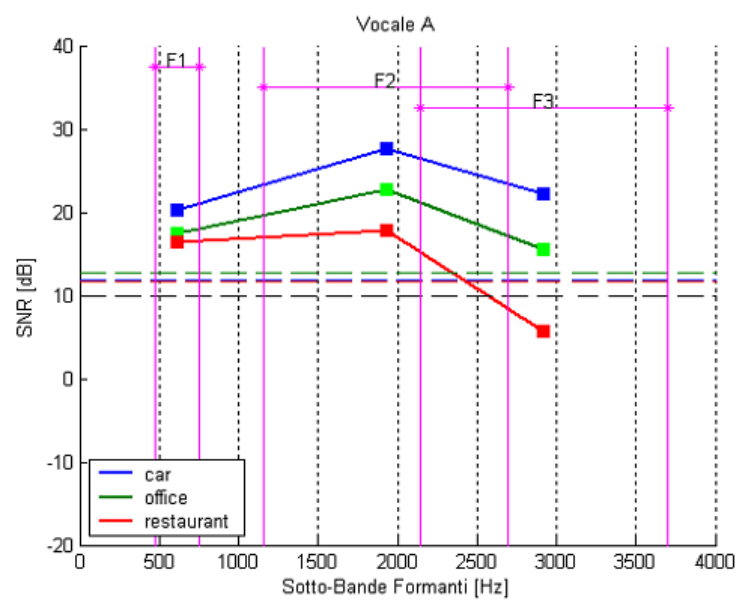


Fig. 4: SNR per un set di 25 esempi di vocali (rumore RESTAURANT, 10 dB)

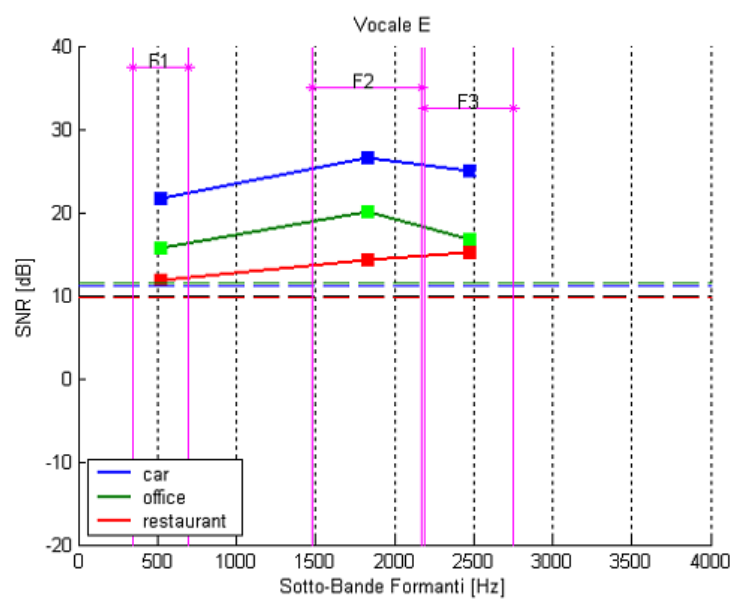
Da questa analisi si deduce che è opportuno effettuare la misura dell'SNR per ogni singola sottobanda di interesse relativa alle tre formanti in quanto si ha una migliore valutazione della bontà del dato biometrico.

La Fig. 5 (a,b,c,d) illustra i valori medi di SNR, espressi in dB, calcolati nelle sottobande di interesse degli indici biometrici nel caso di SNR medio calcolato su tutta la sequenza pari a 10dB e al variare delle quattro vocali e del tipo di rumore. In particolare le curve blu rappresentano il caso con rumore *car*, quelle verdi il caso con rumore *office* e infine le curve in rosso il caso di rumore *restaurant*.

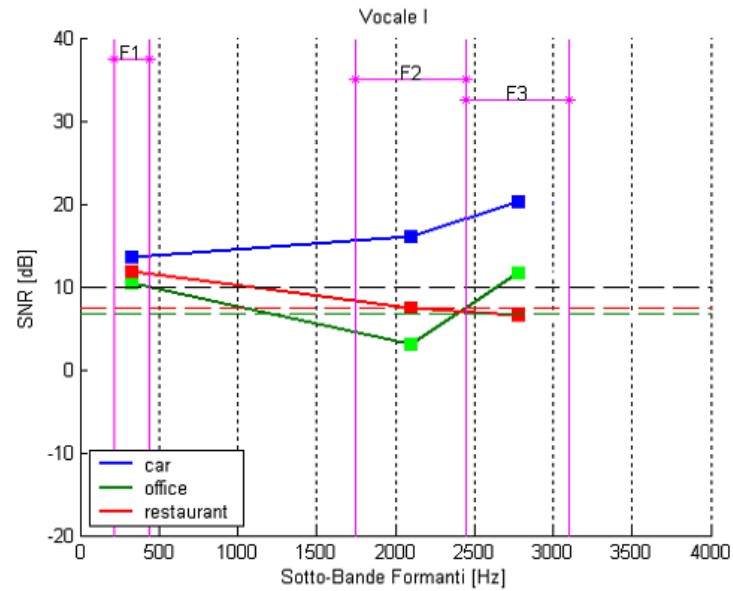
Si nota che uno stesso SNR medio su tutta la conversazione porta a un diverso impatto del rumore a livello di sottobande di interesse degli indici biometrici sia in funzione del tipo di rumore che della specifica sottobanda e vocale.



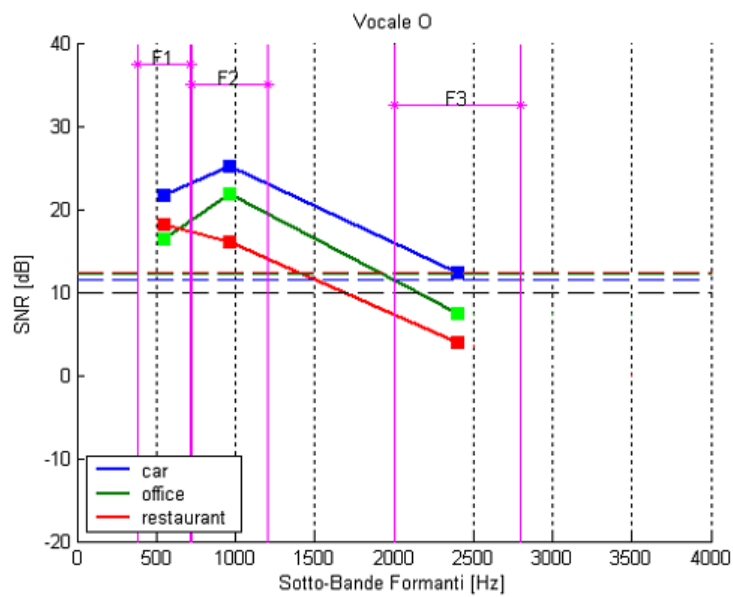
(a)



(b)



(c)



(d)

Fig. 5: Valori medi dell'SNR (in dB) nelle sottobande di interesse degli indici biometrici nel caso di SNR medio su tutta la banda pari a 10dB (linea tratteggiata)

In generale è possibile verificare che l'effetto del rumore di tipo *car* non altera il segnale nelle sottobande di interesse come evidenziato in figura 6. La figura, in particolare, riporta, nel caso della vocale "a", la distribuzione dello scarto delle frequenze formati rispetto al valore nominale del caso clean, una volta aggiunto rumore *car* a 10 dB di SNR. Come si evince le variazioni dei valori di frequenza sono minimi per tutte le 4 formanti e questo conferma i risultati mostrate in figura 5 (a) relativi alle misure di SNR per la vocale "a" che superano sempre i 20 dB.

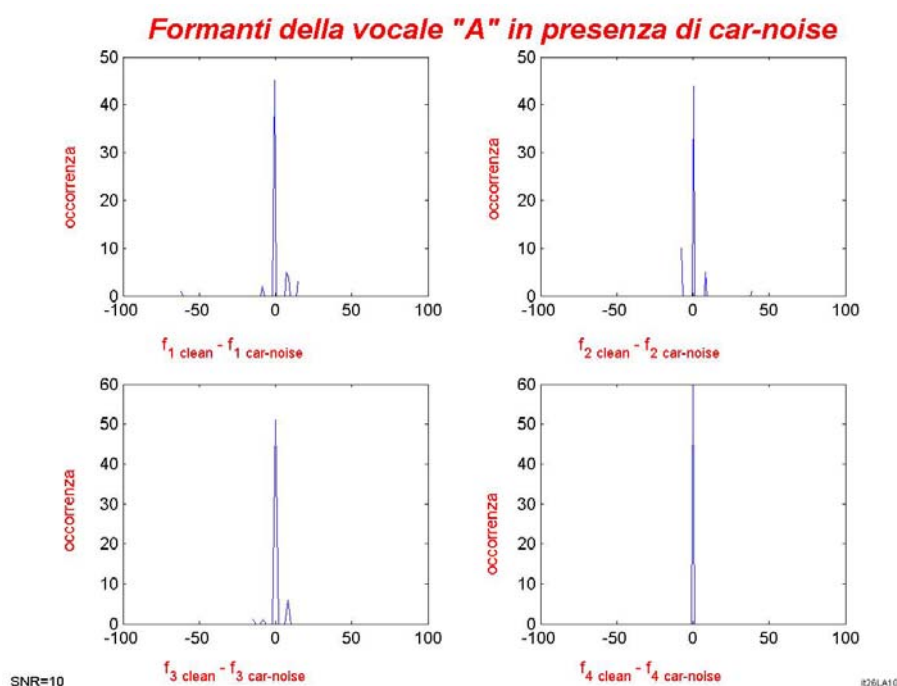


Fig. 6: Sensibilità delle formanti al rumore *car*: distribuzione delle variazioni di frequenza subite a causa del rumore

La figura 7 illustra le distribuzione dello scarto dei valori delle formanti nel caso di rumore *restaurant* a 10 dB di SNR nel caso della vocale "i". In questo caso si ha una certa sensibilità alla presenza del rumore e si notano delle variazioni di frequenza principalmente di piccola entità ma che possono superare in alcuni casi i 100 Hz. Tra le 4 formanti sicuramente la seconda è la più sensibile al disturbo in quanto presenta una più ampia varianza della distribuzione. Anche in questo caso, guardando la figura 3 nel caso della vocale "i" possiamo dedurre che esiste un certo legame tra entità delle variazioni subite dalle frequenze formanti e l'SNR misurato in quanto l'SNR è intorno ai 10 dB se misurato nelle bande di interesse della prima e terza formante e vale circa 0 dB nel caso in cui viene calcolato nella sottobanda di interesse della seconda formante.

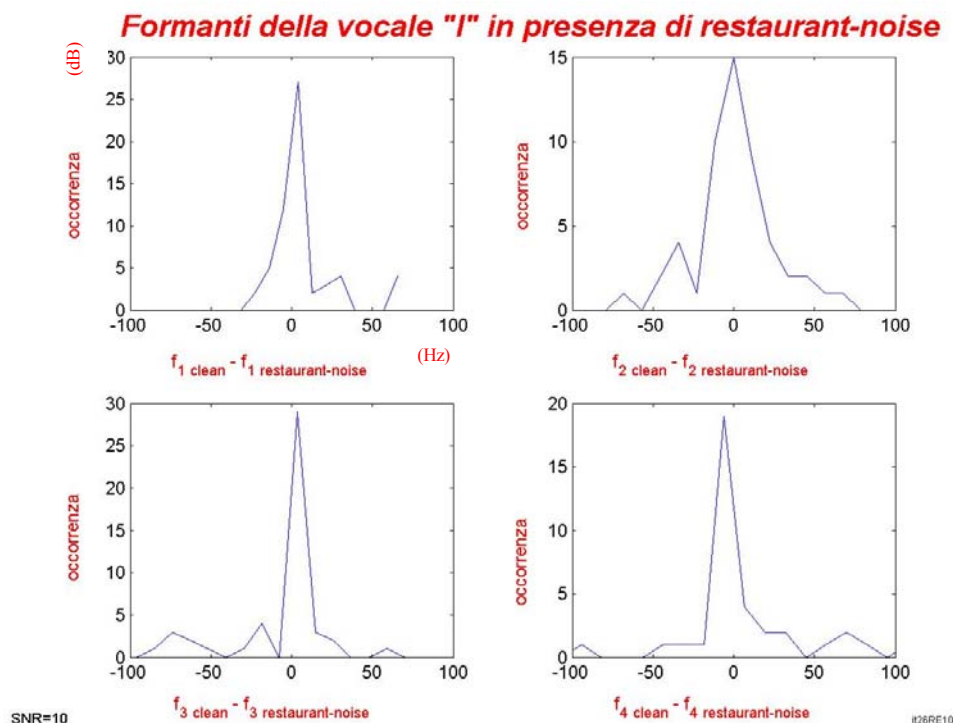


Fig. 7: Sensibilità delle formanti al rumore *restaurant*: distribuzione delle variazioni di frequenza subite a causa del rumore

L'analisi della distribuzione delle variazioni dei valori formatici di sequenze rumorose rispetto al caso clean, suggerisce un criterio per la determinazione degli soglie critiche di SNR al di sotto delle quali è opportuno scartare la singola misura biometrica. Per ogni tipo di rumore, per ogni vocale, e per ogni formante si impone il valore della deviazione standard massima che può subire una formante a causa del rumore. In corrispondenza di tale valore si determina la corrispondente soglia critica.

## 5. CONCLUSIONI

In conclusione, questo lavoro ha messo in evidenza il diverso effetto del rumore ambientale sul calcolo dell'SNR nelle sottobande di interesse degli indici biometrici adottati nelle comparazioni foniche. I risultati ottenuti indicano un diverso impatto del rumore ambientale al variare della vocale e della singola formante. La stima dell'SNR nella sottobanda di interesse di ogni indice biometrico permette una più accurata selezione dei dati biometrici da utilizzare nelle successive fasi relative al test di identità e di verosimiglianza per il calcolo dell'errore di falsa identità o di falsa esclusione.

## BIBLIOGRAFIA

- Hollien, H. (2001), *Forensic Voice Identification*, London: Academic Press.
- Nolan, F. (1997), Speaker Recognition and Forensic Phonetics, in *A Handbook of Phonetics Science* (W. Hardcastle and J. Laver, editors), Oxford: Blackwell, 146-153.
- Falcone, M., Paoloni, A., De Sario, N. (1995), IDEM: A Software Tool to Study Vowel Formant in Speaker Identification, in *Proceedings of the ICPhS '95*, Stoccolma, 145-150.
- Paoloni, A. (2003), Note sul riconoscimento del parlante nelle applicazioni forensi con particolare riferimento al metodo parametrico IDEM, *Rivista Italiana di Acustica*, 27, n. 3-4, 113-128.
- Reynolds, D.A. (1995), Automatic Speaker Recognition Using Gaussian Mixture Speaker Model, *The Lincoln Laboratory Journal*, 8, 173-191.
- Rose, P. (2002), *Forensic Speaker Identification*, London: Taylor and Francis.

# SVILUPPO DI UN SISTEMA DI *KEYWORD SPOTTING* PER L'INDICIZZAZIONE AUTOMATICA DEI DOCUMENTI AUDIO

Graziano Tisato\*, Piero Cosi\*, Isabella Gagliardi°

(\*) Istituto di Scienze e Tecnologie della Cognizione del CNR  
Via Martiri della Libertà, 2 - 35127 Padova, Italia  
[tisato@pd.istc.cnr.it](mailto:tisato@pd.istc.cnr.it), [cosi@pd.istc.cnr.it](mailto:cosi@pd.istc.cnr.it)

(°) Istituto per le Tecnologie della Costruzione del CNR  
Via Bassini 15 - 20133 Milano, Italia  
[gagliardi@itc.cnr.it](mailto:gagliardi@itc.cnr.it)

## 1. SOMMARIO

L'applicazione del *Keyword Spotting* (KWS) presentata in questo lavoro rientra nel campo più generale dell'*Information Retrieval* (IR), e in quelli più specifici dello *Spoken Document Retrieval* (SDR), dell'*Automatic Speech Recognition* (ASR) e del *Large Vocabulary Continuous Speech Recognition* (LVCSR).

Il *Keyword Spotting* è sostanzialmente un processo di *speech-to-text* del tutto simile al riconoscimento del parlato, in questo caso limitato solo all'individuazione di determinate parole chiave all'interno di un flusso audio continuo.

Il campo applicativo del KWS va dall'indicizzazione dei documenti audio, alla loro categorizzazione, ai sistemi di comando vocale, al rilevamento di particolari eventi, alla consultazione vocale delle basi dati, ad esempio per i motori di ricerca del Web, ecc. L'insieme delle applicazioni del KWS è anche sommariamente definito come *Speech Analytics*.

In generale, il KWS può rappresentare un valido aiuto nell'interazione uomo-macchina, permettendo l'uso del linguaggio naturale nella comunicazione.

Lo sviluppo e il miglioramento delle tecniche di riconoscimento del parlato e del KWS, assieme con la riduzione del tempo di elaborazione, sceso ormai al di sotto del tempo reale, ha esteso il campo applicativo dell'IR, in passato limitato ai documenti testuali, anche ai documenti audio.

Per dare un'idea del progresso ottenuto, mentre nel 1997, con un database di addestramento di 150 ore di parlato, si riusciva ad ottenere un errore sul riconoscimento delle parole cercate del 22%, già nel 2004 con *corpora* molto più grossi si scendeva al 9-10%.

La sfida attuale riguarda le tipologie di parlato che tradizionalmente ottengono i risultati peggiori, e cioè:

- Parlato rumoroso (telefonia, conferenze, ecc.).
- Parlato conversazionale.
- Variabilità di stili e accento dei parlatori.

Per quanto riguarda il *Keyword Spotting*, l'estrazione di determinate parole chiave può essere il primo passo di procedure di elaborazione tipiche dell'IR, che sono tradizionalmente basate sul testo, per ottenere le informazioni volute.

In certe applicazioni, in effetti, può essere di maggior utilità estrarre la presenza di parole significative dal punto di vista semantico, piuttosto che ricavare l'intera sequenza del

parlato, in modo da lanciare una azione appropriata. In questi casi, l'interesse è dato dalla velocità con cui si ottiene la risposta ad una interrogazione. Si tenga comunque presente che la velocità di elaborazione del *KWS* non è attualmente molto diversa da quella del riconoscimento del parlato continuo.

Una obiezione a questo approccio potrebbe essere quella che gli *ASR* non sono infallibili e sfornano un gran numero di parole errate, che possono avere una qualche parentela fonetica con l'originale. Si potrebbe pensare che lo scambio di parole nel processo di *Keyword Spotting* possa pregiudicare il funzionamento di un sistema di *IR*, che dipende forzatamente dalla correttezza delle stesse.

La scoperta, per certi versi sorprendente, fatta in questi ultimi anni è che l'influenza di questi errori sulle prestazioni complessive di un sistema di *IR* è molto limitata per la naturale ridondanza delle parole chiave relative ad un certo argomento. In effetti, è molto improbabile che tutte le occorrenze di una certa parola o dell'insieme delle parole chiave siano contemporaneamente scambiate con parole errate o semplicemente ignorate.

Ad esempio, con una percentuale di parole errate (*WER*) che passi dallo 0% al 40%, l'efficacia del sistema *IR* nell'individuare un documento secondo certi criteri diminuisce solo del 10% [Ng, 2000], [Allan, 2002]: si veda, ad esempio, gli esperimenti fatti dal 1997 (*TREC-6*) al 2000 (*TREC-9*) dalla *NIST Text REtrieval Conference (TREC)* oppure nel 1998 dalla *Topic Detection and Tracking (TDT)*.

Questo spiega l'interesse che può rivestire l'utilizzo del *Keyword Spotting* nel campo dell'*Information Retrieval*.

La relazione si articola in questi argomenti:

- Introduzione al *Keyword Spotting* (Cap. 2)
- Possibili applicazioni (Cap. 3)
- Architetture implementate in questo lavoro e che si basano sull'azione contemporanea di due canali di riconoscimento (Cap. 4):
  - Il primo è un tipico *ASR* basato su un Modello Acustico (*AM*) e su un Modello Statistico del Linguaggio (*LM*).
  - Il secondo implementa una Grammatica a Stati Finiti (*GSF*), che non necessita della modellazione di un *LM* e permette la ricerca di una parola qualsiasi.
- Misure di valutazione della performance di un sistema di *WKS* (Cap. 5).
- Le caratteristiche dell'interfaccia grafico realizzato, per permettere la configurazione in una forma interattiva e rapida dei parametri dell'*ASR* utilizzato (Sonic – CSLR dell'Università del Colorado), la visualizzazione e la verifica immediata dei risultati della ricerca delle parole (Cap. 6).
- La valutazione dei risultati ottenuti, che nel caso del *Keyword Spotting* presenta un certo grado di complessità, dal momento che dipendono dai documenti scelti per il test e dalla velocità che si vuole imporre all'*ASR*. Sui test disponibili, i risultati della precisione delle parole riconosciute correttamente (60%) va giudicato con una certa indulgenza, considerando che si è utilizzato in questa prima fase del lavoro un Modello Acustico, *speaker independent*, ricavato da un corpus (*APASCI*) non conversazionale, e dunque non adatto ai documenti analizzati (Cap. 7-8).
- Le prospettive future che prevedono fra l'altro l'addestramento di un Modello Acustico su parlato conversazionale, l'utilizzo di parser semantici, l'uso di tecniche di adattamento (*Vocal Tract Length Normalization*, *Structured Maximum a Posterior*



*Linear Regression*, ecc.), dovrebbero contribuire a migliorare sensibilmente le prestazioni dell'ASR (Cap. 9).

## 2. INFORMATION RETRIEVAL E KEYWORD SPOTTING

Lo sviluppo scientifico e tecnologico ha reso evidente l'importanza della conoscenza e del trattamento della conoscenza nel mondo contemporaneo.

La quantità, la complessità e la varietà della conoscenza umana aumenta esponenzialmente assieme ai documenti, archivi, *database*, ecc. che la contengono a vari livelli gerarchici.

Ora, mentre ci si rende conto facilmente della difficoltà, ma, diciamo anche, dell'impossibilità e dell'inutilità di memorizzare ogni cosa nella propria testa, diventa invece ogni giorno più vitale sapere dove trovare e come accedere alle diverse forme della conoscenza disponibile.

L'approccio meno difficoltoso e costoso è quindi quello di trovare mezzi veloci ed efficaci per il recupero dell'informazione nel momento in cui si rende necessaria, piuttosto che cercare di stipare *una tantum* nel cervello una quantità di nozioni immensa.

In definitiva, l'aspetto rilevante della questione è il raggiungimento dello scopo (ottenere l'informazione voluta), e non il contenitore o la modalità (con cui l'informazione è ricavata).

L'azione combinata di queste esigenze e dei progressi delle scienze e delle tecnologie informatiche e di telecomunicazione ha portato alla attuale civiltà dell'informazione.

In particolare, per quanto riguarda l'*Information Retrieval*, si è assistito fin dal primo dopoguerra (vedi il *Memex* di *Bush* del 1945 [Buckland, 1992]) ad un proliferare degli studi teorici e delle implementazioni pratiche relative al recupero di informazioni da fonti testuali. I motori di ricerca sul *Web* sono la punta dell'*iceberg* di questa attività.

Oltre a questa tipologia statica di documenti, esiste però una miniera di informazioni, potenzialmente inesauribile, costituita dai documenti multimediali (archivi audio e video, trasmissioni radio e TV, lezioni accademiche, ecc.) che si sono accumulati negli anni passati e che aumentano vertiginosamente di giorno in giorno.

Una ulteriore categoria (storicamente la prima) candidata alla ricerca di informazioni è costituita dalle intercettazioni delle conversazioni telefoniche per scopi militari, di sicurezza nazionale e di spionaggio industriale (leggi *Echelon*) [Cernocky et al., 2007], ma anche per la supervisione delle conversazioni cliente-azienda, ecc.

Rispetto ai documenti testuali, i documenti multimediali pongono grossi problemi legati alla:

- Difficoltà di indicizzarli, classificarli, riassumerli, ecc., data la quantità enorme che è prodotta ogni giorno.
- Difficoltà per la ricerca e la consultazione, dato che per loro natura questi documenti hanno un accesso sequenziale.

Il *Keyword Spotting (KWS)* è una prima risposta all'esigenza di estrazione di informazioni da documenti multimediali e consiste nell'individuazione di un certo insieme di parole (*Keywords*) rilevanti dal punto di vista semantico [Higgins et al., 1985], [Rose et al., 1990], [Foote et al., 1995].

Essendo l'informazione ricavata da documenti audio il *KWS* può anche essere chiamato con il nome di *Acoustic Keyword Spotting*.

Come si è detto nel sommario, il *KWS* è un settore di ricerca tipicamente interdisciplinare che fa riferimento a:

- *Information Retrieval (IR)* (che a sua volta si ispira alle scienze cognitive, alla [linguistica](#), alla [semiotica](#), alla [scienza dell'informazione](#), *data mining*, ecc.)
- *Spoken Document Retrieval (SDR)*
- *Automatic Speech Recognition (ASR)*
- *Large Vocabulary Continuous Speech Recognition (LVCSR)*.  
In effetti il *KWS* può essere realizzato in due processi distinti:

- *Speech-to-text*: Questa prima fase è simile al riconoscimento automatico del parlato per cui dalla voce in ingresso si ottiene una sequenza di parole, con la sostanziale differenza che nel *KWS* la ricerca può essere limitata solo a determinate parole chiave (o ad una loro combinazione). Generalmente dunque il vocabolario di un sistema di *KWS* consiste di un numero limitato di parole.
- *Information Retrieval*: Nella seconda fase si sfruttano tecnologie di *Information Retrieval* già sviluppate da alcuni decenni per ricavare le informazioni volute dalla sequenza delle parole in uscita dall'*ASR*.

Gli sviluppi del *KWS* sono relativamente recenti, dal momento che dipendono strettamente dalle prestazioni dei sistemi di riconoscimento automatico del parlato e che solo recentemente sono arrivati a *performance* e a tempi di risposta soddisfacenti.

Un contributo fondamentale al *KWS* basato su *Hidden Markov Model (HMM)* è dovuto al lavoro di Rose e Paul, fra i primi ad utilizzare l'algoritmo di Viterbi [Rose et al. 1990]. Questo approccio di calcolo della probabilità delle parole in uscita garantiva che le ipotesi generate per le varie parole non si sovrapponevano, ed limitava la generazione dei cosiddetti *falsi allarmi*, e cioè l'individuazione errata di parole chiave [James, 1995].

Assieme ai lavori teorici e pratici che si sono succeduti negli anni, si deve ricordare l'indispensabile attività dovuta al *National Institute of Standards and Technology (NIST)*, che è un'agenzia del governo degli [Stati Uniti d'America](#), alla *Defence Advanced Research Projects Agency (DARPA)* e all'*Advanced Research and Development Activity (ARDA)* del Dipartimento della Difesa degli USA. A partire dal 1992 il *NIST* sponsorizzò una serie di conferenze (*Text Retrieval Conference - TREC*) fondamentali per stabilire i criteri di valutazione oggettiva dei sistemi di *Information Retrieval*. In particolare, fra il 1997 e il 2000, la conferenza si dedicò alla valutazione dei documenti di parlato. Fra il 1998 e il 2004, il *NIST* sviluppò un programma di ricerca intitolato *Topic Detection and Tracking (TDT)* che faceva parte del *Translingual Information Detection, Extraction, and Summarization (TIDES)*, sponsorizzato dal *DARPA*. Un altro importante progetto, intitolato *Spoken Term Detection (STD)*, fu lanciato nel 2006 per la ricerca su archivi audio eterogenei in tre lingue (Arabo, Inglese, e Cinese), che non sia dipendente dal vocabolario (le keyword della ricerca non sono note a priori dal riconoscitore).

### 3. APPLICAZIONI DI KEYWORD SPOTTING

Come si è detto nei capitoli precedenti, il *KWS* ha un suo ambito preciso nell'*Information Retrieval*, per cui non interessa trascrivere automaticamente l'intera sequenza del documento parlato, ma recuperare solo le parole che lo possano qualificare in base a criteri voluti.

Il vantaggio non trascurabile di questo approccio è dato dalla velocità con cui si ottiene il risultato, visto che può essere di vari ordini di grandezza inferiore ad una completa decodifica con un normale *ASR*. In effetti, i tempi di esecuzione di un *ASR* crescono (anche

se non-linearmente) con l'aumentare del numero di parole cercate, per cui con un *KWS* di qualche centinaio di parole il tempo può essere di molto inferiore alla durata del brano in analisi (vedi cap. 7-8), mentre un riconoscitore che usi un vocabolario di decine di migliaia di parole può impiegare un tempo decine di volte il tempo reale, rendendo almeno per il momento improponibile questa strada di fronte alla enorme quantità dei documenti candidati ad una trascrizione.

I campi di applicazione del *KWS* sono i seguenti:

- Indicizzazione dei documenti audio, per cui si individuano le parole rilevanti in un documento e si ricava la loro posizione temporale.
- Categorizzazione (catalogazione, classificazione) dei documenti audio, per cui determinate parole chiave servono a collocare opportunamente il documento secondo criteri e algoritmi tipici del *NLP* (*Natural Language Processing*), e dell'analisi statistica e semantica.
- Sistemi di comando vocale, per cui il sistema risponde a determinate parole o gruppi di parole con l'azione appropriata (ad esempio nei servizi di risposta automatica delle banche, ecc.).
- Rilevamento di particolari eventi, simile al punto precedente, ma senza implicazione di interattività con l'utente.
- Intercettazioni su conversazioni telefoniche e comunicazioni radio e video.
- Consultazione delle basi dati (ad es. per i motori di ricerca del *Web*).
- Supervisione di chiamate ai *Call Center*, per cui il verificarsi di determinate circostanze (parole o numero di ripetizioni di una parola) in una conversazione fra cliente e assistente del centralino di una ditta provocano l'intervento di un responsabile addetto a questo compito.

#### 4. ARCHITETTURA DEL KEYWORD SPOTTING

Qualunque sia il disegno implementativo di un *KWS*, il compito è solo apparentemente più semplice di un *ASR* normale. In effetti le parole da cercare non sono isolate, ma immerse in una catena fonica continua che subisce profonde modificazione per il contesto, per i fenomeni di coarticolazione, e per le caratteristiche individuali del parlante (età, sesso, varianti dialettali, ecc.).

A questo si deve aggiungere il rumore introdotto dal mezzo di comunicazione e dall'ambiente.

I problemi e le soluzioni adottate sono del tutto simili a quelle del normale riconoscimento, e cioè:

- Identificare inizio e fine delle parole.
- Estrarre quelle che con la probabilità più alta corrispondono alle parole volute.

Formalmente si tratta di un processo di decodifica, per cui data la sequenza vettoriale *O* dei coefficienti *HMM* corrispondenti al parlato analizzato, si calcola la probabilità che una parola *W* (anch'essa rappresentata da una *HMM*) produca la sequenza osservata *O*, e cioè  $P(W|O)$ .

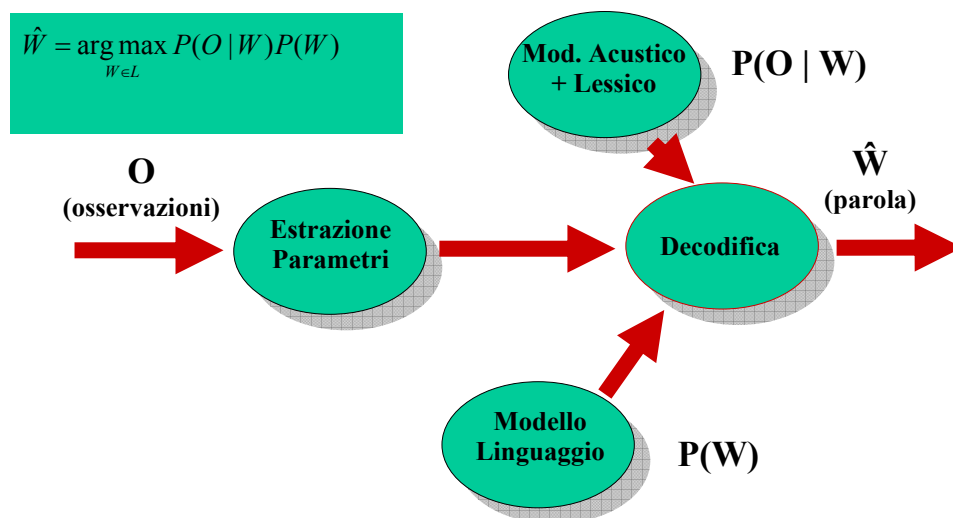


Fig. 1 – Schema di ASR con Modello Acustico e Modello del Linguaggio.

Una stima diretta di  $P(W|O)$  è in realtà poco praticabile, per cui si ricorre ad una riformulazione del problema seconda la regola di Bayes:

$$\hat{W} = \operatorname{argmax}_W P(W|O) = \operatorname{argmax}_W P(O|W) P(W) / P(O) = \operatorname{argmax}_W P(O|W) P(W)$$

dove  $P(W)$  è la probabilità a priori di osservare la parola indipendentemente dal reale input vocale che si stia analizzando e corrisponde ad una modellazione opportuna di quella particolare tipologia di linguaggio che si vuole analizzare (Fig. 1).

Nella seconda parte dell'equazione,  $P(O|W) / P(O)$  si riduce al solo termine  $P(O|W)$ , dal momento che  $P(O)$  non dipende dalla variabile  $W$ .  $P(O|W)$  calcola la probabilità di osservare un certo insieme di vettori cepstrali  $O$  all'occorrenza di una parola  $W$ . Questa quantità deriva dall'addestramento di un Modello Acustico su un certo insieme di campioni del parlato.

Il *Keyword Spotting* consiste dunque nel cercare la sequenza di stati di una certa parola chiave  $\hat{W}$ , che con maggior probabilità produca il parlato osservato  $O$ . La sequenza più probabile si ottiene massimizzando l'equazione vista.

Nel *KWS* sono definite due classi distinte di modelli [Higgins et al., 1985], [Rose et al., 1990], [Manos, 1996], [Pellom, 2001], [Pellom et al., 2003]:

Il primo modello riguarda le *keywords* che sono rappresentate dalla loro sequenza fonetica. Questa modellizzazione è unica e caratteristica, come si vede nella Fig. 2, ad esempio, per la parola inglese /find/.

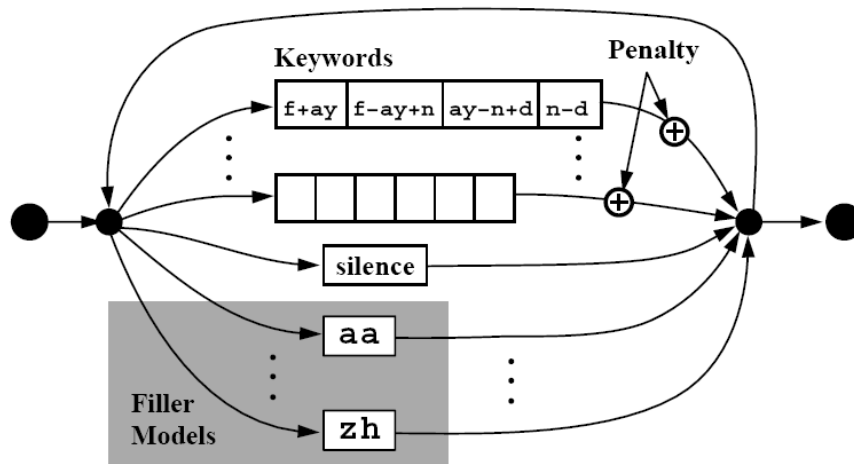


Fig. 2 – Modellazione delle parole chiave e dei *filler/garbage* nel KWS [Foote et al., 1995].

Il secondo modello riguarda i *filler/garbage*, di tutto ciò che deve essere eliminato come riempimento o spazzatura, e tutto ciò che non è una *keyword* compresa nel vocabolario delle parole chiave. Possono comparire in qualsiasi momento nel corso di una frase e sono assoggettati solo ad un parametro che ne controlla il comportamento (ad esempio, in Sonic, si tratta del *filler\_penalty*).

È importante disporre di una qualche metodologia per la modellazione acustica di questi oggetti, in modo da ridurre i falsi allarmi. Nel caso più elementare questi modelli si riducono a una serie di singoli foni (Fig. 2). In altri casi si può volere la modellazione esplicita di tutte le parole da escludere. Anche fenomeni paralinguistici come la risata, il respiro, la tosse, ecc., possono essere modellati opportunamente nello stesso modo.

Sono stati tentati approcci diversi per la modellazione dei *filler/garbage*: ad esempio, nello *sliding window* si utilizzano tecniche di programmazione dinamica per estrarre le parole cercate [Wilpon et al., 1990], [Junkawitsch et al. 1996], [Silaghi et al. 2005]. Ogni percorso dinamico è considerato come l'occorrenza di una possibile parola chiave. È possibile rimuovere le sovrapposizioni eventuali di parole uguali e normalizzare in un secondo tempo i punteggi assegnati. Infine, opportune soglie permettono di ricavare le parole volute.

A differenza del modello *filler/garbage*, questa tecnica non richiede l'esplicita modellizzazione delle parti del parlato che non interessano la ricerca. Si calcola invece con Viterbi un allineamento che comincia in ogni finestra, dal momento che la parola chiave può essere presente in un punto qualsiasi. Si ricava il punteggio di tutti gli allineamenti che iniziano e finiscono in quel *frame*, e si assegna alla Misura di Confidenza per una certa *keyword*, il punteggio dell'allineamento migliore [Hacioglu et al., 2002].

A questo punto, si considera che una certa parola chiave sia effettivamente presente, se la misura della confidenza del modello di quella parola indica una alta probabilità di allineamento.

L'architettura di un sistema di KWS può essere classificata secondo tre principali categorie:

- *KWS* basato su *Large Vocabulary Continuous Speech Recognition (LVCSR)* (cap. 4.1).
- *KWS* basato su un riconoscimento di tipo fonetico (cap. 4.2).
- *KWS* basato su una Grammatica a Stati Finiti (cap. 4.3).

#### 4.1 - *KWS* basato su *LVCSR*

Come risulta evidente dai capitoli precedenti, l'approccio più diretto al *KWS* che si possa pensare è quello di ricorrere ad un normale *ASR* del tipo *Large Vocabulary Continuous Speech Recognition*, tradizionalmente dotato di un Modello Acustico e di un Modello di Linguaggio, secondo lo schema di Fig. 1. [Rohlicek, 1989]. La soluzione banale è quella di analizzare l'uscita del riconoscitore per scoprire le eventuali parole chiave presenti.

Nella prima fase, che chiameremo di *speech-to-text*, un riconoscitore di parlato continuo ricava l'informazione testuale dai documenti (Fig. 3).

Nella seconda fase, si elabora l'uscita testuale del riconoscitore con i metodi tipici di *Information Retrieval* (Fig. 3).

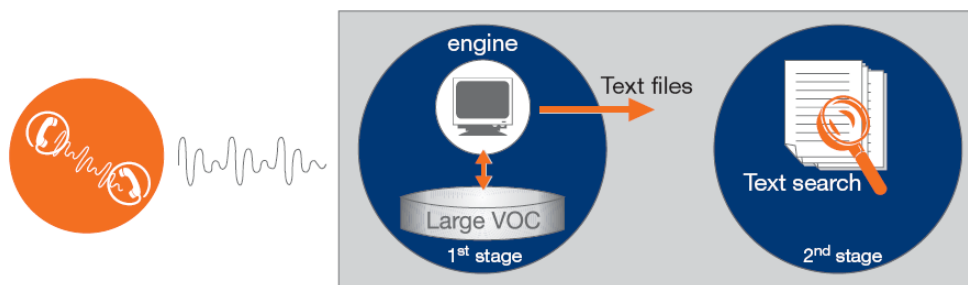


Fig. 3 – Sistema di *Keyword Spotting* basato su un *LVCSR* [Alon, 2005].

I vantaggi di questo approccio, nell'ipotesi che l'errore totale sulle parole (*WER*) sia basso, sono molti:

- Si può ottenere, volendo, la trascrizione completa dei documenti, e dunque una maniera economica per l'archiviazione.
- Le ricerche successive, che sfruttano il testo decodificato, sono ovviamente molto più veloci.
- Non c'è necessità di determinare a priori le parole chiave (tutte quelle presenti nel vocabolario sono candidate possibili).
- Si possono cercare nuove parole (sempre che siano presenti nel vocabolario) senza dover ricorrere ad un'altra fase di elaborazione.
- Con questa soluzione si ottengono i risultati migliori rispetto alle altre soluzioni, dal momento che il Modello del Linguaggio provvede ad eliminare molti errori.

Gli svantaggi sono dovuti alla poca flessibilità del sistema per le parole non comprese nel vocabolario:

- Il *LVCSR* comporta la necessità di creare un Modello del Linguaggio adatto al dominio di competenza (e questo può essere penalizzante).
- Il vocabolario previsto per il *LVCSR* non riesce a esaurire tutte le parole chiave che l'utente può richiedere (problema del *out of vocabulary words* – *OOV*). Quando si presenta questa necessità è necessario ridefinire il Modello del Linguaggio e rielaborare i documenti su cui si fa la ricerca.
- Come si era detto nel cap. 3, quando ci sia l'esigenza di risposte rapide e si voglia trattare documenti di grandi dimensioni (decine di migliaia di parole) e in grande quantità, l'utilizzo di un *ASR* completo è penalizzante per il carico computazionale (ma ormai non più così proibitivo con le macchine attuali).

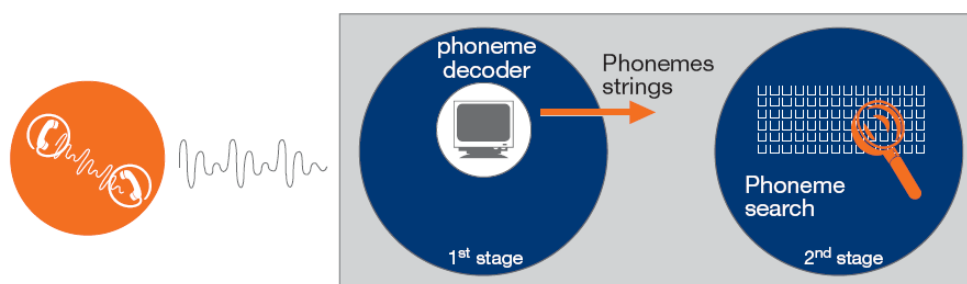


Fig. 4 – *Keyword Spotting* basato su riconoscimento fonetico [Alon, 2005]

#### 4.2 - *KWS con Riconoscimento Fonetico*

Anche in questo caso, si tratta di un sistema con due processi distinti.

Nel primo lavora un riconoscitore fonetico che fornisce in uscita il flusso dei fonemi riconosciuti, senza tentare l'individuazione delle parole (Fig. 4). L'*ASR* non ha quindi necessità di un Modello del Linguaggio. L'uscita dell'*ASR* è una stringa fonetica generalmente soggetta ad errori (con l'effetto di propagarsi nella successiva conversione a testo) (Fig. 4).

Il secondo passo tenta il recupero delle parole chiave la cui descrizione fonetica coincide in qualche punto con la sequenza ottenuta dall'*ASR* e prosegue con l'elaborazione di *Information Retrieval*.

Il vantaggio più rilevante di questa tipologia di *KWS* è che le parole chiave non devono essere predefinite: si può cercare un qualsiasi termine (nomi, parole straniere, ecc.) di cui si dia la trascrizione fonetica.

Gli svantaggi sono dovuti alla mancanza di un Modello del Linguaggio nella fase di decodifica che provveda ad eliminare in partenza molte delle possibili alternative errate della stringa fonetica.

Anche questo approccio, come quello basato su *LVCSR*, non è adatto nel caso si vogliano tempi di risposta rapidi e nel caso di grande quantità di documenti, poiché la ricerca sulla stringa fonetica è più complessa e pesante dal punto di vista computazionale che quella fatta su un testo di parole.

#### 4.3 - *KWS Basato su Grammatica GSF*

In questo caso l'*ASR* restringe la sua ricerca ad una lista limitata di parole (o una loro combinazione). Il Modello del Linguaggio vero e proprio è sostituito da un modello a uni-

grammi (le parole chiave hanno tutte la stessa probabilità) o ad una Grammatica a Stati Finiti che stabilisce le possibili relazioni fra le parole ammesse (Fig. 5).

Vantaggi:

- Le parole sono fissate a priori e senza restrizioni: si possono inserire parole non comuni, parole straniere, ecc. Non si pone il problema *OOV*. Si può lanciare una ricerca contemporaneamente su parole in lingue diverse.
- Non c'è alcuna dipendenza dal contesto come avviene in un *ASR* normale e non c'è necessità di modellazione del linguaggio.

Si ottengono *performance* migliori rispetto agli altri approcci. La velocità è più elevata di un normale *ASR* e dipende dal numero di parole cercate (può essere inferiore di molte volte al tempo reale anche su un normale *PC*).

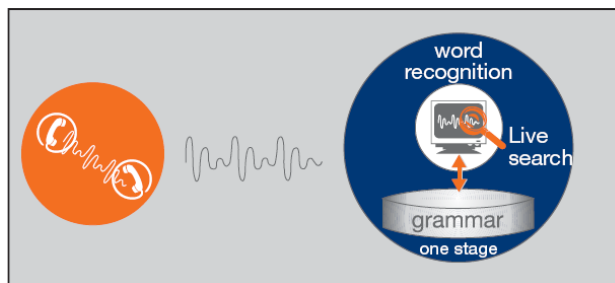


Fig. 5 – *Keyword Spotting* basato con una Grammatica a Stati Finiti [Alon, 2005]

Gli svantaggi dipendono anche in questo caso dall'assenza di *LM*, ma, come si è detto nel sommario, gli errori incidono in maniera poco rilevante sulle *performance* totali, essendo compensati spesso dalla ridondanza naturalmente presente nei documenti.

## 5. MISURE DI VALUTAZIONE DEL KWS

Questo è un aspetto critico per il *Keyword Spotting*, poiché le prestazioni di un sistema dipendono dalla scelta delle parole (ci sono parole più o meno facili da riconoscere), dal numero di parole chiave presenti e dalla tipologia dei campioni analizzati.

Come si è anticipato, il *NIST* ha compiuto sforzi notevoli per la formalizzazione di questo aspetto rilevante del *KWS*.

Le performance del *KWS* possono essere definite da varie misure:

**Precisione** o **Word Correct Rate (WCR)**: percentuale delle parole individuate correttamente rispetto a tutte quelle effettivamente rilevate, oppure in altre parole la *proporzione rilevante delle parole trovate* (Fig. 6).

**Word Error Rate (WER)**: percentuale delle parole errate (Sub+Del+Ins) rispetto a tutte quelle effettivamente rilevate, oppure in altre parole la *proporzione errata delle parole trovate* (Fig. 6).

Le parole errate comprendono:

Sub - % di parole sostituite con altre (*False alarm*)

Del - % di parole cancellate erroneamente (*False reject*)

Ins - % di parole inserite per errore (*False alarm*)



|                         | <i>Words</i> | <i>Keys</i> | <i>Dur(s)</i> | <i>Elap.(s)</i> | <i>RT Ratio</i> | <i>WCR</i> | <i>Sub</i> | <i>Del</i> | <i>Ins</i> | <i>WER</i> |
|-------------------------|--------------|-------------|---------------|-----------------|-----------------|------------|------------|------------|------------|------------|
| <i>Results with LM</i>  | 829          | 150         | 379           | 428.27          | 1.13            | 95.4       | 0.0        | 4.6        | 0.0        | 4.6        |
| <i>Results with GFS</i> | 829          | 150         | 379           | 152.83          | 0.40            | 60.9       | 14.9       | 24.1       | 8.0        | 47.1       |

Fig. 6 – *WCR* e *WER* per un test di *Keyword Spotting* con Modello del Linguaggio rispetto ad una Grammatica a Stati Finiti.

Fra le altre misure adottate nel *KWS*:

**Accuratezza:**  $(\text{Corr} - \text{Ins}) / (\text{Corr} + \text{Sub} + \text{Del})$

e cioè la percentuale fra la differenza parole\_corrette e parole\_inserite rispetto a tutte le parole chiave effettivamente presenti nel documento.

**Figura di Merito:** Un differente approccio per la misura della performance di un *KWS* è basato sulla Figura di Merito (FOM), e cioè la media di parole correttamente rilevate per un certo numero di falsi allarmi per ora.

**Recall:**  $\text{Corr} / (\text{Corr} + \text{Sub} + \text{Del})$

e cioè la percentuale fra parole corrette e tutte le parole chiave effettivamente presenti nel documento. In genere si usa plottare i valori della Precisione rispetto al *Recall* (Fig. 7).

**Mean Average Precision (mAP)** si calcola facendo la media aritmetica dei valori di Precisione Media rispetto ai punti individuati dal *recall* per tutti i test compiuti sui documenti (conversazioni, ecc.) (Fig. 8).

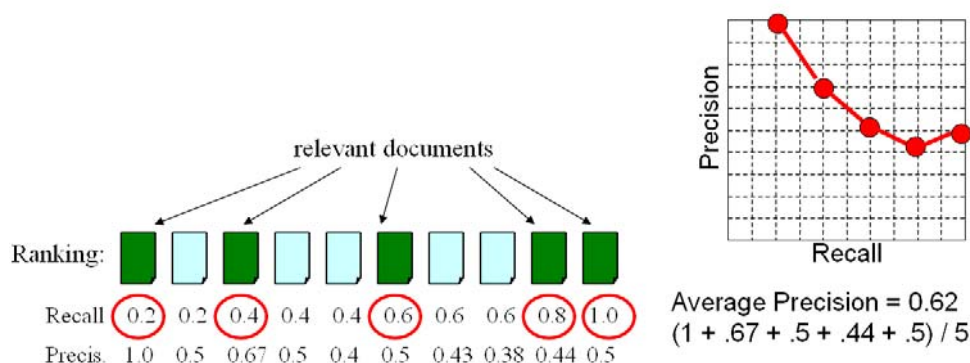


Fig. 7 – Grafico della Precisione rispetto al *Recall* e calcolo della Precisione Media [Lavrenko, 2005].

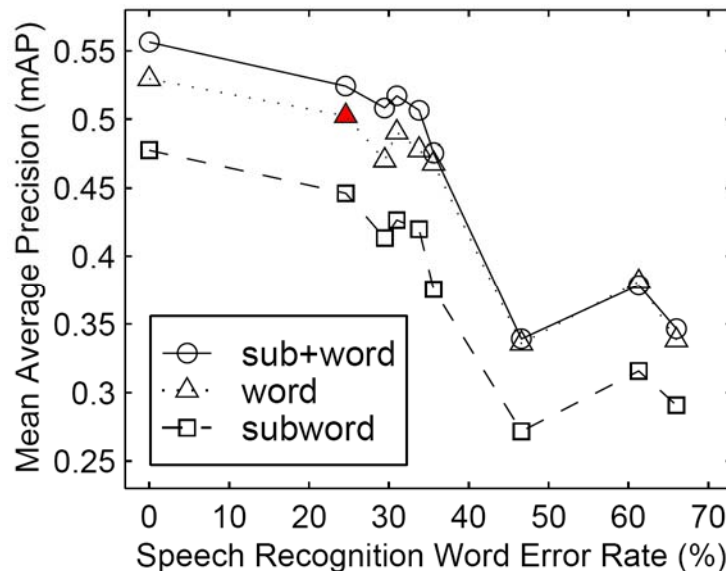


Fig. 8 – *Mean Average Precision (mAP)* rispetto a *Word Error Rate (WER)*: si vede che anche con indici di errore *WER* abbastanza alti la *performance* media di precisione cala in maniera minima. Ad es., si passa da  $mAP = 0.5295$  per le trascrizioni di riferimento (per definizione corrette cioè  $WER = 0\%$ ) ad un  $mAP = 0.5025$  per un  $WER = 24.6\%$  nel punto contrassegnato con un triangolo rosso [Ng, 2000].

Questo parametro è importante nella valutazione di vari sistemi di *KWS*: in effetti, se si plottano i dati di *mAP* rispetto al *Word Error Rate (WER)*, si scopre in che modo varia la efficienza del sistema nell'individuare documenti, conversazioni, ecc. contenenti le parole desiderate. Nella Fig. 8 si può notare come si passa da  $mAP = 0.5295$  per le trascrizioni di riferimento (per definizione corrette  $WER = 0\%$ ) ad una *Mean Average Precision* = 0.5025 per un  $WER = 24.6\%$  nel caso di test su documenti reali (punto contrassegnato con un triangolo rosso), con un peggioramento di solo il 5% circa. Questo sta ad indicare come malgrado l'aumento notevole del *WER*, i documenti siano ancora individuati con successo [Ng, 2000].

La spiegazione della scarsa sensibilità del *KWS* agli errori del motore di riconoscimento va cercata, come si era visto nel sommario, nella naturale ridondanza di informazione presente in un documento [Allan, 2002].

## 6. CARATTERISTICHE DEL SISTEMA REALIZZATO

L'interfaccia grafica, realizzata per questo lavoro, *Keyword Spotting Interface*, è stata progettata con lo scopo di facilitare le operazioni di impostazione dei parametri e la loro variazione, e soprattutto di consentire una rapida e efficace verifica dei risultati ottenuti (Fig. 9).

Il programma è scritto in **TCL/Tk** (<http://www.activestate.com/Products/ActiveTcl>) per compatibilità con vari sistemi operativi ed anche per la rapidità di sviluppo consentita.

Le caratteristiche principali del sistema sono:

- *Speaker Independent*: il *training* del Modello Acustico è stato fatto sul corpus APASCI elaborato dall'ITC-IRST e distribuito da Elda [www.elda.org/catalogue/en/speech/S0039.html](http://www.elda.org/catalogue/en/speech/S0039.html)
- *Keyword Spotting su parlato continuo*: non c'è alcuna condizione restrittiva nell'estrazione delle parole (ad esempio che debbano essere isolate o che debbano essere limitate ad un certo numero) e sull'introduzione di nuove parole (*out of vocabulary words* - *OOV*).
- Trascrizione fonetica automatica delle parole da ricercare (da un database oppure con una conversione *letter-to-sound* per gli *OOV*).
- Il motore di riconoscimento utilizzato è Sonic [Pellom, 2001], [Pellom et al., 2003] [cslr.colorado.edu/beginweb/speech\\_recognition/sonic.html](http://cslr.colorado.edu/beginweb/speech_recognition/sonic.html)
- *Compatibilità con Sonic*: Si è mantenuta la completa compatibilità con Sonic, per cui il funzionamento è garantito anche in assenza di interfaccia.
- *Compatibilità con vari Sistemi Operativi*: La nuova interfaccia può girare su tutti i sistemi operativi su cui già funzionava Sonic (in realtà finora è stato testato solo su Linux e Microsoft Windows XP).
- *Indipendenza da Sonic*: L'interfaccia grafica non dipende dal motore di riconoscimento scelto in questo caso (e cioè Sonic), ma può funzionare in modo trasparente con un altro motore *ASR* (ad es. Sphinx).
- *Interfaccia user-friendly*: L'interfaccia grafica realizzata facilita notevolmente il lavoro di test e modifica dei parametri dell'*ASR* (Fig. 10), e in modo particolare la visualizzazione e la verifica dei risultati con *ScLite* (Fig. 11). È anche possibile una verifica manuale, in assenza di trascrizione, dal momento che si può ascoltare sequenzialmente, o isolatamente, nel file sonoro i punti corrispondenti alle parole individuate, e constatarne la correttezza (Fig. 9). È previsto in futuro anche la contemporanea visualizzazione dei *frame* video relativi ad un eventuale filmato su cui si faccia il *Keyword Spotting*.

Il sistema si compone di due parti:

- *Sonic\_batch*: Il riconoscitore vero e proprio, che sfrutta una versione opportunamente modificata del programma scritto in **c** da B. Pellom per poter estrarre le parole chiave volute secondo determinati criteri [Pellom, 2001].
- *Keyword Spotting Interface*: L'interfaccia di cui si è appena parlato (Fig. 9-11).

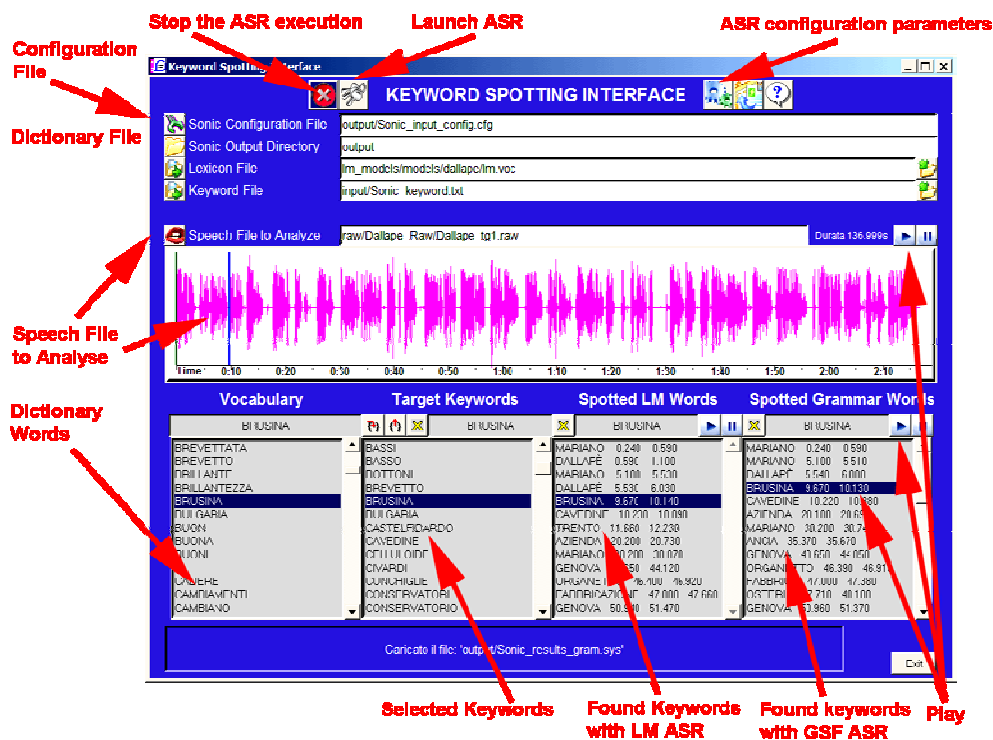


Fig. 9 – Keyword Spotting Interface: Finestra Principale

### Parametri di Configurazione

I parametri di configurazione dell'ASR sono assegnati in fase di inizializzazione di Sonic, ma possono essere modificati dinamicamente nella finestra visibile in Fig. 10.

Il loro aggiustamento è molto utile per ottenere il miglior compromesso possibile fra risultati e tempo di esecuzione.

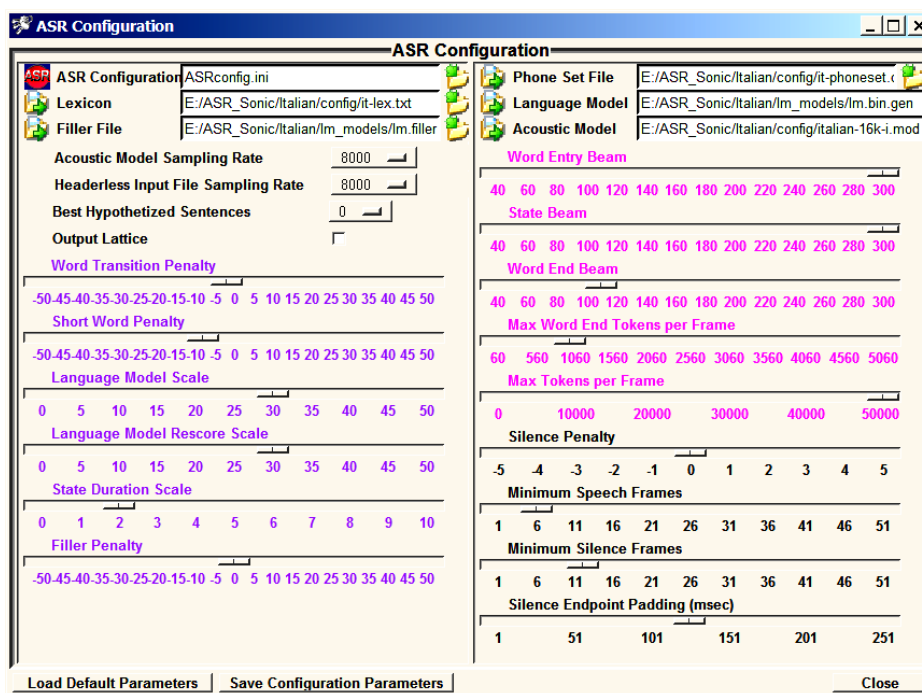


Fig. 10 – Keyword Spotting Interface: Configurazione dei parametri di riconoscimento.

I parametri più rilevanti rispetto alle prestazioni dell'ASR sono i seguenti:

**Word Transition Penalty:** Penalizzazione nella transizione da una parola ad un'altra. Con valori decrescenti si aumentano gli errori di cancellazione di parole, mentre si diminuiscono le parole inserite per errore.

**Short Word Penalty:** Fattore di penalizzazione applicato alle parole corte (tre fonemi o meno). Con valori che diminuiscono, si privilegiano l'individuazione di parole corte.

**Language Model Scale:** Peso assegnato al Modello del Linguaggio nella valutazione probabilistica delle parole. Diminuendo i valori di questo parametro, aumentano in modo rilevante i tempi di esecuzione e gli errori di rilevamento delle parole chiave.

**Max Active States:** Numero massimo di stati contemporaneamente attivi in ogni finestra di analisi. Aumentando il suo valore migliorano le prestazioni del riconoscitore, ma cresce (di molto) il tempo di esecuzione e l'occupazione di memoria.

**Word Entry Beam:** Soglia di potatura dei rami ammessi nella ricerca al primo stato di ogni parola. Valori bassi migliorano la velocità di esecuzione, ma peggiorano il grado di accuratezza del rilevamento delle parole.

**State Beam:** Soglia di potatura dei rami attivi nella ricerca per gli stati successivi al primo (ed escluso l'ultimo). Valori bassi migliorano

la velocità di esecuzione e deprimono l'accuratezza dei risultati ottenuti.

**Word End Beam:** Soglia di potatura dei rami attivi nella ricerca per l'ultimo stato di una parola. Valori bassi migliorano la velocità di esecuzione a scapito dei risultati ottenuti.

**Max Word End Tokens Per Frame:** Massimo numero di parole considerate nel passaggio da un *frame* all'altro. Valori limitati restringono la ricerca e migliorano i tempi di esecuzione, ma peggiorano i risultati.

**Keyword Threshold:** Soglia di confidenza che controlla l'accettazione delle parole chiave nel *Keyword Spotting*.

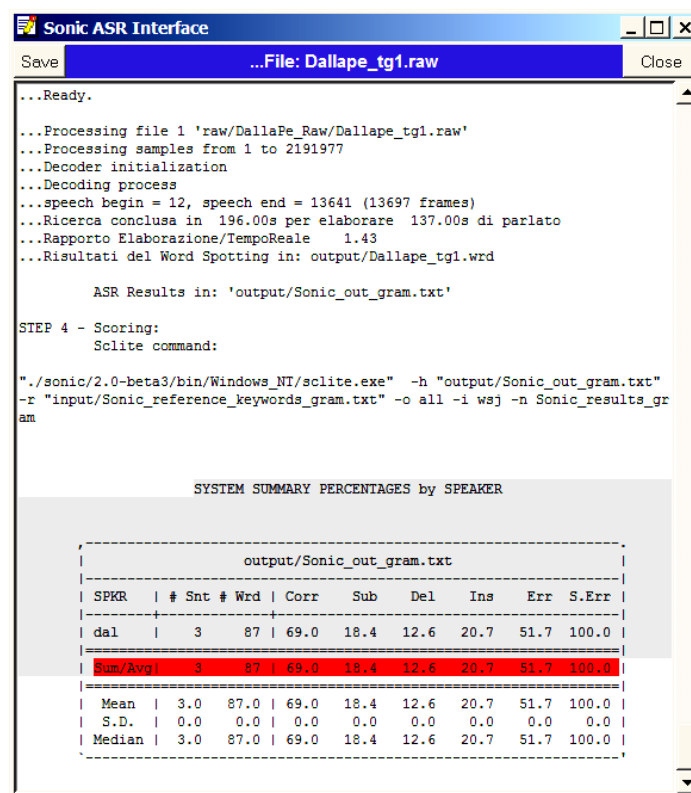


Fig. 11 – Keyword Spotting Interface: Risultati del decodificatore con Scilite.

## 7. TEST GSF SU CAMPIONI CLIPS TV E RADIO

Il sistema descritto in precedenza è stato testato su 10 brani estratti (a caso) dal database CLIPS (*Corpora e Lessici di Italiano Parlato e Scritto* [www.clips.unina.it](http://www.clips.unina.it)) e relativi a parlato televisivo e radiofonico (divulgazione, cultura, intrattenimento, ecc.) per complessivi 35 m. di parlato.

Il Modello Acustico utilizzato proviene invece dal corpus *APASCI* che è parlato microfonico su frasi lette, tipologicamente abbastanza diverso dal CLIPS.

Il criterio per la scelta delle parole chiave da cercare è stato quello di selezionare tutti i sostantivi presenti nei vari brani, che non fossero verbi, aggettivi, articoli, ecc. per un totale di 856 keywords, e cioè il 14% del totale di 5936 parole.

Il test è stato fatto solo con una Grammatica a Stati Finiti comprendente le parole selezionate per ogni brano. Non si è potuto ricavare un Modello di Linguaggio adeguato per la disparità e la limitata disponibilità di materiale trascritto.

I risultati sono i seguenti (Fig. 12):

- *Word Correct Rate (WCR)* = 60 %
- *Word Error Rate (WER)* = 47.3 %
- *Recall* = 74.2 %
- *Real-Time Ratio* = 0.3 %

| <i>Speaker</i>       | <i>Words</i> | <i>Keys</i> | <i>Dur(s)</i> | <i>Elap.(s)</i> | <i>RT Ratio</i> | <i>WCR</i> | <i>Sub</i> | <i>Del</i> | <i>Ins</i> | <i>WER</i> |
|----------------------|--------------|-------------|---------------|-----------------|-----------------|------------|------------|------------|------------|------------|
| LeccecorpusRDit14L   | 294          | 40          | 144.81        | 37.13           | 0.26            | 52.50      | 10.00      | 37.50      | 0.00       | 47.50      |
| MilanocorpusTVit02M  | 453          | 43          | 145.49        | 38.70           | 0.27            | 72.10      | 14.00      | 14.00      | 7.00       | 34.90      |
| NazcorpusRDde02Z     | 1091         | 102         | 317.00        | 101.70          | 0.32            | 64.70      | 10.80      | 24.50      | 10.80      | 46.10      |
| NazcorpusRDis01Z     | 717          | 33          | 306.27        | 72.05           | 0.24            | 51.50      | 15.20      | 33.30      | 6.10       | 54.50      |
| ParmacorpusTVit02E   | 616          | 79          | 221.36        | 61.36           | 0.28            | 60.80      | 24.10      | 15.20      | 11.40      | 50.60      |
| PerugiacorpusRDis020 | 388          | 108         | 135.27        | 45.84           | 0.34            | 87.00      | 2.80       | 10.20      | 8.30       | 21.30      |
| PerugiacorpusTVis040 | 446          | 125         | 158.12        | 60.36           | 0.38            | 60.80      | 13.60      | 25.60      | 10.40      | 49.60      |
| PerugiacorpusTVit040 | 454          | 76          | 149.78        | 43.75           | 0.29            | 36.80      | 11.80      | 51.30      | 3.90       | 67.10      |
| TorinocorpusTVit02T  | 452          | 59          | 191.31        | 53.02           | 0.28            | 52.50      | 23.70      | 23.70      | 6.80       | 54.20      |
| VeneziacorpusRDit01V | 1025         | 191         | 329.18        | 123.08          | 0.37            | 53.40      | 17.30      | 29.30      | 4.70       | 51.30      |
|                      | <i>Words</i> | <i>Keys</i> | <i>Dur(s)</i> | <i>Elap.(s)</i> | <i>RT Ratio</i> | <i>WCR</i> | <i>Sub</i> | <i>Del</i> | <i>Ins</i> | <i>WER</i> |
| <b>Total</b>         | 5936         | 856         | 2098.5        | 636.99          | 0.30            | 60.00      | 14.10      | 25.80      | 7.40       | 47.30      |

Fig. 12 – Risultati del sistema di KWS con Sclite su 10 brani CLIPS.

Come si vede, i valori di *WCR* e *WER*, pur non entusiasmanti, sono molto promettenti nella prospettiva di poter modellare opportunamente il linguaggio. L'indice di *Recall*, invece, e cioè la percentuale fra parole corrette e tutte le parole chiave effettivamente presenti nel documento è abbastanza elevato (74.2 %).

## 8. TEST SU BRANO CONVERSAZIONALE CON LM + GSF

Il secondo esperimento è stato condotto su un parlato conversazionale abbastanza rumoroso (intervista a Amleto Dallapè).

Il Modello Acustico, come nel caso precedente, è stato allenato su *APASCI* (quindi non adatto a questo brano).

Il Modello del Linguaggio è stato addestrato su circa 50 m del brano disponibile, lasciando al test del riconoscimento 7 m circa.

Come parole chiave da ricercare sono state scelte 150 sostantivi presenti nel brano (trascurando verbi, aggettivi, articoli, ecc.) e ritenuti più rilevanti per il contenuto dell'intervista. Queste parole corrispondono al 12.8% circa del totale di 1176 parole.

I risultati con il Modello del Linguaggio sono stati i seguenti (Fig. 13):

- *Word Correct Rate (WCR)* = 32.0 %
- *Word Error Rate (WER)* = 68.1 %
- *Recall* = 32.0 %
- *Real-Time Ratio* = 0.95 %

I risultati con la Grammatica a Stati Finiti sono stati i seguenti (Fig. 13):

- *Word Correct Rate (WCR)* = 24.0 %
- *Word Error Rate (WER)* = 76.0 %
- *Recall* = 31.6 %
- *Real-Time Ratio* = 0.45 %

|                         | <i>Words</i> | <i>Keys</i> | <i>Dur(s)</i> | <i>Elap.(s)</i> | <i>RT Ratio</i> | <i>WCR</i> | <i>Sub</i> | <i>Del</i> | <i>Ins</i> | <i>WER</i> |
|-------------------------|--------------|-------------|---------------|-----------------|-----------------|------------|------------|------------|------------|------------|
| <i>Results with LM</i>  | 1176         | 150         | 437           | 415.00          | 0.95            | 32.0       | 16.0       | 52.0       | 0.0        | 68.1       |
| <i>Results with GFS</i> | 1176         | 150         | 437           | 197.35          | 0.45            | 24.0       | 60.0       | 16.0       | 0.0        | 76.0       |

Fig. 13 – Risultati del sistema di *KWS* con *Scilite* su una intervista di A. Dallapè.

Inutile dire che i risultati così modesti sono dovuti ad un Modello Acustico inadeguato alla tipologia del brano e che il test va ripetuto non appena addestrato l'AM su un corpus conversazionale.

Se questo esperimento è ripetuto con una registrazione microfonica, adatta all'AM allenato su *APASCI*, in cui si ripete più o meno con le stesse parole una parte (6 m. circa) dell'intervista, allora i risultati migliorano in maniera più soddisfacente (Fig. 6):

I risultati con il Modello del Linguaggio sono stati i seguenti (Fig. 6):

- *Word Correct Rate (WCR)* = 95.4 %
- *Word Error Rate (WER)* = 4.6 %
- *Recall* = 95.4 %
- *Real-Time Ratio* = 1.13 %

I risultati con la Grammatica a Stati Finiti sono stati i seguenti (Fig. 6):

- *Word Correct Rate (WCR)* = 60.9 %
- *Word Error Rate (WER)* = 47.1 %



- *Recall* = 72.6 %
- *Real-Time Ratio* = 0.4 %

## 9. CONCLUSIONI E SVILUPPI FUTURI

È stato implementato un sistema di *KWS* che sfrutta un doppio canale di riconoscimento:

- Il primo è un *Large Vocabulary Continuous Speech Recognition* basato su un *AM* e su un *LM*.
- Il secondo implementa una *GSF*, che non necessita della modellazione di un *LM* e permette la ricerca di una parola qualsiasi.

È stata creata una interfaccia che facilita l'impostazione dei parametri e la visualizzazione e la verifica dei risultati dei due canali del *KWS*.

I primi risultati ottenuti sono incoraggianti e in prospettiva molto promettenti con l'*AM* adeguato. I miglioramenti che si pensa di apportare sono i seguenti:

- Addestramento con modelli acustici conversazionali.
- Integrazione nel *KWS* di parser semantici del tipo di Phoenix (per i quali Sonic è già predisposto e che dovrebbe migliorare notevolmente le *performance*).
- Convergenza automatica (fatta ora fuori linea) dei parametri alla prestazione ottimale (ad es. per ridurre i falsi allarmi o aumentare la correttezza delle parole).
- Visualizzazione sincrona dei frame video relativi in un eventuale filmato su cui si faccia il *Keyword Spotting*.

## RIFERIMENTI

Allan J. (2002), "Perspectives on Information Retrieval and Speech", *Information Retrieval Techniques for Speech Applications*, Coden, Brown and Srinivasan, editors. Springer-Verlag Lecture Notes in Computer Science, Vol. 2273, pp. 1-10.

Alon G. (2005), *Key-Word Spotting-The Base Technology for Speech Analytics*, White Paper, NSC - Natural Speech Communication Ltd.

Buckland, M. (1992), "Emanuel Goldberg, Electronic Document Retrieval, And Vannevar Bush's Memex". *Journal of the American Society for Information Science* 43, n. 4, pp. 284-294.

Cernocky J. et al. (2007), "Search in speech for public security and defense", *Proc. IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, SAFE '07, pp. 1-7

Foote J., Jones G., Jones K., Young S. (1995), "Talker- Independent Keyword Spotting for Information Retrieval", *Proc. Eurospeech* 1995, Vol. 3, pp. 2145-2149.

Garofolo J., Voorhees E., Stanford V., Sparck Jones K. (1998), "TREC-6 1997 spoken document retrieval track overview and results". *Proc. of TREC-6* (1997), pp. 83-92, NIST special publication 500-240.

Garofolo J., Voorhees E., Auzanne C., Stanford V., Lund B. (1999), "1998 TREC-7 spoken document retrieval track overview and results". *Proc. TREC-7* (1998), pp. 79-89, NIST special publication 500-242.

- Garofolo J., Auzanne C., Voorhees E. (2000), "The TREC spoken document retrieval track: A success story". *Proc. of TREC-8* (1999). NIST special publication 500-246.
- Garofolo J., Lard J., Voorhees E. (2001), *2000 TREC-9 spoken document retrieval track*, Powerpoint presentation [rec.nist.gov](http://rec.nist.gov)
- Gelin P. (1997), Détection de mots clés dans un flux de parole : application à l'indexation de documents multimédia, Thèse EPFL, n. 1658.
- Higgins A., Wohlford R. (1985), "Keyword Recognition Using Template Concatenation", *Proc. ICASSP 1985*, pp. 1233- 1236.
- James D. (1995), The application of classical information retrieval techniques to spoken documents, Ph.D. Thesis, University of Cambridge.
- Junkawitsch J., Neubauer L., Hoge H., Ruske G. (1996), "A new keyword spotting algorithm with pre-calculated optimal thresholds", *Proc. ICSLP* Vol. 4, pp. 2067-2070.
- Lavrenko V., (2005), "Information Retrieval", Powerpoint presentation [www.clsp.jhu.edu/ws2005/calendar/documents/LavrenkoJuly5.PPT](http://www.clsp.jhu.edu/ws2005/calendar/documents/LavrenkoJuly5.PPT)
- Manos A. S. (1996), A Study on Out-of-Vocabulary Word Modeling for a Segment-Based Keyword Spotting System, M.S. Thesis, Massachusetts Institute of Technology.
- Myers C.S., Rabiner L.R., Rosenberg A. (1980)., "An Investigation of the Use of Dynamic Time Warping for Word Spotting and Connected Word Recognition", *Proc. ICASSP 1980*, pp. 173-177.
- Ng K. (2000), "Information Fusion For Spoken Document Retrieval", *Proc. International Conference on Acoustics, Speech, and Signal Processing*, pp. 2405–2408
- Pellom B. (2001), *SONIC: The University of Colorado Continuous Speech Recognizer*, University of Colorado, *Tech Report #TR-CSLR-2001-01*
- Pellom B., Hacıoglu K. (2003), "Recent Improvements in the CU SONIC ASR System for Noisy Speech: The SPINE Task", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2003*, Vol. 1, pp. I 4-7.
- Rohlicek J., Russel W., Roukos S., Gish H. (1989), "Continuous Hidden Markov Modeling for Speaker-Independent Word-Spotting", in *Proc. ICASSP 1989*, pp. 627-630.
- Rose R., Paul D. (1990), "A Hidden Markov Model Based Keyword Recognition System", in *Proc. ICASSP 1990*, pp. 129-132.
- Silaghi M., Vargiya R. (2005): "A new evaluation criteria for keyword spotting techniques and a new algorithm", in *Proc. Interspeech 2005*, pp. 1593-1596.
- Szöke I., Schwarz P., Matejka P., Burget L., Fapso M., Karafiát M., Cernocký J. (2005), "Comparison of Keyword Spotting Approaches for Informal Continuous Speech", in *Proc. Eurospeech 05*.
- Wilpon J., Rabiner L., Lee C., Goldman E, (1990), "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models", *IEEE Trans. ASSP*, Vol138. No. 11, pp. 1870-1878.

# SPERIMENTAZIONE DI SISTEMI SENZA ADDESTRAMENTO PER L'ESTRAZIONE DI FONEMI VOCALICI

Francesco Borchì, Monica Carfagni, Matteo Nunziati  
Università degli Studi di Firenze, Dipartimento di Meccanica e Tecnologie Industriali (DMTI)  
{francesco.borchì,monica.carfagni,matteo.nunziati}@unifi.it

## 1. SOMMARIO

L'analisi di corpora vocali è una delle attività più importanti sia per il campo della ricerca che per quello delle attività commerciali legate all'analisi della voce. In entrambi i casi, la possibilità di analizzare ed impiegare database sufficientemente ampi (tali da essere considerate rappresentative dei fenomeni di interesse) è un tassello fondamentale per la correttezza dei risultati prodotti. Tuttavia, il trattamento dei dati contenuti nei corpora vocali, compiuto al fine di estrarre tutte e sole le informazioni di interesse per una data attività, richiede molte ore di lavoro.

Nell'arco degli anni sono stati sviluppati sistemi automatici di segmentazione, che fanno uso di modelli matematici complessi (vedi i sistemi HMM in Huang *et alii* (2001)): tali modelli richiedono, a loro volta, una mole considerevole di dati per la loro messa a punto ed addestramento e si presenta il problema di reperire strumenti intermedi, che aiutino nello sviluppo.

Nell'ambito di un progetto di ricerca relativo al riconoscimento automatico del parlante, il Dipartimento di Meccanica e Tecnologie Industriali dell'Università di Firenze (di seguito DMTI) ha iniziato lo sviluppo e la sperimentazione di un sistema specializzato per la catalogazione di fonemi vocalici accentati. Uno degli obiettivi per lo sviluppo di tale sistema è stato quello di realizzare uno strumento totalmente automatico, che richiedesse costi di taratura minimi, possibilmente nulli. Un altro obiettivo dello studio è da ricercarsi nella possibilità di trasferimento tecnologico dello strumento, che dovrà risultare impiegabile da chiunque anche a scopo commerciale, senza la necessità di adottare un sistema di segmentazione dai costi elevati. A tal fine, il lavoro si è orientato sullo sviluppo di due strumenti diversi: il primo, derivante dalla generalizzazione del metodo proposto da Così *et alii* (1995), impiegato per definire la posizione delle vocali all'interno dell'enunciato; il secondo, caratterizzato dall'utilizzo del piano formantico F1-F2 (in Hz), impiegato per definire il triangolo vocalico di un parlante generico e per classificare le diverse vocali accentate individuabili nel parlato.

Allo stato attuale, un prototipo funzionante di tale sistema è sotto analisi presso il dipartimento. Il prototipo, nella configurazione attuale, è stato testato solo in condizioni testo-dipendenti su parlato letto in laboratorio.

## 2. IL METODO IBT-DESF

L'idea di fondo del sistema, che verrà descritto di seguito, è scaturita dalla lettura del lavoro condotto da Così *et alii* (1995), in cui è stato sviluppato un sistema automatico di individuazione di fonemi vocalici. Le vocali oggetto di studio erano quelle contenute nel corpus AIDA. Gli enunciati presi in considerazione in tale lavoro erano costituiti da bisillabi di tipo CV'CV, per tale motivo erano note agli autori sia le posizioni delle vocali accentate all'interno delle singole sequenze di dati, sia la loro classificazione. Il metodo presentato al convegno AIA del 1995 era basato sull'analisi dell'andamento dell'energia a

breve termine del segnale. L'idea di fondo era la seguente: dal momento che le vocali sono realizzate ostacolando in maniera minima il flusso d'aria, mentre caratteristica delle consonanti è la forte occlusione dell'apparato fonatorio, il livello di energia associato al segnale deve subire un incremento in prossimità delle suddette vocali.

A partire da questa considerazione, di volta in volta, veniva definita una soglia decrescente, il cui scopo era quello di isolare le due aree di massima energia associate alle vocali dei bisillabi. Individuate tali aree, era immediata la loro classificazione: sia l'accento sia il tipo di vocale erano deducibili dalla trascrizione associata al file audio. Una volta raccolti i dati sui fonemi, veniva valutata la capacità di discriminazione fornita da diverse rappresentazioni su piani formantici. A partire dalle intuizioni di tale lavoro, nel nostro studio abbiamo tentato di estendere il campo di impiego della metodologia al fine di individuare le aree di massima intensità associate alle vocali presenti in un testo generico.

Il metodo indicato da Cosi *et alii* (1995) si basava su un'ipotesi ben precisa: ogni bisillabo doveva contenere due picchi di energia, ognuno dei quali associato ad una vocale. Tale ipotesi decade nel contesto della presente ricerca, dal momento che il parlato non viene considerato noto a priori e dal momento che eventuali picchi di energia possono essere dovuti anche ad eventi eccezionali. Tale incremento rimane comunque una condizione necessaria alla presenza di fonemi vocalici: a partire da questa considerazione è stato sviluppato un selettore qui definito IBT-DESF (Intensità a Breve Termine – Durata, Esistenza e Stabilità delle Frequenze caratteristiche). Nei paragrafi che seguono verranno descritte le caratteristiche di tale soluzione.

### 2.1 Il selettore IBT

Lo scopo del selettore IBT è quello di individuare il punti di massimo dell'intensità a breve termine (IBT) ed isolarle dal contesto. In questo caso, è stata definita una soglia  $I_{min}$ , tramite la quale discriminare ciò che è considerabile come un fonema vocalico accentato e ciò che non lo è (vedi figura 1). Ovviamente in casi estremi come in chiusura di frase, ove spesso l'intensità ha un decremento generalizzato, le vocali sono inesorabilmente scartate, in quanto non riescono a superare la soglia minima di intensità imposta dal sistema. Tuttavia, questo comportamento non è stato considerato scorretto: fenomeni in chiusura di frase, associati a cali diffusi di intensità, sono spesso associati anche ad una fonazione non standard da parte del parlatore. Dal momento che lo scopo del metodo è quello di raccogliere informazioni sul contenuto formantico dei fonemi vocalici, si è ritenuto corretto escludere dalle statistiche i casi estremi.

Risulta evidente che la risposta del metodo IBT dipende dalla soglia di intensità scelta dall'operatore. A sua volta tale valore deve essere associato al livello del segnale acquisito. Al fine di verificare il sistema nel suo complesso, si è preferito, in questa fase, limitare le variabili in gioco. Per tale motivo, si sono realizzate delle registrazioni pilota di livello pressoché omogeneo (descritte nel dettaglio in seguito) e si è imposta una soglia fissa di 60 dB, che, in questa sede, non entra nel problema come variabile ma come costante nota. In termini generali si può comunque dire che: dato l'insieme  $M$  dei possibili massimi locali di intensità, il metodo IBT è in grado di isolare tra questi un sottoinsieme  $N$  di picchi aventi intensità maggiore di quella imposta dall'operatore.

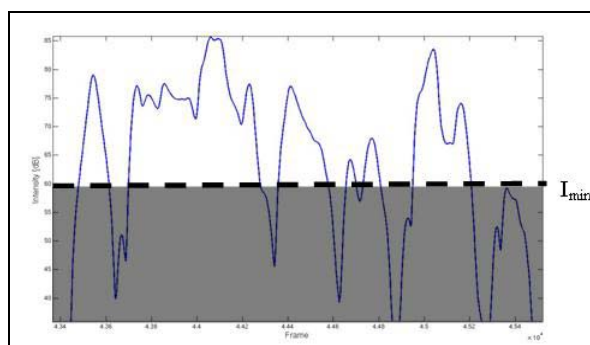


Figura 1: schema esemplificativo del comportamento selettivo della soglia  $I_{min}$ . I massimi locali giacenti nella zona annerita sono scartati dal selettore.

## 2.2 Il selettore DESF

Dal momento che il valore di IBT non è da solo sufficiente a rivelare la presenza di una vocale, il metodo è stato esteso includendo altri parametri di selezione. Prima fra tutti la durata ( $D$ ) del picco di intensità. Dal momento che possibili eventi eccezionali possono inserirsi nell'insieme  $N$ , la valutazione della durata del picco permette una prima discriminazione. Anche in questo caso si ricorre a condizioni necessarie ma non sufficienti: di fatto, viene imposta una durata minima  $D_{min}$ , che, se non rispettata, porta allo scarto dei massimi locali. Risulta evidente come una soglia di questo tipo sia necessariamente legata al ritmo di eloquio di ogni soggetto. Tuttavia la sperimentazione ha dimostrato che esiste una fascia di pochi millisecondi (30 ms.), indipendente dal parlante e sufficientemente ampia da garantire un'eliminazione dei fenomeni spuri. Le prove effettuate hanno dimostrato che al di sotto di tale soglia si presentano fenomeni di massimo locale relativi ad eventi spuri o, al più, massimi locali relativi ad eventi fonatori non sufficientemente intensi da poter essere associati a vocali accentate. In questo caso, infatti, la selezione IBT porta a conservare solo una parte ridotta (quella più ricca di energia) della fase di tenuta di possibili vocali, riducendo di molto la durata del picco in esame (al fine di comprendere meglio il comportamento combinato delle due soglie si faccia riferimento allo schema esemplificativo di figura 2). In questa fase viene tra l'altro scartata una parte rilevante dei fonemi nasali rimanenti dopo la prima selezione. Data la fase di sviluppo del sistema, non è possibile al momento quantificare con una percentuale precisa questo fenomeno, che risulta tuttavia evidente in tutte le tracce sin'ora analizzate.

Individuato così un nuovo sottoinsieme  $V$ , incluso in  $N$ , sono state introdotte altre grandezze di controllo, relative alle caratteristiche fonetiche dei picchi. Ogni vocale è classificabile come un fonema sonoro, perciò uno degli elementi di rilievo per la discriminazione è senza dubbio la presenza nel segnale della frequenza fondamentale (di seguito  $f_0$ ). Le vocali accentate sono inoltre caratterizzate da una durata tendenzialmente maggiore rispetto agli altri fonemi vocalici, come confermato da Alfano (2005): data la maggiore durata del fonema, si è ritenuto che la fase di tenuta consentisse una maggiore stabilizzazione dei valori in frequenza. Questo ha condotto ad ipotizzare che fosse possibile discriminare le vocali tramite vincoli di stabilità imposti sia alla  $f_0$  sia alle frequenze formanti<sup>1</sup> (d'ora in avanti  $f_1$  ed  $f_2$ ). In sintesi, si è tentato di discriminare i fonemi vocalici

<sup>1</sup> Si è limitato lo studio alle prime due formanti.

accentati tramite l'analisi dell'esistenza e della stabilità delle frequenze caratteristiche (ESF).

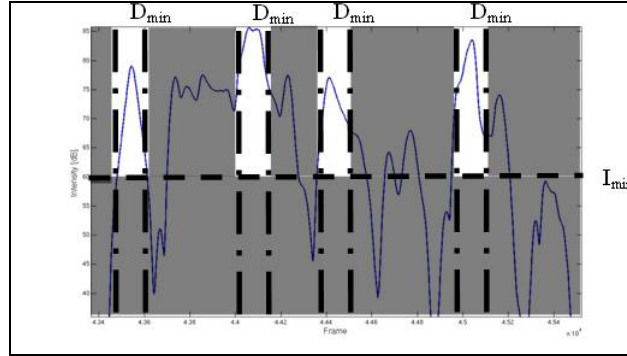


Figura 2: schema esemplificativo del comportamento selettivo delle soglie  $I_{min}$  e  $D_{min}$ . La zona annerita della figura rappresenta i massimi locali di intensità scartati dai selettori.

Il limite di questo approccio consiste nella necessità di accumulare un set  $Z$  di dati, sufficiente a definire in termini numerici il concetto di stabilità. In questa sede, al fine di costituire tale set, è stata presa in considerazione una piccola quantità di vocali tratta da dati segmentati manualmente. Tale quantità è associata ad una porzione dei fonemi, presenti in quattro tracce appartenenti a parlanti diversi e ricavate dal piccolo corpus realizzato per il progetto pilota qui descritto (il corpus sarà descritto dettagliatamente più avanti).

Per il selettore non è rilevante conoscere il valore esatto delle frequenze caratteristiche, quanto la loro variazione del tempo. Per definire il livello di stabilità delle frequenze viene considerato un intervallo  $\Delta T$  di 30 ms., centrato rispetto al picco di intensità individuato dall'IBT. Per ogni singolo fonema, all'interno dell'intervallo  $\Delta T$ , vengono estratte  $f_0$ ,  $f_1$  ed  $f_2$ .

Ottenuti i dati grezzi, per ogni campione  $z_h$  appartenente a  $Z$ , è stata calcolata la media  $\mu_{ih}$ , dove  $i=0,1,2$  indica una delle frequenze caratteristiche. È stata inoltre calcolata la media sull'intero set  $Z$ , indicata con  $\mu_i$ , e la deviazione standard, indicata con  $\sigma_i$ . In altre parole, i dati in frequenza  $f_{ihj}$ , dove  $j=1, \dots, 30$  indica l'istante di calcolo, sono stati considerati come possibili valori di una stessa variabile aleatoria  $F_i$  con distribuzione normale  $N_i=N(\mu_i, \sigma_i)$ . In questa sede sono stati trascurati per semplicità i possibili fenomeni di covarianza.

Successivamente, sono state valutate le differenze istantanee tra i tracciati  $f_{ihj}$  e le medie  $\mu_{ih}$  di ogni singolo fonema  $z_h$ . Ognuna delle vocali è associata quindi a 90 valori istantanei di variazione (uno per ogni millisecondo e per ogni frequenza). Tali valori saranno di seguito indicati con  $\Delta f_{ihj}$ , mentre con  $|\Delta f_{ihj}|_{max}$  si farà riferimento al massimo del loro valore assoluto. Infine, è stata definita una soglia di stabilità convenzionale.

La soglia è valutata a partire da un coefficiente reale positivo  $\gamma$ . Tale coefficiente rappresenta il livello di stabilità imposto dall'operatore al fine di poter includere un fonema tra i possibili candidati delle vocali accentate. Per la stima di  $\gamma$  sono state considerate le seguenti relazioni fortemente semplificate:

$$(1a, 1b, 1c) \quad \left\{ \begin{array}{l} P_{tot,h} = \theta \\ P_{i,h} = P(|\Delta f|_{ih,max} \leq \gamma \sigma_i) \\ P_{tot,h} = \prod_{i=0}^2 P_{i,h} \text{ con } P_{i,h} = P_{i+1,h} \quad \forall (i,h) \end{array} \right.$$

Dove  $\theta$  è una variabile di progetto, che indica la percentuale di fonemi appartenenti a  $\mathbf{Z}$  che si vuol conservare. Come conseguenza diretta delle precedenti relazioni, si deduce che è possibile ricavare  $\gamma$ , imponendo:

$$(2) \quad P_{i,h} = \sqrt[3]{\theta}$$

In questa sede,  $\theta$  è stato posto uguale a 0.99, al fine di preservare la quasi totalità dei campioni del set  $\mathbf{Z}$  (d'altro canto tali fonemi erano interamente costituiti da vocali accentate segmentate manualmente).

Fissata la soglia di selezione tramite la (2), il selettore ripete l'analisi su i generici picchi di intensità inclusi nell'insieme  $\mathbf{V}$ : vengono calcolate quindi le differenze istantanee  $\Delta f_{ilj}$  tra i tracciati di ogni singolo fonema  $v_l$  dell'insieme  $\mathbf{V}$  e le relative medie  $\mu_{il}$ . In accordo con l'equazione (2) e con quanto detto in precedenza relativamente alla natura dei fonemi vocalici, i fonemi dell'insieme  $\mathbf{V}$  sono accettati se e solo se risultano verificati i seguenti vincoli:

$$(3) \quad \left\{ \begin{array}{l} \exists f_{ilj} \quad \forall (i,j) \\ |\Delta f|_{il,max} = \max_{j=1,\dots,30} |f_{ilj} - \mu_{il}| \leq \gamma \sigma_i \quad \forall (i,j) \end{array} \right.$$

A titolo illustrativo, si riporta in figura 3, l'interpretazione grafica dei vincoli ESC: le diverse curve rappresentano i tracciati di f0, f1 ed f2, mentre i rettangoli colorati indicano le aree di variabilità imposte dal sistema. I rettangoli sono centrati verticalmente rispetto alle medie  $\mu_{il}$  ed hanno altezza  $2\gamma\sigma_i$  (ove possibile le medie sono state indicate con una linea orizzontale tratteggiata): appare chiaro che, a parità di campioni di riferimento del set  $\mathbf{Z}$ , modificando  $\gamma$  è possibile restringere o ampliare l'area di tolleranza, accettando un numero rispettivamente minore o maggiore di fonemi. La lunghezza delle frecce, presenti nella figura in basso, indica il valore  $|\Delta f|_{il,max}$  per f1 ed f2.

Compiuta l'ultima analisi, il selettore è in grado di definire un sottogruppo  $\mathbf{W}$  di fonemi sonori, sufficientemente stabili nel tempo da poter essere considerati candidate vocali accentate. Per ogni elemento  $w_k$  dell'insieme  $\mathbf{W}$ , con  $k$  minore o uguale a  $l$ , vengono memorizzati i valori medi  $\mu_{1k}$  e  $\mu_{2k}$ , perciò ogni picco è univocamente individuabile su un piano formantico (f1, f2). Termina in questa fase la prima parte della metodologia proposta.

L'insieme  $\mathbf{W}$  sarà d'ora in avanti considerato esclusivamente come un insieme di punti ( $\mu_{1k}$ ,  $\mu_{2k}$ ) giacenti sul piano formantico. Tale insieme di punti sarà di seguito esplorato tramite algoritmi di ottimizzazione con il duplice scopo di classificare le diverse vocali accentate e di scartare gli elementi spuri, comunque presenti nell'insieme. Una rappresentazione grafica dell'insieme di punti  $\mathbf{W}$  è riportata in figura 4 sinistra.

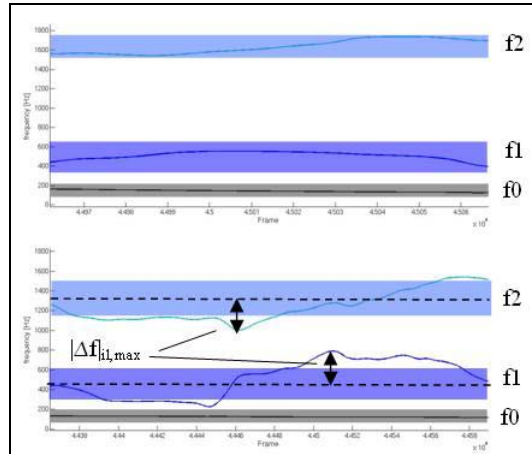


Figura 3: effetti del selettore ESC. Il fonema in alto viene accettato in quanto sufficientemente stabile, quello in basso viene scartato per l'eccessiva instabilità.

### 3. STIMA DEL TRIANGOLO VOCALICO

Durante le fasi di ricerca, i dati dell'insieme  $\mathcal{W}$  sono stati confrontati con la segmentazione manuale, riscontrando un elemento comune in tutti i parlanti: il triangolo definito dai punti  $m_k = (\mu_{1k}, \mu_{2k})$  risultava tendenzialmente sovrapposto con il triangolo vocalico dei diversi soggetti. In particolare le vocali risiedevano costantemente sul bordo di tale triangolo, mentre gli altri fonemi rimanevano all'interno di esso, come indicato in figura 4 centro. A partire da questa osservazione è stato messo a punto un sistema che stimasse le rette di tendenza del triangolo vocalico di ogni parlante. Per raggiungere tale scopo si è deciso di raggruppare i fonemi in tre grandi *cluster* (vedi figura 4 destra): un primo *cluster A* associato alla stima della retta delle vocali anteriori (retta  $a$ ), un secondo *cluster P* associato alla stima della retta delle vocali posteriori (retta  $p$ ) ed infine un terzo *cluster C* di fonemi centralizzati, associati alla bisettrice dell'angolo acuto  $\beta$  formato da  $a$  e  $p$  (retta  $c$ ). Dal momento che la retta  $c$ , per costruzione, è vincolata alla posizione dalle altre due, ci si soffermerà, nel seguito, sul metodo impiegato per definire le rette del triangolo vocalico.

Le rette vengono definite in automatico, tramite un metodo autoassociativo ottenuto modificando il ben noto algoritmo k-means (di cui si parlerà nel dettaglio nel prossimo capitolo – dove verrà impiegato). A titolo introduttivo, si definisce con autoassociativo: un metodo iterativo in grado di adattarsi automaticamente ai dati, con lo scopo di raggrupparli secondo regole e vincoli imposti dall'utente, impiegando un numero di passi finito.

Tecnicamente, le regole prendono il nome di funzione obiettivo e il numero di passi è definito secondo i così detti criteri di arresto. Il metodo viene implementato in forma di problema di ottimizzazione, così che l'algoritmo si riconduca alla minimizzazione o massimizzazione vincolata (a seconda dell'esigenza) della data funzione obiettivo.



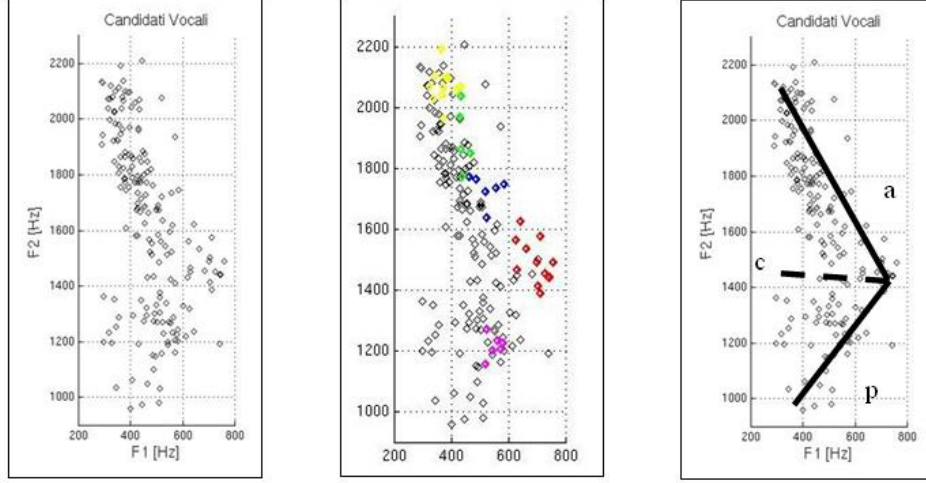


Figura 4: a sinistra rappresentazione grafica dell'insieme  $\mathcal{W}$ , al centro l'insieme sovrapposto alle vocali del parlante, a destra grafica delle rette  $a$ ,  $p$  e  $c$  (tratteggiata).

Nello specifico, questo significa che l'algoritmo adottato deve essere in grado di creare tre gruppi di fonemi, tali per cui sia minimizzata la somma  $S$  delle distanze dei singoli punti appartenenti ad  $\mathcal{A}$  e  $\mathcal{P}$ , rispetto alle rette  $a$  e  $p$ . In maniera compatta, il problema può essere formalizzato secondo la seguente scrittura:

$$(4) \quad \arg \min_{\alpha_1, \alpha_2 \in \mathbb{R}^2} S = S(\alpha_1, \alpha_2) = \sum_{i=1}^2 \sum_{t=1}^{n_i} \frac{|\mu_{2t} - \alpha_{i1}\mu_{1t} - \alpha_{i2}|}{\sqrt{1 + \alpha_{i1}^2}}$$

Con:

$$\alpha_1 = \begin{bmatrix} \alpha_{11} & \alpha_{12} \end{bmatrix}$$

$$\alpha_2 = \begin{bmatrix} \alpha_{21} & \alpha_{22} \end{bmatrix}$$

Dove  $\alpha_1$  e  $\alpha_2$  rappresentano i coefficienti delle rette  $a$  e  $p$ , mentre  $n_i$  indica il numero di punti appartenenti rispettivamente ad  $\mathcal{A}$  e  $\mathcal{P}$ . La funzione obiettivo non è altro che l'equazione della distanza euclidea di un punto da una retta.

Le rette  $a$  e  $p$ , vengono ricalcolate ad ogni passo, imponendo un classico problema ai minimi quadrati: rappresentano, dunque, le linee di tendenza degli ipotetici gruppi anteriore e posteriore. Definita con  $\Delta S$  la differenza tra due successive catalogazioni, il criterio di arresto prevede che l'iterazione termini quando  $\Delta S$  è nullo.

Il problema può essere spiegato in maniera intuitiva in questo senso: viene tentato un raggruppamento con tre rette "di primo tentativo". Ogni punto, rappresentante un fonema, viene associato alla retta più vicina. In questo modo tutti i punti sono inseriti in uno dei tre cluster. I punti del cluster  $\mathcal{C}$  sono messi da parte (sono foni fortemente centralizzati) e viene calcolata la distanza di ogni punto dei cluster  $\mathcal{A}$  e  $\mathcal{P}$  dalle rette associate. Per entrambi i cluster viene calcolata la somma di tutte le distanze, ottenendo due numeri reali positivi.

Infine viene calcolata la somma di questi due valori, ottenendo un unico dato complessivo, indicato con  $S$ .

A questo punto l'algoritmo sostituisce le rette di primo tentativo con le rette di tendenza dei punti inseriti nei *cluster*  $A$  e  $P$ . Essendo mutate  $a$  e  $p$ , viene ricalcolata di conseguenza anche la retta  $c$ . Ottenute delle nuove rette, diverse dalle prime, l'algoritmo annulla il precedente raggruppamento e ne realizza uno nuovo, che darà risultati in genere diversi dal primo: saranno ottenuti dei nuovi cluster  $A$  e  $P$  e muterà di conseguenza anche  $S$ . Confrontando la nuova classificazione con la vecchia, è possibile ottenere la variazione  $\Delta S$ .

L'algoritmo prosegue con nuove iterazioni, fino a che è possibile osservare una stabilizzazione dei *cluster*. Di fatto, all'aumentare del numero di iterazioni diminuisce la differenza  $\Delta S$  tra due catalogazioni successive, sino ad arrivare ad un punto in cui l'algoritmo non adduce alcun cambiamento ad una data configurazione. Questo stato rappresenta il così detto criterio di arresto dell'algoritmo.

Una volta terminato il processo, si è in grado di analizzare le diverse configurazioni ottenute. Dal momento che interessa la configurazione associata al valore minimo di  $S$ , indicato con  $S^*$  (vedi figura 5 destra), si considera le rette ottimali  $a^*$  e  $p^*$  ad esso associate, come la miglior stima possibile del triangolo vocalico del parlante (un esempio grafico è riportato in figura 5 centro).

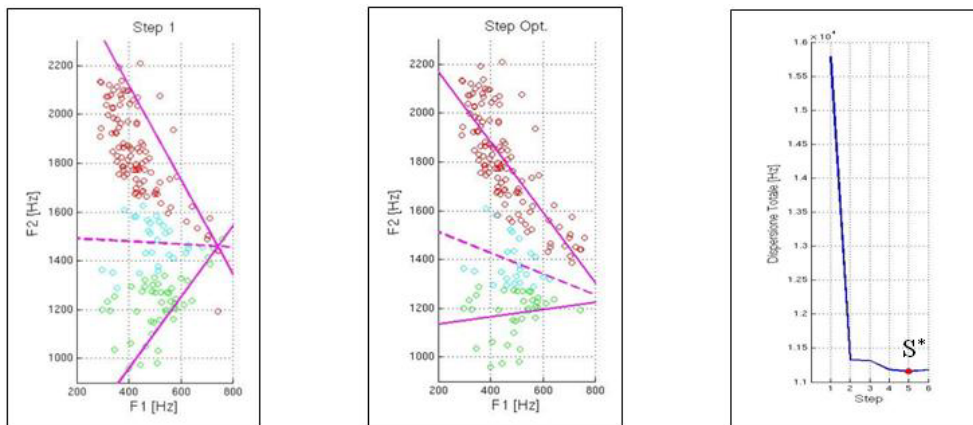


Figura 5: i cluster  $A$  (rosso),  $P$  (verde) e  $C$  (ciano) e la curva  $S$ . A sinistra: primo tentativo; al centro: configurazione ottima; a destra: indicazione di  $S^*$ .

Per quanto riguarda le rette di primo tentativo, esse sono definite a partire da tre punti generici ubicati ai presunti estremi del triangolo. In riferimento alle convenzioni adottate per la rappresentazione del piano formantico, vengono scelte le rette passanti: per il punto posto all'estremo in alto a sinistra del set di dati, per il punto posto all'estremo in basso a sinistra del set, per il punto più a destra in prossimità della frequenza  $f_2$  associata al baricentro del set (vedi figura 5 sinistra).

In ambito strettamente matematico, occorre rilevare che la convergenza dell'algoritmo di impiegato per l'individuazione del triangolo vocalico non è ancora stata propriamente provata. Conseguentemente, un ulteriore elemento da inserire nell'ambito dello studio riguarda proprio la dimostrazione di convergenza, che in questa fase è sempre stata raggiunta ma non assicurata a priori.

#### 4. STIMA DEI CENTRI VOCALICI E SELEZIONE DEI CAMPIONI

Ottenuta la stima del triangolo vocalico, questa può essere impiegata per approssimare la posizione dei centri vocalici del parlante. Con centri vocalici, si intende qui la media dei valori di  $(\mu_{1k}, \mu_{2k})$  associati ai campioni di una certa vocale. In maniera più formale, dato un generico *cluster*  $R_i$  di  $n_i$  punti  $P_t = (\mu_{1t}, \mu_{2t})$ , con  $t=1, \dots, n_i$ , si definisce centro vocalico  $v_j$ , con  $j=1, \dots, 7$  per l'italiano, il punto di coordinate:

$$(5) \quad v_j = (v_{1j}, v_{2j}) = \left( \frac{1}{(d-s+1)} \sum_{j=s}^d \mu_{1j}, \frac{1}{(d-s+1)} \sum_{j=s}^d \mu_{2j} \right)$$

Con  $1 \leq s \leq d \leq n_i$ . Nel nostro caso i *cluster* cui si fa riferimento sono quelli delle vocali anteriori e posteriori: in altre parole, si calcolano i centri vocalici suddividendo i punti di ogni *cluster* in ulteriori sottoinsiemi, ciascuno associato ad una delle vocali dell'italiano.

A tal fine è possibile impostare un ragionamento analogo a quello proposto per le rette del triangolo vocalico: partendo da dei centri di tentativo, si cerca di definire i reali centri vocalici tramite iterazioni successive. La metodologia proposta è una basata su una versione leggermente modificata del problema noto come k-means, definito da J.B. MacQueen nel 1967. Il problema, si prefigge lo scopo di raggruppare un set di dati in un numero  $n$  prefissato di *cluster*, definito dall'utente. Tra i possibili *cluster* con cui si può raggruppare il set di dati, il problema vuol definire una configurazione ottimale, caratterizzata da particolari proprietà di compattezza.

Definiti con  $c_j$  i baricentri dei cluster, con  $j=1, \dots, n$ , ed indicati con  $P_i = (\mu_{1i}, \mu_{2i})$  i punti appartenenti ad ogni singolo *cluster*  $C_j$ , il problema può essere definito secondo la scrittura:

$$(6) \quad \arg \min_{c_j \in \mathbb{R}^2} S = S(c_1, \dots, c_n) = \sum_{j=1}^n \sum_{P_i \in C_j} \|P_i - c_j\|$$

Dove  $c_j$ , in accordo con l'equazione (5), rappresenta la media delle coordinate dei singoli punti  $P_i$  appartenenti al cluster  $C_j$ . Nel caso in questione, una volta raggiunta la convergenza, i baricentri  $c_j$  rappresentano una stima ottima dei centri vocalici  $v_j$ . Il criterio di arresto è nuovamente dato dall'annullamento di  $\Delta S$  tra due catalogazioni successive.

Uno dei più comuni metodi di risoluzione del problema, noto come algoritmo di Lloyd, prevede la definizione di un set di  $n$  baricentri di tentativo (spesso indicati con il termine "semi"), attorno ai quali aggregare i vari dati in gruppi, come descritto in Huang *et alii* (2001). L'algoritmo si sviluppa esattamente come descritto in precedenza per l'individuazione del triangolo vocalico (in effetti, l'algoritmo di Lloyd è il metodo da cui si è partiti per individuare il criterio di stima delle rette  $a$  e  $p$ ). Dunque, si esegue una prima iterazione (vedi figura 6 sinistra), associando ogni elemento del *cluster*  $A$  ad un possibile baricentro delle vocali anteriori ed associando ogni elemento del *cluster*  $P$  ad un possibile baricentro delle vocali posteriori. I dati del *cluster*  $C$  sono considerati spuri ed esclusi a priori, data l'evidente centralizzazione. Conclusa la prima catalogazione, l'algoritmo calcolerà i baricentri dei *cluster* vocalici, ottenendo sette punti distinti. In seguito valuterà la somma  $S$  di tutte le distanze tra punti e baricentri ad essi associati. Eseguita questa operazione, la catalogazione viene annullata. I nuovi baricentri sono impiegati come nuovi semi, perciò verrà compiuta una nuova catalogazione e verrà computata un'ulteriore stima

di  $S$ . Anche in questo caso, catalogazioni diverse saranno associate a valori diversi di  $S$ , tuttavia tali differenze si assottiglieranno con l'aumentare delle iterazioni, sino ad annullarsi. In questo caso è possibile dimostrare che la configurazione più compatta è sempre quella ottenuta con l'ultima iterazione (vedi figura 6 destra). Dunque, terminate le successive catalogazioni, l'algoritmo è in grado di stimare i centri vocalici del parlante.

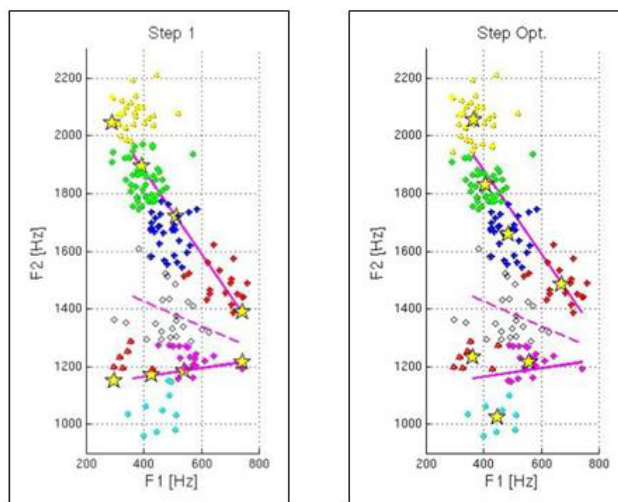


Figura 6: stima dei centri vocalici. A sinistra catalogazione iniziale, a destra catalogazione ottima (le stelle indicano le posizioni dei centri vocalici).

Le modifiche apportate al sistema originale di Lloyd sono due. In primo luogo esiste un problema relativo alla configurazione dei *cluster* vocalici: la vocale 'a' si presenta sia nel *cluster A* sia nel *cluster P* (in effetti, la sua presenza è possibile anche nel *cluster C*, al momento trascurato), per tale motivo vengono definiti 8 semi, assegnando alla vocale 'a' due diversi *cluster* vocalici tra loro indipendenti – uno per il ramo anteriore ed uno per il posteriore. Terminate le iterazioni, i due *cluster* vengono uniti ed il centro vocalico viene stimato a partire dalla fusione dei dati, impiegando la (5). Tale soluzione non è probabilmente la più corretta formalmente ma ha dimostrato di essere efficace nei risultati.

In secondo luogo, esiste la possibilità che una vocale non sia presente tra i dati ricavati sino a questo punto (senza entrare nel merito dei sistemi non standard quali alcuni pentavocalici, si pensi all'effettiva assenza nel parlato di particolari vocali accentate). In questo caso è stata modificata la procedura di *bootstrap* dell'algoritmo. In pratica, dopo la prima catalogazione viene valutata l'esistenza o meno di dati all'interno dei *cluster* vocalici: se alcuni *cluster* rimangono vuoti alla prima iterazione, i semi ad essi associati sono eliminati e il primo passo viene ripetuto nuovamente con i semi rimanenti. È bene precisare che l'assenza di dati nella prima iterazione non è condizione né necessaria né sufficiente per dimostrare l'effettiva assenza di determinate vocali, tuttavia le prove sperimentali hanno indicato che l'errore indotto da questa semplificazione non è rilevante (quantomeno in relazione a questa fase di avanzamento del progetto, dove, si indicherà più avanti, esistono limitazioni più importanti).

Le condizioni iniziali per il posizionamento dei semi attingono nuovamente al lavoro di Cosi *et alii* (1995). All'interno dello studio sono riportate le statistiche relative alle vocali presenti nei bisillabi di AIDA. Tra i vari dati di interesse, sono presenti i valori medi di  $f_1$

ed  $f_2$  sia per gli uomini che per le donne. Tali valori possono essere assunti come i centri vocalici  $v_{mj}$  di un parlante italiano medio. A tali dati sarà dunque associato il triangolo vocalico medio dell'italiano (o una stima verosimile dello stesso). È possibile elaborare questi punti per ricavarne i semi utili per il *bootstrap* dell'algoritmo di catalogazione.

Innanzitutto, i centri  $v_{mj}$  devono essere traslati in maniera tale da far coincidere il centro vocalico medio della 'a' con il punto di intersezione delle rette  $a^*$ ,  $p^*$  e  $c^*$  di ogni parlante. Di seguito il ramo anteriore e posteriore del triangolo medio vengono scalati linearmente e indipendentemente sia lungo la direzione di  $f_1$  sia lungo  $f_2$ . Le operazioni di traslazione e scalatura non fanno altro che spostare e deformare il triangolo del parlante medio sino a sovrapporlo con quello di un parlante generico. Compiute queste operazioni, i centri individuati da Cosi *et alii* (1995), si trovano giacenti sul triangolo del parlante oggetto di studio. Ovviamente, il posizionamento di tali punti non rappresenta l'individuazione di veri centri vocalici (basti pensare alle distanze relative tra le vocali), tuttavia è un primo passo per inizializzare la ricerca dei veri centri vocalici.

Una rappresentazione grafica di riferimento è riportata in figura 7: a sinistra sono indicati il triangolo vocalico di un parlante ed i centri vocalici medi dell'italiano (cerchi blu), al centro è visibile la loro proiezione sul triangolo del parlante (stelle).

Una volta concluse le operazioni di stima ottima dei centri vocalici, i *cluster* associati a tali centri raccolgono tutti i foni appartenenti al ramo anteriore e posteriore del triangolo vocalico. In tale configurazione sono catalogati sia gli effettivi fonemi vocalici accentati sia una parte di dati spuri, relativamente distanti dai centri  $c_j$ . Al fine di affinare la catalogazione ed eliminare i dati spuri, viene imposta una soglia di accettazione regolabile dall'utente. Tale soglia è stimata in maniera dinamica al variare della covarianza dei *cluster* vocalici.

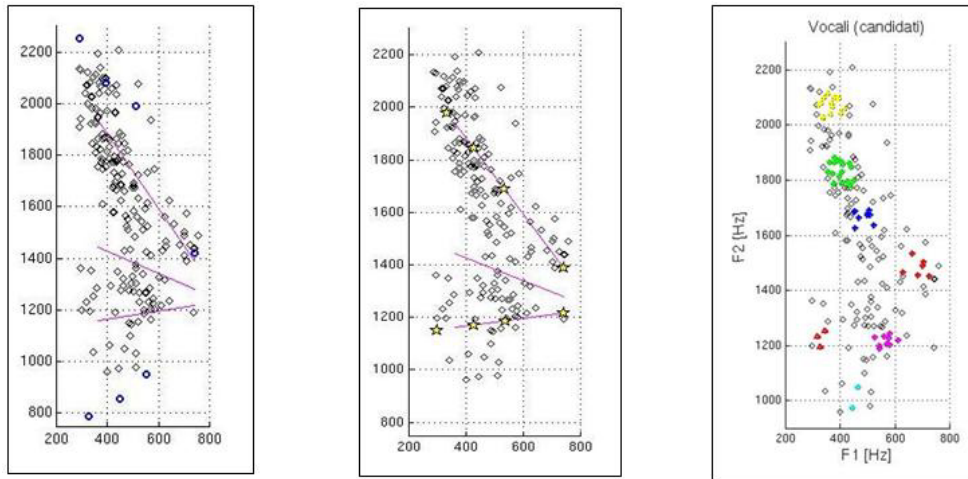


Figura 7: adattamento dei centri vocalici medi (Cosi *et alii*, 1995) al triangolo vocalico di un parlante casuale (sinistra e centro) e presunti fonemi vocalici dopo la sogliatura imposta dall'operatore (destra).

In maniera più precisa, si è assunto in questa sede che i *cluster* vocalici fossero ben modellabili con distribuzioni gaussiane multivariate a 2 dimensioni. Le medie stimate di tali distribuzioni coincidono con i centri  $c_j$  dei *cluster*, mentre le matrici di covarianza  $\Sigma_j$

sono stimate a partire dai dati (spuri o meno) presenti nei *cluster* stessi. Tali parametri permettono di individuare un'ellisse sul piano formantico in prossimità di ogni *cluster*: i centri vocalici coincidono con i centri delle ellissi, mentre gli assi delle stesse e la loro rotazione sono definiti in base a  $\Sigma_j$ .

La soglia di filtraggio è regolata in funzione della matrice  $\Sigma_j$ : l'utente indica un coefficiente positivo  $\zeta$  con cui scalare le dimensioni delle ellissi associate ai *cluster*. Valori di  $\zeta$  maggiori di 1 aumentano l'area dell'ellisse, mentre valori minori di 1 la riducono. Il sistema di filtraggio trattiene solo i punti interni all'ellisse di ogni vocale, scartando i dati esterni. In questo modo è possibile limitare la selezione dei fonemi, trattenendo solo i presunti fonemi vocalici accentati (figura 7 destra).

## 5. SPERIMENTAZIONE

Una volta ipotizzato un possibile metodo di ricerca delle vocali accentate, il primo passo della sperimentazione è stato quello di definire un parlato di natura generica ma ripetibile, che fosse noto a priori durante i test. A tale scopo ci si è orientati sul parlato letto. Il testo scelto è stato liberamente tratto da Georges (1994). Data la brevità, la porzione impiegata è riportata di seguito in maniera integrale:

*“I primi esseri viventi hanno cominciato a esprimersi a gesti che, per chi stava loro accanto, sono diventati segni, tipici del linguaggio orale. I segni della gestualità precisano la parola, la enfatizzano e spesso ne prendono il posto. Il corpo parla al di là della parola e della scrittura; consente di esprimere i propri sentimenti e di comunicare con i propri simili anche senza menzionare alcuna parola. I segni del corpo sono la forma prima del linguaggio. I neonati, ben prima di aver imparato a parlare, si esprimono con il corpo e con diverse manifestazioni fonetiche: strilli, gemiti e risa. Per chi li deve interpretare questi segnali sono solo indizi. A poco a poco però il bambino imparerà non solo il sistema fonetico della lingua madre ma anche il linguaggio somatico, cioè del corpo.”*

### 5.1 Costituzione della base dati

Sono stati acquisiti 16 parlanti maschili di età compresa tra i diciannove ed i trenta anni, principalmente studenti dei corsi di laurea in ingegneria tenuti presso l'Ateneo. Nove di essi sono stati acquisiti per tre volte, a distanza di circa dieci giorni tra una sessione e l'altra; tre di essi sono stati acquisiti per due volte, secondo le medesime modalità; i rimanenti parlanti sono stati acquisiti una sola volta. In totale sono state acquisite 37 tracce. Ad ogni soggetto è stata fatta raccomandazione di leggere in maniera spontanea, senza preoccuparsi dei possibili difetti di dizione, trascurando gli sbagli e, qual'ora si sia presentata qualche manifestazione spontanea estranea al testo (interruzione per risate, per perdita del segno nel testo e quant'altro), è stato chiesto di non interrompere la sessione, riprendendo la lettura dal punto che si ritenesse più opportuno. Nel caso in cui dette manifestazioni abbiano permesso di acquisire enunciati estranei al testo (eventi piuttosto rari), tali enunciati sono stati esclusi nelle fasi di elaborazione.

Le acquisizioni si sono tenute in un locale dell'università adiacente alle aule presso cui si tenevano le lezioni, non dunque in cabine appositamente preparate. Di conseguenza nel locale, arredato a sala conferenze, era presente del riverbero. Inoltre, pur limitando gli eventi accessori, le registrazioni contengono altri segnali oltre a quello vocale. Principalmente si tratta di eventi generati dal lettore stesso, come ad esempio la respirazione o lo stropiccio dei fogli di lettura contro il microfono.

La catena di acquisizione è stata realizzata impiegando un microfono a doppio condensatore, normalmente impiegato per registrazioni musicali. La scelta del doppio condensatore è stata dettata dalla possibilità di disattivare uno dei due elementi, al fine di limitare l'area di acquisizione nella sola direzione del parlante (la sensibilità del microfono era ridotta nelle altre direzioni). Il microfono, preamplificato, trasmetteva il proprio segnale ad una scheda audio munita di porta usb, che scaricava i dati direttamente sul pc. Al fine di mantenere un livello di segnale relativamente costante fra le varie tracce, si è agito sul guadagno della strumentazione per compensare la fonazione dei diversi parlanti. Per la registrazione dei dati è stato utilizzato il software *audacity*, rilasciato su licenza GPL. Il formato wave è stato scelto per la memorizzazione dei dati. Il segnale è stato codificato in PCM mono con frequenza di campionamento pari a 44100Hz e quantizzazione a 16 bit. La codifica scelta è in realtà ridondante rispetto alle effettive necessità del progetto, tuttavia si è preferito acquisire i dati in alta qualità, avendo comunque la possibilità di alterare la stessa in un secondo tempo (ad esempio per produrre ulteriori test con codifica telefonica).

Data la necessità di verificare le capacità discriminatorie e classificatorie del sistema, le tracce sono state innanzi tutto elaborate manualmente tramite *Praat*, isolando i fonemi vocalici di interesse.

## 5.2 Analisi dei risultati

Al fine di valutare l'efficacia dei diversi passaggi costituenti il metodo proposto, si riportano di seguito alcuni dati relativi ai risultati di classificazione. I dati fanno riferimento a due diverse configurazioni: la prima è priva di soglia di filtraggio finale ( $\zeta=+\infty$ ), la seconda è stata ottenuta imponendo una soglia di filtraggio  $\zeta=+2$ . La seconda scelta è dovuta al fatto che tale soglia ha rappresentato con buona approssimazione la configurazione di equal error rate (EER). Per il calcolo dell'EER si è fatto riferimento alla percentuale di dati spuri trattenuti dal sistema ed a quella di vocali accentate respinte; entrambe le percentuali sono state calcolate al termine della catena di selezione, catalogazione e filtraggio.

La tabella sottostante riporta il numero totale di vocali accentate campionate manualmente da tutte le tracce:

| a   | ε   | e   | i   | ɔ   | o   | u |
|-----|-----|-----|-----|-----|-----|---|
| 336 | 222 | 143 | 280 | 203 | 135 | 0 |

Innanzitutto si è analizzato l'efficacia del selettore IBT-DESF ed il contributo prodotto dal metodo di filtraggio posto in coda alla procedura (tramite la variabile  $\zeta$ ). La loro resa è ricavabile dall'osservazione degli errori di Tipo I e II. Relativamente alle cause di tali errori, si ricorda che tutte le condizioni imposte dal selettore IBT-DESF sono strettamente necessarie ma non sufficienti, per tale motivo è congenita la presenza di errori di tipo I (falsa accettazione). Parimenti, data la natura delle soglie di filtraggio, risulta inevitabile anche un errore di tipo II (falsa reiezione). Infine c'è da precisare che la costante  $\gamma$ , così come è stata calcolata, causa un limitato errore di tipo II. Le percentuali di tali errori, nelle due configurazioni, sono risultate pari a:

|                 | Tipo I | Tipo II |
|-----------------|--------|---------|
| $\zeta=+\infty$ | 62.5   | 5.9     |
| $\zeta=+2$      | 39.7   | 39.0    |

La tabella indica che la capacità di selezione del metodo IBT-DESF non è particolarmente performante: una gran parte dei foni spuri continua a permanere al momento di individuare triangolo e centri vocalici. Al fine di raggiungere una configurazione di equilibrio (EER) si è costretti ad intervenire con un filtraggio molto selettivo, causando l'innalzamento del punto di EER. Scendendo nel dettaglio, le percentuali di rumore residuo (foni spuri) e di vocali non catalogate sono così ripartite:

|                 |         | a    | ε    | e    | i    | ɔ    | o    | u   |
|-----------------|---------|------|------|------|------|------|------|-----|
| $\zeta=+\infty$ | Tipo I  | 9.0  | 13.8 | 17.6 | 10.9 | 5.2  | 3.5  | 2.4 |
|                 | Tipo II | 7.4  | 0.5  | 0    | 0    | 16.8 | 17.5 | -   |
| $\zeta=+2$      | Tipo I  | 5.4  | 8.9  | 11.3 | 7.1  | 3.1  | 2.2  | 1.6 |
|                 | Tipo II | 38.3 | 57.7 | 35.0 | 21.8 | 42.4 | 48.2 | -   |

Relativamente alla capacità discriminatoria del sistema, ovvero alla capacità di individuare il triangolo ed i centri vocalici, sono state calcolate le matrici di confusione nelle due configurazioni. I valori sono espressi in riferimento al numero di vocali campionate manualmente, di conseguenza, sono affetti dall'errore di tipo II indotto sia dal selettore sia dalla soglia. Si ricorda che le colonne indicano il tipo di vocale, mentre le righe indicano la classificazione eseguita dal sistema. Matrice di confusione nel caso  $\zeta=+\infty$ :

|   | a           | ε           | e           | i           | ɔ           | o           | u |
|---|-------------|-------------|-------------|-------------|-------------|-------------|---|
| a | <b>89.3</b> | 1.4         | 0           | 0           | 6.4         | 5.9         | - |
| ε | 3.3         | <b>71.6</b> | 5.6         | 0           | 0           | 0           | - |
| e | 0           | 25.7        | <b>68.5</b> | 6.4         | 0           | 0           | - |
| i | 0           | 0.9         | 25.9        | <b>93.6</b> | 0           | 0           | - |
| ɔ | 0           | 0           | 0           | 0           | <b>50.7</b> | 17.8        | - |
| o | 0           | 0           | 0           | 0           | 21.7        | <b>41.5</b> | - |
| u | 0           | 0           | 0           | 0           | 4.4         | 13.3        | - |

Matrice di confusione nel caso  $\zeta=+2$ :

|   | a           | ε           | e           | i           | ɔ           | o           | u |
|---|-------------|-------------|-------------|-------------|-------------|-------------|---|
| a | <b>61.6</b> | 0.5         | 0           | 0           | 3.5         | 2.4         | - |
| ε | 0           | <b>37.4</b> | 3.5         | 0           | 0           | 0           | - |
| e | 0           | 4.5         | <b>53.1</b> | 3.6         | 0           | 0           | - |
| i | 0           | 0           | 8.4         | <b>74.6</b> | 0           | 0           | - |
| ɔ | 0           | 0           | 0           | 0           | <b>37.0</b> | 9.4         | - |
| o | 0           | 0           | 0           | 0           | 13.8        | <b>30.6</b> | - |
| u | 0           | 0           | 0           | 0           | 3.5         | 9.4         | - |

Le matrici di confusione indicano che le vocali vengono in buona parte riconosciute (le percentuali sulla diagonale sono molto superiori a quelle fuori diagonale), se ne deduce, quindi, che i sistemi di catalogazione riescono comunque ad effettuare delle buone stime



del triangolo e dei centri vocalici. Si tenga conto che i parametri fuori diagonale, relativi alle coppie  $\epsilon/\epsilon - \text{ɔ}/\text{o}$ , sono comunque affetti da un inevitabile errore, dovuto alla necessità di porre un confine netto tra fonî, che spesso sfumano in maniera continua da un dominio all'altro. Per tale motivo sono stati indicati in corsivo.

Le conclusioni che si possono dedurre dai due esperimenti pilota, indicano che i principali limiti della tecnologia riguardano il selettore IBT-DESF. In ogni caso si è dimostrato poco capace di filtrare i fonî spuri, specie quelli relativi alle vocali posteriori. In molti casi la stima del triangolo vocalico tendeva a evidenziare un ramo posteriore più centralizzato di quello reale, con ovvie imprecisioni che si ripercuotevano sull'individuazione dei centri vocalici. Uno degli effetti più evidenti di questo fenomeno è l'impropria individuazione di u. Come si evince anche dalla figura 6 destra, la presenza di fonî spuri provoca spesso un impreciso posizionamento dell'ipotetico centro vocalico delle u, che si trova addirittura a valori di  $f_1$  minori del centro vocalico della i. Altre volte, la presenza di fonî spuri si ripercuote addirittura sull'individuazione di tutte le vocali posteriori, la cui catalogazione risulta leggermente traslata rispetto alla reale posizione. Al contrario, le vocali anteriori risultano poco influenzate dai fonî spuri. È altresì vero che il selettore risulta relativamente più efficace sul ramo anteriore, riducendo i fenomeni di rumorosità. Una delle possibili cause della differente capacità di filtraggio può essere associata ad una carente od erronea raccolta di informazioni riguardo alla stabilità delle vocali posteriori, mentre la carenza generale di selettività potrebbe essere parzialmente compensata con regolazioni migliori dei coefficienti empirici.

Le conseguenze sulla fase di catalogazione dei fonî sono ovvie: data l'impossibilità di definire la differenza tra un fono accentato o meno, a partire dalla mera posizione sul piano  $f_1$ - $f_2$ , la classificazione finale conduce a trattenere alcuni fonî erronei, escludendone altri invece rilevanti. Risulta importante, quindi, migliorare il selettore.

Come nota a margine, si fa notare che la segmentazione manuale di confronto, eseguita da personale non specializzato, è sì un primo utile termine di paragone, ma è risultata anch'essa non priva di imperfezioni. Questo ci spinge a ritenere che alcuni dati delle matrici di confusione potrebbero risultare qualitativamente migliori, se paragonati con una segmentazione più accurata.

## 6. CONCLUSIONI E SVILUPPI FUTURI

Il presente articolo descrive la struttura e la prima sperimentazione di un sistema specializzato per l'individuazione e la catalogazione di fonemi vocalici. Il metodo è stato ipotizzato nell'ambito di un progetto di ricerca relativo al riconoscimento automatico del parlante, in essere presso il DMTI. Uno degli obiettivi per lo sviluppo di tale sistema è stato quello di realizzare uno strumento totalmente automatico, che richiedesse costi di taratura minimi. Allo stato attuale della ricerca è stato sviluppato un prototipo funzionante di tale sistema. Il prototipo, nella configurazione attuale, è stato testato solo in condizioni testodipendenti su parlato letto in laboratorio.

La tecnologia impiegata è relativamente semplice e si basa su principi matematici elementari come le rette di regressione (analisi dei dati ai minimi quadrati) e la catalogazione tramite k-means (algoritmo di classificazione di tipo deterministico). Più in specifico, la procedura di classificazione si appoggia su parametri quali l'intensità della traccia audio, la frequenza fondamentale e le prime due formanti. Tale strumento ha lo scopo di eseguire una prima selezione grossolana, che consenta di isolare segmenti classificabili come fonî sonori.

Gli elementi così isolati, rappresentati su un piano formantico F1-F2, appaiono racchiusi in un triangolo, limitato, appunto, dalle vocali accentate. Tramite il posizionamento ottimale delle rette di regressione, è possibile suddividere iterativamente i fonemi in tre gruppi: posteriore, anteriore e dei fonemi centralizzati. Al termine dell'iterazione si ottiene una stima approssimata del triangolo vocalico sottostante i fonemi.

Infine, il sistema di riconoscimento prevede la classificazione delle sette vocali tramite il noto algoritmo k-means e la selezione dei fonemi più rappresentativi tramite l'analisi della covarianza dei singoli *cluster*. Quest'ultima fase è stata parametrizzata, così da regolare il dettaglio del sistema in base alle proprie esigenze. Tramite questa accortezza il prototipo potrà essere impiegato in maniera flessibile sia come strumento di screening che come catalogatore vero e proprio.

La sperimentazione pilota mostra che la classificazione prodotta, anche tenendo conto di un possibile affinamento, risulta sicuramente meno accurata di quella ottenibile con mezzi più sofisticati proposti dalla recente letteratura (vedi Huang *et alii* (2001)). Tuttavia, se talune ricerche richiedono un'analisi estremamente accurata dei dati, dall'altro lato esistono svariate altre applicazioni, specie nel campo della biometrica, per le quali lo strumento, al termine del suo sviluppo, risulterà adeguato. Inoltre, anche nelle applicazioni più esigenti, lo strumento può risultare un ottimo mezzo per lo screening dei corpora con cui un fonetista deve confrontarsi giornalmente.

Lo sviluppo previsto per il progetto è attualmente articolato in due fasi: la prima mira ad affinare la precisione di selezione, migliorando le caratteristiche del selettore IBT-DESF o, nel caso si dimostri necessario, valutando la possibilità di integrare tale selettore.

La seconda fase di sviluppo mira ad applicare il sistema a corpora di parlato spontaneo o semispontaneo acquisiti sia in laboratorio che telefonicamente. L'obiettivo finale, al momento ancora distante, è quello di implementare uno strumento in grado di individuare i fonemi di interesse su acquisizioni del mondo reale (ad esempio conversazioni telefoniche vere) così da facilitare ed automatizzare le procedure di preprocessamento di un sistema di riconoscimento vocale.

## BIBLIOGRAFIA

Alfano, I. (2005), La percezione dell'accento lessicale: un test sull'italiano a confronto con lo spagnolo, in *Atti del II Convegno Nazionale AISV 2005*, Salerno, 30 Novembre – 2 Dicembre.

Così, P., Ferrero, F. & Vaggies K. (1995), Rappresentazioni acustiche e uditive delle vocali italiane, in *Atti del XXIII Congresso Nazionale AIA 1995*, Bologna, 12-14 Settembre, pp. 151-156.

Georges, J. (1994), Il linguaggio dei segni, la scrittura e il suo doppio, Electa/Gallimard.

Huang, X., Acero, A. & Hon, H. (2001), *Spoken language processing: a guide to theory, algorithm and system development*, Washington: Prentice Hall.

# SVILUPPO DI UN SISTEMA DI RICONOSCIMENTO PER L'ARABO: PROBLEMI E SOLUZIONI

Piero Cosi, Mauro Nicolao, Giacomo Somnavilla, Graziano Tisato

Istituto di Scienze e Tecnologie della Cognizione, Sede di Padova “Fonetica e Dialettologia”,  
Consiglio Nazionale delle Ricerche  
via Martiri della Libertà, 2 - 35137 Padova, Italia  
{cosi, nicolao, somnavilla, tisato} @ pd.istc.cnr.it

## 1. INTRODUZIONE

L'Arabo è attualmente una delle lingue più parlate nel mondo. Il numero di parlanti arabi è valutato intorno a 325 milioni, di cui approssimativamente 225 milioni sono parlanti L1 e 100 milioni sono parlanti L2. L'Arabo è la lingua ufficiale in più di 22 paesi e la lingua ufficiale per l'istruzione religiosa dell'Islam, di conseguenza si stima che anche molti altri parlanti ne abbiano una conoscenza almeno passiva.

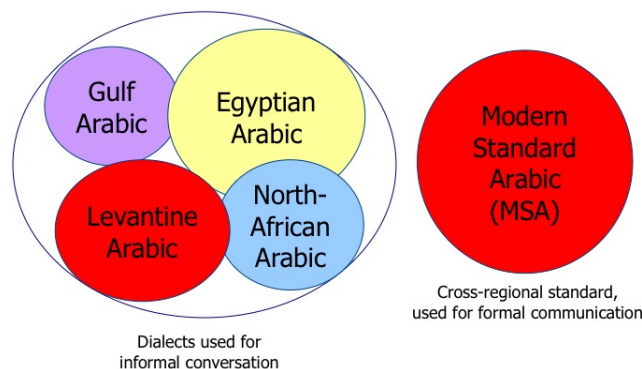


Figura 1: Le varietà linguistiche dell'Arabo (Kirchhoff et alii, 2002).

### *Varietà linguistica*

Prima di tutto, è importante rendersi conto che quando ci si riferisce alla lingua Araba non s'intende una singola varietà linguistica, ma, com'è illustrato in Figura , più che altro, ad una collezione di dialetti linguisticamente anche molto differenti tra loro.

Esiste l'Arabo Classico, che è una lingua antica, letteraria, importante principalmente perché usata nel Corano.

Con il termine Arabo invece, normalmente, s'intende il cosiddetto Arabo Standard Moderno (MSA<sup>1</sup>) che è una versione dell'Arabo Classico con un vocabolario modernizzato. MSA è convenzionalmente adottata come lingua comune a tutti i paesi di cultura araba ed è utilizzata dai mezzi di informazione (giornali, radio, TV), nei discorsi ufficiali e in tutti i principali canali di comunicazione internazionali.

Tuttavia, MSA non è quasi mai usato per la comunicazione quotidiana e informale, per la quale invece vengono impiegati i vari dialetti locali. I dialetti dell'Arabo possono approssimativamente essere divisi in due gruppi:

---

<sup>1</sup> MSA: Modern Standard Arabic

- Arabo Occidentale (Western Arabic), che include i dialetti parlati nel Marocco, in Algeria, in Tunisia ed in Libia,
- Arabo Orientale (Eastern Arabic), che può ulteriormente essere suddiviso in Egiziano, Levantino e Arabo del Golfo Persico.

| Cambiamenti         | MSA           | ECA       |
|---------------------|---------------|-----------|
| /θ/ → /s/, /t/      | /θala:θa/     | /tala:ta/ |
| /ð/ → /z/, /d/      | /ðahab/       | /dahab/   |
| /ay/ → /e:/         | /saif/        | /se:f/    |
| Inflessioni         | yatakallam(u) | yitkallim |
| Vocabolario         | tawila        | tarabeeza |
| Ordine delle parole | VSO           | SVO       |

Tabella 1: Alcune differenze tra MSA e ECA.

Questi dialetti differiscono considerevolmente sia tra di loro e sia rispetto al MSA dal quale tutti sono egualmente distanti. Le differenze interessano tutti i livelli della lingua, cioè fonetica, fonologica, lessicale, morfologica e di sintassi. La Tabella illustra, ad esempio, alcune differenze tra MSA e un dialetto, ECA<sup>2</sup>.

#### *La struttura della lingua araba*

L'Arabo è una lingua scritta da destra a sinistra, ha un alfabeto di ventotto lettere, con venticinque consonanti e tre lettere restanti che rappresentano le vocali cosiddette “lunghe” (/i:/, /a:/, /u:/) oppure, ove necessario, le semivocali (/y/ e /w/). In Arabo, il termine “vocale lunga” indica una vocale pronunciata con particolare intensità, di solito il termine viene utilizzato in contrapposizione con il termine “vocale breve”, che indica invece un suono vocalico di breve durata e intensità. Un’analogia con questa funzione fonologica della durata, si può riscontrare, ad esempio nella lingua inglese, tra le parole *bit* e *beat* o tra *put* e *book*. Una corrispondenza con la lingua italiana è di più difficile individuazione poiché in essa le vocali di lunga o breve durata, non introducono una differenza fonologica. Per semplificare, si potrebbe identificare un’analogia nella differenza tra vocali accentate e non accentate.

In Arabo, la scrittura delle lettere non è univoca. Ogni grafema può apparire in quattro differenti forme a seconda che sia isolato o che occupi una posizione iniziale, centrale o finale all’interno di una parola. Inoltre, non esiste differenziazione tra lettere maiuscole e minuscole.

Una caratteristica distintiva del sistema arabo di scrittura è che le vocali brevi non sono rappresentate da lettere dell’alfabeto, grafemi veri e propri, ma sono contrassegnate dai cosiddetti segni *diacritici* che vengono scritti sopra o sotto la consonante che precede il suono vocalico. Altri diacritici sono presenti nella lingua Araba e sono utilizzati per contrassegnare altri fenomeni di pronuncia, quali ad esempio il raddoppiamento della consonante o l’assenza di suono vocalico.

<sup>2</sup> ECA: Egyptian Colloquial Arabic

Un elenco delle lettere della lingua Araba sono elencati in Tabella 2, mentre la Tabella 3 riporta i diacritici e le combinazioni di lettere particolari. Fatta esclusione per questi ultimi casi particolari, la corrispondenza tra grafemi e fonemi è uno a uno.

| Isolated | Beginning | Middle | End | Name  | Phoneme |
|----------|-----------|--------|-----|-------|---------|
| ا        | ا         | ا      | ا   | 'alif | /a:/    |
| ب        | ب         | ب      | ب   | baa'  | /b/     |
| ت        | ت         | ت      | ت   | taa'  | /t/     |
| ث        | ث         | ث      | ث   | thaa' | /θ/     |
| ج        | ج         | ج      | ج   | gym   | /dʒ/    |
| ح        | ح         | ح      | ح   | Haa'  | /h/     |
| خ        | خ         | خ      | خ   | khaa' | /x/     |
| د        | د         | د      | د   | daal  | /d/     |
| ذ        | ذ         | ذ      | ذ   | dhaal | /ð/     |
| ز        | ز         | ز      | ز   | zayn  | /z/     |
| ر        | ر         | ر      | ر   | raa   | /r/     |
| س        | س         | س      | س   | syn   | /s/     |
| ش        | ش         | ش      | ش   | shyn  | /ʃ/     |
| ص        | ص         | ص      | ص   | Saad  | /s/     |
| ض        | ض         | ض      | ض   | Daad  | /d/     |
| ط        | ط         | ط      | ط   | Taa'  | /t/     |
| ظ        | ظ         | ظ      | ظ   | Zaa'  | /z/     |
| ع        | ع         | ع      | ع   | 'ayn  | /ʕ/     |
| غ        | غ         | غ      | غ   | ghayn | /ɣ/     |
| ك        | ك         | ك      | ك   | kaaf  | /k/     |
| ق        | ق         | ق      | ق   | qaaf  | /q/     |
| ف        | ف         | ف      | ف   | faa'  | /f/     |
| ل        | ل         | ل      | ل   | laam  | /l/     |
| ن        | ن         | ن      | ن   | nuwn  | /n/     |
| م        | م         | م      | م   | mym   | /m/     |
| ه        | ه         | ه      | ه   | haa'  | /h/     |
| و        | و         | و      | و   | waaw  | /u:/    |
| ي        | ي         | ي      | ي   | yaa'  | /i:/    |
| ء        | أ         | ؤ      | ئ   | hamza | /ʔ/     |

Tabella 2: Lettere dell'alfabeto arabo<sup>3</sup>. Il nome delle lettere è dato secondo la romanizzazione Qalam<sup>4</sup> e i fonemi corrispondenti sono in notazione IPA.

<sup>3</sup> Fonte (Kirchhoff et alii, 2002).

<sup>4</sup> Qualam, <http://eserver.org/langs/qalam.txt>

È da segnalare comunque che i testi in lingua Araba non sono quasi mai diacriticizzati: è prassi utilizzare i diacritici solo nei casi ambigui. La mancanza di segni diacritici può portare a un numero considerevole di ambiguità del lessico. Queste devono essere risolte attraverso informazioni contestuali che presuppongono la conoscenza della lingua; di conseguenza, senza questa conoscenza, è impossibile determinare l'effettiva pronuncia di un testo non vocalizzato.

| Example  | Symbol Name     | Meaning                          |
|----------|-----------------|----------------------------------|
| أ        | fatHa           | /a/                              |
| إ        | kasra           | /i/                              |
| أ        | Damma           | /u/                              |
| ر        | shadda          | consonant doubling               |
| دزس      | sukuwn          | absence of vowel after consonant |
| أ        | tanwyn al-fatHa | /an/                             |
| إ        | tanwyn al-kasr  | /in/                             |
| أ        | tanwyn aD-Damm  | /un/                             |
| ى        | 'alif maqsuwa   | /a:/ sound, historical           |
| هذه      | dagger 'alif    | /a:/ sound, historical           |
| آ        | madda           | double alif                      |
| في البيت | waSla           | on 'alif in <i>al</i>            |
| لا       | laam 'alif      | combination of laam and 'alif    |
| ة        | taa marbuwa     | morphophonemic marker            |

Tabella 3: Diacritici della lingua Araba<sup>5</sup>.

Per fare un esempio facilmente comprensibile, sarebbe come se la parola italiana *stucco*, fosse pronunciata normalmente, ma scritta *stcc*. Si vede immediatamente che essa sarebbe equivalente ad altre parole quali: *stucca*, *stacco*, *stacca*, *stecco*, *stecca*, *etc...*

Nel caso di una frase in italiano, si potrebbe dire che la frase pronunciata come:

*“L'Arabo scritto non è semplicemente leggibile poiché è una lingua senza vocali brevi”*

verrebbe scritta come:

*“L'Arb scritt non è smplcemnt lggibl pché è un ling senz vcal brev”*

Un esempio in lingua Araba, la parola **كتب** (ktb) ha 21 potenziali diacriticizzazioni.

Il discorso appena fatto vale per MSA, se ci si sposta nell'ambito dei dialetti arabi, le cose si possono complicare ulteriormente. Il principale motivo è che i dialetti sono delle lingue quasi esclusivamente orali e quindi non esiste uno standard di scrittura con uniformità di convenzioni.

<sup>5</sup> Fonte (Kirchhoff et alii, 2002).

## 2. ROMANIZZAZIONE AUTOMATICA

Dal momento che i grafemi arabi non sono facilmente comprensibili al lettore non esperto e che la loro elaborazione, tramite processi automatizzati di elaborazione dei dati, è piuttosto complicata, quando si studia la lingua araba, nasce presto l'esigenza di tradurre i grafemi con dei simboli romanizzati. Questi tentativi di semplificare la scrittura araba si chiamano *traslitterazioni* o *trascrizioni*.

Con il termine *traslitterazioni* s'intendono tutti quei sistemi di romanizzazione dei grafemi arabi che hanno come scopo mantenere una corrispondenza uno a uno tra i caratteri arabi e l'insieme dei simboli romanizzati (normalmente ASCII standard). Molto meno importanza è data alla coerenza tra simbolo usato e l'effettivo suono che dovrebbe essere pronunciato. Compaiono quindi anche simboli come i segni di interpunzione o le parentesi.

Allo scopo di mantenere la corrispondenza tra suono e simbolo sono utilizzate le *trascrizioni*. Queste possono essere di due tipi:

- *fonemico*, che rende il lettore che non ha familiarità con la lingua originale in grado di pronunciarla con una ragionevole accuratezza;
- *fonetico*, in cui si cerca di essere ancora più precisi nel descrivere, quanto più esattamente possibile il suono che viene pronunciato nella lingua originale, utilizzando simboli fonetici (ad esempio IPA).

Un esempio di romanizzazione di una parola araba è riportato in Tabella 4 e un elenco dei principali sistemi di romanizzazione sono illustrati in Tabella 5. Il metodo che comunemente viene utilizzato in questo ambito è Buckwalter (Buckwalter).

|                              |                                     |                                                                              |
|------------------------------|-------------------------------------|------------------------------------------------------------------------------|
| <b>Arabico</b>               | القاهرة                             | [originale]                                                                  |
| <b>Traslitterazione</b>      | AlqAhrp                             | [buckwalter]                                                                 |
| <b>Trascrizione Fonemica</b> | il Cairo<br>al-Qahira<br>al-Qāhirah | [Trascrizione Primaria]<br>[Trascrizione Standard]<br>[Trascrizione Diretta] |
| <b>Trascrizione Fonetica</b> | al kæhirah                          | [IPA]                                                                        |

Tabella 4: Esempi di romanizzazione della lingua Araba.

I fonemi della lingua araba sono in tutto trentuno, quelli illustrati in Tabella 2 più le tre vocali brevi (a, i, u) che sono la controparte delle vocali lunghe (a:, i:, u:); è da notare che la lunghezza delle vocali è fonemica.

| letter | Qalam | CAT     | LDC | Karboul | Buckwalter |
|--------|-------|---------|-----|---------|------------|
| 'alif  | aa    | aa      | aa  | A       | A          |
| baa'   | b     | b       | b   | B       | b          |
| taa'   | t     | t       | t   | TH      | t          |
| thaa'  | th    | ĉ/th   | th  | Tt      | v          |
| gym    | j     | j       | j   | J       | j          |
| Haa'   | H     | h/H     | h   | H       | H          |
| khaa'  | kh    | k/kh/K  | kh  | Kk      | x          |
| daal'  | d     | d       | d   | D       | d          |
| dhaal' | dh    | zzh     | dh  | Dd      | P          |
| raa'   | r     | r       | r   | R       | r          |
| zayn   | z     | z       | z   | Z       | z          |
| syn    | s     | s       | s   | S       | s          |
| shyn   | sh    | ŝ/sh   | sh  | Sc      | \$         |
| Saad   | S     | s/S     | S   | Ss      | S          |
| Daad   | D     | d/D     | D   | Sd      | D          |
| Taa'   | T     | t/T     | T   | Td      | T          |
| Zaa'   | Z     | z/Z     | Z   | Dt      | Z          |
| 'ayn   | '     | @       | c   | Ar      | E          |
| ghayn  | gh    | ĝ/gh   | R   | G       | g          |
| faa'   | f     | f       | f   | F       | f          |
| qaaf   | q     | q       | q   | Q       | q          |
| kaaf   | k     | k       | k   | K       | k          |
| laam   | l     | l       | l   | L       | l          |
| mym    | m     | m       | m   | M       | m          |
| nuwn   | n     | n       | n   | N       | n          |
| haa    | h     | h       | h   | h       | h          |
| waaw   | w     | w/uu/oo | w   | W       | w          |
| yaa'   | y     | y/ii    | y   | Y       | y          |
| hamza  | '     | '       | C   | unclear | '          |

Tabella 5: Vari schemi di traslitterazione per l'Arabo<sup>6</sup>.

<sup>6</sup> Fonte (Kirchhoff et alii, 2002).



### 3. DIFFICOLTÀ NELLA CREAZIONE DEI SISTEMI AUTOMATICI DI RICONOSCIMENTO

Da queste brevi note introduttive appare evidente come le peculiarità di questa lingua possano tradursi in grosse difficoltà qualora si decida di creare un sistema di riconoscimento automatico (ASR<sup>7</sup>).

Le caratteristiche fonetiche sono modellabili, come per le lingue maggiormente studiate, attraverso modelli statistici complessi quali le *catene di Markov nascoste dipendenti dal contesto* (CDHMM<sup>8</sup>).

Il passaggio che rende veramente complicata la creazione di un ASR per l'Arabo è la raccolta dell'insieme di dati necessari per l'addestramento del sistema. Questi dati sono costituiti da file audio e da file di testo che riportano quanto più fedelmente possibile ciò che viene pronunciato. Il reperimento di questo materiale trascritto e, in particolare, della sua versione diacriticizzata è molto complicato e oneroso. Molto spesso è necessario accontentarsi di trascrizioni non complete oppure bisogna provvedere a trascrivere manualmente ex-novo i file audio con l'aiuto di operatori madrelingua.

Se l'obiettivo poi è creare un ASR in grado di riconoscere il parlato continuo, spontaneo e basato su un vocabolario abbastanza grande, bisogna tener conto della presenza nel parlato spontaneo arabo delle forme dialettali. Questa variabilità aumenta enormemente la complessità del problema.

Un'ulteriore complicazione deriva dalla creazione del modello del linguaggio (LM). Infatti, l'assenza di una trascrizione diacriticizzata, con la conseguente equivalenza ortografica di molte forme in contesti diversi, determina un aumento dell'incertezza con cui le parole vengono contestualizzate. Questo implica una sostanziale diminuzione del potere discriminativi per i modelli del linguaggio. I LM non diacriticizzati saranno quindi molto meno predittivi di quelli diacriticizzati. Per quanto riguarda invece il caso di parlato spontaneo conversazionale si deve considerare la presenza, nel materiale da riconoscere, di molte parole non inserite nel vocabolario (OOV - Out Of Vocabulary).

| Tipo di ASR                        | Nome ASR                                                        | Performance (WER) |
|------------------------------------|-----------------------------------------------------------------|-------------------|
| Dettatura                          | IBM ViaVoice for Arabic                                         | -                 |
| Broadcast                          | BBN Arabic Broadcast News Recognizer                            | 15.3%             |
| Broadcast                          | Arabic GALE transcription system (Soltau et alii, 2007)         | 18.3%             |
| Parlato Conversazionale Dialettale | 1996/1997 NIST CallHome Evaluations (Zavagliakos et alii, 1997) | 55.8%             |

Tabella 6: Alcuni dati sulla qualità del riconoscimento per la lingua Araba presenti in letteratura (Kirchhoff et alii, 2002).

<sup>7</sup> ASR: Automatic Speech Recognition.

<sup>8</sup> CDHMM: Context Dependent Hidden Markov Model.

Le difficoltà di creazione dei modelli acustici e dei LM si traduce perciò in una perdita di accuratezza del ASR e, a parità di condizioni, in un livello di errore sulla parola maggiore che in altre lingue.

#### *Stato dell'arte negli ASR per l'Arabo*

In letteratura si trovano numerosi studi su ASR che si focalizzano, in virtù di quanto detto in precedenza, sulla lingua MSA, tralasciando le componenti dialettali. Molte applicazioni derivano dallo sviluppo di sistemi per la dettatura automatica, come, ad esempio, il sistema IBM ViaVoice.

Un breve panorama dello stato dell'arte nel campo del riconoscimento automatico della lingua Araba è illustrato in Tabella 6.

#### **4. IL SISTEMA DI ASR**

Lo scopo di questo lavoro è quello di analizzare le problematiche e descrivere le soluzioni che permettono lo sviluppo di un sistema di riconoscimento per la lingua Araba.

Si è proceduto quindi ad analizzare ed eseguire tutti i passaggi concernenti la realizzazione del progetto.

Sono state individuate tre fasi:

- *pre-trattamento del testo*: trasformazione delle trascrizioni, secondo i criteri precedentemente enunciati, per romanizzare e fonetizzare i lemmi in lingua Araba;
- *sviluppo* di un'architettura per l'addestramento ed il test di un sistema ASR, che permetta di sfruttare due motori di riconoscimento (SONIC<sup>9</sup> e SPHINX<sup>10</sup>);
- *valutazione* dei risultati ottenuti.

#### *Corpus*

Come punto di partenza per lo sviluppo del nostro progetto, abbiamo scelto la lingua Araba Standard, il MSA. Il relativo materiale vocale è stato ricavato dal corpus, denominato “*West Point Arabic Speech Corpus*” (WPA), distribuito dal LDC (Linguistic Data Consortium<sup>11</sup>).

Il corpus è composto da 8516 file audio, per un totale di circa 12 ore di parlato. Il segnale audio è stato registrato a 22,05 kHz e poi convertito in formato *NIST sphere*<sup>12</sup>. È costituito da parlato controllato, microfonico, letto in ambiente poco rumoroso. Le frasi e i lemmi pronunciati sono limitati (258 frasi e 1130 lemmi), ma l'audio è interamente e dettagliatamente trascritto. Le trascrizioni sono romanizzate con il sistema ArabTex e sono, inoltre, interamente diacriticizzate.

Con il corpus è fornito anche un dizionario con una trascrizione fonetica dei lemmi molto precisa.

I parlatori sono principalmente di madrelingua araba, ma è presente anche una piccola porzione di non madrelingua:

- 7200 delle registrazioni sono state pronunciate da madrelingua, per un totale di 10,5 ore
- 1200 file sono stati prodotti da parlatori non madrelingua, per un totale di poco più di 1 ora.

---

<sup>9</sup> [http://cslr.colorado.edu/beginweb/speech\\_recognition/sonic.html](http://cslr.colorado.edu/beginweb/speech_recognition/sonic.html)

<sup>10</sup> <http://cmusphinx.sourceforge.net>

<sup>11</sup> <http://www ldc.upenn.edu/>

<sup>12</sup> <http://www.nist.gov>

### *Preparazione del corpus*

Il materiale audio è stato adeguato agli standard richiesti per la realizzazione del modello acustico. Sono stati, prima di tutto, convertiti dal formato NIST sphere al normale PCM a 16 bit e successivamente sono stati sotto campionati da 22.05 kHz a 8kHz. Da segnalare che sono stati eliminati i file audio in cui le frasi da pronunciare sono state prodotte in modo incompleto o con errori macroscopici di riproduzione.

Le trascrizioni sono state convertite dalla codifica ArabTex alla romanizzazione Buckwalter per poter ottenere una corrispondenza biunivoca tra grafema arabo e simbolo romanizzato, peculiarità che l'originale sistema di trascrizione non garantisce.

Da queste trascrizioni, inoltre, sono state estratte le singole parole, in modo da creare un dizionario di lemmi ammissibili che costituisce l'insieme di parole che il sistema ASR può riconoscere. Per ogni parola del dizionario, è stata fornita una descrizione fonemica della pronuncia, ricavata da quella originale presente nel corpus.

Infine si è provveduto ad isolare singoli eventi non testuali, detti *filler*, che servono a descrivere le parti di audio che non possono essere catalogate come parlato. Un esempio di eventi extratestuali sono eventi come: risata, rumore, musica o il semplice silenzio.

### *Architettura del sistema di riconoscimento.*

Sono stati utilizzati due sistemi di riconoscimento automatico del parlato: SONIC ver. 2 beta 3 e CMU SPHINX 3 ver. 0.7. Sono stati scelti questi due sistemi perché sono alcuni tra i software che meglio rappresentano lo stato dell'arte nel campo del riconoscimento automatico del parlato, tra quelli che permettono il libero utilizzo per scopi di ricerca.

I due sistemi si basano entrambi sul riconoscimento di sequenze di fonemi tramite l'utilizzo di un Modello Acustico statistico (AM) basato su CDHMM (Context Dependent Hidden Markov Models) e di un Modello del Linguaggio a trigrammi (LM), basato su normali catene di Markov.

Il risultato del riconoscimento è la successione di parole che ha la probabilità più alta di essere stata pronunciata. Questa si ricava calcolando, tramite l'algoritmo di Viterbi, quale sia il percorso con punteggio più alto tra tutti quelli possibili, all'interno di una struttura detta *lattice* (reticolo o traliccio). Questa contiene tutti le possibili successioni fonetiche e, quindi, di parole che, con probabilità superiore ad un certo valore di soglia, possono essere state pronunciate nel file audio analizzato. Essa viene creata attraverso i valori risultanti dalla combinazione delle probabilità provenienti dal AM e dal LM.

Nonostante le numerose caratteristiche comuni, i due sistemi differiscono parecchio (Pellom & Hacıoğlu, 2003; Chan et alii, 2007). Le principali differenze sono:

- *Estrazione delle features.* Le features sono dei vettori di coefficienti che identificano l'andamento dell'involuppo spettrale del segnale audio da riconoscere. SONIC utilizza un metodo di calcolo di questi vettori più sofisticato rispetto a SPHINX. Quest'ultimo, infatti, utilizza i comuni MFCC (Mel Frequency Cepstrum Coefficients), mentre il primo adotta i coefficienti PMVDR (Perceptual Minimum Variance Distortionless Response). Il calcolo di questo tipo di vettori prevede un prefiltraggio del segnale per meglio modellare la curva percettiva dell'orecchio umano rispetto al sistema basato sulla scala Mel.
- *Adattamento del riconoscimento.* Nei processi di ASR, dopo un primo passaggio di riconoscimento, detto *baseline*, si utilizzano le informazioni così ricavate per reiterare il processo applicando delle tecniche di adattamento del modello statistico al segnale audio in esame. Ci sono varie strategie che possono essere applicate. In

SONIC sono implementate il MLLR<sup>13</sup>, il VTLN<sup>14</sup> e SAT<sup>15</sup>, mentre SPHINX permette solamente di utilizzare il metodo MLLR e la stima dei parametri MAP<sup>16</sup>. Questo si traduce, dopo l'adattamento, in un miglioramento delle prestazioni di SONIC nei confronti di quelle di SPHINX anche quando le percentuali della baseline sono sfavorevoli.

Il processo di riconoscimento in un ASR è composto di varie fasi, illustrate in Figura 2:

1. il file audio da riconoscere viene preelaborato per renderlo compatibile con il modello statistico creato, cioè PCM a 16 bit, campionato a 8kHz;
2. da questo file trasformato vengono estratte i vettori di features (MFCC o PMVDR) che saranno la base per il riconoscimento vero e proprio;
3. con le informazioni statistiche ricavate dal modello acustico e dal modello del linguaggio si cerca di catalogare i vettori provenienti dal passaggio precedente per associarli ai fonemi della lingua. Dalla successione di tali fonemi tramite il dizionario dato, si risalirà alle parole pronunciate;
4. Il risultato di questo processo è la sequenza di parole che con maggior probabilità, secondo i modelli dati, corrisponde all'audio in ingresso. Normalmente l'output è composto di una sola ipotesi; all'occorrenza, è possibile estrarre anche le N migliori ipotesi (*N-best*) o addirittura l'intero *lattice* di riconoscimento. Quest'ultimo può essere usato come ingresso per ulteriori passaggi basati su altre strategie di riconoscimento più ad alto livello (ad esempio un riconoscimento semantico, o le *confusion networks*, (Bertoldi & Federico, 2005).

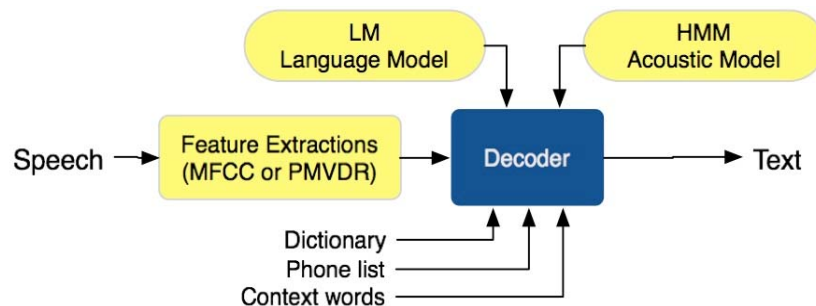


Figura 2: schema del sistema di riconoscimento di un file audio, comune ai software SONIC e SPHINX.

#### Architettura del sistema di addestramento

Per poter utilizzare la struttura ASR precedentemente illustrata, è necessario creare i modelli acustici e linguistici adatti. Questa è la parte più delicata di tutto il processo e anche quella che permette più margini di sviluppo.

I passaggi necessari per la creazione del modello acustico sono illustrati in Figura 3. L'input del sistema è costituito, come detto in precedenza, da un insieme di audio

<sup>13</sup> MLLR: Maximum Likelihood Linear Regression, adattamento senza supervisione.

<sup>14</sup> VTLN: Vocal Tract Length Normalization, normalizzazione rispetto alla lunghezza del tratto vocale del parlatore.

<sup>15</sup> SAT: Speaker Adaptive Training, adattamento ad un parlatore specifico

<sup>16</sup> MAP: Maximum A Posterior, adattamento con supervisione.

omogeneamente organizzato e dalle relative trascrizioni. Normalmente, è necessaria una ingente mole di dati, oltre le 10 ore di parlato, per ottenere un modello abbastanza stabile.

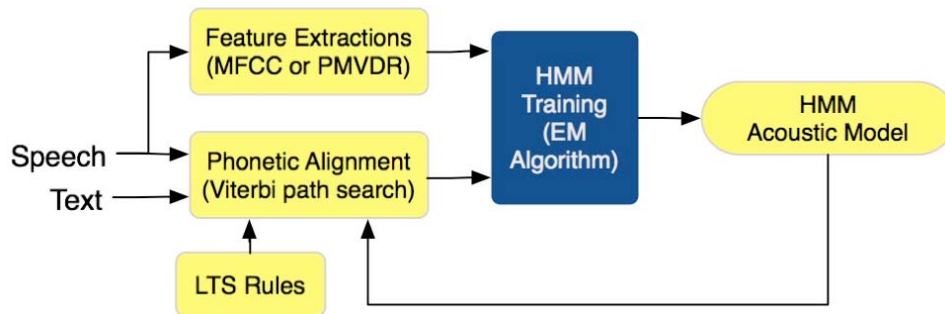


Figura 3: schema del sistema di addestramento comune ai software SONIC e SPHINX.

L'audio e le trascrizioni costituiscono quindi l'input del primo passaggio dell'addestramento: il cosiddetto *allineamento*. In questa fase il sistema, sulla base delle informazioni provenienti dalle trascrizioni, unite con quelle provenienti dal dizionario fonetizzato, cerca di far corrispondere, minimizzando la probabilità di errore, ogni segmento di file audio ad un fonema.

Anche in questo caso vi è una differenza tra SONIC e SPHINX. Il primo crea questa segmentazione sulla base anche di informazioni ricavate dal modello acustico di un'altra lingua (di solito l'inglese) e di corrispondenze tra i fonemi delle due lingue, che vengono decise manualmente in fase di preparazione della procedura di addestramento. Utilizzando il secondo sistema, invece, si parte da una segmentazione uniforme del file audio di ingresso e con successivi passaggi di allineamento basati sui modelli acustici così ottenuti, viene perfezionata la segmentazione.

Si è potuto constatare che, nel nostro caso, le segmentazioni risultanti sono grossomodo equivalenti.

Per quanto riguarda la creazione del modello del linguaggio è stato utilizzato un *toolkit* sviluppato presso la Carnegie Mellon University, il CMU-Cambridge Statistical Language Modeling toolkit<sup>17</sup>. Con esso è possibile estrarre, da grandi moli di testo scritto, le informazioni statistiche che permettono di creare la struttura a trigrammi, bigrammi e unigrammi che il sistema ASR utilizzerà per il riconoscimento. Nel nostro caso abbiamo utilizzato per creare il modello del linguaggio l'insieme delle trascrizioni di addestramento.

Questo crea un LM abbastanza polarizzato verso il tipo di conversazioni che dovranno essere riconosciute in fase di test.

## 5. RISULTATI SPERIMENTALI

Le performance del sistema con queste specifiche sono molto elevate, si ha infatti meno del 2% di errore di riconoscimento sulla singola parola (WER) e sono in linea con i risultati ottenuti in altri laboratori. In (Alotaibi et alii, 2007) le prestazioni di un sistema di riconoscimento, bastato sul sistema HTK (Young et alii, 2005) e addestrato su parlatori madrelingua del WPA, che processa file provenienti da madrelingua, utilizzando un

<sup>17</sup> [http://www.speech.cs.cmu.edu/SLM\\_info.html](http://www.speech.cs.cmu.edu/SLM_info.html)

modello del linguaggio a bigrammi, raggiunge una correttezza del 99,05%. Mentre lo stesso sistema allenato anche con parlatori non madrelingua riconosce l'audio dei madrelingua con una correttezza del 93,98%.

I nostri test sono stati svolti utilizzando diverse configurazioni:

- con entrambi i sistemi di riconoscimento;
- con l'addestramento fatto su tipologie diverse di parlatore:
  - solo parlatori di madrelingua araba
  - sia parlatori madrelingua che non madrelingua;
- modificando i vari parametri di regolazione dei sistemi: ampiezza del *lattice*, peso del LM, etc;

I test sono stati effettuati sullo stesso insieme di file di test: circa mezz'ora di parlato continuo controllato, pronunciato da 5 parlatori diversi di madrelingua araba, proveniente dal WPA, ma non utilizzato per l'addestramento dei modelli.

I risultati ottenuti sono stati tutti soddisfacenti e non si differenziano tra loro in modo eclatante. Per chiarezza, illustriamo i due migliori risultati ottenuti per i due diversi software di riconoscimento denominati test A e test B.

La Figura 4 e la Tabella 7 mostrano i risultati finali del riconoscimento.

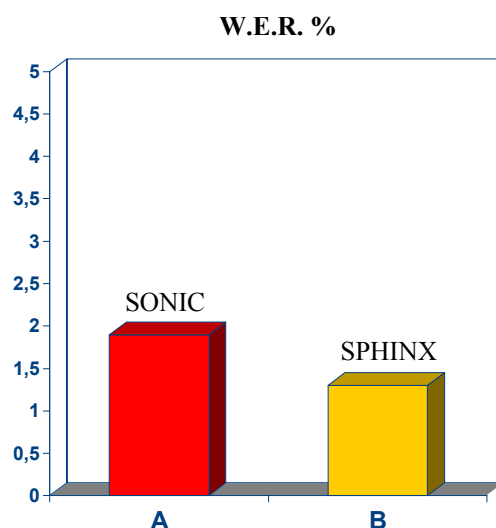


Figura 4: Grafico di confronto tra la percentuale di Word Error Rate (WER) per il sistema di riconoscimento basato su SONIC (A) e su SPHINX (B)

|   | Correttezza | Sostituzione | Cancellazione | Inserimento | WER   |
|---|-------------|--------------|---------------|-------------|-------|
| A | 98.4 %      | 0.9 %        | 0.7 %         | 0.3 %       | 1.9 % |
| B | 98.9 %      | 0.5 %        | 0.6 %         | 0.2 %       | 1.3 % |

Tabella 7: Percentuali di riconoscimento sull'insieme di file audio di test.

Il test A è stato svolto facendo riconoscere i file di test dal sistema ASR SONIC allenato con un insieme di file audio di soli parlatori madrelingua. Per il test B, invece è stato utilizzato il riconoscitore SPHINX addestrato con tutto il materiale del corpus WPA meno i file di test.

Le elevate prestazioni di questi test di riconoscimento sono spiegate facilmente attraverso l'analisi del problema che stiamo affrontando. Si tratta infatti di:

- un riconoscimento di parlato controllato, microfonico e non rumoroso, quindi contenente pochi elementi difficilmente catalogabili dal AM;
- il vocabolario è interamente noto e limitato (1130 lemmi), quindi l'ambito semantico è molto focalizzato e l'incertezza sui vocaboli è molto bassa;
- il modello del linguaggio è anch'esso molto semplice e al suo interno presenta molti schemi che si trovano anche nelle conversazioni di test.

Un elemento interessante è che il sistema non viene per nulla peggiorato se si introducono, in fase di addestramento, anche dei "disturbi" derivanti da pronunce non perfette di parlatori non madrelingua. Questo indica che il modello del linguaggio, quando è molto focalizzato, è in grado di supplire alle carenze di modellazione acustica dei fonemi.

## 6. CONCLUSIONI E SVILUPPI FUTURI

Per quanto riguarda i possibili sviluppi futuri, è nostra intenzione considerare l'estensione di quanto sperimentato in questo contesto semplificato in un ambito più complesso come il parlato arabo spontaneo e conversazionale.

Il riconoscimento automatico in queste condizioni rappresenta un *task* molto più difficile. Il vocabolario utilizzato è molto più ampio, di solito è nell'ordine delle 40.000 parole. La costruzione della frase non è sempre regolare, sono presenti false partenze, interruzioni e ripetizioni. La qualità del segnale audio di solito è peggiorata dalla presenza di disturbi quali il rumore di fondo, altre voci o espressioni extra testuali.

Altro problema poi sta nel reperimento del materiale per creare l'allenamento dei modelli statistici. Generalmente questi corpora sono forniti esclusivamente con trascrizioni non diacriticizzate e non contengono informazioni sulle vocali brevi. Di conseguenza in una prima fase, per l'addestramento del sistema, sarà necessario basarci soltanto sull'informazione ortografica non vocalizzata, su cui potremo applicare eventualmente semplici regole di sostituzione grafema-fonema.

Nonostante l'oggettiva difficoltà di addestrare un sistema di riconoscimento sulla base delle sole trascrizioni ortografiche invece delle complete trascrizioni fonetiche, i risultati ottenuti in letteratura sono molto promettenti e consentono ampi margini di miglioramento soprattutto considerando il fatto che qualora si riesca a disporre della diacriticizzazione o della trascrizione fonetica precisa di una buona parte del corpus l'addestramento del sistema risulterà sicuramente più preciso ed affidabile.

## RINGRAZIAMENTI

Parte di questo lavoro è stato possibile grazie ad una collaborazione con la ditta Expert System Spa<sup>18</sup> di Modena.

---

<sup>18</sup> [www.expertsystem.it](http://www.expertsystem.it)

## BIBLIOGRAFIA

Alotaibi, Y. A., Selouani, S., O'Shaughnessy, D., (2007) *Experiments on Automatic Recognition of Nonnative Arabic Speech*, accepted for EURASIP Journal on Audio, Speech, and Music Processing, Hindawi Publishing Corporation

Chan, A., Gouvea, E., Singh, R., Ravishankar, M., Rosenfeld, R., Sun, Y., Huggins-Daines, D., Seltzer, M., (2007) *The Hieroglyphs: Building Speech Applications Using CMU Sphinx and Related Resources*, March 11, 2007, <http://www.cs.cmu.edu/~archan/>.

Bertoldi, N., Federico, M., (2005) *A new decoder for spoken language translation based on confusion networks*, Automatic Speech Recognition and Understanding, IEEE Workshop, 27 Nov-1 Dec, 86-91.

Buckwalter, <http://www.xrce.xerox.com/competencies/content-analysis/arabic/info/buckwalter-about.html>; <http://www.qamus.org/transliteration.htm>.

Kirchhoff, K., Bilmes, J., Henderson, J., Schwartz, R., Noamany, M., Schone, P., Ji, G., Das, S., Egan, M., He, F., Vergyri, D., Liu, D., Duta, N., (2002) *Novel Speech Recognition Models for Arabic*, Final Report, Johns-Hopkins University Summer Research Workshop, <http://www.clsp.jhu.edu/ws02>.

Kirchhoff, K., Vergyri, D., Bilmes, J., Duh, K., Stolcke, A., (2004), *Morphology-based language modeling for conversational Arabic speech recognition*, INTERSPEECH - ICSLP, International Conference on Spoken Language Processing, Korea, October 4-8, 2245-2248.

Pellom, B., Hacıoğlu, K. (2003) *Sonic: The University of Colorado Continuous Speech Recognizer - Technical Report TR-CSLR-2001-01*, Center for Spoken Language Research University of Colorado, Boulder.

Soltau, H., Saon, G., Kingsbury, B., Kuo, J.; Mangu, L., Povey, D., Zweig, G., (2007), The IBM 2006 Gale Arabic ASR System, *Acoustics, Speech and Signal Processing 2007. ICASSP 2007*, Volume 4, April 15-20, 349 - 352.

Young, S., Evermann, G., Gales, M. et alii, (2005) *The HTK Book (for HTK Version. 3.3)*, Cambridge University Engineering Department, <http://htk.eng.cam.ac.uk/>

Zavaliagos, G., McDonough, J., Miller, D., El-Jaroudi, A., Billa, J., Richardson, F., Ma, K., Siu, M., Gish, H. (1998), The BBN Byblos 1997 large vocabulary conversational speech recognition system. *Acoustics, Speech and Signal Processing 1997. Proceedings of the IEEE International Conference*, Volume 2, May 12-15, 905-908.



# IMPLEMENTAZIONE DI UN ALGORITMO DI BSS IN FREQUENZA PER LA SEPARAZIONE DI SORGENTI ACUSTICHE

Francesco Nesta<sup>1,2</sup>, Maurizio Omologo<sup>1</sup>, Piergiorgio Svaizer<sup>1</sup>

1) Fondazione Bruno Kessler - IRST, 2) Università di Trento

*{nesta,omologo,svaizer}@fbk.eu*

## 1. INTRODUZIONE

Negli ultimi dieci anni notevole interesse ha ottenuto un emergente ramo di ricerca dell'elaborazione numerica di segnali, la Blind Source Separation (BSS)[13]. Le tecniche di BSS, infatti, sono diventate in questi ultimi anni molto robuste e in un prossimo futuro potranno incrementare sostanzialmente le prestazioni di sistemi reali di riconoscimento automatico del parlato. Lo scopo della BSS è di separare diverse sorgenti avendo come un'unica informazione le misture dei segnali osservate mediante dei sensori. Con il termine "Blind" s'intende la capacità del sistema di effettuare la separazione senza avere nessuna conoscenza a priori sul processo di mixing o sulla natura dei segnali. Le tecniche di BSS sono applicate con successo da diversi anni nel campo della neurologia [7][8] mentre solo recentemente gli algoritmi di separazione per segnali acustici sono diventati sufficientemente robusti per essere usati in ambienti reali.

Nel caso delle applicazioni per segnali neurologici le tecniche di BSS sono basate su algoritmi di Independent Component Analysis (ICA) usati per separare segnali elettrici dell'attività cerebrale ottenuti mediante encefalogramma (EEG). Gli algoritmi di ICA si basano sull'assunzione che le sorgenti siano statisticamente indipendenti e sfruttano quest'ulteriore informazione per stimare un processo di separazione che permetta di ottenere le componenti di segnale indipendenti.

Nelle applicazioni audio un sistema di BSS molto efficace, utilizzato spesso come termine di paragone, è l'apparato uditivo periferico umano e per questo motivo tali tecniche sono inizialmente nate con lo scopo risolvere la separazione nel campo dei segnali acustici. L'uomo, infatti, è capace di discriminare più sorgenti sonore presenti in un ambiente, anche in condizioni molto avverse. Questa capacità è spesso referenziata con l'espressione "Cocktail Party Effect".

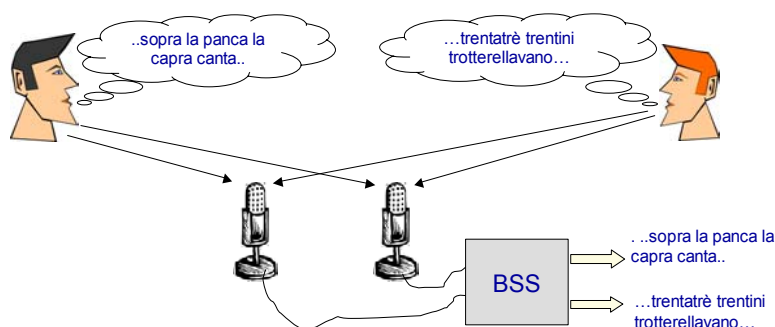


Figura 1: BSS per separazione di sorgenti acustiche

Un sistema di BSS può essere implementato attraverso differenti approcci. Principalmente è possibile individuare due categorie: 1) BSS ad un canale, 2) BSS multicanale.

Nel primo caso le sorgenti sono separate utilizzando come informazione le osservazioni di una sola mistura di segnali ottenuta mediante un sensore. Diverse metodologie sono state esplorate come l'Independent Subspace Analysis (ISA)[5] e la Non-Negative Matrix Factorization (NMF), e combinazioni ibride di differenti metodi per ottenere risultati più robusti. Non tutti questi metodi in realtà sono completamente “blind” poiché le diverse tecniche formulano ipotesi sulla struttura dei segnali di partenza assumendo ad esempio ridondanza di informazione nello spettro. La BSS mono canale può essere considerata allo stato attuale il ramo di ricerca più ambizioso della BSS e nonostante i recenti progressi l'affidabilità e la robustezza di questi metodi sono ancora insufficienti perché possano essere adottati in applicazioni reali.

Nel caso di BSS multicanale le sorgenti di segnale sono osservate utilizzando una schiera di sensori. Tale approccio permette di introdurre ulteriori osservazioni utili alla stima del processo di separazione. In particolare nel caso di BSS multicanale per applicazioni audio si utilizzano implicitamente informazioni di indipendenza spaziale delle sorgenti. Nonostante la BSS multicanale richieda ancora un elevato costo computazionale, i recenti progressi hanno portato a metodi molto affidabili e sufficientemente veloci perché possano essere implementati in real-time.

Comunque anche in questo caso esistono ancora dei problemi, quali ad esempio la separazione di sorgenti spazialmente molto vicine o quando il numero di sorgenti è superiore a quello dei sensori (caso “underdetermined”). Inoltre il tempo di riverberazione è ancora molto influente sulla qualità complessiva del segnale separato.

Oltre alla sopracitata esiste un'ulteriore possibile classificazione: 1) algoritmi per misture istantanee 2) algoritmi per misture convolutive.

Nel primo caso i segnali osservati ai microfoni (nel caso audio) possono essere visti come una combinazione lineare dei segnali originali delle sorgenti. In queste condizioni gli algoritmi di ICA sono sufficienti a separare efficacemente le sorgenti. Nel secondo caso invece i segnali osservati ai microfoni sono una versione dei segnali originali convoluta con la risposta impulsiva dell'ambiente. Questo è il caso di un sistema che cerchi di separare sorgenti acustiche in un ambiente riverberante.

In questo lavoro ci occupiamo di sviluppare un algoritmo di BSS utilizzabile in ambienti reali anche in condizioni di riverberazione molto avverse. Un approccio molto comune, lo stesso da noi adottato, è di eseguire la separazione nel dominio della frequenza. In questo caso, infatti, è possibile ricondurre il problema della separazione di misture convolutive a misture istantanee, permettendo quindi l'utilizzo delle efficienti tecniche di ICA. La separazione dei segnali viene effettuata in modo indipendente per ogni frequenza e i segnali sono successivamente ricostruiti nel dominio del tempo. In questo lavoro proponiamo in particolare un nuovo approccio per risolvere il noto problema delle “permutazioni”, tipico dei metodi basati sul dominio della frequenza.

Prima di entrare nei dettagli della tecnica di BSS proposta, nella sezione 2 e 3 formuliamo il problema di separazione rispettivamente per il caso di misture istantanee e convolutive. Nella sezione 4 viene affrontato il problema dell'ambiguità della soluzione di ICA e di seguito un nuovo schema per la soluzione delle permutazione è presentato nella sezione 5. Infine risultati sperimentali ottenuti su dati reali sono presentati nell'ultima sezione e confermano la validità dell'approccio proposto.

## 2. BSS PER MISTURE ISTANTANEE E ICA

Consideriamo un vettore  $\mathbf{x}(t)$  di  $N$  elementi le cui componenti  $x_i(t)$  rappresentano il valore assunto dal segnale  $i$ -esimo all'istante  $t$ . Se supponiamo che i segnali siano miscelati istantaneamente possiamo indicare con  $\mathbf{H}(t)$  la matrice di mixing  $M \times N$  e con  $\mathbf{y}(t)$  il vettore dei segnali miscelati captati da  $M$  microfoni. Se il sistema di mixing è stazionario, possiamo considerare la matrice di mixing costante e i segnali miscelati  $y_i(t)$  sono rappresentabili come una combinazione lineare dei segnali originali delle sorgenti. Per semplificare la trattazione supponiamo che  $N=M=2$ :

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad (1)$$

Il problema di separazione è ricondotto ad una semplice soluzione del sistema lineare attraverso il calcolo della matrice di separazione  $\mathbf{W}$ , inversa di quella di mixing  $\mathbf{H}$ . Si può subito osservare che perché sia possibile l'inversione,  $\mathbf{H}$  non deve essere singolare. Nel caso semplice in cui  $M = N$  l'inversione è matematicamente possibile e quindi il calcolo di si traduce nell'effettuare una stima della matrice  $\mathbf{H}$ . In ogni modo nel caso "blind" i segnali originali non sono conosciuti a priori e quindi non esiste una soluzione analitica del problema. Infatti si può osservare che il numero di incognite è superiore a quello delle equazioni. La matrice  $\mathbf{W}$  quindi può essere solo stimata utilizzando un processo iterativo che usi un'informazione aggiuntiva.

L'approccio più comune per la separazione di misture istantanee è quello di supporre che le sorgenti siano statisticamente indipendenti. In questo caso la matrice  $\mathbf{W}$  è stimata in modo da massimizzare l'indipendenza statistica tra segnali di output associati alle sorgenti. Questo principio è sfruttato dai metodi di ICA ormai largamente usati nel campo della neurologia per l'analisi, per la decomposizione di segnali elettrici associati all'attività cerebrale.

Per la stima di  $\mathbf{W}$  è quindi necessaria una metrica che valuti la dipendenza statistica tra i segnali. Se consideriamo le osservazioni dei segnali  $y_i(t)$  come realizzazioni di una variabile aleatoria possiamo considerare i segnali solo da un punto di vista statistico, perdendo l'informazione temporale. I segnali osservati possono essere considerati come una combinazione lineare di variabili aleatorie associate ai segnali delle sorgenti.

$$\mathbf{y} = \mathbf{H}\mathbf{x} \quad (2)$$

Supponendo che le sorgenti non siano gaussiane possiamo usare come misura di dipendenza la gaussianità delle componenti ottenute, poiché secondo il teorema del limite centrale la somma di variabili aleatorie indipendenti tende asintoticamente ad assumere una densità di probabilità gaussiana. La matrice  $\mathbf{W}$  può quindi essere stimata attraverso un approccio a gradiente seguendo la direzione che minimizza la gaussianità dei segnali delle sorgenti. La gaussianità può essere valutata mediante misure di kurtosis, negentropy e in generale attraverso statistiche di ordine superiore.

Un altro approccio analogo, proveniente dalla teoria dell'informazione, è quello di utilizzare la distanza statistica tra le sorgenti attraverso il calcolo della mutua informazione.

Algoritmi strettamente legati a questa misura sono l'InfoMax[3] e le sue recenti varianti come lo Scaled Natural Gradient[4]. Ad ogni iterazione (*iter*) un algoritmo di ICA basato sul gradiente naturale stima la matrice di separazione secondo la seguente ricorsione:

$$\mathbf{W}_{iter} \longleftarrow \mathbf{W}_{iter-1} + \eta[\mathbf{I} - \langle \alpha(\bar{\mathbf{x}})\bar{\mathbf{x}}^H \rangle_{\tau}] \mathbf{W}_{iter-1} \quad (3)$$

dove  $\langle \cdot \rangle_{\tau}$  è l'operatore di media su  $\tau$ ,  $\mathbf{I}$  è la matrice unitaria,  $\bar{\mathbf{x}}$  rappresenta il segnale *i-esimo* ottenuto separando le misture  $\mathbf{y}$  con la matrice  $\mathbf{W}$  ottenuta al passo di iterazione precedente:

$$\bar{\mathbf{x}} = \mathbf{W}_{iter-1} \mathbf{y} \quad (4)$$

ed  $\alpha(\bar{\mathbf{x}})$  è una funzione non lineare definita da:

$$\alpha(\bar{x}_i) = -\frac{\delta}{\delta \bar{x}_i} \log p(\bar{x}_i) \quad (5)$$

in cui  $p(\bar{x}_i)$  rappresenta la distribuzione di probabilità di  $\bar{x}_i$ .

### 3. BSS PER MISTURE CONVOLUTIVE

Nel caso reale di segnali acustici osservati mediante microfoni, il modello basato su misture istantanee non è più appropriato. Infatti l'acustica dell'ambiente modifica temporalmente i segnali osservati ai microfoni secondo una propria caratteristica risposta convolutiva.

Supponiamo di avere nuovamente 2 sorgenti e 2 microfoni e indichiamo con  $x_i(t)$  le sorgenti di segnale. Nel caso convolutivo possiamo rappresentare la matrice di mixing come una matrice  $\mathbf{H}$  di elementi  $h_{ij}(l)$  che rappresentano le risposte impulsive (a supporto limitato) tra il microfono *i-esimo* e la sorgente *j-esima*:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} h_{11}(l) & h_{12}(l) \\ h_{21}(l) & h_{22}(l) \end{bmatrix} * \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad (6)$$

dove abbiamo indicato con \* l'operazione di convoluzione. I segnali osservati dai microfoni possono essere rappresentati come una combinazione lineare delle sorgenti convolute con le risposte impulsive  $h_{ij}(l)$ .

$$y_i(t) = \sum_{j=1}^N \sum_{l=0}^{L-1} h_{ij}(l) x_j(t-l), \quad i = 1, \dots, N \quad (7)$$

dove L è la lunghezza delle risposte impulsive (per ipotesi a supporto limitato). Per ottenere una stima dei segnali di partenza  $\bar{\mathbf{x}}$  è necessario determinare una matrice di deconvoluzione in modo che i segnali nel tempo siano statisticamente indipendenti. A differenza del caso di misture istantanee nel caso convolutivo è necessario stimare un numero elevato di parametri dipendente della lunghezza dei filtri di deconvoluzione L. Nonostante la formulazione nel dominio del tempo sia quella più corretta, l'eccessivo costo computazionale dovuto alla congiunta stima dei filtri  $h_{ij}(l)$  rende impraticabile la separazione in tempi di convergenza ragionevoli. Per semplificare la stima dei parametri  $h_{ij}(l)$  un metodo comunemente utilizzato è quello di effettuare la separazione nel dominio delle frequenze. Come ben noto l'operazione di convoluzione nel tempo corrisponde ad un prodotto nel dominio delle frequenze, quindi applicando la trasformata di Fourier alla (7) otteniamo:

$$Y_i(f) = \sum_{j=1}^N h_{ij}(f) X_j(f) \quad (8)$$

E' evidente quindi che la separazione per segnali convolutivi possa essere ricondotta alla separazione di misture istantanee trattando ogni involuppo di frequenza in modo indipendente. Essendo i segnali  $\mathbf{x}(t)$  non stazionari, applicando la Short-Time Fourier Transform (STFT) si ottengono degli involuppi temporali per ogni frequenza. Il segnale osservato ai microfoni può essere rappresentato come:

$$y_i(f, \tau) = \sum_{j=1}^N h_{ij}(f) x_j(f, \tau) \quad (9)$$

Si può quindi notare che le componenti di frequenza osservate ai microfoni sono una combinazione lineare delle componenti associate ai segnali originali. E' evidente quindi che nel dominio delle frequenze si possano applicare gli algoritmi per misture istantanee basati su ICA, effettuando la separazione in modo indipendente lungo le frequenze.

Applicando la separazione agli involuppi si determina una matrice di separazione  $\mathbf{W}(f)$  che separa le componenti di frequenza indipendenti in  $\bar{\mathbf{x}}(f, \tau)$ :

$$\bar{\mathbf{x}}(f, \tau) = \mathbf{W}(f) \mathbf{y}(f, \tau) \quad (10)$$

I coefficienti dei filtri di deconvoluzione  $w_{ij}(l)$  saranno infine ottenuti mediante antitrasformazione degli elementi  $W_{ij}(f)$  in modo da ottenere la stima di  $\mathbf{x}(t)$  come:

$$\bar{\mathbf{x}}(t) = \mathbf{W}(l) * \mathbf{y}(t) \quad (11)$$

dove  $\mathbf{W}(l)$  è una matrice di filtri di deconvoluzione  $N \times N$  parametrizzata in  $l$ .

Sebbene l'approccio in frequenza permetta l'utilizzo di filtri di deconvoluzione molto lunghi ad un basso costo computazionale ci sono diversi problemi che necessitano di essere risolti legati all'ambiguità della soluzione calcolata attraverso l'ICA.

#### 4. AMBIGUITA' NELLA BSS IN FREQUENZA

Nella sezione 2 abbiamo mostrato come la separazione possa essere effettuata massimizzando l'indipendenza statistica delle sorgenti. La soluzione per la matrice  $\mathbf{W}$  in ogni caso non è unica. Infatti definita una matrice diagonale di scalari  $\mathbf{C}$ :

$$\mathbf{C} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \quad (12)$$

Ogni matrice  $\mathbf{CW}$  è sempre una soluzione secondo ICA in quanto l'indipendenza statistica dei segnali è invariante rispetto al prodotto con degli scalari. Nel dominio delle frequenze gli scalari sono dei numeri complessi e per ogni frequenza le matrici di separazione possono essere scalate in modo non coerente. La risoluzione dell'ambiguità del modulo dei coefficienti  $c_i$  permette di non alterare il contenuto spettrale dei segnali originali mentre la corretta stima della fase dei coefficienti  $c_i$  può essere interpretata come una corretta ricostruzione del segnale privo di riverberazione. Un esatto calcolo dei coefficienti non è possibile ma è possibile determinare una matrice di scaling che minimizzi la distorsione sui segnali. Un approccio comunemente adottato è quello del Minimal Distortion Principle (MDP)[9].

Analogamente se supponiamo di avere una matrice di permutazione  $\mathbf{\Lambda}$  definita ad esempio come:

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (13)$$

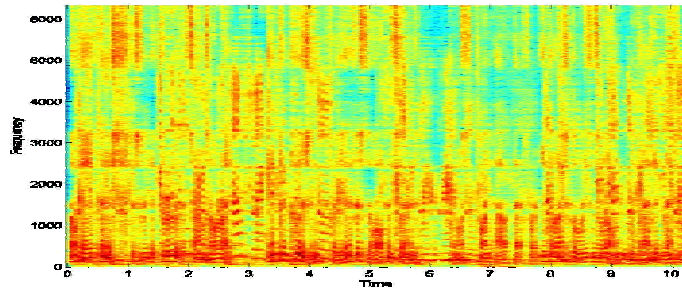


Figura 2: Tipica situazione di disallineamento a blocchi

anche in questo caso ogni matrice  $\mathbf{A}\mathbf{W}$  è sempre soluzione ICA in quanto la permutazione corrisponde solo ad una differente associazione tra le righe di input e di output ma l'indipendenza statistica è sempre verificata. Ricordiamo che nella BSS nel dominio del frequenze la separazione è effettuata in modo indipendente e quindi l'allineamento degli involuppi di frequenza di uscita non è necessariamente costante. In altre parole se non è applicato nessun metodo di allineamento che risolva il problema delle permutazioni, non c'è nessuna certezza che tutte le frequenze del segnale separato in uscita  $\bar{x}_i(t)$  appartengano alla stessa sorgente  $x_i(t)$ . Un esempio tipico di permutazione è mostrato in figura 2. Si può notare che differenti bande dello spettro sono associate evidentemente a segnali differenti.

Nel caso delle permutazioni non esiste una soluzione ampiamente riconosciuta e in questi ultimi anni numerosi approcci sono stati proposti in quanto la strategia di risoluzione delle permutazioni è quella che maggiormente determina le prestazioni totali del sistema. Un ulteriore problema della BSS in frequenza è la presenza di spike dovuti alla circolarità dell'FFT. Essendo infatti i filtri in frequenza a supporto limitato i filtri di deconvoluzione nel tempo dovrebbero essere teoricamente di lunghezza infinita con andamento periodico mentre nella pratica sono troncati ad  $L$ . Per risolvere tale problema generalmente si effettua un filtraggio delle risposte in frequenza basate su finestrazione di hanning[18].

## 5. RISOLUZIONE DELLE PERMUTAZIONI

Mentre i problemi dello scaling e della circolarità sono risolti mediante tecniche largamente riconosciute, per le permutazioni non c'è una soluzione ufficialmente utilizzata e efficace per tutte le situazioni. In generale esistono diversi approcci per risolvere tale ambiguità che possono essere classificati in 3 categorie principali:

- Allineamento basato sulla DOA (direction of arrivals). Gli involuppi di frequenza sono raggruppati secondo informazioni di fase. Infatti la disposizione delle sorgenti in punti differenti dello spazio definisce differenti DOA che possono essere stimate dagli involuppi di frequenza separati o direttamente dalle matrici di separazione  $\mathbf{W}(f)$ [16].

- Allineamento basato su correlazione di ampiezza. Tale metodo utilizza informazioni di correlazione statistica tra l'energia di involuppi adiacenti, nell'ipotesi quindi che i segnali siano correlati in frequenza (come nel caso del parlato) [15][16].

- Coerenza delle matrici di separazione. Suppone che la risposta impulsiva dell'ambiente non presenti discontinuità evidenti e quindi tale continuità si propaghi anche sulle matrici di separazione[14]. Tale metodo è strettamente correlato con quello basato sulla DOA ma non effettua una stima esplicita delle DOA delle sorgenti.

Nel nostro algoritmo di BSS abbiamo sviluppato due tecniche complementari basate sugli ultimi due approcci. In [10] abbiamo mostrato come le matrici di separazione  $\mathbf{W}(f)$  varino con continuità lungo le frequenze. La fase del rapporto degli elementi di una riga di  $\mathbf{W}(f)$  è infatti strettamente connessa con il modello di propagazione dell'onda acustica delle sorgenti e teoricamente varia linearmente lungo le frequenze. In [11] abbiamo mostrato che se ad una determinata frequenza l'ICA è inizializzato con la matrice di separazione  $\mathbf{W}(f-I)$  ottenuta per una frequenza adiacente, ci sarà un'alta probabilità che l'ICA converga verso

una soluzione che ha una permutazione coerente lungo tutte le frequenze. Per evitare le permutazioni l'approccio da seguire potrebbe essere quello di effettuare la separazione a partire dalle frequenze più alte verso quelle più basse operando ad ogni frequenza la ricorsione. In ogni modo le matrici di separazione  $\mathbf{W}(f)$  ottenute da ICA possono essere considerate come un'osservazione rumorosa delle matrici reali:

$$\overline{\mathbf{W}}(f) = \mathbf{W}(f) + \mathbf{N}(f) \quad (14)$$

Tale rumore è principalmente introdotto dalla riverberazione che introduce delle non linearità della fase anche lungo frequenze adiacenti. Per questo motivo metodi di inizializzazione ricorsiva come quello in [14] falliscono in presenza di alta riverberazione o di semplici errori locali di stima della  $\mathbf{W}(f)$ . Per migliorare questo approccio in [10] abbiamo mostrato come la  $\mathbf{W}(f)$  possa essere stimata ricorsivamente mediante un filtraggio Least Mean Square che minimizzi le fluttuazioni del determinante delle  $\mathbf{W}(f)$ . In questo modo l'ICA lungo le frequenze è ricorsivamente inizializzato con una stima "smooth" di  $\mathbf{W}(f)$  evitando quindi che discontinuità locali possano generare errori di allineamento che si propagano lungo le frequenze successive.

Tale metodo si è dimostrato molto efficace e permette di ottenere notevoli risultati anche in condizioni di riverberazione molto critiche e con piccole porzioni di segnale osservato. Ad ogni modo se le sorgenti sono molto vicine e se i microfoni sono distanziati in modo da ottenere aliasing spaziale la stima delle  $\mathbf{W}(f)$  può non essere accurata e può quindi generare disallineamenti a blocchi su bande dello spettro (come quelli mostrati in figura 2). E' necessario quindi risolvere l'ambiguità con una tecnica complementare.

Come tecnica alternativa è possibile utilizzare un approccio basato sulla correlazione dell'energia dello spettro lungo frequenze adiacenti. Infatti per la maggior parte dei segnali acustici osservati e in particolare per il parlato esiste un'alta correlazione interfrequenziale. Un approccio molto comune è quello di calcolare il coefficiente di correlazione tra gli involucri di frequenza adiacenti e di determinare la permutazione che massimizza la correlazione totale. Anche in questo caso però un errore locale può ripercuotersi negativamente verso tutte le frequenze successive. Una soluzione è di effettuare una stima ricorsiva dell'involuppo in frequenza e di utilizzare tale involuppo come base di riferimento per l'allineamento. Tale tecnica presentata in [12] si è mostrata efficace anche per la separazione di sorgenti posizionate lungo una stessa direzione angolare la quale impedisce ovviamente l'utilizzo di tecniche basate sulla stima del modello di propagazione.

Anche in questo caso però tale tecnica può diventare instabile se in una porzione dello spettro ripetuti errori modificano la base di riferimento generando di conseguenza errori di allineamento nelle frequenze successive.

Per quanto siano efficaci entrambi i metodi, essendo basati su di un approccio ricorsivo sono sensibili ad errori locali che possono propagarsi lungo le frequenze successive. E' necessario quindi minimizzare la probabilità di propagazione di un errore locale. Per i succitati motivi proponiamo un nuovo metodo di risoluzione delle permutazioni che combina entrambe le tecniche minimizzando gli errori di decisione. Tale metodo è basato su di uno schema di due "decisori interlacciati" in cui gli errori di decisione sono corretti reciprocamente. In tale schema supponiamo che gli errori di decisione siano incorrelati. In figura 3 mostriamo lo schema interlacciato di decisione delle permutazioni:



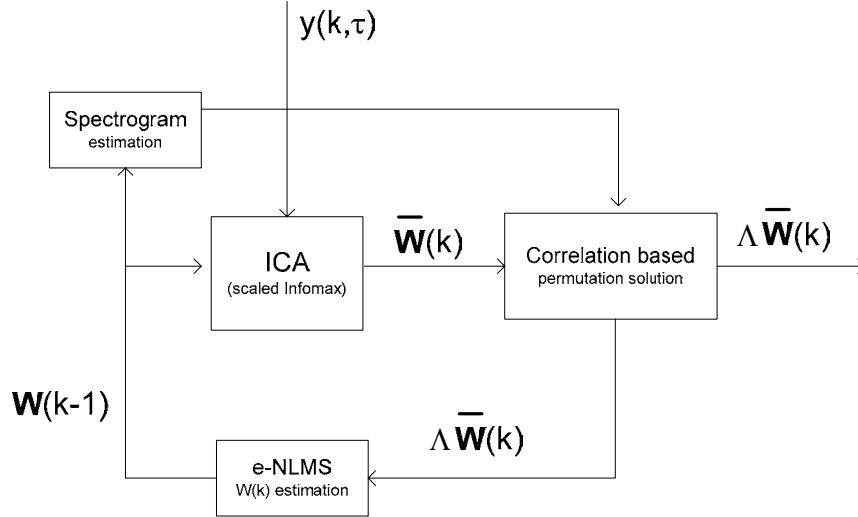


Figura 3: Schema interlacciato di correzione delle permutazioni

Ad ogni bin di frequenza  $k$  l'ICA genera una matrice di separazione  $\bar{\mathbf{W}}(k)$ . Il metodo basato sulla correlazione delle energia tra gli involucri determina la permutazione ottimale  $\Lambda$ . La matrice  $\Lambda \bar{\mathbf{W}}(k)$  è usata come osservazione per l'algoritmo che stima la matrice  $\mathbf{W}(k)$ . In tal modo un errore sulla permutazione è corretto dal blocco basato sulla stima LMS di  $\mathbf{W}(k)$ . La matrice risultante, indicata con  $\mathbf{W}(k-1)$  è usata per inizializzare l'ICA e per ristimare lo spettro. Analogamente se il blocco "e-NLMS" genera un errore, al passo successivo il blocco di decisione "Correlation based" riallineerà le matrici evitando che la permutazione errata si propaghi lungo tutte le frequenze. Di conseguenza è facile comprendere che la probabilità che ci sia una permutazione errata sia:

$$p(e_{\text{permutazione}}) = p(e_1, e_2) = p(e_1) p(e_2) \quad (15)$$

dove  $p(e_1)$  e  $p(e_2)$  sono le probabilità di errore dovute ai singoli metodi di decisione in cui abbiamo supposto che i due metodi generino degli errori incorrelati.

Tale tecnica combina in modo ottimale i due metodi di allineamento i quali essendo complementari utilizzano tutta l'informazione a disposizione.

## 6. RISULTATI SPERIMENTALI

Il sistema è stato implementato per funzionare con un numero generico di sorgenti. In questo lavoro mostriamo i risultati numerici ottenuti nella separazione di due sorgenti in differenti condizioni di riverberazione ( $T_{60} = 160$  ms e  $T_{60} = 700$  ms). La disposizione dei microfoni e delle sorgenti è indicata in figura 4.

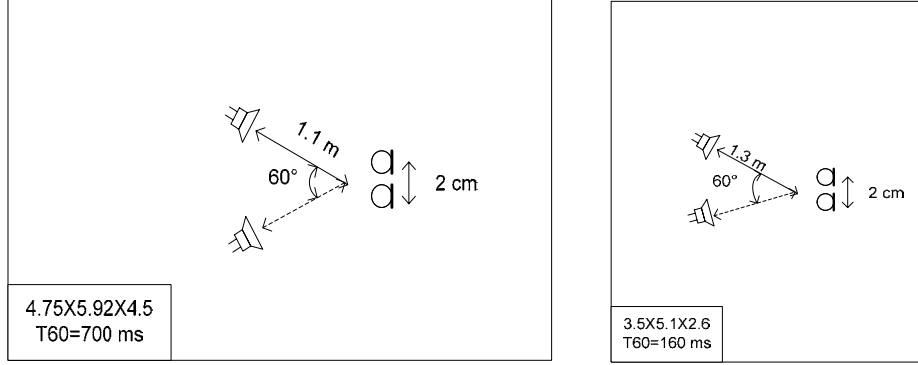


Figura 4: Configurazione dei test reali effettuati in camere con differenti  $T_{60}$

Inoltre l'algoritmo è stato applicato a registrazioni audio di altri ricercatori e disponibili liberamente on-line. Per tali registrazioni non abbiamo valutazioni numeriche dato che i segnali parziali di riferimento non sono disponibili. Ad ogni modo in tutti i casi la separazione è stata efficace e il metodo si è dimostrato robusto anche in casi in cui è presente aliasing spaziale o quando le sorgenti sono posizionate lungo una stessa direzione angolare. Esempi di separazione sono disponibili on-line in [19].

Le prestazioni del sistema sono state calcolate misurando il Signal Interference Ratio (SIR) dell' $i$ -esima sorgente di segnale come:

$$SIR_i : 10 \log_{10} \frac{\langle \left| \sum_l u_{ii}(l) x_i(t-l) \right|^2 \rangle_t}{\langle \left| \sum_{k \neq i} \sum_l u_{ik}(l) x_k(t-l) \right|^2 \rangle_t} \text{ (dB)} \quad (15)$$

$$u_{ik}(l) = \sum_{j=1}^N \sum_{\tau=0}^{L-1} w_{ij}(\tau) h_{jk}(l-\tau) \quad (16)$$

dove indichiamo con  $u_{ik}(l)$  la risposta impulsiva tra la sorgente  $x_k(t)$  e il segnale separato  $\bar{x}_i(t)$ . Il SIR riportato in tabella 1 è quello mediato su tutte le sorgenti.

| Configurazione          | SIR medio dopo la BSS |
|-------------------------|-----------------------|
| 2 sorgenti (T60=160 ms) | 18.1 dB               |
| 2 sorgenti (T60=700 ms) | 12.5 dB               |

Tabella 1 – Prestazioni in termine di Signal Interference Ratio

I segnali utilizzati hanno una durata di 9 secondi e sono stati campionati con  $f_s=16$  kHz. Nel test con  $T_{60} = 160$  ms è stata applicata una la STFT con finestre di hanning di 4096 punti. Nel test con  $T_{60} = 700$  ms si sono utilizzate finestre lunghe 8192 punti per ottenere filtri di deconvoluzione più lunghi in grado di coprire un tempo di riverberazione maggiore. I valori indicati in tabella 1 mostrano come i livelli di separazione siano molto elevati anche nel caso di condizioni di riverberazione molto avverse. Questo indica che, sebbene la BSS in frequenza sia intrinsecamente e asintoticamente limitata dalla riverberazione, è possibile comunque ottenere performance accettabili se il problema delle permutazioni è risolto correttamente.

## 7. CONCLUSIONI E ATTIVITA' FUTURE

In questo lavoro è stato presentato un nuovo metodo per la risoluzione del problema delle permutazioni che emerge nell'approccio in frequenza della blind source separation. E' stato formulato un nuovo schema interlacciato che permette di combinare i vantaggi di due metodi complementari di risoluzione delle permutazioni, presentati in lavori precedenti. Tale schema è stato utilizzato per implementare un algoritmo di BSS in grado di separare sorgenti acustiche in ambienti molto riverberanti e i risultati sperimentali confermano la validità dell'approccio. Avendo osservato che tale tecnica è molto robusta anche per la separazione di segnali acustici molto brevi, stiamo sviluppando un'estensione per il caso di sorgenti in movimento in grado di separare delle sorgenti acustiche in tempo reale.

## BIBLIOGRAFIA

- [1] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.
- [2] S. Araki, S. Makino, R. Mukay, T. Nishikawa, H. Saruwatari "Fundamenta limitation of frequency domain blind source separation for convolved mixture of speech
- [3] Cardoso, J.-F. (1997). "Infomax and maximum likelihood for source separation." *IEEE Letters on Signal Processing*, 4:112–114.
- [4] Douglas, S.C. Gupta, M. "Scaled Natural Gradient Algorithms for Instantaneous and Convolutional Blind Source Separation", *ICASSP 2007*, vol II, pp 637-640, Apr. 2007
- [5] M. A. Casey and A. Westner. "Separation of mixed audio sources by independent subspace analysis." In *Proceedings of International Computer Music Conference*, Berlin, Germany, 2000.
- [6] Lee and H.S. Seung. "Algorithms for non-negative matrix factorization" In *Proceedings of Neural Information Processing System* pages 556-562, Denver, USA 2001

- [7] Makeig S, Bell AJ, Jung T-P, Sejnowski TJ, "Independent Component Analysis of Electroencephalographic data", Advances in Neural Information Processing Systems, 1996
- [8] McKeown MU, Makeig S, Brown GG, Jung PT, Kindermann SS, Bell AJ, Sejnowski TJ (1998): "Analysis of fMRI data by blind separation into independent component analysis", Human Brain Mapping 6: 160-188.
- [9] K.Matsuoka, S. Nakashima, "Minimal distortion principle for blind source separation" in: International Symposium on ICA and Blind Signal Separation, San Diego, CA, USA, December 2001
- [10] Francesco Nesta, Piergiorgio Svaizer, Maurizio Omologo, "Separating short signals in highly reverberant environment by a recursive frequency-domain BSS", in proceedings of HSCMA, Trento, May 2008.
- [11] Francesco Nesta, Piergiorgio Svaizer, Maurizio Omologo, "A BSS method for short utterances by a recursive solution to the permutation problem", in proceedings to SAM, Darmstadt, July 2008.
- [12] Francesco Nesta, Maurizio Omologo, "A recursive high order statistics approach to solve the permutation problem in frequency-domain BSS", in proceedings of Toni Mian Workshop, October 2007
- [13] M. S. Pedersen, J. Larsen, U. Kjems, L. C. Parra, "A Survey of Convolutional Blind Source Separation Methods", Springer Handbook of Speech, Nov. 2007.
- [14] D.-T. Pham, Ch. Serviere, and H. Boumaraf, "Blind separation of convolutional audio mixtures using nonstationarity," in Proceedings of the 4th International Conference on Independent Component Analysis and Blind Signal Separation (ICA '03), pp. 981-986, Nara, Japan, April 2003.
- [15] D.-T. Pham, Ch. Serviere, and H. Boumaraf, "Blind separation of speech mixtures based on nonstationarity," in Proceedings of 7th International Symposium on Signal Processing and Its Applications (ISSPA '03), vol. 2, pp. 73-76, Paris, France, July 2003.
- [16] H. Sawada, R. Mukai, S. Araki, S. Makino, "Frequency-Domain Blind source separation" in Speech Enhancement (J. Benesty, S. Makino J. Cheen), ch. 13, pp. 299-327, Springer, 2005
- [17] H.Sawada, R.Mukai, S.Araki, S.Makino. "A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation." IEEE transaction on speech and audio processing vol.12. no.5 september 2004
- [18] H. Sawada, R. Mukai, S. de la Kethulle de Ryhove, S. Araki, S. Makino "Spectral smoothing for frequency-domain blind source separation", IWAENC 2003, Sept. 2003, Kyoto, Japan
- [19] <http://shine.fbk.eu/people/nesta/>

# **NATURAL LANGUAGE PROCESSING**

# **SPEECH AUTOMATION NEL CONTESTO DELLE ARCHITETTURE DI SERVIZIO DEI CONTACT CENTER**

Alessandro Bruni, Gianpaolo Coro  
Abla s.r.l.  
*{alessandro.bruni, gianpaolo.coro}@abla.it*

## **1. SOMMARIO**

L'obiettivo di questo scritto è duplice:

- si tratta di fornire una collocazione e una tassonomia alle funzionalità di speech automation nel contesto delle architetture di gestione telefonica, cioè nella collocazione funzionale in cui esse hanno sinora trovato più rispondenza da un punto di vista della loro applicazione pratica;
- per poi mettere in evidenza quali possano essere le evoluzioni applicative di tali funzionalità e tecnologie.

Il taglio di questa trattazione non è di carattere scientifico. Il testo dà per scontato che sullo sfondo delle soluzioni applicative di 'speech automation' vi sia una robusta e intensa ricerca di base rivolta

- alla indagine sui fondamenti logici ed epistemologici del linguaggio inteso nel suo complesso; quella che noi chiameremo linguistica nel senso più ampio;
- al linguaggio nella sua fenomenologia fisica ed in particolare acustica, cioè all'insieme delle condizioni che rendono la comunicazione orale materialmente possibile;
- all'insieme delle tecniche matematiche che consentono la trasformazione del segnale acustico in una forma algoritmica che sia manipolabile e rappresentabile all'interno di un sistema di regole applicative-informatiche;
- alla indagine del linguaggio nelle sue forme e articolazioni come fenomeno di tipo neurologico.

## **2. PREMESSA**

La filosofia ben prima di qualsiasi altra disciplina si è occupata del linguaggio umano e lo ha fatto in un contesto 'fondativo' in cui pochi erano gli strumenti concettuali e quelli della indagine scientifica (De Saussure, 2007). La indagine filosofica antica comincia ad occuparsi del linguaggio nel momento 'ideale' in cui si registra il passaggio dalla primitiva oralità alle culture della lingua scritta e proprio per essere maturata in tale contesto così fertile i suoi contenuti restano di interesse e attualità per inquadrare meglio anche l'odierna indagine scientifica e la prassi tecnologica ed orientarle in senso epistemologico e non solo (Gargani, 2007).

Molte delle 'aporie', contraddizioni e punti aperti che ancora oggi contraddistinguono l'indagine sui fondamenti della lingua umana si riflettono in modo puntuale all'altro capo del lungo filo e cioè nel contesto tecnologico e applicativo (Galimberti, 2006). Questa difficoltà a circoscrivere il fenomeno del linguaggio entro ambiti deterministici e assoggettabili ad un'indagine esaustiva persiste e si trasferisce sulle difficoltà che incontra l'evoluzione tecnologica e applicativa. Questo limite si manifesta sia sul versante delle prestazioni tecniche delle applicazioni di riconoscimento che su quello della percezione insoddisfacente che ne ha la stragrande maggioranza degli utenti: fino al punto che questi

due aspetti condizionano fortemente la evoluzione e le aspettative legate allo sviluppo di queste applicazioni.

### 3. MITO, FILOSOFIA E SCIENZA

L'aporia di base che domina l'indagine sul linguaggio ha molteplici facce fra di loro connesse: in prima istanza non è del tutto chiaro quale sia il rapporto e la correlazione fra significante e significato. Inoltre, il linguaggio appare come una pratica in cui è difficile scindere gli elementi deterministici da quelli aleatori, il contesto sociale dalla pratica individuale, l'efficacia dalla precisione, etc.; in generale lo statuto logico che sottende all'atto linguistico ed alla sua comprensione. Intorno al linguaggio restano ampiamente aperti una serie di quesiti di fondo:

- quale sia il rapporto fra linguaggio e realtà (o la nostra immagine mentale della realtà);
- la natura intrinseca dell'atto linguistico come prassi antropologica;
- i meccanismi che consentono l'apprendimento della lingua;
- le leggi che regolano la sua trasformazione nel tempo;
- la sua geografia;
- le leggi che presiedono alla articolazione delle morfologie linguistiche;
- etc.

Il '900 ci ha consegnato due estremi della riflessione sul linguaggio nel pensiero di Wittgenstein (Wittgenstein, 1918) ed in quello di Chomsky (Chomsky, 1975). Il primo, partito da un programma positivistic di stampo logico, approderà dopo il *Tractatus Logico-philosophicus* ad una idea del linguaggio come '*prassi che funziona*'. Il secondo attraverso le categorie duali di '*competenza*' ed '*esecuzione*' tenterà di circoscrivere sia l'aspetto strutturale che quello esecutivo dell'atto linguistico aprendo la strada alla idea che la lingua parlata si fonda su due livelli:

- uno livello più profondo che consiste in una struttura permanente che abilita l'atto linguistico
- e uno più superficiale, è fatto di trasformazioni eseguite con un certo grado di libertà e reso possibile proprio dalla esistenza, solidità e permanenza della struttura profonda su cui la libertà espressiva si appoggia come su una base.

Gli sviluppi più recenti delle indagini tendono sempre di più a fare una distinzione ontologica fra la lingua scritta e quella parlata (Ong, 2007), come due strutture completamente autonome e diverse e non l'una come la 'traduzione' o una 'forma alternativa' dell'altra. Uno degli elementi più interessanti di questa prospettiva è la sua direzione anti-riduzionistica che promette sviluppi non scontati.

Un'altra prospettiva interessante è la possibile convergenza fra linguistica e neuroscienze: la prima sempre meno assillata da necessità riduzioniste e le seconde sempre più inclini al passaggio dal classico paradigma stimolo-risposta a quello comportamentale.

Il dualismo fra oralità e scrittura ci riporta ad un altro dualismo già percepito nella antichità, sia nella cultura ebraica (la Legge viene scritta, scolpita dopo essere stata enunciata), che in quella greca (Platone privilegia la intensità potente del parlato alla praticità 'muta' dello scritto), che in quella romana ('*scripta manent, verba volant*') che nelle tradizioni letterarie medievali che – infine – nelle indagini ottocentesche sulle origini orali delle opere scritte e sui meccanismi con cui tali opere si sono tramandate in epoche diverse.

Esso - peraltro – ha un suo fascino epistemologico e suscita affascinanti interrogativi fino a richiamare per analogia l'idea che la ridondanza abbia di per sé un valore nel contesto dell'antropologia come in altri 'sistemi naturali'.

Il Linguaggio ha - nella nostra cultura - sia un carattere comunicativo che generativo: nel mito (che è tanta parte del nostro apprendere) la narrazione si sposta rapidamente dalla creazione (Logos) al compito dato ad Adamo di nominare le cose attribuendo loro essenza ed esistenza per poi passare ad un contesto di carattere comunicativo. Adamo ed Eva sperimentano e ci mostrano il linguaggio come comunicazione ma anche i suoi esiti ambigui, come esso si presti a finalità soggettive, finanche all'inganno e ciò con conseguenze non trascurabili, stando alla Bibbia.

Fuori dal mito e ormai nell'ambito della 'metafora scientifica' la linguistica moderna, diventata disciplina sistematica, mostra sostanzialmente due facce: da una parte c'è lo strutturalismo (figlio anche del neopositivismo logico) e dall'altra una visione antropologica del linguaggio come prassi che tutto sommato 'tiene'. Con Chomsky sembra riaffacciarsi un certo interesse sull'ambito di generalità delle strutture linguistiche (Bandler & Grinder, 2007) che in una forma più intensa era stato di Aristotele: nel 'De Interpretazione', parte dell'Organon, che significa '*strumento*', lo Stagirita pone per primo il tema della convenzionalità del significante rispetto al significato.

#### 4. TECNOLOGIA

Se abbandoniamo il terreno del passaggio fra filosofia e scienza e ci portiamo nell'ambito della pratica tecnologica vediamo che lo sviluppo dell'automazione del linguaggio riporta in campo molti degli elementi cui abbiamo appena accennato: per esempio, le due tecniche di base che compongono un dialogo automatico (ASR, Automatic Speech Recognition e TTS, Text to speech) mantengono aperti tutti i temi sin qui accennati:

- il segnale fisico come elemento di partenza e la forma d'onda come sua trasformazione in significante;
- la asimmetria di scritto e orale e viceversa (la trasformazione  $S \rightarrow O$  comporta forme di ambiguità diverse da quella  $O \rightarrow S$ , con conseguenze applicative fondamentali);
- il dominio del senso come modello di accoppiamento fra significante e significato;
- il contesto come elemento abilitante/determinante la ri-soluzione dell'atto linguistico;

Le tecnologie informatiche che si occupano di riproduzione dell'atto linguistico procedono per ora attraverso semplificazioni crescenti: i mezzi caratteristici che hanno a disposizione sono essenzialmente

- la **capacità elaborativa** che consente di trasformare i segnali di base in forme altrimenti rappresentabili;
- i **database**, per realizzare anche con tecniche statistiche l'accoppiamento migliore fra il segnale formalizzato e insiemi di potenziali significanti (dizionari);
- la **logica applicativa**, per risolvere la articolazione del dialogo.

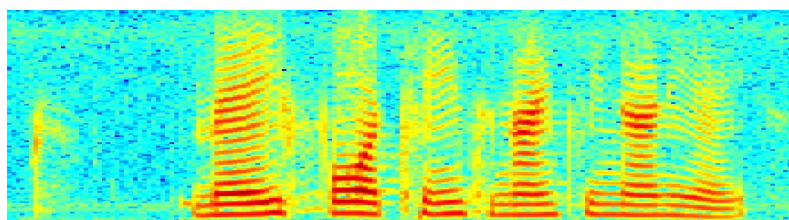
A ben vedere le prassi tecnologiche nel loro insieme sono figlie dello strutturalismo linguistico e anche dei suoi inevitabili riduzionismi; anzi per certi versi sembra che lo strutturalismo sia una sorta di preliminare 'palestra dell'ardimento' epistemologica su cui si basa lo sviluppo tecnologico.

Nel caso del riconoscimento vocale ad oggi la tecnologia informatica ci consente la rappresentazione di una forma d'onda provocata da una 'espressione vocale' (vocal utterance) in una forma adeguata alla sua correlazione con uno o più elementi (fonemi,



morfemi) appartenenti ad un dizionario. In più, tecniche statistiche ci consentono di restringere a piacimento l'insieme degli elementi del dizionario che possano con probabilità essere accoppiati alla forma d'onda che ci viene rappresentata. Questa 'metafora', semplice o complessa che sia, è ad oggi ciò che noi chiamiamo 'riconoscimento vocale' o Automatic Speech Recognition.

Inoltre, la tecnologia informatica ci consente di corredare questa funzione di riconoscimento con altre funzioni che supportano il 'dialogo' fra uomo e macchina: p.e. tecniche di scelta multipla, strategie di recupero degli errori ed anche – ultima ratio – la possibilità di passare ad una modalità diversa di gestione della transazione. Tutte queste tecniche 'al contorno' fanno parte dell'approccio applicativo alla tecnologia di riconoscimento vocale, nel senso che sono state sviluppate per essere utilizzate in contesti industriali e commerciali, di supporto a transazioni fra imprese, organizzazioni e singoli utenti o clienti.



Fonemi: - m ey f ao r t iy th n ayt iyn ay dx iy ax t

➡ "14 Maggio 1998"

Figura 1: segnale audio corrispondente al significato 'May 14th 1994'.

Dunque, riassumendo:

1. allo stato attuale noi possiamo analizzare e manipolare forme d'onda caratteristiche di un atto linguistico (vedi figura qui sopra) e provare ad associarle ad un dizionario;
2. intorno alla pure funzione di riconoscimento si sono sviluppate tecniche applicative anche sofisticate che cercano di emulare non solo il riconoscimento vocale ma l'intero contesto di un dialogo fra uomo e macchina;
3. altri elementi pur pertinenti ancora ci sfuggono completamente o in buona parte: l'intenzionalità, il contesto, le analogie di significato, la intonazione, le intenzioni retoriche, i paradossi, le metafore, la sinestesia (intesa come 'concorrenza' di più sensi alla percezione del parlato), etc.

Torna in campo ancora una volta la supposta (fin qui esperita) incapacità strutturale delle macchine di gestire in maniera pratica un contesto extra-logico o logico-allargato, dove convergono funzioni selettive della memoria, intelligenza discriminante, stili soggettivi, gerarchia delle priorità, elementi della 'gestalt', nel senso della 'forma di tutto ciò che può essere pertinente alla comprensione'. Ma se (per ora?) le macchine non comprendono, sono altresì in grado di manipolare in termini trasformativi un dominio

determinato e dotato di 'senso logico': non è poco, anche se ai nostri fini si tratta di una piccola parte della prassi linguistica.

In generale tutte le prassi umane (ivi compresa la tecnologia) sono in grado di svolgersi ed evolvere anche se e quando siano temporaneamente prive di un fondamento. Alcune di queste prassi adottano magari uno statuto logico molto semplificato o limitano il loro campo applicativo. Per esempio l'economia ha rinunciato da tempo ad una teoria unificante (o al proprio compito originale di risolvere il problema della distribuzione della ricchezza) e si accontenta di articolare la propria prassi in ambiti di per sé limitati: economia aziendale, scienza della finanza, etc. Anche la politica sembra aver rinunciato alla teoria del perseguimento del bene comune e si accontenta di perseguire un insieme di prassi che di per sé hanno una pretesa di utilità. La medicina evolve velocemente seppure non abbia ancora risolto in via sostanziale la contraddizione fra approccio sintomatico e approccio olistico. Conosciamo poco del mondo esterno e di noi stessi umani ma ci orientiamo a risolvere molti problemi: a volte ci riusciamo, a volte no. Nondimeno tentiamo nel sottofondo sempre implicitamente di ricondurre prassi e formalismi ad un dominio di senso più ampio. La scienza moderna in generale è interessata da questa difficoltà di 'tenere insieme' i 'fondamenti' con le necessità della prassi e probabilmente assisteremo nel tempo ad una sorta di inasprimento del conflitto fra queste due istanze. Di questo conflitto la tecnologia è parte inerente in quanto lo rappresenta nella pratica. Per quanto ci riguarda, anche le tecnologie della voce sono su questa linea di frizione con la loro 'scienza di riferimento' e cioè la linguistica. Noi pensiamo che questo conflitto sia illuminante per la evoluzione di entrambe.

## 5. HIC SALTUS

Qui vi chiediamo di fare un salto, di lasciarvi alle spalle le questioni aperte e di immaginare le capacità applicative legate al riconoscimento ed alla sintesi vocale all'opera in un contesto organizzativo per molti versi complesso quale è quello di un call center. Ma prima di addentrarci in questo contesto applicativo daremo brevemente conto della sua storia ed evoluzione. I call-center sono una naturale evoluzione organizzativa dei centralini telefonici, cioè di quelle organizzazioni rese celeberrime dal cinema in cui uno o più operatori (in realtà ce le ricordiamo come gentili operatrici, inappuntabili e ben truccate seppur presenti al pubblico con la sola voce) mettevano in contatto un utente chiamante con un destinatario. La funzione svolta era una funzione pubblica di commutazione (switching). Nel tempo la evoluzione delle reti di telefonia fissa e dei sistemi di numerazione ha consentito di svolgere tale funzione di 'connessione' in maniera automatica.

Ma il servizio di 'commutazione' umana non è mai morto: si è spostato dalla rete pubblica ai centralini delle imprese e delle organizzazioni (ormai anche degli studi professionali), dando – attraverso un operatore umano – molte funzioni: ricerca e messa in contatto di persone; informazioni di ogni tipo relative a prodotti, servizi e orari; appuntamenti; prenotazioni e ordinativi; pagamenti; disdette, reclami, etc.

Per una serie di considerazioni di opportunità, negli ultimi 10-15 anni tali servizi sono stati offerti prevalentemente in modalità di 'outsourcing' da società specializzate che rendono il servizio alle imprese attraverso strutture organizzative, tecnologiche e di personale proprie.

I call-center rivestono un ruolo fondamentale nelle strategie e operazioni di contatto delle imprese con il loro pubblico, i loro clienti, fornitori e reti commerciali e operative. Essi rappresentano uno dei canali fondamentali di relazione e convivono con gli sportelli, le agenzie, il web e altri strumenti di contatto. Attraverso i call center delle grandi imprese

passano milioni di telefonate e spesso essi sono un canale esclusivo di contatto col pubblico.

Per altri versi i call-center sono anche uno dei principali motivi di lagnanza da parte della clientela e spesso percepiti come un ostacolo posto ad arte fra l'impresa e il pubblico.

D'altra parte se una organizzazione decide di fare del call-center un elemento di differenziazione del suo servizio deve dedicarvi grande attenzione.

I call-center sono destinati (pur nella loro inevitabile evoluzione) a continuare ad avere un ruolo fondamentale nella relazione uno-molti che caratterizza il rapporto fra utenti e imprese. Anzi, non solo le transazioni telefoniche sono in costante aumento in tutte le grandi imprese e in tutti i paesi evoluti ma addirittura il telefono si sta imponendo come strumento di supporto al canale web che sembrava a freddo doverne decretare il declino. Vedremo.

Se dovessimo farci una idea 'dal vivo' delle condizioni in cui opera un call-center nei suoi orari di apertura possiamo immaginare:

- persone (clienti) che chiamano chiedendo ogni sorta di informazioni (prezzi, orari, funzionalità, prodotti, condizioni contrattuali, etc.) e volendo svolgere al telefono transazioni (acquisti, prenotazioni, cambiamenti di ordine, pagamenti, lamentele, rimborsi, cambi di indirizzo e di destinazione, etc.);
- operatori che rispondono avendo a disposizione competenze più o meno adeguate rispetto al dominio del problema nonché un insieme composito di strumenti di ausilio;
- le richieste di informazioni e di soluzioni di problemi i più disparati si riferiscono ad una gamma di servizi e di prodotti sempre più crescente e complessa;
- anche il sistema delle aspettative di chi chiama è sempre più complesso: ad esso contribuiscono contesti magari sconosciuti all'operatore come per esempio la pubblicità, la stampa, informazioni date dai diversi canali distributivi e di assistenza ma anche dal passa-parola, dalla comunità degli utenti, dal web;
- esiste un dominio semantico del 'già detto' del 'non detto' del 'detto già ad altri' sullo sfondo della relazione telefonica e anche insito nel 'vissuto' dei prodotti e dei servizi: un vissuto esperienziale degli utenti ma anche degli operatori, che si muove su piattaforme diverse con punti di contatto e divergenze.
- le storie e psicologie personali e esigenze materiali e psicologiche scandite da tempi, costi, vincoli organizzativi e tecnici...

Questa elencazione può espandersi a piacimento a seconda del contesto di cui si parla. Essa vuole mettere in evidenza una serie di elementi pertinenti alla relazione ma per definizione problematici da gestire attraverso la pura tecnologia.

## **6. COMPONENTI ARCHITETTURALI DI UN DIALOGO TELEFONICO E COLLOCAZIONE DEL RICONOSCIMENTO E DELLA SINTESI VOCALE**

Eppure tutto questo contesto così complesso può e deve ai nostri fini essere ridotto e rappresentato in un flusso telefonico che porta un segnale audio dalla rete pubblica e che viene instradato attraverso un PBX verso un operatore (uomo o macchina). Vediamo qui di seguito ad uno ad uno questi componenti (la figura qui sotto mostra il quadro di insieme della architettura). L'insieme di questi elementi contenuti in un'architettura logicamente coerente rappresenta un contesto che abilita un dialogo linguistico in un ambito dove sono attuate le restrizioni di spazio e tempo.

Le tecnologie che abilitano l'utilizzo del linguaggio automatico risiedono in maniera eminente nel componente tecnologico denominato **IVR**. Da questo punto di vista e per molti versi l'IVR è in questa fase della evoluzione della telefonia il cuore applicativo di queste architetture. Gli altri elementi hanno una funzione specifica nell'abilitare e rendere materialmente possibile il dialogo telefonico e di corredarlo di quante più risorse di tipo informativo possibile.

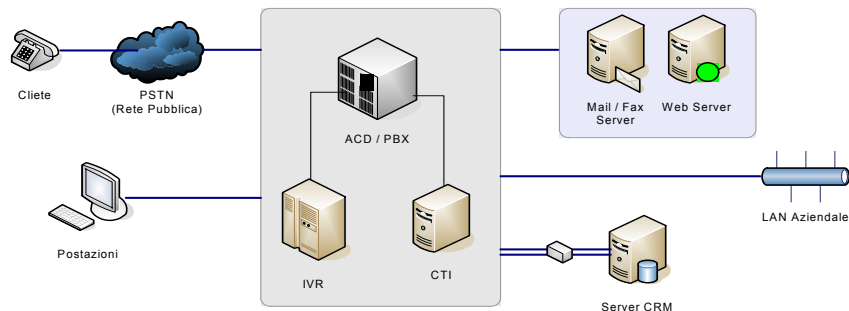


Figura 2: esempio di architettura telefonica.

#### **PBX** - Private Branch eXchange

Il Centralino telefonico è l'elemento di ingresso della struttura di un Contact center. Gestisce le chiamate in ingresso ed in uscita, consentendo a tutti gli utenti interni al Contact center di condividere differenti risorse esterne (es. linee telefoniche) ed interne (es. postazioni operatore).

Un centralino può incorporare la tecnologia **ACD** che consente di gestire la distribuzione e l'accodamento delle chiamate entranti rispetto ai gruppi di Operatori secondo diversi parametri configurabili.

#### **IVR** - Interactive Voice Response

Il risponditore automatico è la componente che consente l'erogazione di servizi informativi "**automatici**" (senza coinvolgimento di Operatori), riducendo anche il carico di chiamate in entrata verso gli operatori del Contact Center. Un sistema IVR permette la creazione di "alberi telefonici" preimpostati che guidano l'utente verso operazioni di self-service o verso gli Operatori stessi. In unione con la tecnologia CTI permette inoltre di effettuare la "identificazione" del Cliente e del Servizio, prima di trasferire la telefonata verso il gruppo di Operatori più adeguato.

- **Routing:** Funzionalità che permette di smistare in modo "intelligente" le varie interazioni (fonia, e-mail, fax, web, etc.) associandole a differenti livelli di priorità in base al loro valore (in termini di business), al livello di servizio richiesto o tenendo conto di necessità specifiche, indirizzandole agli operatori più idonei.
- **Alcuni esempi di possibili opzioni di instradamento**

##### *In base al livello di servizio*

La soluzione adegua automaticamente il dimensionamento del gruppo di operatori e le loro attività (front e back office) al fine di garantire strategie di business e livello di servizio predefinito con il Partner (SLA).

*In base ai dati ed al contenuto delle interazioni*

La soluzione raccoglie dati e informazioni da diverse fonti esterne (databases Partner, sistemi legacy, etc.) e “modula” il routing in base a queste informazioni.

La soluzione “indaga” il contenuto delle interazioni (ad es. il testo delle mail ricevute) e basandosi sull’analisi di parole e testi chiave, indirizza l’interazione e la conseguente attività al gruppo più idoneo.

*In base al profilo dell’operatore*

La soluzione è in grado di associare, per ogni determinato tipo di interazione, l’operatore con le caratteristiche (esperienza, specializzazione, etc.) più adeguate per rispondere al Cliente.

*In base all’ultimo operatore*

La soluzione è in grado di indirizzare le interazioni ad una persona specifica, per esempio all’ultimo operatore che si è occupato del Cliente (consigliata per Clienti di alto valore).

*All’interno di un Contact Center Virtuale*

La soluzione consente alle aziende di gestire tutti i centri e gli operatori disponibili come appartenenti ad un unico Contact Center virtuale, indipendentemente dalla loro ubicazione (telelavoro).

**CTI - Computer to Telephony Integration**

Il CTI è una tecnologia che permette di interfacciare un sistema telefonico con un sistema informatico aziendale. Nello specifico consente di riconoscere il chiamante dal suo numero telefonico, di riconoscere il servizio di cui ha bisogno e di fornire delle informazioni pertinenti alla conversazione, prelevandole da un database aziendale. Le informazioni sul chiamante possono essere di tipo anagrafico, relative ai servizi attivati o al profilo generale del Cliente rispetto ai servizi offerti dall’azienda. Tali informazioni possono essere anche visualizzate (tramite script o pop-up) sul computer di un operatore telefonico in un Contact center, consentendo di fornire informazioni o dare eventuale supporto e assistenza.

IVR e CTI insieme, data la loro capacità di

- elaborare input diversi provenienti da porzioni diverse dei sistemi informativi,
- di adottare e far sua la logica delle transazioni web,
- di fornire supporto strutturato alla gestione della telefonata in quanto tale,
- di fornire al singolo operatore ed alla organizzazione complessiva del call center funzioni di ottimizzazione,
- di ‘collaborare’ in maniera interattiva con gli operatori,
- di ‘riconoscere’ in senso logico le competenze degli operatori da una parte e le motivazioni della chiamata dall’altra e accoppiarle in maniera ottimale,

costituiscono una sorta di espansione del dominio linguistico pertinente al dialogo fra due persone, anche se (diversamente da quanto accade nel dialogo umano tout-court) in un ambito di senso ancora fortemente pre-determinato e limitato.

### **Postazioni Operatori e Supervisor**

L'insieme di strumenti telefonici e non a disposizione degli operatori del Contact center e dei supervisor del gruppo. Visualizzano tutte le informazioni ricevute dal CTI e dal CRM aziendale, e gestiscono le chiamate tramite un applicativo client server o web based, denominato 'barra telefonica'.

### **Server CRM – Customer Relationship Management**

E' il modulo informatico che supporta le funzionalità applicative del Contact center; la piattaforma CRM rende disponibili al CTI tutte le informazioni riguardanti l'utente chiamante. E' solitamente integrato con il CTI tramite una interfaccia dedicata.

- CRM operativo: soluzioni metodologiche e tecnologiche per automatizzare i processi di business che prevedono il contatto diretto con il cliente.
- CRM analitico: procedure e strumenti per migliorare la conoscenza del cliente attraverso l'estrazione di dati dal CRM operativo, la loro analisi e lo studio revisionale sui comportamenti dei clienti stessi.
- CRM collaborativo: metodologie e tecnologie integrate con gli strumenti di comunicazione (telefono, fax, e-mail, etc.) per gestire il contatto con il cliente.

## **7. AMBITI APPLICATIVI PER ASR/TTS**

Nel contesto delle architetture applicative le funzioni di riconoscimento (ASR) e sintesi (TTS) sono una forma come altre di input/output inserite in una architettura applicativa complessiva composta da diverse funzionalità assolutamente fondamentali e senza le quali le funzionalità di ASR e TTS di per sé non hanno alcuna rilevanza applicativa. Alcune di queste funzioni sono:

- la struttura dell'albero di ingresso che consente di instaurare un primo livello di senso fra il chiamante ed il sistema, instaurando un primo livello di scelta semantica e di ambito della conversazione;
- Il call-flow, come struttura sequenziale che contiene il dialogo possibile fra uomo e macchina;
- Il dialog controller che consente di mantenere un controllo sullo stato della transazione
- Le error-recovery strategies che consentono la trattazione della deviazione dal sentiero del dialogo;
- Le exit strategies che istruiscono l'uscita dal dialogo automatico e le sue regole;
- La corretta integrazione con i sistemi legacy e i data base che consentono di reperire dati e informazioni pertinenti alla riuscita del dialogo;
- La corretta integrazione con l'operatore umano che consente di proseguire il dialogo fuori dal rapporto uomo-macchina e nel contesto della relazione più naturale fra due umani (Norman, 2007);
- La qualità intrinseca della transazione, che detta le condizioni dello svolgimento del dialogo dentro un ambito semantico preciso;
- La certezza del risultato, cioè il successo della transazione linguistica;
- I canali telefonici il cui corretto dimensionamento garantisce la disponibilità delle risorse necessarie ad accogliere il flusso telefonico;

- Le prestazioni tecniche complessive della architettura che – correttamente dimensionata – garantisce il successo della transazione stessa;

Qui di seguito alcuni esempi di transazioni anche complesse che si possono proficuamente e anche con relativa soddisfazione dell'utente svolgere completamente in automatico:

- **Pagamenti** : è possibile gestire in forma completamente automatica transazioni di prenotazione ed acquisto nonché pagamenti effettuati tramite carta di credito, tutto in ambito interamente vocale. Queste transazioni consentono di
  - Selezionare l'oggetto dello acquisto, sia esso un prodotto o un posto a teatro o allo stadio (in questo caso scegliendo anche giorno e ora),
  - Declinare eventuali titolarità a sconti,
  - Effettuare il pagamento tramite carta di credito.
- **Posto operatore Automatico (Directory Assistant)**: queste applicazioni sono in grado di sostituire un operatore di reception telefonica a tutti gli effetti (vedi figura sotto). Si può:
  - Chiedere di parlare con una persona connotandola attraverso il nome e/o cognome o il ruolo,
  - Risolvere eventuali omonimie,
  - Effettuare il passaggio all'interno telefonico della persona,
  - Gestire la richiesta di ricerca multipla del destinatario (su numeri diversi da quello digitato in prima istanza),
  - Gestire messaggi.
- **Localizzazione**: informazioni relative ai punti vendita o di servizio anche in modalità multilingua e geograficamente distribuiti. Attraverso la pronuncia del nome della località si agganciano una serie di informazioni di riferimento
- **Voice-Mail**: si può 'leggere' la propria casella di mail dando comandi vocali di 'navigazione' e rispondere alle mail attraverso semplici comandi vocali.
- **Banking and Trading Info**: si può
  - controllare e gestire il conto corrente,
  - avere informazioni sul proprio portafoglio titoli e sui movimenti,
  - effettuare operazioni bancarie,
  - ricevere informazioni sulle quotazioni di Borsa.
- **Controllo accessi**: è possibile gestire contemporaneamente funzioni di verifica della impronta vocale e di password di accesso.
- **Informativa su orari**: è possibile ottenere informazioni su treni, aerei, spettacoli, etc. dando informazione sulla data e ora richiesta.

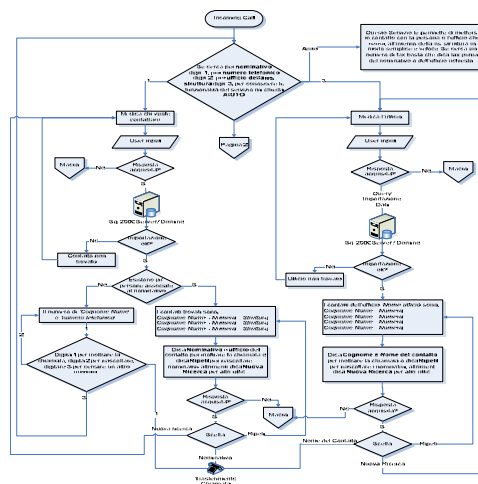


Figura 3: esempio di flusso di dialogo di un servizio di ‘directory assistance’.

L’elenco esemplificativo si riferisce non più ad applicazioni sperimentali bensì a transazioni funzionanti quotidianamente a beneficio a volte di migliaia di utenti.

Viene naturale pensare che l’ambito di applicazione più promettente sia quello che vede la automazione della voce come concorrente di internet, in virtù della maggiore ‘mobilità’ della telefonia rispetto al pc. Sicuramente il successo di tali applicazioni sarà progressivamente determinato dalla capacità della tecnologia di evolvere e di rendere la transazione telefonica automatica molto più ‘confortevole’ (non ‘natutale’) di quanto sia oggi.

## 8. ESEMPI DI APPLICAZIONI FUNZIONANTI IN AMBITO AZIENDALE

Quella che segue è una rassegna di progetti svolti negli ultimi sei anni ed effettivamente andati in produzione. In ciascuno di essi ci sono funzionalità IVR avanzate con uso di tecniche di ASR e TTS. Tutte le applicazioni sono state o sono ancora pienamente utilizzate da un pubblico vasto e disomogeneo di utenti. I ‘success-rate’ di tali transazioni sono superiori all’ordine dell’85% complessivamente e comunque tali da non aver mai ingenerato problematiche di insoddisfazione sensibile da parte della clientela. I casi sono citati per gentile disponibilità delle imprese che hanno commissionato ad Abil il progetto e ne sono destinatarie. Come si potrà evincere si tratta essenzialmente di aziende operanti nel settore dei servizi (ed in particolare negli ambito bancario, assicurativo, media, entertainment, utilities, telco) caratterizzate da una base di clienti molto vasta ed eterogenea. la tipologia dei progetti spazia dalla realizzazione di servizi di tipo informativo ad altri di tipo dispositivo.

In alcuni casi l’utilizzo di ASR è estremamente limitato e – anzi – quasi dispensabile; a testimonianza che nella esperienza pratica le capacità applicative complessive che intervengono nella gestione complessiva del dialogo sono altrettanto importanti e anzi rappresentano una linea evolutiva altrettanto forte di quella rappresentata dalla capacità intrinseca di riconoscimento.



### *8.1 Caso della Banca Telefonica di Fineco.*

Fineco è la principale Banca On-Line italiana. Nata come società di trading-online, nel tempo ha iniziato a fornire servizi bancari più tradizionali attraverso diversi canali: web, telefono, sportello, agenzia, rete di promotori. Per il suo utilizzo spinto della multicanalità essa rappresenta un paradigma che va al di là del semplice concetto di on-line, ponendosi come esempio e modalità di gestione della relazione col pubblico in una ottica molto articolata.

Il progetto ha realizzato in ASR combinato con operatore di call center, servizi di orientamento e informativa relativi a transazioni di home-banking, trading e gestione delle carte di credito. Il servizio funziona da anni e raccoglie milioni di utenti con soddisfazione della banca e dei clienti stessi. È un esempio di come le tecnologie vocali si prestino molto bene alla gestione

- di transazioni informative anche complesse,
- con un target di clienti di estrazione tecnologica (provengono dal web),
- molto fidelizzati ed esigenti,
- realizzando un grande risparmio di tempo operatore secondo la logica di dedicare l'operatore alle parti più complesse della relazione,
- in coordinamento logico con altri canali.

### *8.2 Caso del Pagamento dei Canoni e Tributi di Esatri/Banca Intesa*

Esatri è la società di esattoria del Gruppo Banca Intesa (prima della recente ristrutturazione del settore con la costituzione di un'unica società a livello nazionale). Essa riscuote per conto di enti terzi, i canoni per i servizi resi da questi ultimi, fra cui il canone della RAI e tributi locali per conto di Comuni. La applicazione consente il pagamento a comandi vocali di canoni e tributi con carta di credito. Il clou della applicazione consiste nell'identificare l'utente chiamante, controllare la sua posizione (cartella), proporre il pagamento ed eseguirlo riconoscendo gli estremi della carta di credito e veicolando la transazione su un circuito interbancario per la esecuzione.

Da un punto di vista tecnologico il riconoscimento opera in un ambito molto ristretto la cui parte più complessa è l'acquisizione della stringa numerica caratteristica del numero di carta di credito. Il contesto però è delicatissimo perché eventuali errori o il fallimento della transazione hanno conseguenze serie sulla tenuta e percezione del servizio. Inoltre gli obiettivi ascritti al servizio sono particolarmente ambiziosi perché ci si aspetta che consenta di operare con un numero di operatori di call-center costante a prescindere dai picchi che le scadenze connesse ai pagamenti determinano.

### *8.3 Caso del Portale Vocale di WIND*

Nonostante il sistema non sia più in produzione, si tratta del caso più appassionante dal nostro punto di vista. Si tratta di una sorta di 'traduzione editoriale' di servizi web-like in tecnologia vocale: chiamando un unico numero a pagamento il cliente viene autenticato automaticamente ed è in grado di accedere a diversi servizi anche molto diffusi fra loro: lettura della sua posta elettronica, lettura i notiziari, lettura delle notizie di Borsa, Info-traffico, racconti e storielle, etc.

In particolare la gestione della Posta Elettronica consente di leggere le proprie mail e di rispondere alle stesse sempre a comandi vocali.

Il modello di business che sottende alla applicazione è anch'esso singolare: un service provider fornisce la funzionalità della piattaforma; il carrier telefonico (in questo caso WIND) fornisce (e attira) traffico voce; i content-provider forniscono contenuti 'editoriali'.

Da un punto di vista tecnologico l'ASR fornisce funzionalità di word-spotting al fine per l'utente di orientarsi facilmente sul menù principale e funzioni di Barge-in (scorciatoie) per passare da un servizio all'altro qualora sia già esperto.

#### *8.4 Caso del Portale Vocale di AEM/A2A*

L'aspetto rilevante di questo caso è l'eshaustività dei servizi resi ed un uso dell'ASR completamente integrato con l'operatore: p.e., nel caso del servizio di lettura automatica del contatore dei consumi di Luce e Gas, se la transazione automatica si trova in uno stato di impasse dovuta a qualsiasi tipo di errore o incomprensione, la telefonata viene passata ad un operatore che ha a disposizione una visione precisa e dettagliata di dove si trovasse la conversazione fra utente e sistema automatico. Questo per evitare la antipatica circostanza (spesso diffusa) di dover ri-spiegare tutto da capo.

Inoltre le telefonate vengono automaticamente dirottate sull'operatore che meglio può rispondere alla tipologia di chiamata e motivazione della stessa.

#### *8.5 Caso dell'informativa orari di Trenitalia (sviluppata insieme a TI e Avaya)*

Si tratta di un'applicazione molto semplice ma di notevole effetto: l'utente chiama il numero a pagamento di Trenitalia e con un dialogo strutturato viene messo a conoscenza degli orari di partenza e arrivo su qualsiasi tratta nazionale indicando a sua volta una partenza, una destinazione, un giorno e un'ora del giorno. Il naturale sviluppo di codesta applicazione è la capacità (già tecnicamente possibile) di acquistare il biglietto del treno. L'applicazione è semplice e studiata in maniera che l'utente riceva indicazione di un certo numero di treni 'a partire' dall'orario richiesto e che possa indicare il giorno preferito in guisa diversa (data precisa oppure con un avverbio tipo 'ieri' oppure un giorno della settimana, etc.). L'applicazione rappresenta un naturale complemento all'uso del web e costituisce il classico 'trade-off' virtuoso che consente di risparmiare tempo dell'operatore per dedicarlo a compiti di relazione più delicati e complessi.

#### *8.6 Caso del Trading Banca Sella*

Diversa nel dominio di applicazione ma simile nella logica alla transazione precedente: all'utente (cliente registrato della Banca) vengono date in automatico informazioni sulle quotazioni di Borsa e sulla composizione del proprio portafoglio (che costituiscono più del 70% delle telefonate) per essere – invece – indirizzato ad un operatore per svolgere transazioni (ben più impegnative) di compra-vendita.

#### *8.7 Caso dei localizzatori di ATM di Banca Popolare di Novara*

Si tratta di una applicazione che ha qualche anno di vita e che consente ad un utente chiamante di ottenere informazioni relativamente agli ATM (Bancomat) disponibili e funzionanti della Banca nel raggio di tre distanze da un certo comune indicato nella telefonata. L'applicazione, questo il suo punto di interesse tecnico, gestisce il riconoscimento di tutti e più di 8.000 comuni italiani ed è in grado di risolvere la ambiguità determinata da eventuali omonimie in maniera molto semplice ed efficace.

In tutti i casi esposti il successo della applicazione è dipeso da tre fattori assolutamente non disgiunti:

- capacità di riconoscimento efficaci ed efficienti;
- strategie applicative di gestione del dialogo semplici, prevedibili e lineari,
- capacità di sviluppare correttamente il modello di interfaccia ai sistemi 'core'.

Non dimentichiamo – infatti – che in contesti applicativi è la rilevanza dei contenuti specifici della chiamata a costituire l'elemento più interessante per l'utente e per l'azienda.

## 9. EVOLUZIONE PREVEDIBILE

Perché possa diffondersi in maniera significativa l'uso della Voice Automation dovrà subire un'evoluzione impegnativa: oggi le transazioni, anche quelle di maggior successo, si svolgono in ambiti semantici estremamente ristretti e risentono di forti limitazioni e rigidità da un punto di vista del comfort complessivo e delle performance.

Qui di seguito indichiamo due plausibili linee di sviluppo.

Uno è costituito dal 'puntare' sulla emulazione del linguaggio naturale: questa ipotesi si basa sulla possibilità di far evolvere le prestazioni tecnologiche dei sistemi a tal punto da poter simulare il comportamento umano sia nelle prestazioni di riconoscimento che in quelle di sintesi.

L'ipotesi alternativa è quella di accettare il dato di fatto che il linguaggio umano è una pratica troppo complessa per essere gestita da una tecnologia e lavorare il più possibile al miglioramento delle prestazioni in contesti più ristretti e specializzati, magari rendendo esplicito questo aspetto: in questa ipotesi si assisterebbe - per fare un esempio – alla ricerca del compromesso fra ampiezza del dizionario e prestazioni del riconoscimento.

Entrambe le linee evolutive potrebbero temporaneamente prendere la piega di un approccio riduzionistico sia sul lato dell'utenza ('invitare' gli utenti ad adattarsi alle macchine) che su quello delle realizzazioni (restringere l'ambito e la qualità delle applicazioni).

In ognuno dei due casi si possono immaginare alcune linee evolutive, classificabili per comodità a seconda che riguardino le pure funzionalità di speech recognition (A), gli aspetti di architettura (B) oppure le applicazioni (C):

### A) Evoluzione delle funzioni ASR-Automatic Speech Recognition:

- sviluppo di grammatiche specialistiche a supporto di ambiti applicativi più ristretti ma con maggiori capacità di riconoscimento all'interno del dominio circoscritto (p.e. gerghi specialistici, scientifici, professionali, dialetti),
- sviluppo del parlato spontaneo ("say anything"), anch'esso focalizzato in ambiti prevedibili,
- capacità di gestire in maniera più selettiva i disturbi ambientali (rumori, sottofondi, pluralità di voci, etc.),
- evoluzione della analisi semantica del parlato (già perseguita con successo nello scritto): è un campo di evoluzione molto interessante perché consentirebbe lo sviluppo di applicazioni in ambiti come il marketing, permettendo di riconoscere e discriminare 'pattern' significativi relativi al profilo dell'utente, ai suoi orientamenti alla sua 'experience' proprio e anche attraverso le transazioni telefoniche.

### B) Evoluzione delle architetture:

- la integrazione con altri canali non telefonici come il web, la mail ma anche il video è sicuramente uno dei campi più promettenti perché in linea con la esigenza già molto sentita di incoraggiare la convivenza fra i canali come elemento utile a migliorare il contatto fra utenza e aziende;
- tecnicamente si intravede la tendenza all'uso di architetture clusterizzate, con uso di Web Service e forte integrazione con contenuti web. La correlazione fra Voice Automation e web è molto forte proprio nella misura in cui la prima rappresenta

una sorta di 'ritorno di fiamma' dell'uso del telefono in una epoca di predominio del web. Questo rende inevitabile la ricerca dell'integrazione tecnico-funzionale fra questi due canali, essendo spesso i contenuti 'editoriali' comuni.

### **C) Applicazioni**

- i cosiddetti Portali Vocali possono affiancare in maniera efficace i portali web, trasformando il telefono in un vero e proprio strumento editoriale;
- applicazioni legate alla mobilità (Work Force Management);
- il potenziamento dei servizi Self Service in ambiti come quello bancario, assicurativo, dei trasporti, degli spettacoli, etc.;
- il supporto che tecniche ASR daranno agli operatori umani in un quadro di cooperazione spinta;
- i sistemi di sicurezza, cioè accessi, ma anche sistemi di alert, sono un altro campo promettente;
- i call center multietnici in grado di accogliere senza discriminazione molte lingue;
- lo sviluppo di agenti interattivi in grado di orientare il dialogo in virtù di una serie di circostanze e informazioni, fino ad un vero e proprio adattamento dinamico all'utente;
- funzionalità editoriali come per esempio la lettura di notiziari, la trascrizione e analisi di stream-audio, la lettura di rassegne stampa, la analisi e classificazione semantica di voice-streaming.

Per quanto aleatoria e deprecabile possa essere ritenuta la pratica della predizione del futuro, non ci siamo sottratti perché pensiamo che evocare il futuro possa essere un'assunzione di impegno a renderlo migliore nella prassi professionale di tutti i giorni.

### **RINGRAZIAMENTI**

Vorremmo ringraziare l'organizzazione del convegno AISV; gli studenti dell'università della Calabria per la loro cortesia e attenzione; i professori Francesco Cutugno dell'Università di Napoli e Piero Così di Padova per la occasione offertaci di parlare ed ascoltare di cose che ci stanno a cuore professionalmente e i nostri clienti che ci consentono di sperimentare nella pratica ed in contesti sfidanti le tecnologie e le applicazioni.

### **BIBLIOGRAFIA**

- Bandler R. & Grinder J. (2007), *La struttura della magia*, Ed. Astrolabio.
- Chomsky N. (1975), *Grammatica trasformativa*, Ed. Bollati-Boringhieri.
- De Saussure F. (2007), *Corso di linguistica generale*, Ed. Laterza.
- Galimberti U. (2006), *Psiche e Techne*, Ed. Feltrinelli.
- Gargani A. G. (2007), *Il sapere senza fondamento*, Ed. Feltrinelli.
- Norman D. (2007), *Il design del futuro*, Ed. Apogeo.
- Ong W. J. (1986), *Oralità e scrittura*, Ed. Il Mulino.
- Wittgenstein L. (1918), *Tractatus Logico-philosophicus*, Ed. Einaudi.

# **ACUSTICA & RESTAURO**

# UN PROTOCOLLO DELLE PROCEDURE DI RESTAURO ALL'INTERNO DELL'ARCHIVIO SONORO CALABRESE<sup>1</sup>

Luciano Romito, Vittorio Scullari  
Laboratorio di Fonetica, Università della Calabria  
[luciano.romito@unical.it](mailto:luciano.romito@unical.it), [yik77kd@yahoo.it](mailto:yik77kd@yahoo.it)

## SOMMARIO

In questo lavoro vengono presentate alcune tecniche di elaborazione del segnale audio, con le procedure e le applicazioni di plug-in di Noise Reduction utilizzati da alcuni software in commercio (Sound Forge, Cakewalk ProAudio, Steinberg WaveLab, IQS SAWPro e Vegas).

Il lavoro presenta inoltre un esame delle caratteristiche dell'*Archivio Sonoro Calabrese*, delle procedure di intervento e dell'analisi del segnale di un campione rappresentativo di registrazioni con lo scopo di giungere alla stesura di un protocollo comune all'interno dell'archivio.

## 1. RESTAURO DIGITALE AUDIO

Le registrazioni audio, soprattutto se riguardanti incisioni datate, su supporti mal conservati o usurati, possono presentare dei disturbi che compromettono la qualità del segnale. I disturbi più diffusi nelle registrazioni sonore sono i rumori impulsivi, come i "clicks", i rumori a banda larga, quali il fruscio dei nastri magnetici e i rumori di tipo armonico come ad esempio i 50 Hz della portante.

Per migliorare il segnale audio di queste registrazioni sono stati elaborati degli algoritmi che consentono di intervenire adeguatamente in base alle caratteristiche del rumore.

Se l'approccio al materiale da restaurare è legato a esigenze archivistiche, si vorrà conservare un suono più vicino possibile all'originale, con lo scopo di ottenere una copia esatta, generalmente in digitale, del segnale presente sul supporto analogico.

Nel caso delle copie d'archivio non si deve parlare di vero e proprio restauro, ma piuttosto di riversamento. Con questa procedura tutti i rumori, i sibili, i fruscii, i *crackles*, i ronzii presenti sul supporto analogico originale verranno mantenuti anche nella copia digitale.

Se, invece, si vuole ripulire la registrazione da questo tipo di disturbi occorre utilizzare delle tecniche digitali di elaborazione del segnale. Rispetto alle tecniche analogiche, l'elaborazione digitale consente di ottenere risultati di gran lunga superiori e un alto grado di flessibilità di lavoro, ma un loro utilizzo in maniera indiscriminata può produrre risultati disastrosi.

La maggior parte degli algoritmi di restauro audio sono basati sulla progettazione di un modello matematico che rispecchi le caratteristiche del segnale. I migliori risultati si otterranno, dunque, da quegli algoritmi che utilizzino un modello più vicino possibile al segnale audio.

---

<sup>1</sup> L'archivio sonoro calabrese è frutto di un progetto di ricerca del Laboratorio di Fonetica dell'UNICAL dal titolo: *I corpora vocali come patrimonio culturale: archiviazione, conservazione, restauro di materiale sonoro calabrese* meglio dettagliato in Romito, Lio 2006.

### *1.1 Intervento sul rumore a banda larga: il Noise Reduction*

Il sistema di *Noise Reduction* è progettato per analizzare e rimuovere dalle registrazioni il rumore di background causato dal motore elettrico del sistema di registrazione, dall'ambiente di registrazione e dal deterioramento del supporto.

La procedura per l'eliminazione di un rumore a banda larga le cui caratteristiche sono "statiche" e note a priori, viene denominata "*DeNoise*". Diversamente da un normale filtro, con questo sistema si può intervenire senza rimuovere parte del segnale originale. Tutto ciò grazie alla possibilità di separare l'audio nelle sue componenti frequenziali e di usare un "noiseprint" (*impronta del rumore*) per distinguere tra rumore indesiderato e segnale da conservare.

Analizzando una porzione della registrazione in una zona di silenzio, il software è in grado di immagazzinare informazioni sulla frequenza e sull'ampiezza del rumore. L'operatore potrà determinare quale zona dello spettro di frequenza sia necessario eliminare.

Dunque, la procedura è in due fasi, la "cattura" dell'impronta del rumore e la sottrazione dello spettro del rumore dal segnale originale.

I tre principali strumenti di processamento del plug-in testato, legati al *Noise Reduction*, sono il *Reduce Noise by*, il *Reduction Type* e il *Noise Bias*. Nella maggior parte dei casi, usando le opzioni predefinite e aggiustando leggermente questi tre strumenti si ottengono eccellenti risultati

#### *1.1.1 Reduction Type (Mode 0, Mode 1, Mode 2 o Mode 3)*

Lo strumento di *Reduction Type* consente di scegliere tra quattro diversi algoritmi che agiscono sulla rimozione del rumore di fondo in maniera leggermente differenziata.

Il Mode 0 è indicato per la rimozione di rumori di livello molto basso ed è in grado di eliminare grandi quantitativi di rumore. Tuttavia, se si riduce troppo rumore o si regola lo strumento *Noise Bias* a un valore troppo basso, in molti casi si creeranno degli artefatti udibili. Il Mode 1 è simile al precedente. La sua applicazione riduce il rischio di artefatti, ma consente una eliminazione del rumore inferiore. Questa considerazione è valida per ogni altro Mode: il 2 e il 3 generano meno artefatti ma eliminano meno rumore. La modalità predefinita (la migliore per la maggior parte dei casi) è la 2. Ricordiamo però che la riduzione del rumore dipende dalla qualità del segnale originale.

#### *1.1.2 Reduce Noise By*

Questo è lo strumento principale da regolare, è quello che controlla la "quantità" di rumore da rimuovere. Una regolazione di questo strumento a valori elevati consente l'eliminazione di gran parte del rumore, ma bisogna procedere con cautela, perché è facile introdurre artefatti indesiderati. Il settaggio migliore nella maggior parte dei casi va da 10 a 20 dB.

Nel caso di rumore ostinato è consigliabile procedere con passaggi a 10 e 20 dB, "ricattare" una nuova impronta del rumore e quindi riprocessare il file. Due passaggi di 20 dB solitamente danno migliori risultati che un singolo passaggio a 40 dB.

#### *1.1.3 Noise Bias (-20 to +20 dB)*

Il settaggio predefinito a 0 dB di questo strumento è generalmente efficace. Una regolazione a valori più elevati rischierà di creare artefatti udibili, un suono attenuato,

ovattato, mentre valori inferiori potrebbero non essere sufficienti a eliminare il rumore o, se si utilizza il Mode 0 del *Reduction Type*, si possono introdurre artefatti striduli.

#### 1.1.4 Keep Residual Output

Il Keep Residual Output inverte il processo di *Noise Reduction*, consentendo di ascoltare il rumore che si sta rimuovendo. Può essere uno strumento prezioso perché permette di sentire se si sta eliminando parte del segnale originale che in realtà si vorrebbe tenere.

Per esempio, se si sta riducendo del rumore di sottofondo su delle registrazioni di parlato e utilizzando questo strumento si percepisce la presenza di una parte di parlato nel segnale in uscita, vorrà dire che si sta compromettendo la qualità del segnale originale. Quindi bisognerà modificare i settaggi in modo tale che il segnale in uscita non contenga o contenga in piccole quantità, il segnale che si vuole conservare.

Altri strumenti del *Noise Reduction* sono l'*Attack Speed* e il *Release Speed*, che determinano la velocità con cui l'algoritmo reagisce ai cambiamenti del livello del rumore.

Il segnale restaurato con il "*DeNoise*" appare generalmente meno "brillante" (questo a causa del taglio delle alte frequenze). I software che abbiamo testato posseggono uno strumento (*Lost High Frequencies*) capace di compensare questa perdita. Si tratta di un filtro high-shelf che ridà "brillantezza" al segnale trattato.

Un altro utile strumento è l'*FFT Size*, che determina la dimensione (in campioni) della finestra di analisi e il numero di frequenze discrete usate durante la riduzione del rumore. Il valore predefinito di 2,048 è consigliato per la maggior parte dei casi. Con un settaggio a valori più elevati si ottiene una risoluzione migliore, ma sono necessari tempi di processamento più lunghi. Strumento legato all'*FFT Size* è l'*Overlap*, col quale si può stabilire l'ammontare di overlap (sovrapposizione) tra le finestre di analisi dell'*FFT Size*.

Nel caso in cui non fosse possibile disporre di un'impronta del rumore, qualora ad esempio le caratteristiche del rumore fossero variabili nel tempo, per eliminare il rumore a banda larga, si può utilizzare un algoritmo basato su un modello standard del rumore stesso, il "*DeHiss*".

Non essendo possibile conoscere a priori l'entità di questo tipo di rumore presente ogni istante in ognuna delle bande di frequenze occupate dal segnale, non è possibile eliminare il rumore per semplice sottrazione. Per questo, in base a modelli standard di rumore, si applica questa procedura di ottimizzazione.

Da notare che l'algoritmo di *Noise Reduction* funziona bene per rimuovere rumore di sottofondo costante. Cambiamenti rapidi o rumore "irregolare" (come suoni del traffico) non sono indicati per questo algoritmo. Inoltre, nei casi in cui il livello di volume del rumore sia uguale o maggiore del suono, può essere difficile, se non impossibile, ripulire il segnale senza che vi sia l'introduzione di artefatti udibili.

Altro limite del "*DeNoise*" è che questa procedura spesso introduce un disturbo di tipo musicale (*musical noise*) che compromette la qualità del segnale. Per ovviare a questo problema, sono stati sviluppati degli innovativi algoritmi capaci di ridurre sensibilmente il rumore musicale. In particolare sono degni di nota il *Filtro Ephraim-Malah* (EMSR) e le sue successive evoluzioni.



### *1.2 Intervento sui disturbi impulsivi: “Click and Crackle Removal”*

La maggior parte delle vecchie registrazioni, soprattutto su disco, a causa di graffi causati ad esempio dalla polvere, spesso sono afflitte da disturbi di tipo impulsivo. I clicks generano discontinuità nella forma d'onda. Il disturbo colpisce una sequenza di campioni, con inizio e durata casuale, e può avere un'ampiezza molto variabile anche all'interno della stessa registrazione.

Il processo di eliminazione dei disturbi impulsivi si divide in due fasi: l'algoritmo è in grado di individuare i fenomeni non “in linea” con le normali caratteristiche del segnale audio e una volta individuati, li elimina progressivamente. L'algoritmo riconosce la posizione del click, analizza lo sviluppo dell'onda prima e dopo il click e quindi, in base alle informazioni ricavate da questa analisi, ricostruisce la forma d'onda.

La funzione di *Click and Crackle Removal* del plug-in di *Noise Reduction* dispone di diversi settaggi predefiniti che, nella maggior parte dei casi, sono sufficienti a ripulire il segnale adeguatamente. Se si interviene manualmente sui singoli click, è necessario non selezionare solo il disturbo, ma lasciare un margine di qualche millisecondo attorno al fenomeno impulsivo, in modo tale da permettere all'algoritmo di ricostruire il più precisamente possibile la forma d'onda originale.

Il *Click and Crackle Removal* consente di regolare diversi controlli.

#### *1.2.1 Sensitivity (1 to 20)*

Questo strumento regola la sensibilità con cui l'algoritmo individua i disturbi impulsivi.

Con settaggi al di sotto di 10, solo i clicks particolarmente significativi verranno rimossi. Tra 10 e 17 saranno rimossi anche i clicks più piccoli. Sopra i 17 saranno rimossi anche i minuscoli clicks che causano crepitio.

Le registrazioni su LP in vinile solitamente possono essere ripulite in maniera soddisfacente con settaggi tra 12 e 16.

#### *1.2.2 Click Shape (1 to 8)*

Il controllo determina il tipo di click che l'algoritmo individua.

La regolazione consigliata nella maggior parte dei casi è quella tra 4 e 6. Settaggi più elevati sono indicati per individuare piccoli clicks ad alte frequenze, (per esempio in registrazioni su LP), mentre settaggi inferiori al 4 sono indicati per registrazioni su vecchi 78 giri, dove il contenuto spettrale è maggiormente distribuito sulle basse frequenze.

#### *1.2.3 Max Click Size (0.1 to 3.0 milliseconds)*

Con il *Max Click Size* si determina la lunghezza massima del click che l'algoritmo individuerà. Prima di utilizzare questo strumento è importante tener presente il contenuto registrato che si sta per trattare. Se si mettono a confronto le forme d'onda dei clicks con quelle di suoni percussivi (come piatti, nacchere, ecc.) si noterà una certa somiglianza, anche se quest'ultimi hanno generalmente una durata maggiore. Per evitare che l'algoritmo identifichi questi suoni come disturbi impulsivi, è consigliabile regolare il *Max Click Size* su valori non superiori agli 0.8 millisecondi e successivamente rimuovere eventuali disturbi residui.

#### 1.2.4 Noise Level (low, medium, high)

Utilizzando questo strumento il segnale in entrata viene diviso in porzioni di rumore e non rumore e l'algoritmo eliminerà solo i clicks presenti nelle zone di rumore. C'è da dire però che l'uso di questi controlli spesso causa l'introduzione di artefatti udibili.

Se si attiva la funzione *Remove Low-Frequency Rumble* le frequenze molto basse (sotto i 30 Hz) saranno rimosse. Questo strumento può aiutare l'algoritmo a individuare i clicks, ma può modificare la forma d'onda, soprattutto in vecchie registrazioni dove la presenza di rumore a frequenze molto basse è elevata. Quando si rimuovono clicks da brevi porzioni di segnale (come un secondo) è consigliabile disattivare questa funzione, per preservare la continuità della forma d'onda nei punti finali.

Come abbiamo visto per il "DeNoise", anche il *Click and Crackle Removal* possiede lo strumento *Keep Residual Output*, che consente di ascoltare solo il suono che è stato rimosso.

#### 1.3 Clipped Peak Restoration

Talvolta, per errori di disposizione e regolazione della strumentazione, il segnale acustico satura gli amplificatori e viene dapprima deformato e poi tagliato di netto agli apici. All'ascolto questo tipo di distorsione produce una grave compromissione della qualità sonora.

Il "DeCrackle" (utile per affrontare questo problema) usa una funzione specifica che consente di interpolare le zone danneggiate e ricostruirle con un processo di re-sintesi.

Il plug-in da noi testato contiene questo algoritmo nel **Clipped Peak Restoration**.

Il *Clipped Peak Restoration* consente di recuperare il segnale perso a causa dei tagli e riduce la distorsione, "arrotondando" l'estremità dei picchi tagliati e applicando un limite di picco all'area immediatamente circostante il taglio. Le parti di audio non interessate dalla saturazione non vengono coinvolte nel processo. L'intervento è diviso in due fasi: per prima cosa si deve garantire al sistema uno spazio (*headroom*) sufficiente a ricostruire la forma d'onda. Dopodiché, se il segnale è saturato, bisogna ricampionarlo a un livello di input inferiore (da -6 a -10 dB) e quindi testare la qualità di ricostruzione, impostando i vari parametri presenti nel programma.

Il *Clipped Peak Restoration* dovrebbe essere usato solo su materiali in cui la saturazione riguarda piccole parti della registrazione. Materiale pesantemente distorto, con aree dove quasi tutti i picchi sono tagliati, è spesso irrecuperabile.

I settaggi predefiniti disponibili nel *Clipped Peak Restoration* sono spesso sufficienti a migliorare il segnale (ricordiamo però sempre che il risultato finale dipende dalla qualità di partenza della registrazione).

Il controllo *No attenuation, limit clips* attenua solo i picchi tagliati; il *-3 dB with limiter* riduce il volume complessivo del file audio di 3 dB e applica una compressione ai picchi tagliati che evita di introdurre ulteriori tagli; il *-6 dB, no limiter* riduce il volume complessivo del file audio di 6 dB.

Il controllo *Attenuation* consente di regolare manualmente il livello di volume da attenuare da 0 a -24 dB. Se si vuole applicare manualmente anche una compressione ai picchi tagliati prima del restauro bisogna attivare la funzione *Enable Post Limiter*.

Per evitare il rischio di introdurre clicks udibili durante il restauro con *Clipped Peak Restoration* attivare la funzione *Crossfade Edges*.

#### *1.4 Altre tecniche di elaborazione del segnale*

##### *1.4.1 Controllo automatico del guadagno*

Grazie alla possibilità di stimare la potenza media del segnale nelle regioni di “non-silenzio”, questa procedura consente di mantenere ottimale il livello d’ampiezza del segnale audio target, grazie alla possibilità di modificarne i valori d’intensità.

##### *1.4.2 Compressione ed espansione della dinamica del segnale*

Questa funzione consente di differenziare le regioni del segnale in base alle caratteristiche d’ampiezza. Così facendo è possibile modificare selettivamente l’ampiezza, aumentandola nelle regioni più deboli e mantenendola invariata in quelle più forti. È uno strumento utile, ad esempio, nel caso di registrazioni di parlato, per migliorare l’intelligibilità del segnale vocale, dove vi sia presente un elevato rumore di sottofondo.

##### *1.4.3 Intervento sui disturbi armonici: “DeBuzz”*

Molta parte del rumore presente nell’ambiente è causato da motori, ventilatori o altro dispositivo che produce un tipo di rumore continuo o a tratti, caratterizzato spettralmente da un gran numero di armoniche. È il caso, ad esempio, della portante a 50 Hz. Questo tipo di disturbi possono essere trattati con dei filtri a pettine (*comb filters*) o con algoritmi che catturano la frequenza fondamentale e cancellano le armoniche mediante iniezione di componenti armoniche in opposizione di fase.

##### *1.4.4 Speech Enhancement*

Le tecniche digitali fin qui analizzate consentono di migliorare la qualità delle registrazioni audio (anche se non in modo miracolistico). A meno che il segnale non sia eccessivamente compromesso, il degrado raramente è irreversibile. L’operazione di filtraggio può migliorare il rapporto segnale-rumore, ma non in tutti i casi può far ottenere i risultati sperati. Se, ad esempio, l’obiettivo è il miglioramento dell’intelligibilità dell’informazione, e si ha a che fare, ad esempio, con una registrazione in cui il rumore di fondo è composto dalla sovrapposizione di più voci, i filtri sono del tutto inutili.

Con lo scopo di migliorare specificatamente il segnale vocale, sono stati sviluppati degli algoritmi di *Speech Enhancement*. Questo software è in grado di modificare selettivamente l’ampiezza di sezioni di segnale. Amplificando le regioni del segnale più deboli (come i suoni nasali o fricativi), si differenziano le varie categorie fonetiche, aumentando così l’intelligibilità complessiva del messaggio verbale.

Le tecniche qui descritte sono quelle più diffusamente utilizzate nel restauro audio, e derivano da un adattamento del filtro di Wiener nel dominio della frequenza.

Negli ultimi anni sono stati sviluppati altri algoritmi utilizzati nel settore, come le tecniche che utilizzano un modello psicoacustico dell’orecchio umano (che hanno lo scopo di rimuovere solo le componenti udibili di rumore, lasciando così il segnale più inalterato possibile), e tecniche che agiscono nel dominio del tempo, come il *Filtro di Kalman Esteso* (EFK).

## 2. RIVERSAMENTO E RESTAURO: L'ARCHIVIO SONORO CALABRESE

### 2.1 *Composizione dell'Archivio Sonoro Calabrese*

L'Archivio Sonoro Calabrese, che si trova presso il Laboratorio di Fonetica del Dipartimento di Linguistica dell'Università della Calabria, è costituito da centinaia di registrazioni su audiocassette, microcassette, bobine e minibobine magnetiche.

Questi documenti sonori provengono in buona parte dalla cattedre di Antropologia Culturale, di Etnolinguistica, di Storia delle Tradizioni Popolari, Dialettologia, Fonetica e Sociolinguistica dell'Unical, ma anche da registrazioni provenienti da singoli privati, comuni o fondazioni. Il contenuto di questi documenti è composto da interviste, canti religiosi, racconti popolari, studi linguistici e sociolinguistici, lettura di brani, racconti e liste di parole, ecc., tutto materiale registrato a partire dalla fine degli anni '60 in buona parte della Calabria, che riguarda sia i dialetti della regione sia delle minoranze linguistiche come quelle arbëreshe, grecaniche e provenzali.

Queste registrazioni sono rappresentative di testimonianze di usi, costumi e tradizioni calabresi e ciò ne rileva la grande importanza storica, culturale e scientifica. La mancata conservazione o restauro causerebbero una grave perdita sul piano del patrimonio culturale calabrese.

#### 2.1.1 *Procedure d'intervento*

Prima di procedere al riversamento è necessario seguire una serie di operazioni preliminari, quali il restauro del supporto (pulitura, riparazioni, avvolgimenti dei nastri) e la verifica dei sistemi e delle attrezzature utilizzate per la registrazione.

Con lo scopo di preservare l'identità storica dei documenti, si è scelto di riversare la registrazione presente sui supporti analogici in formato digitale, senza modificare il contenuto originale.

Il restauro digitale viene effettuato solo su una copia del file riversato, previa un'approfondita analisi acustica del segnale.

Dunque, il nostro archivio conterrà una copia filologicamente aderente all'originale, comprensiva di rumori e alterazioni dovute sia al tempo che allo strumento di registrazione e se necessario, una copia restaurata.

La copia conservativa è stata effettuata con una attrezzatura professionale, compatibile con la strumentazione originale, con lo scopo di minimizzare la perdita di informazione accusata durante il processo di riversamento.

Nonostante la gran parte del materiale sia in buono stato, in alcuni casi ci si è imbattuti in alcuni supporti mal conservati o usurati. Il processo di restauro, in questi casi, dovrà seguire un iter in due distinte fasi: restauro fisico del supporto danneggiato e successivamente restauro del contenuto, tramite le tecniche digitali.

Ogni intervento restaurativo verrà dettagliatamente descritto, assieme a ogni informazione concernente le operazioni effettuate sul documento sonoro.

#### 2.1.2 *Analisi del segnale*

Per esemplificare le condizioni delle registrazioni presenti nell'*Archivio Sonoro Calabrese*, si è scelto di analizzare dei segnali che potessero rappresentare lo stato generale del materiale in nostro possesso.

Per i materiali rappresentativi dell'intero archivio si sono scelte delle registrazioni effettuate in epoche diverse, su supporti differenti (bobine e audiocassette).

L'attrezzatura utilizzata per la lettura, il riversamento e l'analisi è quella in dotazione presso il Laboratorio di Fonetica del Dipartimento di Linguistica dell'Università della Calabria. Per la lettura delle bobine dell'intero archivio si sono utilizzati i registratori Uher modello RT 2000 e l'AKAI 4000 DS Mk-II. Le audiocassette sono state riversate utilizzando un registratore TEAC W-860R.

Per ciò che concerne la digitalizzazione abbiamo usato come convertitore analogico-digitale, la scheda audio esterna Edirol USB Audio Capture UA-25 a 24 bit e 96 kHz.

L'analisi del segnale è stata condotta grazie all'utilizzo del software Praat sviluppato da Paul Boersma e David Weenink, nella versione 5.0.19.

Il primo file preso in considerazione riguarda una registrazione su bobina effettuata nel 1967. La bobina è una BASF speciale G (900 feet) (Fig. 1) e la velocità di registrazione è a 9,53 cm/s.



Figura 1: Esempio di bobina presente nell'Archivio Sonoro Calabrese

Lo stato di conservazione del supporto è ottimo. Il file è stato campionato a una velocità di 1058 kbps, a 24 bit, canale mono, 44 kHz.

L'analisi è stata effettuata su una larghezza di banda di 16000 Hz, su una porzione di registrazione di circa 5 secondi, un range dinamico di 60 dB. (Figura.2)

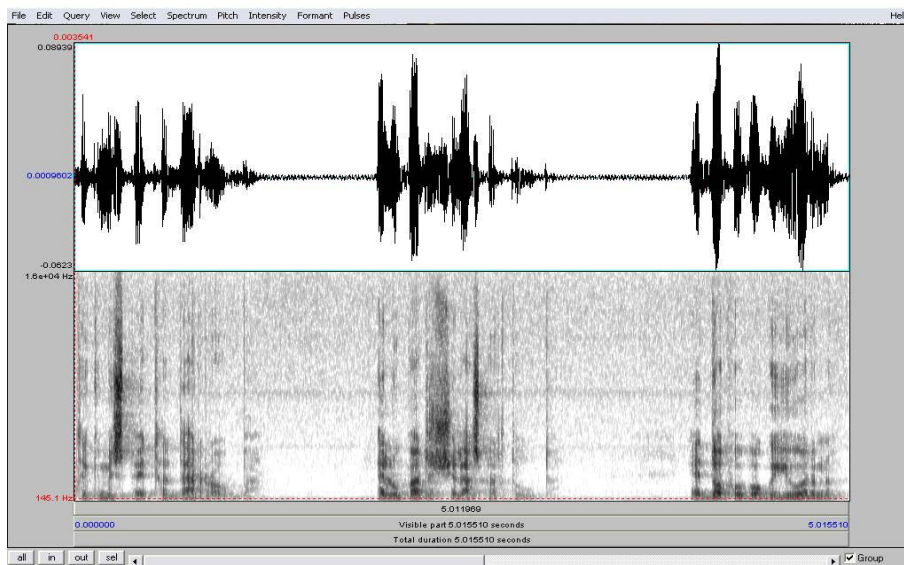


Figura 2: Spettro di un estratto della registrazione “La Parabola del Figliol Prodigo” (segnale 1)

Il segnale presenta un rumore di fondo additivo, stazionario e non correlato al segnale. Il disturbo presenta intensità significativa intorno ai 120, ai 3700 e ai 7500 Hz. (Figura 3)

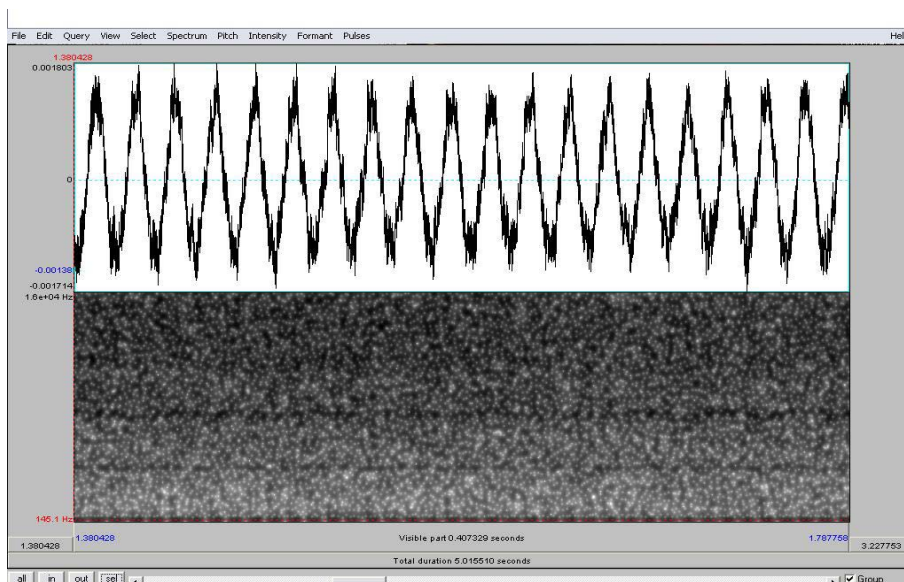


Figura 3: Spettro del rumore del segnale 1

Il secondo caso è quello di una registrazione riversata da una bobina Scotch Magnetic Type 262, 1200 ft. La registrazione risale al 1979. Lo stato del supporto è discreto. Il riversamento è stato effettuato con il registratore AKAI 4000 DS Mk-II e il segnale è stato campionato a una velocità di 1058 kbps, 24 bit, canale mono, 44 kHz.

L'analisi è stata effettuata su una larghezza di banda di 16000 Hz, su una porzione di registrazione di circa 5 secondi, range dinamico 80 dB. (Figura 4)

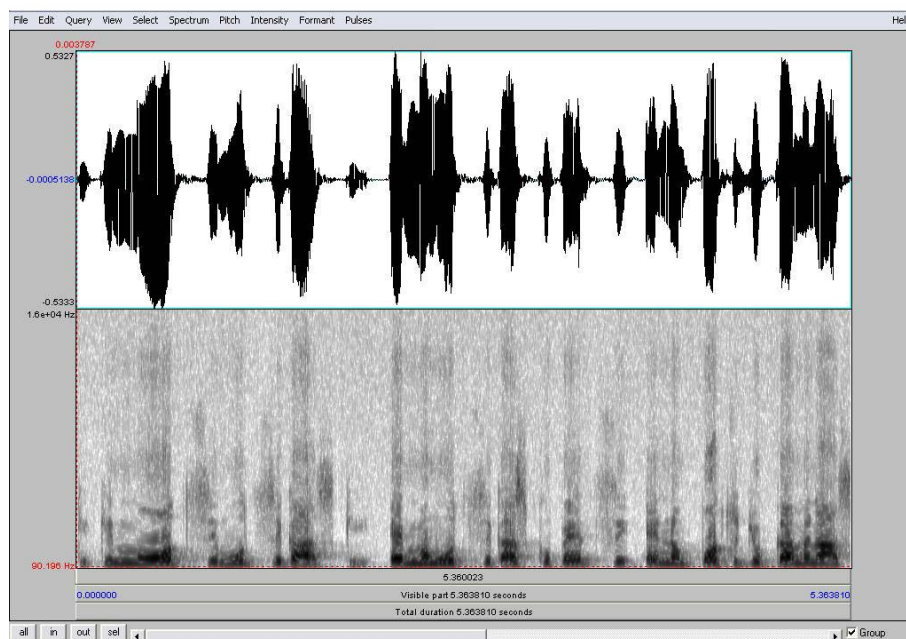


Figura 4: Spettro di un estratto della registrazione “n°2 – ‘79” (segnale 2)

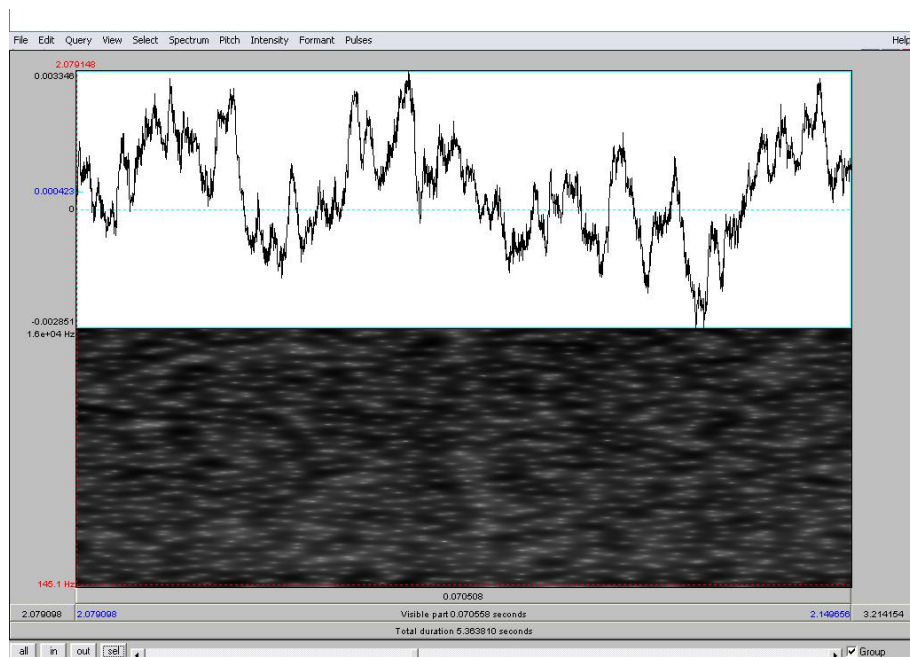


Figura 5: Spettro del rumore del segnale 2

La terza registrazione presa in analisi proviene da un riversamento di una audiocassetta TDK 90 Normal Position IEC I/TYPE I. Lo stato di conservazione del supporto è ottimo. La registrazione è avvenuta nel 2002. Il file è stato campionato a una velocità di 352 kbps, 16 bit, canale mono, 22 kHz.

L'analisi è stata effettuata su una larghezza di banda di 11000 Hz, su una porzione di registrazione di circa 5 secondi, range dinamico 60 dB. (FIG.7)



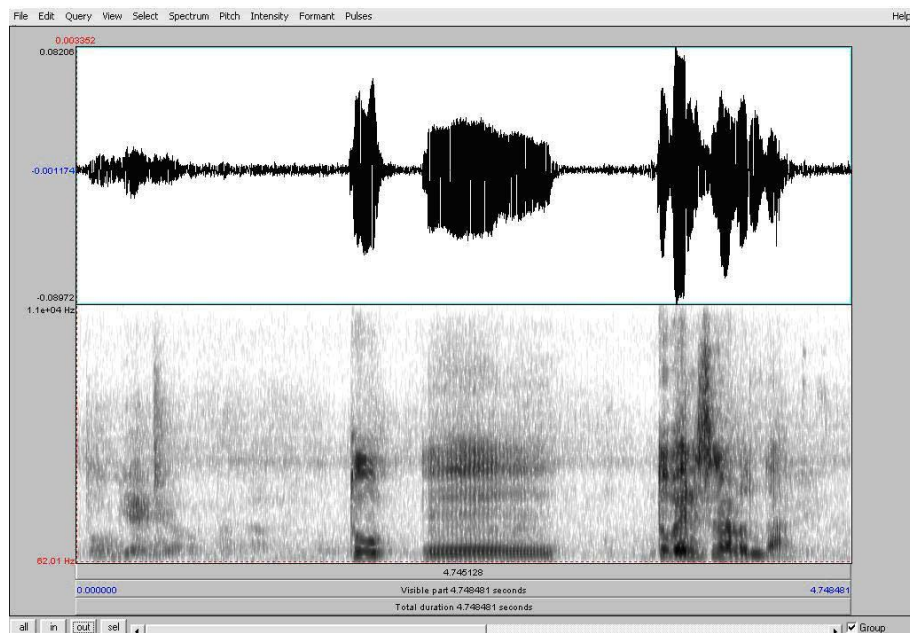


Figura 6: Spettro di un estratto della registrazione “San Nicola Arcella, Santa Domenica Talao, Papisidero, Orsomarso, Santa Maria del Cedro, Scalea” (segnale 3)

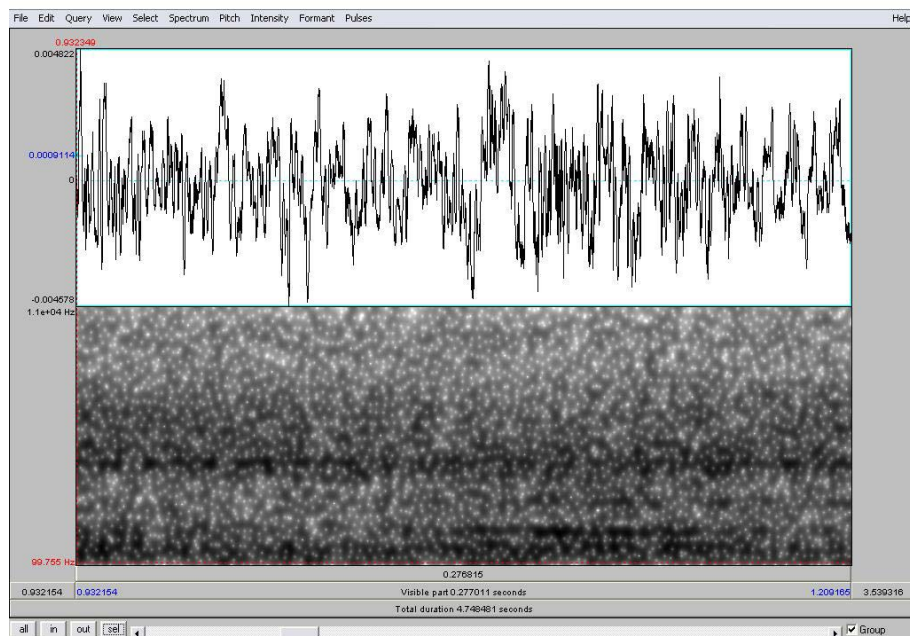


Figura 7: Spettro del rumore del segnale 3

### 3. CONCLUSIONI

L'*Archivio Sonoro Calabrese* consta di un insieme di materiali dalla natura molto varia, sia sotto l'aspetto dei supporti, della registrazione, che della conservazione e dei contenuti. Le registrazioni vanno dalla fine degli anni '60 fino ai giorni nostri, la strumentazione usata per la registrazione comprende apparecchiature sia professionali che commerciali, i supporti sono per lo più in buono stato, ma in alcuni casi i nastri sono risultati mal conservati, presentando ondulazioni, sovrapposizioni irregolari delle spire, ecc.). Questa complessità strutturale necessita, dunque, di un accurato studio preventivo di analisi e verifiche sui formati, sulle apparecchiature, sulle condizioni di conservazione.

Un primo problema nasce proprio da qui. Molti supporti non riportano una adeguata documentazione a corredo della registrazione (molto spesso mancano dati relativi all'equalizzazione, alla velocità di registrazione, al numero delle tracce, ecc.). Una lettura di questi supporti con apparecchiature tarate con un formato errato, porterebbe all'introduzione di artefatti nel segnale. Per questo motivo è necessario determinare il formato tramite l'analisi del supporto e del segnale riversato.

Prima di procedere all'operazione di riversamento è stato necessario accertarsi se si trattasse di una copia o di un originale. Una copia potrebbe essere stata oggetto di manipolazioni, ragion per cui il riversamento dovrebbe avvenire solo da un originale. Il rischio è quello di archiviare informazioni falsificate. In questo lavoro, in quei casi in cui non si era in possesso dell'originale, ma solo di copie apparentemente identiche, il riversamento è stato effettuato dalla copia che presentava il migliore stato di conservazione.

La conservazione e il restauro del segnale vocale si identifica con il restauro del supporto fisico su cui è registrato. In realtà la questione è più problematica. Questo lavoro è svolto per lo più da ingegneri del suono, ma le competenze tecnologiche e applicative riguardano esclusivamente la salvaguardia materiale, ma non sono sufficienti a garantire la conservazione culturale. Sono necessarie delle competenze linguistiche, storiche, ecc.

È necessario procedere con cautela sia durante il riversamento che durante il restauro, interventi indiscriminati o avventati possono causare danni irreversibili a documenti storici di valore.

Inoltre, diversamente da quanto avviene con i processi di conservazione e archiviazione, regolati da un'etica e da procedure standard che garantiscono una certa oggettività, il restauro audio manca di una regolamentazione relativa alle prassi di intervento.

L'analisi effettuata sui campioni di segnale ha evidenziato dei disturbi tipici legati all'epoca della registrazione, ai supporti, alle attrezzature e allo stato di conservazione. A partire da questa considerazione, si può procedere alla stesura di un protocollo comune di intervento relativo all'*Archivio Sonoro Calabrese*, contenente le procedure standard di restauro.

### BIBLIOGRAFIA

Adamo G., Debiasi G. B., De Poli G., Giua P., Mian G. A., Re G., Sotgiu M. C., Vidolin A., (1979), Problemi di Conservazione e Restauro di Archivi Sonori, in *Atti del Congresso Associazione Italiana di Acustica*, Perugia, Maggio.

Canazza S., (1998) Restauro Materiale Audio, in [www.dei.unipd.it/english/csc/people/all/Canazza.html](http://www.dei.unipd.it/english/csc/people/all/Canazza.html), CSC, Udine.

- Canazza S., De Mezzo G., Orcalli A., (2002), Conservazione e Restauro dei Documenti Sonori al Laboratorio MIRAGE, in *Suoni in Corso: Percezione ed Espressione dell'Uomo Tecnologico*, Mittelfest Editore, Cividale del Friuli, Febbraio.
- Canazza S., (2006), Conservazione Attiva e Restauro Audio dei 78 Giri. In *Ri-mediazione dei Documenti Sonori*, Canazza, S. e Casadei Turronei Monti, M. (a cura di), pp. 695-715, Udine.
- Canazza S., (2007), *Noise and Representation Systems: A Comparison among Audio Restoration Algorithms*, Lulu.com.
- Così P., (2002), *Il Restauro Digitale del Segnale Audio*, in De Dominicis A., *La Voce Come Bene Culturale*, Carocci Editore, Roma.
- Cossetti L., (2004), *Conservazione Attiva e Restauro dell'Archivio Audio di Fernanda Pivano*, Università degli Studi di Udine, Dicembre.
- De Mezzo G., Orcalli A., (2003), Creating a Digital Archive of Analogue Recordings: Technological Aspects and Musicological Implications in *XIV CIM Proceedings*, Firenze.
- Edmonson R., (1991), Etica e Principi del Restauro, in *Cinegrafie* n°4: La Guerra Giusta, Bologna.
- Ferrero F., Genre A., Boe L. J., Contini M., (1979), *Nozioni di Fonetica Acustica*, Ed. Omega, Torino.
- Godsill S.J., Rayner P.J.W., Godsill S.H., (1998), *Digital Audio Restoration: a Statistical Model Based Approach*, Springer Verlag, Berlin.
- Lucchesi R., (1992), La Soluzione Digitale al Problema del Rumore, in *Audioreview* n° 112, Technimedia, Gennaio.
- Orcalli A., (2006), Orientamenti Ai Documenti Sonori. In *Ri-mediazione dei Documenti Sonori*, Canazza S. e Casadei Turronei Monti, M. (a cura di), pp. 15-94, Udine.
- Rothenberg J., (1995), La Conservazione dei Documenti Digitali, in *Le Scienze* n° 315, Marzo.
- Saraceno C., Rumore, (1991), No Grazie, in *Audioreview* n°109, Technimedia, Ottobre.
- Schuller D., (1991), The Ethics of Preservation, Restoration and Re-issues of Historical Sound Recordings, *J. Audio Engineering Society*, n°12.
- Schuller D., (1994), Informazione Audio e Video. Dalla Preservazione dei Supporti Fisici alla Preservazione delle Informazioni, in *L'eclisse delle memorie*, a cura di Gregory T., Morelli M., Laterza, Bari.

# **LOGOPEDIA, FONIATRIA & AUDIOLOGIA**

## EVALUATION OF ELECTRO-ACOUSTIC PARAMETERS OF THE VOICE PRE AND POST ADENO-TONSILLECTOMY IN CHILDREN

Nicola Lombardo, Eugenia Allegra\*, Alessandro La Boria, Aldo Garozzo  
ENT Unit, Catanzaro University, Italy - \* ENT Unit, Garibaldi-Nesima Hospital, Catania, Italy  
*garozzoa@unicz.it*

### SOMMARIO

Il vocal tract, l'insieme delle cavità situate al di sopra del piano glottico, serve ad arricchire il tono fondamentale laringeo (F0), prodotto dalla vibrazione della mucosa delle corde vocali, di un insieme di suoni multipli dello stesso: le armoniche. Gruppi di armoniche con maggiore contenuto energetico rappresentano le formanti vocaliche. Esse sono in numero variabile ma le più significative sono le prime tre (F1, F2, F3). Modificazioni anatomiche del vocal tract determinano una alterazione dei valori formantici. L'intervento di adeno-tonsillectomia dovrebbe teoricamente comportare un cambiamento delle formanti. Scopo del nostro lavoro è stato quello di verificare il reale impatto che l'intervento di adeno-tonsillectomia determina sul prodotto vocale ed in particolare sui parametri elettroacustici della voce. Il nostro campione è costituito da 28 bambini di età compresa tra 5 e 14 aa di cui 13 maschi e 15 femmine con un'età media di 8,6 aa. Tutti i soggetti sono stati sottoposti ad una registrazione delle cinque vocali della lingua italiana /a/, /e/, /i/, /o/, /u/ prima e trenta giorni dopo l'intervento di adeno-tonsillectomia e ad una valutazione psicoperceptiva della voce mediante la scala GIRBAS.

I campioni vocali sono stati analizzati mediante l'utilizzo di due software di analisi vocale: il WinPitch e lo SpeechViewer IBM. In particolare sono stati valutati i valori della frequenza fondamentale F0, e delle prime tre formanti (F1, F2, F3). I risultati ottenuti hanno evidenziato l'assenza di alterazioni a carico della F0, per la F1 un lieve aumento per la vocale /o/, per la F2 un moderato incremento per la /o/ e lieve per /a/ ed /i/, per la F3 moderati incrementi per tutte le vocali tranne che per la /a/. L'analisi statistica condotta con il T-test ed il Mann Whitney test ha evidenziato l'assenza di qualunque significatività delle variazioni riscontrate. I nostri risultati sono in linea con quelli presentati da altri autori per cui è ragionevole concludere che l'intervento di adeno-tonsillectomia non comporta modificazioni dei parametri elettroacustici della voce, e la valutazione clinica del prodotto vocale mediante l'utilizzazione della scala GIRBAS rappresenti un test valido per la monitorizzazione della voce. D'altra parte le piccole modificazioni riscontrate, seppure di nessun significato, statistico impongono una maggiore attenzione nei soggetti professionisti della voce che si sottopongono ad intervento di adeno-tonsillectomia.

### SUMMARY

The vocal tract, that is all cavities placed above the glottis, is meant to enrich the larynx fundamental frequency (F0), produced by the vibration of the vocal folds, through a series of accessory sounds multiples of the fundamental tone: the harmonics. Groups of harmonics with an energetic enrichment from the supraglottals resounding form the vowel formants. These are variable in number but the most important ones are F1, F2, F3, that is the first, the second and the third of the formants. Anatomical changes of the vocal tract produce a modification of the formants values. Therefore, adeno-tonsillectomy might change the formants. The aim of this study was to evaluate the real impact that adeno-

tonsillectomy determines on the vocal performance and, also, on the electro-acoustic parameters of the voice. The sample we utilized consists in 28 children aged between 5 and 14, of whom 13 males and 15 females (the average is 8,6 years). All patients underwent the registration of the five Italian vowels /a/, /e/, /i/, /o/, /u/ before and 30 days after the adeno-tonsillectomy and a psycho-perceptive evaluation of the voice by using GIRBAS scale. Vocal samples were analyzed by two software of vocal analysis: the WinPitch and the Speech Viewer IBM. Specifically, we evaluated the values of the fundamental frequency (F0) and of the first three formants (F1, F2, F3).

The data we obtained revealed the absence of variations for F0, a moderate increase of /o/ for F1, the same moderate increase of /o/ and a minor increase of /a/ and /i/ for F2, finally, a moderate increase of all vowels except /a/ for F3. T-test and Mann Whitney statistical analysis have attributed no significance to these variations. Our results are in line with those carried-out by other authors, therefore we can conclude that adeno-tonsillectomy cannot modify the electro-acoustic parameters of the voice and that only the psycho-perceptiveness evaluation of the voice by GIRBAS scale represents a valid test for its evaluation. On the other hand, although the observed little modifications do not have a statistical importance, they ask for particular care to the voice of professional people (i.e. singers, actors etc.) that want to undergo to the adeno-tonsillectomy.

## 1. INTRODUCTION

The vocal tract, beginning at the glottis and extending to the lips, is considered to be the acoustic resonator of the voice. It is a very complex anatomical tract in which the fundamental tone, generated by the vibrations of the vocal folds is enriched, by a series of accessory sounds that are multiples of the fundamental tone: harmonics.

Groups of them acquire an energetic enrichment from the supraglottic resounding and form the vowel formants.

For this reason the vocal tract may be assimilated to the sound box of a musical instrument. The difference between a musical instrument and the human voice consists in the fact that our vocal apparatus is susceptible to huge changes and therefore any change in its structure may cause different kinds of voice modifications (Ferrero F., Genre A., Boe L. J., Contini M., 1979).

Adeno-tonsillar hypertrophy determines tightening of the vocal tract because of the total or partial isolation of the oro-nasal cavity. Consequently, this modifies the voice reducing rhino-pharyngeal and oro-pharyngeal spaces. Studies on the true impact that adeno-tonsillectomy has on the vocal production are very few (Chuma et al., 1999; Ilk et al., 2002; Saida H., Hirose H., 1996; Neri et al., 2002).

Chuma et al. (1999) have not encountered substantial vocal modifications following adeno-tonsillectomy with the exception of the second vowel formant F2 for the /i/ and /a/ which increase in tone and the first vowel formant F1 for the /o/ which appears to be reduced following surgery.

Ilk et al (2002) have found changes in vowel formant F3 for the /o/ which tends to reduce either for adequate hearing feed-back or for the changes following surgery. Saida et al (1996) did not find any variation in F0, F1, F2 following surgery but only a little decrease of F3. Neri et al (2002) have found statistical significant variations of the F1 for the vowels /a/, /o/, /u/, and less significance for /i/ and /e/. F2 was statistically increased for /e/ and /u/ and less for /i/. Finally they noted, three months after the surgery, significant decrease of F3 in relation to an improvement of the orbicular muscle tone of the lips and therefore in consequence of the recovery of the nasal airways.

The aim of this study was to evaluate if there exist any alterations of the electro-acoustic parameters following adeno-tonsillectomy verifying:

- if and how the values of the fundamental tone (F0) and vowel formants (F1, F2, F3) of speech could modify, in relation to the changes of the vocal tract resonance;
- if there is any use of electro-acoustic evaluation of the voice versus psycho-perceptive and therefore clinical assessment.

## 2. MATERIALS AND METHODS

A total of 28 patients were investigated between the age of 5 and 14 of whom 13 males and 15 females (mean age 8.6 years). The subjects were mixed with the patients who were being treated in the ENT Clinic (in the last two years) at the University of Catanzaro and who were subjected to adeno-tonsillectomy according to the normal clinic protocol.

Everyone underwent an objective ENT exam using the flexible or rigid 2.7 mm Explorent paediatric fiberoendoscope.

In all the patients, the following conditions were found: nasal obstruction, breathing obstruction during sleep, recurrent inflammation, nasal language, breath control through the mouth due to the lacking of a better breath control following repeated cycles of medical therapy.

All patients affected by head, facial and hearing malformations are not included in this study. Due to the difficulty in adequate recording, patient below the age of 5 were excluded. None of the patients underwent speech therapy prior to surgery. All patients underwent an evaluation of their voice using GIRBAS method, a reliable and trustworthy instrument with a rapid perceptual analysis of the speech production. The GIRBAS scale, which was proposed by Hirano (1981), integrated by Dejonckere (1998) and recommended by European Laryngological Society (2001) is a perceptual evaluation that refers to the ability of the clinic, independent of the acoustic measurement. For each parameter a point system is used, score from 0 to 3, where 0 represents normal condition, while 1, 2, 3 correspond to the grades: light, moderate and heavy.

The subjects chosen were in a normal condition (0 points). They were subjected to a voice recording test one day before surgery and 30 days after, when the recovery was complete. In each phase the patient was invited to pronounce the vowels /a/, /e/, /i/, /o/, /u/ prolonged for 5-6 seconds.

A microphone (Model TM 333, a unidirectional dynamic type, TEV Co Taejon) was used placed at a standard distance of about 20 cm from the mouth in a silent surrounding, respecting the procedures recommended by the European Society of Laryngology.

The voice samples at each stage were made digitally, using a PC that analyzed the data using two different software: Win Pitch and Speech Viewer III.

Win Pitch is a vocal spectrograph programme. It is a non invasive method and able to analyze the psycho-acoustic parameter of the voice (Fig. 1)

Based on the harmonic structure and the location of the aperiodic sign within the spectre we can distinguish the various grades of alteration from the vocal product (classified from Yanagihara, 1967). In a scale from 1 to 4, where 4 indicates maximum grade of alteration, the use of this programme allows us to obtain the values of the parameter F0.

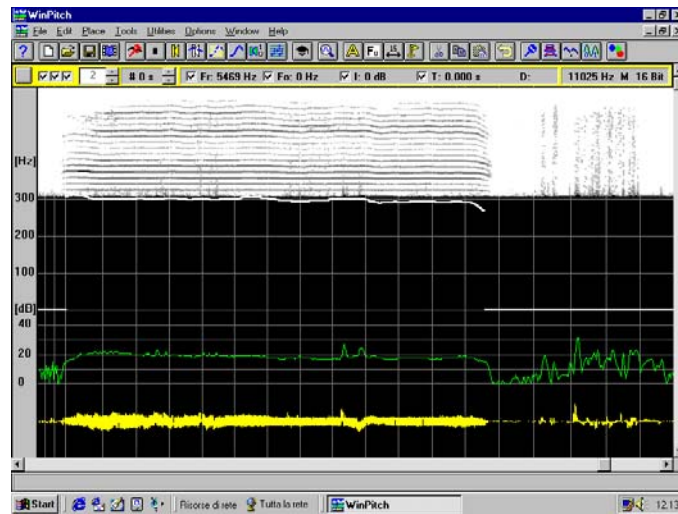


Fig. 1 An example of the spectrogram as appears with Win –Pitch software.

The F0 or the fundamental frequency, that corresponds to the number of vocal folds vibrations per second, depends on the anatomical and physiological characteristics of the individual and to his particular vocal tract. It is directly linked to the vocal folds and therefore to their length, mass and tension.

With the use of Speech Viewer III it is possible to calculate the values of other acoustic parameters, which are the object of our work: the vowel formants.

These are groups of harmonics with a greater content of energy and a multiple of the basic frequency. The formants reflect the characteristics of the resonance through the upper laryngeal vocal tract, associated with the articulation of the tongue with the other articulated structures which are both fixed and moveable and reflect in their value the deadening sound and energy absorbed by the vocal tract during each production of every vowel.

The formant F1, between 250 and 1050 Hz, depending on the vowel pronounced, is created in the oral cavity and is directly correlated with the major or minor opening of the mouth and the relationship the tongue encounters with the mandibular arch. The formant F2 is between 600 and 3200 Hz and is originated by the position of the tongue. The formant F3 is between 2400 and 3750 Hz and is caused by the movement of the orbicular muscle of the lips. (Fig. 2)



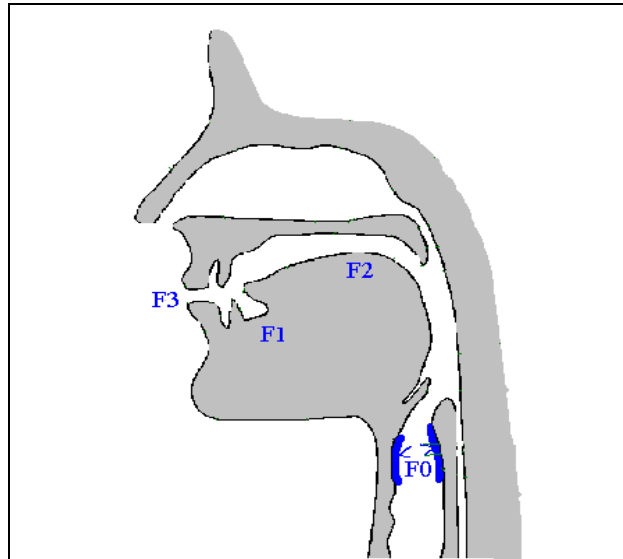


Fig. 2: Side of production of the fundamental frequency (F0) and vowel formants (F1, F2, F3) in the vocal tract.

In the spectrogram these are identified by a series of peaks that are shown on a long frequency field (Fig. 3). It is their different disposition on these frequencies that allows to distinguish one vowel sound from another.

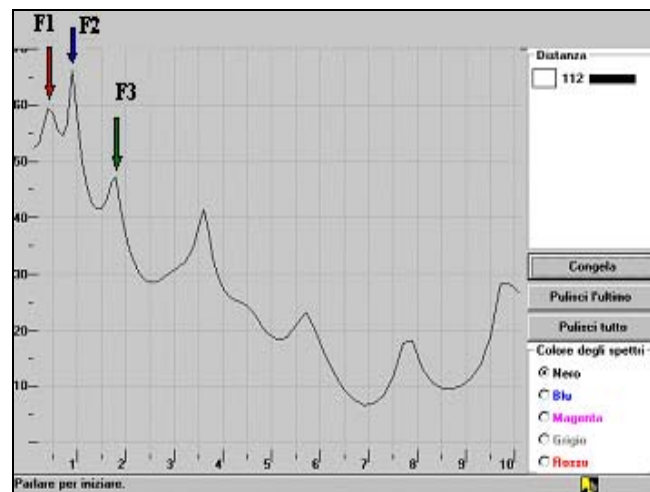


Fig. 3: An example of the vowel formants (F1, F2, F3) view with speech viewer software.

The registered spectres have been suspended in about three seconds from the beginning of the phonation, which is the time considered to be perfect and also more manageable.

Statistical analysis was performed using Mann-Witney unpaired tests to assess pre and post-surgical changes in speech rating data and Paired t-test for dependent means to value pre and post-operative significant variations in vowel formant structure.

### 3. RESULTS

The quantitative analysis of the voice has been derived from the values obtained for the fundamental frequency (F0) and for the formants F1, F2, and F3. Those were drawn from the registration in the period immediately preceding surgery and 30 days after.

Regarding the calculation of the formant values, an inclusive media of the singular parameter F1, F2 and F3 ( all subjects of whom were included in the registration of the vowel /a/, /e/, /i/, /o/, /u/) during the period of examination have been made. In particular the mean values of F0, F1, F2 and F3 derived from the recording of the vowels are indicated in Table I.

| Vowel | F0 pre | F0 post | F1 pre | F1 post | F2 pre | F2 post | F3 pre | F3 post |
|-------|--------|---------|--------|---------|--------|---------|--------|---------|
| /a/   | 288    | 278     | 415    | 430     | 1184   | 1254    | 2219   | 2242    |
| /e/   | 276    | 288     | 395    | 400     | 995    | 1000    | 2322   | 2461    |
| /i/   | 298    | 286     | 380    | 380     | 1268   | 1338    | 2809   | 2995    |
| /o/   | 291    | 296     | 525    | 608     | 1139   | 1273    | 2532   | 2652    |
| /u/   | 294    | 284     | 409    | 448     | 852    | 920     | 2739   | 2830    |

Tab. I: Mean values obtained for the fundamental frequency (F0) and for the formants (F1-F2-F3) regarding all patients for each vowel.

The results obtained from t-test and Mann-Witney test revealed no significant changes in vowel structure for all of the vowels (Tab. II).

| <i>Vowel a</i> |        |                  |
|----------------|--------|------------------|
|                | T-test | Mann Witney test |
| F0             | 0,106  | >0,10            |
| F1             | 0,8917 | >0,10            |
| F2             | 0,791  | >0,10            |
| F3             | 0,8965 | >0,10            |
| <i>Vowel e</i> |        |                  |
|                | T-test | Mann Witney test |
| F0             | 0,396  | >0,10            |
| F1             | 0,4657 | >0,10            |
| F2             | 0,7762 | >0,10            |
| F3             | 0,2883 | >0,10            |
| <i>Vowel i</i> |        |                  |

|                | T-test | Mann Witney test |
|----------------|--------|------------------|
| F0             | 0,0382 | >0,10            |
| F1             | 1,000  | >0,10            |
| F2             | 0,7486 | >0,10            |
| F3             | 0,5086 | >0,10            |
| <i>Vowel o</i> |        |                  |
|                | T-test | Mann Witney test |
| F0             | 1,000  | >0,10            |
| F1             | 0,6163 | >0,10            |
| F2             | 0,5185 | >0,10            |
| F3             | 0,6400 | >0,10            |
| <i>Vowel u</i> |        |                  |
|                | T-test | Mann Witney test |
| F0             | 0,5830 | >0,10            |
| F1             | 0,7007 | >0,10            |
| F2             | 0,7846 | >0,10            |
| F3             | 0,5316 | >0,10            |

Tab. II: Statistical analysis performed to assess pre and postoperative vowels changes.

#### 4. DISCUSSION

According to the results we have drawn some considerations about the modifications of the acoustic parameters of the voice following adeno-tonsillectomy that allows an enlargement of the oro-rhino-pharynx cavities.

By definition, the acoustic resonance is the enrichment of the fundamental tone (F0) from the larynx with a series of accessory sounds, the so called harmonics, produced by the multiples of the fundamental frequency, that are moved by supralaryngeal cavity.

In particular, the electro-acoustic study of the vocal resonance is done using some parameters including the study of formants. These are harmonic groups with major energy peaking in a different way and with a special location for each vowel which are also at the base of the differential and therefore the recognition of such. The most important are the first two F1 and F2, while F3, F4 and F5 have a minor importance.

The analysis of the results of component F1 and F2 have shown no variation of F1 with the exception of a moderate elevation for /o/ and the modification of the values of F2. Still a slight increase for /o/ and an even more modest one for the vowels /a/ and /i/, while there was no change in the remaining vowels 30 days after surgery in respect to the initial values

(prior surgery). Such change however, that is statistically evaluated by t-test and Mann Witney test, have not revealed any significance. On the other side, we can say that modifications most consistently verified for the vowel /o/ (increase of F1 and F2) even if without significant statistics, do not determine any change in the difference of the interformantic F2-F1 and this parameter is known to be the most significant data for the vocal perception more than absolute value of the single vowel formant.

In regards to the parameter F3, there have been increases for all vowels with the exception of /a/ but here also the statistic analysis have not revealed great significance.

Our data are in line with what can be found in literature (2, 3, 4, 5) where only small post-surgery initial variations of these parameters (F1 and F2) have been kept in a variable period of time (from 90-120 days).

Return to normality has been attributed to an acoustic feed-back from the patient, the acoustic memory and more to the post surgical reparation of the tonsil fossae. Given the lack of statistical significance of the changes in F0 and in vowel formants F1, F2 and F3, we can conclude that psycho-perceptive analysis of the voice, using GIBBAS scale, allows by itself, to express an exact evaluation of the acoustic modification parameters of the voice after adeno-tonsillectomy. Not with standing this, we maintain that the minor differences we have already noted and they were also pointed out by various authors, should be taken into careful account whenever this surgical operation is proposed to professionals of the voice (actors, singers).

## 5. CONCLUSION

This study started with the intention to evaluate the real impact of adeno-tonsillectomy on the acoustic characteristics and perception of the voice.

From the analysis of the results, it is evident that the values of F0 result unchanged after surgery, as F0 reflects the condition of the vocal folds, which are not involved in the surgery.

For vowel formants however, multiple modifications have appeared although statistical analysis has revealed not significant; particularly there was no modification of the parameter F1 of all the vowels with the exception of /o/, moderate increase of F2 in the vowel /o/ and minimal in the vowel /a/ and /i/, whereas in the remaining vowels this parameter was not modified. The evaluation of the interformant difference F2-F1 does not undergo substantial modification and this parameter is regarded as the most important for the vocal perception, more than the single values.

Moreover, the results have shown that F3 increased in all the vowels with the exception of /a/ but it is also true that the importance of this at the end of the vocal perception it is absolutely minimal and in time tends to become normal.

Our results are therefore in line with those carried out from other authors and we can conclude that given the low evidence of modifications shown by the instrumental analysis, only the psico-perceptiveness of the voice carried out by speech therapists could be useful to show the possible initial changes in the voice characteristics.

Furthermore, these changes that are almost always minimal, allow the patient to undergo the adeno-tonsillectomy in absolute tranquillity, especially when we consider the voice professionals (singers, actors) who are particularly careful to minimal changes of their voice.

## BIBLIOGRAPHY

- Chuma A. V., Cacace A. T., Rosen R., Feustel P., Koltai P. J. (1999), Effects of tonsillectomy and/or adenoidectomy on vocal function: laryngeal, supralaryngeal and perceptual characteristics, *International Journal of Paediatric Otorhinolaryngology*, 47:1.
- Dejonckere P. H., Bradley P., Clemente P., Cornut G., Crevier Buchman L., Friedrich G., et al (2001), A basic Protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical ) treatments and evaluating new assessment techniques, *Eur Arch Otorhinolaryngol*, 258:77-82.
- Dejonckere P., Remacle M., Fresnel-Elbaz E., Woisard V., Crevier-Buchman L., Millet B. (1998), Reliability and clinical relevance of perceptual evaluation of pathological voices, *Rev. Laryngol Otol. Rhinol*, 119:247-8.
- Ferrero F., Genre A., Boe L. J., Contini M. (1979), *Nozioni di fonetica acustica*, Torino, Ed. Omega.
- Hirano M. (1981), Psycho acoustic evaluation of voice, in: Hirano M., *Clinical examination of voice*, New York: Springer-Verlag.
- Ilk H. G., Erogul O., Satar B., Ozkaptan Y. (2002), Effects of tonsillectomy on speech spectrum, *J Voice*, 16(4):580
- Neri G., Chiri Z. M., Ballone E., Falcone G., Di Mascio R., Croce A. (2002), Effects of adenotonsillectomy on acoustic and perceptive voice quality, *Riv. Orl. Aud. Fon*, Vol. 22(4), pp. 131-136.
- Saida H., Hirose H. (1996), Acoustic changes in voice after tonsillectomy, *Acta Otolaryngol. Suppl.*, 523:239.
- Yanagihara N. (1967), Significance of harmonic changes and noise components in hoarseness, *J Speech Hear Res*, 10:531-541.

## **CARATTERISTICHE DI PERSONALITA' E MODALITA' DI UTILIZZO DELLA VOCE NEI SOGGETTI CON PATOLOGIA VOCALE**

Nicola Angelillo, Brigida Di Costanzo, Assunta Fiorillo, Umberto Barillari  
Servizio di Foniatria ed Audiologia, Dir. Prof. Umberto Barillari  
Seconda Università degli Studi di Napoli  
[Nicola.angelillo@unina2.it](mailto:Nicola.angelillo@unina2.it); [b.dicostanzo@tiscali.it](mailto:b.dicostanzo@tiscali.it); [susyff@supereva.it](mailto:susyff@supereva.it);  
[umberto.barillari@unina2.it](mailto:umberto.barillari@unina2.it)

### **1. SOMMARIO**

Le patologie a carico della voce possono essere suddivise in disfonie organiche, causate da alterazioni primitive della laringe, e disfonie disfunzionali in cui l'alterazione qualitativa e/o quantitativa della voce si verifica in assenza di alterazioni strutturali primitive della laringe. Ancora molto controversa è l'eziologia della disfonia disfunzionale. Da un lato questa forma di disfonia viene considerata di origine psicogena, come manifestazione di un'alterazione dell'equilibrio psicologico interno; dall'altra parte, l'abuso ed il cattivo uso della voce sono chiamati a svolgere un ruolo importante nella genesi della disfonia disfunzionale.

Dalla nostra esperienza, confermata anche da numerosi lavori presenti in letteratura, i pazienti con disfonia disfunzionale presentano sia un'alterata autopercezione della voce sia peculiari caratteristiche psicologiche e di personalità, con prevalenza di ansia, insicurezza e bassa autostima.

Partendo da queste osservazioni abbiamo ipotizzato che il soggetto disfonico abbia una alterata capacità di decodificare e codificare in maniera corretta il livello emozionale veicolato attraverso il canale vocale.

Per verificare la nostra ipotesi abbiamo realizzato un test che ci consentisse di valutare la capacità da parte dei pazienti disfonici di riconoscere il contenuto emozionale veicolato da frasi sintatticamente neutre, interpretate da un attore professionista, che esprimessero non solo tre diverse emozioni primarie (rabbia, gioia, tristezza) ma anche tre diversi livelli di intensità per ciascuna emozione ed un test in grado di individuare soggetti che tendono ad utilizzare la voce come veicolo per esprimere le proprie frustrazioni e come modalità di affermazione sugli altri.

Tali test sono stati somministrati sia a soggetti affetti da disfonia disfunzionale che a soggetti non disfonici al fine di valutare l'esistenza di una relazione tra disfonia disfunzionale e capacità di decodifica e codifica delle emozioni.

I risultati di questi test sono stati comparati sia con quelli ottenuti al Voice Handicap Index (VHI), test per l'autovalutazione vocale, che allo State Trait Anxiety Inventory (STAI), test per la valutazione dell'ansia.

Le risposte ai test hanno evidenziato che i soggetti disfonici presentano significative difficoltà sia nel decodificare le emozioni che nel dosare in maniera adeguata l'intensità emozionale della propria voce e soprattutto di adattarla ai differenti contesti relazionali in cui vengono a trovarsi quotidianamente.

Appare chiaro che la disfonia disfunzionale è una patologia complessa e multidisciplinare, per la quale occorre stilare trattamenti riabilitativi integrati, in quanto in molti casi non è facilmente correggibile con la sola terapia logopedica convenzionale.

## 2. INTRODUZIONE

La voce è espressione dell'individuo e della sua personalità, nonché specchio e risultato della sua integrazione nella società e del suo ruolo in essa, in quanto possiede la capacità di trasmettere in modo autonomo precise e distinte informazioni circa gli stati affettivi ed emotivi dell'individuo indipendentemente dagli aspetti linguistici dell'enunciato. La voce è a sua volta influenzata dallo stato emotivo-affettivo del soggetto. E' stato infatti ampiamente dimostrato che le emozioni sono in grado di influenzare diversi parametri vocali (quali media, variabilità e gamma della frequenza fondamentale, intensità e tono della voce, ritmo di articolazione, profilo di intonazione) e che esistono profili acustici associati a ciascuna emozione in misura decisamente superiore al livello di casualità (Anolli & Ciceri, 1997; Erickson *et alii*, 2006; Monnot *et alii*, 2003; Raithel & Hielscher-Fastabend, 2004; Sobin & Alpert, 1999; Toivanen *et alii*, 2006; Whiteside, 1999). Ad esempio la collera è caratterizzata da un aumento di media, variabilità e gamma della frequenza fondamentale, da un aumento dell'intensità, dalla presenza di pause molto brevi o dalla loro assenza, da un ritmo elevato e da un profilo di intonazione caratterizzato da frequenti variazioni.

L'esistenza di pattern acustici specifici di ciascuna emozione è inoltre confermata dalla possibilità di decodificare il contenuto emozionale di un messaggio verbale prestando attenzione esclusivamente alle sue caratteristiche vocali. Dagli studi presenti in letteratura è emersa una capacità di riconoscimento pari al 60%, un valore al di sopra delle percentuali di riconoscimento delle emozioni attraverso la mimica facciale. Questo dato può essere interpretato come un sostegno all'ipotesi dell'esistenza di regole universali di inferenza delle emozioni sulla base delle proprietà acustiche della voce, confermando l'ipotesi dell'autonomia semantica dei tratti paralinguistici, ovvero di una relazione tra significato emozionale e parametri acustici della voce.

La voce rappresenta quindi un importante canale attraverso cui comunicare le proprie emozioni in maniera più o meno consapevole e può a sua volta essere influenzata dallo stato emozionale del soggetto più o meno significativamente in relazione alle sue caratteristiche psicologiche e di personalità.

Un'alterazione qualitativa e/o quantitativa della voce (disfonia) si può verificare o in conseguenza di lesioni primitive della laringe (disfonia organica) o anche in assenza di lesioni della laringe (disfonia disfunzionale) (Le Huche, 1992). Dal punto di vista epidemiologico, nell'ambito delle disfonie, le forme disfunzionali risultano significativamente prevalenti rispetto a quelle organiche, rappresentando circa il 70-80% di tutte le forme di disfonia. L'eziologia della disfonia disfunzionale è ancora poco chiara ed oggetto di numerosi studi finalizzati a comprenderne i meccanismi etiopatogenetici. Alcuni ricercatori hanno evidenziato particolari caratteristiche di personalità nei pazienti affetti da disfonia disfunzionale, riscontrando la prevalenza in tali soggetti di una personalità caratterizzata da elevato nevroticismo e bassa estroversione (Di Costanzo *et alii*, 2007; Gerritsma, 1991; Roy *et alii*, 1997; Roy & Bless, 2000; Roy *et alii*, 2000a; Roy *et alii*, 2000b). Da una precedente sperimentazione del nostro gruppo di ricerca (Barillari *et alii*, 2003) è inoltre emerso che i soggetti affetti da disfonia disfunzionale presentano una maggiore instabilità emotiva ed ansia sia di stato che di tratto. Questi studi supportano

l'ipotesi che un ruolo importante nella genesi della disfonia disfunzionale possa essere svolto da fattori psicologici, per cui la disfonia disfunzionale potrebbe essere la manifestazione di una situazione di squilibrio interno.

Considerando le caratteristiche di personalità riscontrate nei soggetti con disfonia disfunzionale e la stretta relazione esistente tra manifestazione delle emozioni e canale vocale (Angelillo *et alii*, in press), è possibile ipotizzare che il soggetto disfonico non sia in grado di decodificare in maniera corretta il livello emozionale veicolato dal messaggio vocale, sia qualitativamente che quantitativamente, e di conseguenza non sia nemmeno in grado di codificare correttamente, attraverso la voce, il proprio stato d'animo e di adattare la voce ai differenti contesti relazionali in cui viene a trovarsi quotidianamente. Se infatti il soggetto ha difficoltà nella decodifica di una emozione, avrà anche difficoltà a "settare" correttamente tutti i parametri vocali relativi a quell'emozione. Questa situazione si ripercuoterà, a lungo andare, sulla qualità della voce.

Nella genesi della disfonia disfunzionale vengono spesso chiamati in causa l'abuso ed il cattivo uso della voce, quest'ultimo secondario ad un'alterata dinamica respiratoria e all'assenza di un buon accordo tra il respirare ed il parlare (accordo pneumofonico). Parte del trattamento logopedico delle disfonie disfunzionali prevede infatti l'esecuzione di esercizi di respirazione e rilassamento, finalizzati a migliorare la dinamica respiratoria e l'accordo pneumofonico. Va tuttavia sottolineato che i soggetti affetti da disfonia disfunzionale, frequentemente, pur avendo adeguatamente appreso ed acquisito, attraverso la terapia logopedica convenzionale, tutte le norme di igiene vocale, finalizzate a ridurre la tendenza all'abuso vocale, e gli esercizi e le tecniche per una adeguata respirazione costo-diaframmatica e per un corretto accordo pneumofonico, al fine di correggere il cattivo uso della voce, non riescono a risolvere la disfonia, in quanto non sono in grado di applicare quanto appreso nelle situazioni comuni della vita quotidiana: il volume della voce viene gestito dal soggetto come una possibilità di maggiore incisività negli eventi e la voce viene utilizzata come materia per dare valenza alla propria persona (ad esempio in una discussione, in un rimprovero o in un litigio). Nella vita quotidiana una voce gridata è una voce che si vuole imporre, che colpisce per attirare l'attenzione, per sovrastare l'altro in una discussione dimostrando, come in una lotta, la propria superiorità fisica e di personalità (la voce individua la persona e il suo essere). La vita ci pone frequentemente di fronte a soprusi, contrarietà, discussioni spiacevoli e, in generale, situazioni nelle quali è necessario far valere le proprie ragioni. In questi casi un comportamento adeguato consisterebbe nell'esprimere le proprie idee con calma, con argomentazioni valide, senza puntare eccessivamente sul volume della propria voce. Un soggetto con personalità insicura, che non riesce ad argomentare le proprie ragioni in maniera valida, che si lascia sopraffare dalla propria emotività, può arrivare alla conclusione che non è importante il contenuto di ciò che si vuole esprimere per far valere la propria opinione, ma piuttosto il modo, convincendosi che urlare è la chiave per essere ascoltati, avere ragione e non venire contraddetti.

L'influenza dell'assetto emotivo-relazionale di un individuo sulle sue modalità di comunicazione in generale e sul canale vocale in particolare, spinge a porre l'attenzione sull'importanza di considerare questi aspetti nella valutazione e riabilitazione dei soggetti con disfonia disfunzionale. A tale scopo il nostro gruppo di ricerca ha realizzato una serie di strumenti finalizzati a valutare in tali soggetti sia la capacità di decodificare il contenuto ed il livello emozionale del messaggio vocale sia la tendenza ad utilizzare abitualmente un volume eccessivo della voce ed in particolare se per il soggetto "urlare" rappresenta uno strumento utile ad attirare l'attenzione e a risolvere i problemi, consentendo di dare maggiore incisività alle proprie ragioni e di affermarsi sugli altri.



### 3. MATERIALI E METODI

Per valutare l'ipotesi da noi formulata abbiamo analizzato 3 gruppi di soggetti (**Tabella 1**):

- **gruppo I:** 50 soggetti non affetti da disfonia (25 di sesso femminile e 25 di sesso maschile, di età compresa tra i 25 e i 45 anni);
- **gruppo II:** 61 soggetti affetti da disfonia disfunzionale (48 di sesso femminile e 13 di sesso maschile, di età compresa tra i 25 e i 45 anni), che non hanno mai praticato terapia logopedica;
- **gruppo III:** 20 soggetti (14 di sesso femminile e 6 di sesso maschile, di età compresa tra i 25 e i 45 anni), che pur presentando una corretta respirazione costo-diaframmatica e un adeguato accordo pneumofonico dopo diversi cicli di terapia logopedica, continuano a presentare disfonia.

| Gruppo I                                                   |    | Gruppo II                                                                                                        |    | Gruppo III                                                                                                                                                                                               |    |
|------------------------------------------------------------|----|------------------------------------------------------------------------------------------------------------------|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>50</b><br>soggetti non affetti da disfonia<br>Età 25-45 |    | <b>61</b><br>soggetti affetti da disfonia funzionale, che non hanno mai praticato terapia logopedia<br>Età 25-45 |    | <b>20</b><br>soggetti affetti da disfonia funzionale, che presentano una corretta respirazione e un adeguato accordo pneumofonico, avendo già effettuato diversi cicli di terapia logopedia<br>Età 25-45 |    |
| M                                                          | F  | M                                                                                                                | F  | M                                                                                                                                                                                                        | F  |
| 25                                                         | 25 | 13                                                                                                               | 48 | 6                                                                                                                                                                                                        | 14 |

Tabella 1: Reclutamento pazienti.

Sono stati esclusi dallo studio i pz con disfonia organica (ossia in cui la disfonia è dovuta ad alterazioni primitive della laringe).

A questi soggetti abbiamo somministrato i seguenti test:

- **Test per valutare la predisposizione all'abuso vocale;**
- **Voice Handicap Index (VHI)** per l'autovalutazione vocale;
- **State Trait Anxiety Inventory (STAI)** per la valutazione dell'ansia.

Ai soggetti di controllo e ai soggetti del **gruppo III** abbiamo somministrato un ulteriore test, il **Test di decodifica delle emozioni**, al fine di indagare le motivazioni della mancata risoluzione della patologia vocale nonostante una corretta terapia logopedica.

#### 3.1 Test per valutare la predisposizione all'abuso vocale

La realizzazione di questo test nasce dall'ipotesi che i pazienti disfonici tendano ad urlare facilmente e ad utilizzare tale modalità di comportamento vocale per far valere le proprie ragioni e affermarsi sugli altri.

Il test (Angelillo *et alii*, 2007a; Pezone, 2006) consta di 40 affermazioni suddivise in 4 sezioni, ciascuna costituita da 10 affermazioni:

- le 10 affermazioni della **I sezione** esplorano la tendenza ad arrabbiarsi facilmente:
  - Mi arrabbio facilmente
  - Perdo facilmente la pazienza
  - I contrattempi della vita quotidiana mi innervosiscono molto
  - Mi arrabbio molto quando gli altri non mi capiscono al volo
  - Non tollero di essere contraddetto
  - Non sopporto che le persone si distraggano mentre sto parlando
  - Per me arrabbiarsi è una cosa istintiva
  - Per me urlare è una cosa istintiva
  - Le persone che mi conoscono dicono che mi arrabbio facilmente
  - Le persone che mi conoscono dicono che urlo anche per delle sciocchezze
- le 10 affermazioni della **II sezione** esplorano la tendenza ad urlare per esprimere la propria rabbia:
  - Urlo sempre
  - Quando mi innervosisco urlo facilmente
  - La modalità migliore per far capire di essere arrabbiati è urlare
  - Non credo che si possa essere arrabbiati senza urlare
  - Se non si urla la gente non capisce che sono arrabbiato
  - Per far capire che si è veramente arrabbiati è necessario urlare
  - Urlando mi sfogo
  - Se urlo poi mi sento meglio
  - Se non urlo la mia ansia aumenta
  - Quando sono molto arrabbiato urlo per evitare di usare le mani
- le 10 affermazioni della **III sezione** esplorano la tendenza ad urlare per attirare l'attenzione:
  - Alla gente non importa se sono arrabbiato
  - Urlando la gente capisce che sono arrabbiato e mi aiuta
  - Urlando gli altri mi ascoltano più facilmente
  - Quando parlo con gli altri parlo con un tono della voce alto perché altrimenti si distraggono
  - Per mantenere viva l'attenzione bisogna parlare a voce molto alta
  - Se parlo a voce bassa le persone si distraggono
  - Quando parlo voglio essere guardato
  - Se quando parlo le persone non mi guardano o si distraggono, aumento il volume della voce
  - Parlare a voce alta è la maniera migliore per essere considerati sicuri di sé
  - In una discussione riesco a farmi notare soltanto se aumento molto il volume della voce
- le 10 affermazioni della **IV sezione** esplorano la tendenza ad urlare per risolvere i problemi:
  - Quando parlo a voce alta mi sento più sicuro
  - Urlando mi sfogo e riacquisto la padronanza di me
  - Urlando mi sento più sicuro di me
  - Se urlo le persone mi temono

- Urlo per zittire gli altri
- In una discussione spesso ha la meglio chi urla di più
- Spesso chi urla di più in una discussione appare sicuro di sé, deciso e dalla parte della ragione anche se non lo è
- Se non urlo sto male
- Se urlo poi mi sento più calmo
- Se urlo mi sembra di ragionare meglio

A lato di ciascuna affermazione vi è una casella con la scritta **VERO** ed una casella con la scritta **FALSO**.

Si invita il paziente a leggere le 40 affermazioni che esplorano la tendenza ad arrabbiarsi facilmente, la tendenza ad urlare per esprimere la propria rabbia, la tendenza ad urlare per attirare l'attenzione, la tendenza ad urlare per risolvere i propri problemi. Va sottolineato che nella versione del test presentata al paziente non è visibile questa suddivisione in sezioni, ma le affermazioni sono presentate in successione continua. Il paziente è quindi invitato a sbarrare, per ogni affermazione, una delle due possibilità (vero o falso).

Ultimata la somministrazione si congeda il soggetto e si passa allo studio ed elaborazione dei dati così ottenuti. Nel calcolo del punteggio è stato attribuito il valore 1 per ogni risposta VERO e 0 per ogni risposta FALSO. Si è quindi proceduto a stabilire la media del vero per ogni sezione del test.

### 3.2 Voice handicap index (VHI)

È la versione italiana del test proposto da Jacobson e colleghi nel 1997 e modificato da Luppi (2002), per l'autovalutazione vocale. La valutazione è basata su 30 domande che indagano tre ambiti definiti:

- **sfera funzionale**, corrispondente all'impatto delle problematiche vocali sulle normali attività quotidiane (10 domande);
- **sfera emozionale**, corrispondente all'impatto psicologico del problema vocale e quindi alla risposta dell'individuo al disagio (10 domande);
- **sfera fisica**, corrispondente alla percezione del soggetto delle caratteristiche della sua emissione vocale (10 domande).

Il soggetto è tenuto a rispondere alle domande con un valore da 0 a 4 in relazione al livello di difficoltà percepito (dove 0 indica assenza di difficoltà e 4 il massimo disagio possibile).

Il punteggio totale è dato dalla somma dei punteggi ottenuti in ciascun ambito.

### 3.3 State Trait Anxiety Inventory (STAI)

È un test che consente di valutare l'ansia (Spielberger, 1980). Si compone di due brevi subtest di 20 item ciascuno, ai quali si risponde con una scala a 4 livelli d'intensità.

Il test fa riferimento alla distinzione concettuale tra ansia intesa come stato transitorio ed ansia intesa come tratto relativamente stabile di personalità. Il primo subtest si riferisce allo stato d'ansia al momento della somministrazione (ansia di stato). L'ansia di stato è una condizione transitoria che si presenta in associazione a stimoli specifici e persiste qualora le condizioni evocative non cessino. Lo stato d'ansia può variare d'intensità e fluttuare nel tempo in funzione della minaccia percepita. Il secondo subtest valuta l'ansia di tratto che è un aspetto relativamente durevole della personalità nella propensione all'ansia e consiste nella tendenza a percepire una vasta gamma di condizioni di vita come minacciose e a reagire ad esse con un'intensità elevata di ansia. Questa tendenza rimane latente finché non viene attivata da stress associati a reali o presunti pericoli particolari. Più alti sono i livelli

d'ansia di tratto più è probabile che un individuo possa sperimentare punte elevate d'ansia di stato in situazioni percepite come minacciose, specialmente in quelle che comportano rapporti interpersonali nei quali è ravvisabile una minaccia all'autostima del soggetto.

#### *3.4 Test di decodifica delle emozioni*

Il test (Angelillo *et alii*, 2007b; Fiorillo, 2005) consta nell'ascolto di una frase, semanticamente neutra, "Paolo legge tutto il giorno", interpretata da un soggetto di sesso maschile esprimendo tre diverse emozioni: Rabbia, Gioia, Tristezza, a tre diversi livelli d'intensità.

Inizialmente verranno presentate le frasi senza fornire informazioni circa la presenza di livelli di intensità differenti. Ad ogni frase ascoltata il soggetto dovrà indicare l'emozione che percepisce. Se non riesce a riconoscere alcuna emozione deve rispondere all'esaminatore che la frase non gli ha suscitato emozioni e la definisce quindi NEUTRA. Successivamente si cerca di capire se il soggetto si è reso conto di aver ascoltato diversi livelli di intensità per ciascuna emozione. Quindi si informa il soggetto circa l'esistenza di tre livelli di intensità per ciascuna emozione e gli si chiede di riascoltare le frasi, assegnando alle intensità uno dei tre livelli.

Le risposte del soggetto vengono annotate su un foglio di registrazione del test. Ultimata la somministrazione si congeda il soggetto e si passa allo studio ed elaborazione dei dati così ottenuti.

Una prima somministrazione pilota del test ha consentito al nostro gruppo di ricerca di rielaborare il test stesso, al fine di far emergere eventuali difetti di costruzione delle prove e di accertare il comportamento effettivo dei soggetti di fronte alle prove, per ottimizzarne le condizioni di somministrazione.

Il test è stato quindi somministrato ad un numero considerevole di soggetti di controllo di età compresa tra i 25 e i 50 anni, di ambedue i sessi.

Tali soggetti hanno risposto alle consegne richieste riportando soddisfacenti risultati che hanno permesso di verificare l'attendibilità delle prove.

Successivamente, il test revisionato è stato sottoposto al gruppo di pazienti in studio.

## **4. RISULTATI**

Per quanto riguarda il **Test per valutare la predisposizione all'abuso vocale**, va specificato che nel **gruppo II** è stata evidenziata una significativa disomogeneità nelle risposte, per cui i pazienti sono stati suddivisi in 2 sottogruppi in base alla media delle risposte VERO per ciascuna sezione del test: sottogruppo **A**, formato da 37 soggetti, in cui la media dei VERO per ciascuna sezione del test risultava inferiore a 5 e sottogruppo **B**, formato da 6 soggetti, in cui la media dei VERO risultava uguale o superiore a 5.

Nel dettaglio, abbiamo avuto i seguenti risultati (**Tabella 2**):

| I sezione<br>tendenza ad arrabbiarsi<br>facilmente |              |               | II sezione<br>tendenza ad urlare per<br>esprimere la propria rabbia |              |               | III sezione<br>tendenza ad urlare per attirare<br>l'attenzione |              |               | IV sezione<br>tendenza ad urlare per risolvere<br>i problemi |              |               |
|----------------------------------------------------|--------------|---------------|---------------------------------------------------------------------|--------------|---------------|----------------------------------------------------------------|--------------|---------------|--------------------------------------------------------------|--------------|---------------|
| I<br>gruppo                                        | II<br>gruppo | III<br>gruppo | I<br>gruppo                                                         | II<br>gruppo | III<br>gruppo | I<br>gruppo                                                    | II<br>gruppo | III<br>gruppo | I<br>gruppo                                                  | II<br>gruppo | III<br>gruppo |
|                                                    | A B          |               |                                                                     | A B          |               |                                                                | A B          |               |                                                              | A B          |               |
|                                                    | 3 6          |               |                                                                     | 2 7          |               |                                                                | 2 6          |               |                                                              | 3 7          |               |
| 1                                                  | 4,5          | 8             | 1                                                                   | 4,5          | 7             | 0                                                              | 4            | 7             | 1                                                            | 5            | 9             |

Tabella 2: Risultati del “Test per valutare la predisposizione all’abuso vocale”.

- per quanto riguarda la **I sezione**:

**I gruppo**: la media dei VERO è stata 1;

**II gruppo**: la media dei VERO è stata 4,5; all’interno del sottogruppo **A** la media dei VERO è stata 3 e all’interno del sottogruppo **B** è stata 6;

**III gruppo**: la media dei VERO è stata 8;

- per quanto riguarda la **II sezione**:

**I gruppo**: la media dei VERO è stata 1;

**II gruppo**: la media dei VERO è stata 4,5; all’interno del sottogruppo **A** la media dei VERO è stata 2 e all’interno del sottogruppo **B** è stata 7;

**III gruppo**: la media dei VERO è stata 7;

- per quanto riguarda la **III sezione**:

**I gruppo**: la media dei VERO è stata 0;

**II gruppo**: la media dei VERO è stata 4; all’interno del sottogruppo **A** la media dei VERO è stata 2 e all’interno del sottogruppo **B** è stata 6;

**III gruppo**: la media dei VERO è stata 7;

- per quanto riguarda la **IV sezione**:

**I gruppo**: la media dei VERO è stata 1;

**II gruppo**: la media dei VERO è stata 5; all’interno del sottogruppo **A** la media dei VERO è stata 3 e all’interno del sottogruppo **B** è stata 7;

**III gruppo**: la media dei VERO è stata 9.

Il **Voice Handicap Index** non ha evidenziato differenze significative tra i punteggi ottenuti dal **gruppo I** e quelli ottenuti dal **sottogruppo A** del **gruppo II**. Viceversa nel **sottogruppo B** del **gruppo II** e nel **gruppo III** si sono evidenziati punteggi elevati in particolare alle affermazioni che indagano l’impatto della patologia vocale sulla sfera fisica (percezione del paziente delle caratteristiche dell’emissione vocale) ed, in misura minore, alle affermazioni che indagano l’impatto della patologia vocale sulla sfera emozionale (impatto psicologico del sintomo) e sulla sfera funzionale (impatto delle limitazioni vocali sulle normali attività quotidiane). Analogamente, in questi soggetti (**sottogruppo B** del **gruppo II** e **gruppo III**) lo **STAI** ha evidenziato presenza significativa di ansia intesa sia come reazione attivata da condizioni che il soggetto valuta come minacciose sia come tendenza stabile a percepire come minacciose una vasta gamma di condizioni di vita.

Le risposte al **test di decodifica delle emozioni** hanno rilevato un’importante difficoltà nel riconoscimento delle emozioni (**Tabella 3-4**) nei soggetti del **gruppo III**.

|                    | Neut.    | Rabbia     |           |          | Gioia     |           |          | Tristezza |           |          |
|--------------------|----------|------------|-----------|----------|-----------|-----------|----------|-----------|-----------|----------|
|                    |          | I          | II        | III      | I         | II        | III      | I         | II        | III      |
| <b>Disf. Funz.</b> | <b>0</b> | <b>100</b> | <b>80</b> | <b>5</b> | <b>50</b> | <b>10</b> | <b>0</b> | <b>70</b> | <b>50</b> | <b>5</b> |
| <b>Controllo</b>   | <b>0</b> | <b>50</b>  | <b>15</b> | <b>0</b> | <b>10</b> | <b>0</b>  | <b>0</b> | <b>20</b> | <b>0</b>  | <b>0</b> |

Tabella 3: Percentuale di errore nella discriminazione tra frase neutra e frase a contenuto emozionale nei soggetti con disfonia disfunzionale (gruppo III) e nel gruppo di controllo (prima valutazione).

|                    | Neut.    | Rabbia    |           |          | Gioia     |          |          | Tristezza |           |          |
|--------------------|----------|-----------|-----------|----------|-----------|----------|----------|-----------|-----------|----------|
|                    |          | I         | II        | III      | I         | II       | III      | I         | II        | III      |
| <b>Disf. Funz.</b> | <b>0</b> | <b>85</b> | <b>25</b> | <b>5</b> | <b>35</b> | <b>5</b> | <b>0</b> | <b>40</b> | <b>30</b> | <b>5</b> |
| <b>Controllo</b>   | <b>0</b> | <b>35</b> | <b>10</b> | <b>0</b> | <b>5</b>  | <b>0</b> | <b>0</b> | <b>10</b> | <b>0</b>  | <b>0</b> |

Tabella 4: Percentuale di errore nella discriminazione tra frase neutra e frase a contenuto emozionale nei soggetti con disfonia disfunzionale (gruppo III) e nel gruppo di controllo (seconda valutazione).

In particolare alla prima valutazione (**Tabella 3**), in cui al paziente venivano fatte ascoltare le frasi dicendogli esclusivamente che si trattava di frasi emotivamente neutre e frasi che esprimevano diverse emozioni (gioia, tristezza, rabbia), senza specificare l'esistenza di diversi livelli di intensità per ciascuna emozione, un'alta percentuale di soggetti identificava la maggior parte delle frasi come emotivamente neutre e riusciva ad individuare correttamente il contenuto emozionale veicolato dalla frase solo al livello di intensità più elevato. Quando, alla seconda valutazione (**Tabella 4**), veniva spiegata l'esistenza di diversi livelli di intensità per ciascuna emozione, senza precisarne il numero, e veniva chiesto di identificare non solo l'emozione ma anche il livello di intensità, la difficoltà maggiore si riscontrava con la rabbia: infatti l'85% identificava la rabbia al primo livello di intensità come frase neutra, il 25% identificava la rabbia al secondo livello di intensità come frase neutra e il 5% identificava addirittura il terzo livello di intensità come frase neutra. Minori difficoltà si riscontravano con la tristezza e soprattutto con la gioia, anche se vi erano comunque grosse differenze rispetto al gruppo di controllo.

## 5. CONCLUSIONI

L'importante influenza delle emozioni e della personalità sul fenomeno vocale ed il ruolo svolto dalla voce come espressione della personalità e dello stato d'animo, hanno indotto numerosi studiosi a cercare di comprendere le correlazioni intercorrenti tra questi elementi, con particolare riguardo all'ambito dei disordini vocali. Le differenze metodologiche utilizzate rendono difficile effettuare una comparazione tra questi studi; tuttavia alcune caratteristiche comuni sono state riscontrate nei pazienti affetti da disfonia disfunzionale.

Un aumento dell'ansia è uno dei fattori più frequentemente menzionati negli studi effettuati su pazienti con alterazioni vocali di tipo funzionale: i pazienti vengono descritti come socialmente ansiosi, eccessivamente preoccupati per gli eventi della vita quotidiana.

L'ansia potrebbe essere un tratto distintivo, associato ad insicurezza ed eccessiva apprensione. È stata inoltre evidenziata una relazione temporale tra eventi stressanti, conflitti irrisolti o problematiche nelle relazioni familiari e l'insorgenza di un disordine vocale di tipo non organico. Le cause di conflitti di questo tipo sembrano essere fondate non su radicate problematiche psicopatologiche interne quanto piuttosto su quotidiane ansie, insuccessi, prepotenze, contrarietà, insoddisfazioni verso se stessi o gli altri, mancata realizzazione dei propri desideri, sentimenti di inadeguatezza e di perdita di fiducia in se stessi. In definitiva la personalità dei pazienti con disfonia disfunzionale può essere descritta come alquanto ansiosa, tendente a fuggire dai conflitti, con problemi nell'adeguamento emozionale e difficoltà nella manifestazione delle proprie emozioni. Sembra che gli individui con questi tratti siano particolarmente predisposti a sviluppare una disfonia funzionale e che sia da escludere che i cambiamenti di personalità possano essere conseguenza del disordine vocale (Di Costanzo *et alii*, 2007; Gerritsma, 1991; Roy *et alii*, 1997; Roy & Bless, 2000; Roy *et alii*, 2000a; Roy *et alii*, 2000b).

Molto importante potrebbe inoltre essere il valore attribuito dal soggetto al fenomeno vocale, nel senso che il soggetto potrebbe conferire al "volume" della voce un peso ed un valore eccessivi, considerandolo, impropriamente, uno strumento di affermazione sugli altri. Da una precedente sperimentazione del nostro gruppo è emerso che i pazienti con disfonia disfunzionale presentano un'alterata autopercezione della propria voce, cioè tendono a sentire la propria voce più compromessa di quanto sia, tanto da limitare anche le attività che richiedono l'utilizzo della voce. Se il disordine vocale persiste, si possono verificare ripercussioni psicosociali, come la sensazione di essere gravemente ammalato, di essere inadeguato e spesso una apprensione di tipo esistenziale, soprattutto nei soggetti in cui l'attività professionale dipende dalla voce. In questo modo il disordine vocale determinerà conseguenze di tipo psicologico, rinforzando il distress vocale, che, come in un circolo vizioso, condurrà ad un peggioramento della funzione vocale.

Partendo da queste osservazioni, abbiamo ipotizzato, che alla base dell'insorgenza della disfonia disfunzionale, vi possa essere un'incapacità nel decodificare e codificare correttamente le proprie emozioni attraverso il canale vocale. Abbiamo pertanto valutato (**Tabella I**): 50 soggetti (**gruppo I**) non affetti da disfonia, 61 soggetti (**gruppo II**) affetti da disfonia disfunzionale che non avevano mai praticato terapia logopedica e 20 soggetti (**gruppo III**) che, pur presentando una corretta respirazione costo-diaframmatica e un adeguato accordo pneumofonico dopo diversi cicli di terapia logopedica, continuavano a presentare disfonia.

In questi soggetti abbiamo analizzato:

- la tendenza ad arrabbiarsi facilmente e ad urlare come modalità di affermazione sugli altri mediante il **Test per valutare la predisposizione all'abuso vocale**;
- la autovalutazione vocale (mediante il **Voice Handicap Index - VHI**);
- la presenza di ansia (mediante lo **State Trait Anxiety Inventory - STAI**).

Nei soggetti del **gruppo III** abbiamo valutato anche la capacità di discriminare il contenuto emozionale veicolato da una frase semanticamente neutra, recitata da un attore professionista, che esprimesse tre emozioni primarie (rabbia, gioia, tristezza) a tre diversi livelli di intensità (**Test di decodifica delle emozioni**), confrontando i risultati ottenuti da questi soggetti con quelli ottenuti dai soggetti del gruppo di controllo.

L'analisi delle risposte al **Test per valutare la predisposizione all'abuso vocale (Tabella 2)** ha evidenziato una netta prevalenza di risposte VERO nel **gruppo III** e in alcuni soggetti del **gruppo II** che abbiamo definito **sottogruppo B**.

Questi stessi soggetti hanno presentato al VHI una prevalenza del disagio soprattutto a carico della sfera fisica e allo STAI presenza significativa di ansia.

Il test di decodifica delle emozioni (**Tabella 3-4**) ha evidenziato nei soggetti disfonici del **gruppo III** la incapacità di riconoscere correttamente il contenuto emozionale delle frasi proposte se non a livelli elevati di intensità, con difficoltà maggiori per la rabbia.

È possibile ipotizzare che nei pazienti del **gruppo III** la mancata risoluzione della disfonia, nonostante il corretto accordo pneumofonico, potrebbe dipendere dal fatto che tali pazienti pur controllando bene tutti i parametri vocali (attacco, tenuta) in situazioni ottimali, non riescono a gestirli in maniera altrettanto adeguata nelle situazioni comuni della vita quotidiana, probabilmente in quanto tendono a conferire al "volume" della voce un peso ed un valore eccessivi, considerandolo, impropriamente, uno strumento di affermazione sugli altri. Ciò coincide con la netta prevalenza di risposte VERO al **Test per valutare la predisposizione all'abuso vocale**.

Per quanto riguarda il **sottogruppo B del gruppo II** la prevalenza di risposte VERO al nostro test fa supporre che questi soggetti presentino caratteristiche di personalità che potrebbero essere una concausa nella genesi della disfonia. Pertanto questi soggetti potrebbero non risolvere la disfonia nonostante una adeguata terapia logopedica.

I risultati da noi ottenuti confermano l'influenza dell'assetto emotivo-relazionale sulle modalità di comunicazione in genere e sul canale vocale in particolare, sottolineando l'importanza della valutazione di tali aspetti nei soggetti affetti da disfonia disfunzionale. È possibile inoltre ipotizzare che in alcuni pazienti affetti da disfonia disfunzionale sarebbe utile associare al trattamento riabilitativo tradizionale uno specifico protocollo riabilitativo (Angelillo *et alii*, 2007c) al fine di aiutare i pazienti disfonici a prendere consapevolezza della loro difficoltà nella codifica e decodifica delle emozioni e ad esprimere in maniera adeguata il contenuto ed il livello emozionale della propria voce nei vari contesti di vita sociale, familiare e professionale.

## BIBLIOGRAFIA

Angelillo, N., Di Costanzo, B., Pezone, M.R., Fiorillo, A., Costa, G. & Barillari, U. (2007), Proposta di un nuovo test per la valutazione dell'assetto emotivo-relazionale nell'etiopatogenesi della disfonia disfunzionale, *XXXI Congresso Nazionale della Società Italiana di Audiologia e Foniatria "VOCE, DIAPASON E NANOTECH"*, Ferrara, 12-15 settembre.

Angelillo, N., Di Costanzo, B., Fiorillo, A., Pezone, M.R., d'Anna, G. & Barillari, U. (2007), Proposta di un nuovo test nell'inquadramento multidisciplinare della disfonia disfunzionale: "test per la decodifica e codifica delle emozioni", *XXXI Congresso Nazionale della Società Italiana di Audiologia e Foniatria "VOCE, DIAPASON E NANOTECH"*, Ferrara, 12-15 settembre.

Angelillo, N., Fiorillo, A., Napolitano, N., Di Costanzo, B., Tripodi M. & Barillari, U. (2007), Rieducare alle emozioni: il laboratorio teatrale come trattamento integrato nella presa in carico del paziente con alterazioni funzionali della voce, *XXXI Congresso Nazionale della Società Italiana di Audiologia e Foniatria "VOCE, DIAPASON E NANOTECH"*, Ferrara, 12-15 settembre.



Angelillo, N., Di Costanzo, B., Fiorillo, A., Mirra, G. & Barillari U. (in press), Il ruolo dell'assetto emotivo-relazionale nell'etiopatogenesi della disfonia disfunzionale, *Acta Phoniatica Latina*.

Anolli, L. & Ciceri, R. (1997), *La voce delle emozioni*, Milano: Franco Angeli.

Barillari, U., de Miranda, M. ecc. (2003), Uno studio sulla disfonia disfunzionale attraverso l'Illness Behaviour Questionnaire, l'Eysenck Personality Questionnaire, lo State Trait Anxiety Inventory e la Scala di Paykel per gli Eventi Stressanti, *Acta Phoniatica Latina*, 25, 359-369.

Di Costanzo, B., Angelillo, N., Pezone, M.R., Fiorillo, A., Mirra, G. & Barillari U. (2007), Caratteristiche di personalità nei soggetti adulti con disfonia disfunzionale, *XLI Congresso Nazionale S.I.F.E.L. (Società Italiana Foniatria e Logopedia)*, Cervia Milano Marittima, 21-24 novembre.

Erickson, D., Yoshida, K., Menezes, C., Fujino, A., Mochida, T. & Shibuya, Y. (2006), Exploratory study of some acoustic and articulatory characteristics of sad speech, *Phonetica*, 63, 1-25.

Fiorillo, A. (A.A. 2004-05), *Rieducare alle emozioni: il laboratorio teatrale come trattamento integrato nella presa in carico del paziente con alterazioni funzionali della voce*, Tesi di Laurea in Logopedia, Seconda Università degli Studi di Napoli, Relatore Dott. N. Angelillo.

Gerritsma, E.J. (1991), An investigation into some personality characteristics of patients with psychogenic aphonia and dysphonia, *Folia Phoniatica* (Basel), 43, 13-20.

Le Huche, F. & Allali, A. (1992), *La Voce: Anatomia e Fisiologia, Patologia, Terapia*, Milano: Masson.

Monnot, M., Orbelo, D., Riccardo, L., Sikka, S. & Rossa, E. (2003), Acoustic analyses support subjective judgments of vocal emotion, *Annals of the New York Academy of Sciences*, 1000, 288-292.

Pezone, M.R. (A.A. 2005-06), *Nuovo test nell'inquadramento multidisciplinare della disfonia funzionale: Test per la valutazione della tendenza all'abuso vocale in relazione al contesto ed allo stato d'animo*, Tesi di Laurea in Logopedia, Seconda Università degli Studi di Napoli, Relatore Dott. N. Angelillo.

Raithel, V. & Hielscher-Fastabend, M. (2004), Emotional and linguistic perception of prosody, *Folia Phoniatica Logopedica*, 56, 7-13.

Roy, N., McGrory, J.J., Tasko, S.M., Bless, D.M., Heisey, D. & Ford, C.N. (1997), Psychological correlates of functional dysphonia: an investigation using the Minnesota Multiphasic Personality Inventory, *Journal of Voice*, 11, 443-451.

Roy, N. & Bless, D.M. (2000), Personality traits and psychological factors in voice pathology: a foundation for future research. *Journal of Speech, Language and Hearing Research*, 43, 737-748.

Roy, N., Bless, D.M. & Heisey, D. (2000), Personality and voice disorders: a superfactor trait analysis, *Journal of Speech, Language and Hearing Research*, 43, 749-768.

- Roy, N., Bless, D.M. & Heisey, D. (2000), Personality and voice disorders: a multitrait-multidisorder analysis, *Journal of Voice*, 14, 521-548.
- Sobin, C. & Alpert, M. (1999), Emotion in speech: the acoustic attributes of fear, anger, sadness, and joy, *Journal of Psycholinguistic Research*, 28, 347-365.
- Toivanen, J., Waaramaa, T., Alku, P., Laukkanen, A.M., Seppanen, T., Vayrynen E. & Airas, M. (2006), Emotions in [a]: a perceptual and acoustic study, *Logopedics, Phoniatrics, Vocology*, 31, 43-48.
- Whiteside, S.P. (1999), Acoustic characteristics of vocal emotions simulated by actors, *Perceptual and motor skills*, 89, 1195-1208.

## DEFICIT DELLE COMPETENZE SEMANTICO-PRAGMATICHE IN SOGGETTI NORMOUDENTI ED IPOACUSICI

Nicola Angelillo, Anna Vitiello, Brigida Di Costanzo, Umberto Barillari  
Servizio di Foniatria ed Audiologia, Dir. Prof. Umberto Barillari  
Seconda Università degli Studi di Napoli  
[Nicola.angelillo@unina2.it](mailto:Nicola.angelillo@unina2.it); [anna.vitiello1@virgilio.it](mailto:anna.vitiello1@virgilio.it); [b.dicostanzo@tiscali.it](mailto:b.dicostanzo@tiscali.it);  
[umberto.barillari@unina2.it](mailto:umberto.barillari@unina2.it)

### 1. SOMMARIO

La semantica è la disciplina che studia il significato delle parole indipendentemente dal contesto, mentre la pragmatica si occupa degli scopi per cui la lingua viene utilizzata, ossia di come il contesto influisca sull'interpretazione dei significati. Adeguate competenze semantico-pragmatiche sono indispensabili per poter utilizzare la comunicazione verbale e non verbale in modo idoneo agli scopi e alle varie situazioni contestuali. Il deficit di tali competenze può essere da un lato un'entità a se stante, secondaria ad alterazioni organiche o funzionali delle strutture cerebrali preposte all'organizzazione delle suddette competenze linguistiche, in assenza di deficit cognitivi, sensoriali, motori, affettivi o di carenze socio-ambientali importanti; dall'altro, può associarsi a quadri clinici nosograficamente definiti che comportano un deficit delle competenze comunicative a più livelli.

In particolare, i soggetti affetti da ipoacusia severa o profonda insorta in epoca pre-verbale presentano difficoltà semantico-pragmatiche in quanto il canale sensoriale che, per eccellenza, viene utilizzato per accrescere le proprie conoscenze, e cioè "l'udito", è deficitario. Tale deficit comporta un rallentamento nello sviluppo di molti processi integrativi superiori, alcuni dei quali risultano gravemente inficiati. L'entità del deficit è ovviamente correlata alla gravità dell'ipoacusia, alla precocità e al tipo di protesizzazione, alla precocità e alle modalità di riabilitazione. Va sottolineata, a nostro avviso, l'importanza, per lo sviluppo di adeguate competenze semantico-pragmatiche nei soggetti ipoacusici, dell'utilizzo di canali comunicativi vicarianti attraverso cui acquisire informazioni, come la L.I.S. (Lingua Italiana dei Segni), fin dalle prime epoche di vita. Dalla nostra esperienza è infatti emerso che soggetti ipoacusici, appartenenti a famiglie di ipoacusici che utilizzano la L.I.S., adeguatamente stimolati tramite questo canale di comunicazione, presentano competenze semantico-pragmatiche quasi comparabili a quelle dei normoudenti, a prescindere dalle loro abilità fonetico-fonologiche.

Riguardo la valutazione delle competenze semantico-pragmatiche, va sottolineato che, mentre esistono numerosi tests per la valutazione e protocolli per la riabilitazione degli aspetti formali del linguaggio, vi è una carenza di tali strumenti per quanto riguarda le competenze semantico-pragmatiche, che pure sono fondamentali per lo sviluppo di capacità comunicative realmente efficaci.

Per tali ragioni gli autori hanno realizzato una batteria di test finalizzata a valutare le competenze semantico-pragmatiche in soggetti normoudenti ed ipoacusici e specifici protocolli di riabilitazione che agiscano su queste competenze.

### 2. INTRODUZIONE

Il linguaggio è un codice costituito da un complesso armonico di segni in grado di produrre un numero infinito di parole, frasi, discorsi, secondo principi atti ad associare gli elementi della struttura fonologica (i significanti) con i corrispondenti elementi della

struttura semantica (i significati). Il codice linguistico va, dunque, integrato con i processi situazionali in cui la comunicazione avviene (il contesto) e con le capacità di trarre le opportune inferenze da parte dell'interlocutore. Se, da un lato, la semantica analizza il significato delle espressioni linguistiche, dall'altro, la pragmatica ne contestualizza i significati, conseguentemente all'adattarsi della comunicazione ad un determinato contesto situazionale.

La semantica è la disciplina che studia il significato delle parole indipendentemente dal contesto, mentre la pragmatica si occupa degli scopi per cui la lingua viene utilizzata, ossia di come il contesto influisca sull'interpretazione dei significati (McDonald, 1999; Prutting, 1987).

La semantica è infatti il settore che si occupa del significato che noi diamo alle cose, la pragmatica è il settore che si occupa del reale significato che assume quella cosa nel linguaggio che utilizziamo tutti i giorni.

Dire la parola "campana", ci fa immediatamente venire alla mente la forma, il materiale ed il suono di questo oggetto, ma dire "bisogna sempre ascoltare entrambe le campane", vuol dire che prima di decidere tra due litiganti chi abbia ragione, bisogna ascoltare le ragioni di entrambi. Ancora, con l'espressione "stare in campana" si invita qualcuno a stare all'erta, stare sveglio, pronto a tutto, attendersi situazioni da affrontare, pericoli; mentre l'espressione "mettere sotto una campana di vetro" focalizza l'attenzione dell'interlocutore verso un'azione di protezione, di salvaguardare qualcuno o qualcosa dall'ambiente esterno. Ovviamente, il giusto significato da intendere sarà quello legato agli atti che si svolgono nella situazione contestuale in cui sono inseriti i parlanti di quella conversazione. Ecco come tra la semantica e la pragmatica sussiste un rapporto relazionale inscindibile.

Il linguaggio che si utilizza tutti i giorni è pieno di espressioni idiomatiche e sono appunto le adeguate competenze semantico-pragmatiche possedute a farci capire che, spesso, non bisogna fermarsi al significato letterale di un'espressione verbale ma cercare di comprendere e decodificare il suo significato non letterale.

Adeguate competenze semantico-pragmatiche sono indispensabili per poter utilizzare la comunicazione verbale e non verbale in modo idoneo agli scopi e alle varie situazioni contestuali e sono frutto di abilità cognitive superiori, risultato dell'integrazione di più capacità sia innate che acquisite.

### 3. IL DEFICIT DELLE COMPETENZE SEMANTICO-PRAGMATICHE

I deficit a carico delle competenze semantico-pragmatiche che riguardano presumibilmente una alterazione selettiva delle strutture cerebrali preposte all'organizzazione delle suddette competenze linguistiche, in assenza di deficit cognitivi, sensoriali, motori, affettivi o di carenze socio-ambientali importanti, determinano il quadro del **Disturbo semantico-pragmatico puro** – DSP (Fasano, 2007). In letteratura sono presenti numerosi studi sul deficit delle competenze semantico-pragmatiche effettuati su soggetti con lesione cerebrale a carico dell'emisfero destro e su soggetti che avevano subito un intervento chirurgico di split-brain (Angelillo *et alii*, 2005; Baum *et alii*, 1997; Boss, 1996; Brownell & Stringfellow, 1999; Juncos-Rabadan, 1992; Kaplan *et alii*, 1990; Kasher *et alii*, 1999; Petrania, 2004; Shields, 1991; Siegal *et alii*, 1996). Tali studi hanno evidenziato come questi soggetti non presentassero gravi difficoltà in ordine alle competenze fonetico-fonologiche, morfo-sintattiche e semantico-lessicali, ma mostravano evidenti deficit in ordine all'elaborazione degli aspetti prosodici della codifica linguistica (ad esempio difficoltà nel confrontare stati d'animo espressi mediante modulazioni del tono vocale con stati d'animo rappresentati in disegni raffiguranti espressioni facciali, difficoltà

nella identificazione della giusta posizione che deve assumere un accento enfatico, difficoltà nel differenziare una domanda da un'affermazione), all'analisi dei significati non letterali (l'ironia, le metafore, le espressioni idiomatiche, i proverbi), all'elaborazione di aspetti pragmatici e testuali dell'informazione veicolata incentrati sulle nozioni di contesto, inferenze ed intenzioni comunicative, alla capacità di interpretare la mimica facciale, lo sguardo, i movimenti del corpo, i gesti.

D'altro canto, in presenza di quadri patologici conclamati, le competenze semantico-pragmatiche possono risultare inadeguate come conseguenza di un danno generalizzato, secondario a deficit primari (Adams & Lloyd, 2005; Bara, 1999; Campbell & Shriberg, 1982; Carlomagno *et alii*, 2005; Happè, 1993; Monfort, 2004; Zaidel *et alii*, 2000), quali:

- Ritardo specifico del linguaggio
- Deficit sensoriali
- Deficit cognitivi
- Danno cerebrale acquisito o congenito
- Disturbi comportamentali o affettivo-relazionali
- Ipostimolazione ambientale

In tutti questi casi si parlerà di **deficit a carico delle competenze Semantico-Pragmatiche Secondario**, in cui, oltre ai deficit relativi a queste competenze, saranno evidenti tutti gli altri deficit che rappresenteranno la patologia in atto nella sua complessità (Fasano, 2007).

Nella maggior parte dei casi di deficit semantico-pragmatico, in particolare, quando non si riscontra in associazione a casi patologici conclamati, inizialmente si evidenzia un ritardo nelle tappe dello sviluppo del linguaggio, caratterizzato da eloquio poco comprensibile per la frequente presenza di dislalie multiple, difficoltà nella codifica e decodifica fonologica, e, occasionalmente, da ecolalie, associato a scarsa propositività e partecipazione all'ambiente circostante. In seguito, il linguaggio diviene relativamente ben strutturato, sebbene continuano a esserci difficoltà in output verbale. Ciò comporta una serie di problematiche che si presentano man mano che il bambino fa esperienza della realtà. I soggetti con disturbo semantico-pragmatico, in realtà, sono in grado di apprendere numerose parole, anche di uso non comune, senza riuscire ad utilizzarle durante una comunicazione verbale in quanto non sono in grado di applicarle contestualmente alla situazione vissuta. Nel corso del tempo, migliora la comprensione verbale ma si evidenziano difficoltà nelle interazioni sociali: benché questi soggetti presentino un linguaggio sintatticamente ben formato e fonologicamente corretto, il contenuto di ciò che viene espresso non è adeguato al contesto, si passa frequentemente da un argomento all'altro con difficoltà nel sostenere una normale conversazione e nell'impegnarsi in un discorso a scopo comunicativo. Successivamente emergono difficoltà nel comprendere le regole delle situazioni sociali e delle conversazioni e si evidenzia la tendenza ad interpretare il linguaggio in maniera molto letterale con difficoltà nel comprendere il linguaggio non letterale, come le espressioni idiomatiche, i proverbi, l'umorismo, le valenze ironiche, gli elementi sarcastici. Ancora, si rileva difficoltà nell'integrare le differenti informazioni e nell'inferire il significato di un enunciato, nell'uso corretto delle richieste indirette, nel comprendere ed interpretare la prosodia emotiva ed espressivo-situazionale. Tutto ciò pregiudicherà nel suo insieme la capacità comunicativa globale dei soggetti affetti da DSP.

Diversamente avviene quando il deficit di tali competenze è presente in associazione a quadri patologici conclamati, in presenza dei quali, oltre al quadro clinico, come su esposto

circa lo sviluppo delle competenze semantico-pragmatiche, saranno evidenti tutti gli altri deficit che rappresenteranno la patologia in atto nella sua complessità (Fasano, 2007).

#### **4. IL DEFICIT DELLE COMPETENZE SEMANTICO-PRAGMATICHE NEI SOGGETTI IPOACUSICI**

In caso di soggetti affetti da Ipoacusia, lo sviluppo delle competenze semantico-pragmatiche varia a seconda del grado di perdita uditiva, del tipo (protesi acustica, impianto cocleare) e dell'epoca di protesizzazione, del tipo di riabilitazione (metodo oralista, metodo gestuale, metodo bimodale, metodo verbo tonale) realizzato e dalla presenza/assenza di disabilità associate (Angelillo, 2007). In genere, a livello semantico mancheranno i nessi gerarchici che consentono di passare da categorie semantiche generali a categorie semantiche specifiche; l'interazione tra memoria semantica e memoria episodica risulterà completamente assente ed il processo di elaborazione entrate lessicali apparirà molto superficiale e legato all'esperienza diretta del piccolo.

L'incapacità del bambino ipoacusico di trarre significato dall'elaborazione dell'informazione a livello centrale può avere un effetto da lieve a devastante sulle sue competenze pragmatiche: gli aspetti non verbali, come il tono della voce, la prosodia emotiva e l'uso di forme di linguaggio non letterale come battute di spirito, ironia e sarcasmo, decodifica di metafora e proverbi opportunamente usati rispetto ad un evento collocato nello spazio e nel tempo, significano ben poco per questo bambino. Il soggetto ipoacusico, infatti, tende ad interpretare in modo letterale il significato delle parole, laddove invece acquisiscono significati molto diversi. Oltre ad un lessico povero e poco flessibile si riscontra una difficoltà nel formulare ipotesi e soprattutto nel fare inferenze sia lessicali che testuali, traducendosi in una contrazione delle competenze conversazionali, narrative ed in un'incapacità di comprensione del testo, che resta uno dei gravi problemi anche per i soggetti ipoacusici con discreta competenza nell'espressione verbale.

E' possibile affermare, dai dati in nostro possesso (Angelillo, 2007), che la maggior parte dei soggetti affetti da ipoacusia di grado medio che hanno seguito un training logopedico presenteranno uno sviluppo delle competenze semantiche adeguato, mentre mostreranno lievi difficoltà per quanto concerne lo sviluppo delle competenze pragmatiche: in particolare, non saranno in grado di comprendere immediatamente il senso sarcastico durante una conversazione tra più interlocutori, né di inferire i significati situazionali più complessi di un'implicatura linguistica. Caratteristiche simili presentano anche la maggior parte di ipoacusici protesizzati, figli di soggetti ipoacusici che utilizzano come modalità di comunicazione la L.I.S. (Lingua Italiana dei Segni), riabilitati con training logopedico oralista. Per i soggetti ipoacusici che si esprimono esclusivamente attraverso la L.I.S., si evidenzieranno lievi difficoltà nell'estrapolare da una categoria semantica primaria le sottocategorie ad essa riferite, specie se gli elementi appartenenti alla categoria prescelta sono connotati da proprietà di tipo storico, scientifico, tecnico ecc. Mentre a livello pragmatico presenteranno difficoltà simili, e non superiori come ci si potrebbe aspettare, a quelle mostrate dai soggetti affetti da ipoacusia di grado medio che hanno seguito un training logopedico con metodica oralista.

Tutti i soggetti affetti da ipoacusia grave o profonda, protesizzati e sottoposti a training logopedico mostreranno difficoltà di grado variabile, nella codifica e decodifica del linguaggio non letterale della comunicazione. In particolare non saranno in grado di avvalersi dell'uso di espressioni idiomatiche, metaforiche e proverbiali per rendere verbalmente la propria idea in un dato contesto situazionale, non saranno capaci di cogliere il sarcasmo o l'ironia durante un'interazione sociale né di utilizzare battute di spirito per

trasmettere un'informazione cercando di uscire dal proprio *schema mentis*, che risulterà così molto rigido e poco fluido (Angelillo, 2007).

Ancora più complesso risulterà il quadro clinico dei soggetti ipoacusici con disabilità associate, per i quali il deficit semantico-pragmatico risulterà tanto più grave quanto più canali d'ingresso, oltre a quello uditivo, saranno compromessi.

## **5. LA VALUTAZIONE E RIABILITAZIONE DELLE COMPETENZE SEMANTICO-PRAGMATICHE**

Mentre esistono numerosi tests per la valutazione e protocolli per la riabilitazione degli aspetti formali del linguaggio, ossia i livelli fonetico-fonologico, lessicale e morfo-sintattico, vi è una carenza di tali strumenti (McDonald & Pearce, 1995; Zanini *et alii*, 2005) per quanto riguarda le competenze semantico-pragmatiche, che pure sono fondamentali per lo sviluppo di capacità comunicative realmente efficaci.

Per tali ragioni è stata realizzata una batteria di test (Vitiello, 2005, Angelillo *et alii*, in press), articolata in cinque sezioni principali riservate a semantica, umorismo, emozioni, linguaggio non letterale e processi integrativi, che consenta di valutare le competenze semantico-pragmatiche in età evolutiva e di guidare e monitorare l'andamento del programma riabilitativo. Tale batteria di test originaria è stata riadattata in modo tale da poter essere somministrata anche a soggetti ipoacusici (Di Ruocco, 2005; Angelillo, 2007), indipendentemente dall'entità del deficit uditivo, utilizzando una modalità di somministrazione non verbale, eliminando in tal modo le difficoltà di comprensione del messaggio verbale e concentrando l'attenzione sulle consegne.

Il programma riabilitativo da noi elaborato (Vitiello *et alii*, 2007) nei soggetti con deficit delle competenze semantico-pragmatiche, da affiancare ai protocolli convenzionali nei casi in cui siano presenti deficit ad altri livelli della comunicazione (Angelillo *et alii*, 2007), è finalizzato ad educare il soggetto all'uso del linguaggio nelle reali dinamiche comunicative, rendendolo capace di codificare e decodificare il contenuto reale di un enunciato, di attribuire il significato soggettivo della realtà in cui si svolge l'atto comunicativo, definito all'interno della lingua usata, di usare la comunicazione verbale e non verbale in modo adeguato agli scopi e alla situazione contestuale e di studiare il linguaggio in rapporto all'agire del parlante.

Stimolare le competenze semantico-pragmatiche, infatti, significa educare il soggetto all'uso del linguaggio nelle reali dinamiche comunicative. La nostra proposta mira alla stimolazione di tutte quelle abilità non verbali, quali la cinesica, la prossemica, l'aptica e la cronemica, promuovendo l'attivazione costante di quei processi cognitivi superiori che consentono a ciascun individuo di servirsi di forme di linguaggio idiomatico per comunicare, di attivare implicature inferenziali, di contestualizzare rispetto ad una data situazione il proprio agire, di codificare e decodificare i vari modi di dire, gli stati emotivi che muovono i comportamenti altrui, di prevenire i comportamenti altrui comprendendo il contesto in cui essi si svolgono, di intendere il sarcasmo, l'ironia, l'umor, la metafora, il proverbio o una morale come strumenti comunicativi potenti in grado di enfatizzare un'opinione nel momento in cui essa viene espressa. In tal modo si riuscirà ad aiutare il soggetto con Deficit semantico pragmatico a servirsi del linguaggio in tutti i suoi aspetti sia linguistici che extralinguistici e ad annullare false prognosi dovute ad un inquadramento diagnostico poco specifico e settoriale (casi di falsi negativi o falsi positivi)

Il trattamento logopedico per il deficit delle competenze Semantico-Pragmatiche Secondario presuppone un'analisi attenta di tutti i vari aspetti del quadro clinico del paziente e richiede la presa in carico globale del soggetto in modo da realizzare un

intervento riabilitativo personalizzato ed adeguato rispetto al quadro patologico complessivo. In tal senso, il trattamento atto all'abilitazione o alla riabilitazione delle competenze semantico-pragmatiche in quadri patologici conclamati, potrà non solo migliorare ed adeguare le suddette competenze, ma rappresentare un efficace trampolino di lancio attraverso cui si potrà agire su tutti gli altri livelli del linguaggio, garantendo al soggetto un miglioramento soprattutto nell'interazione sociale (Angelillo *et alii*, 2007).

Nel caso di deficit delle competenze semantico-pragmatiche secondario a diagnosi di Ipoacusia la stimolazione uditiva rappresenterà l'aspetto più importante della riabilitazione. Partendo dal concetto che il processo di decodificazione del messaggio si attua grazie alla capacità insita nel soggetto di differenziare il suono dal silenzio, la parola dal rumore, una parola dalle altre parole, la stimolazione uditiva si sviluppa secondo le tradizionali tappe dell'allenamento acustico quali detezione, discriminazione, identificazione, riconoscimento, comprensione. Si parlerà di vera "comprensione" solo quando il soggetto ipoacusico percepisce il valore semantico del discorso e, all'occorrenza, risponde, in modo appropriato. La percezione di altri parametri, quali i caratteri soprasegmentali della catena parlata, aggiunti al carattere ridondante della lingua e alle informazioni fornite da canali paralleli a quello acustico, facilita la decodificazione del messaggio.

La riabilitazione, all'inizio, verterà sullo sviluppo delle prestazioni relative alla percezione di rumori e suoni ambientali, delle informazioni prosodiche del discorso e, in seguito, delle strutture fonetiche. Per il soggetto ipoacusico in età evolutiva accanto all'allenamento acustico andranno parallelamente svolti esercizi tesi allo sviluppo della semantica che consentiranno uno sviluppo in tappe dell'assetto sintattico, del livello morfologico e fonetico fonologico. Inoltre, quando il soggetto si troverà in età scolare, potrà cominciare anche una massiva stimolazione di tutte le abilità extralinguistiche che rientrano nel campo della pragmatica. Per il soggetto ipoacusico in età adulta la riabilitazione delle competenze semantico-pragmatiche riguarderà soprattutto le caratteristiche prosodiche del discorso e la decodifica e codifica del linguaggio non letterale.

L'intervento logopedico in ambito semantico-pragmatico verterà sull'ampliamento del patrimonio semantico mediante esercizi specifici (Di Ruocco, 2005; Fasano, 2007). Successivamente, ci si concentrerà soprattutto sulla decodifica e codifica della Prosodia Linguistica e della Prosodia Emozionale, solo più tardi sarà possibile soffermarsi sulla tappa tesa alla decodifica e codifica di tutti i meccanismi facenti parte della comunicazione non verbale. Infine, si consentirà al soggetto di automatizzare le acquisizioni apprese e di adattarsi coerentemente ad ogni contesto sia esso conversazionale che situazionale. Il lavoro svolto a livello semantico consentirà un ampliamento ed un miglioramento del lessico, mentre il lavoro svolto circa il corretto uso della prosodia linguistica ed emozionale influirà sull'assetto del livello morfosintattico sia in input che in output. Ancora i meccanismi di logica miglioreranno attraverso la stimolazione delle abilità di trarre inferenza ed attivare processi pertinenti ad elaborazioni che implicino un'adeguata Teoria della Mente. Le abilità dialogico-narrative si evolveranno enormemente così come il pensiero logico-matematico, dove il soggetto risulterà più dinamico e pronto a risolvere le consegne presentate mediante attivazione di processi di analisi e logico-deduttivi.

Soltanto agendo sugli aspetti appena esposti sarà possibile dotare il soggetto di competenze comunicative oggettivamente valide nelle reali situazioni della vita quotidiana.



## 6. CONCLUSIONI

Scopo del nostro lavoro è stato quello di evidenziare l'importanza del deficit delle competenze semantico-pragmatiche del linguaggio in età evolutiva, deficit che viene spesso misconosciuto, e che, se non opportunamente diagnosticato e trattato, non consente al paziente di acquisire adeguate competenze nelle reali situazioni comunicative, con grave limitazione delle possibilità di interazione sociale.

Nel caso in cui, inoltre, il deficit delle competenze semantico-pragmatiche sia secondario ad altri quadri patologici, quali il deficit uditivo, l'approccio riabilitativo ad orientamento pragmatico potrà non solo migliorare ed adeguare le competenze semantico-pragmatiche ma rappresenterà un efficace strumento attraverso cui si potrà agire su tutti gli altri livelli del linguaggio, garantendo al soggetto un miglioramento soprattutto delle capacità di interazione sociale (Angelillo, 2007; Di Ruocco 2005; Fasano, 2007).

Abbiamo pertanto realizzato una batteria di test (Vitiello, 2005; Angelillo *et alii*, in press), finalizzata a valutare e quantificare il deficit delle competenze semantico-pragmatiche, in grado inoltre di "guidare" il trattamento riabilitativo e di monitorarne i risultati nel tempo, così da consentire la personalizzazione del progetto riabilitativo individuabile.

## BIBLIOGRAFIA

Adams, C. & Lloyd, J. (2005), Elicited and spontaneous communicative functions and stability of conversational measures with children who have pragmatic language impairments, *International Journal of Language and Communication Disorders*, 40, 333-347.

Angelillo, N., de Miranda, M., Mirra, G., Di Costanzo, B., Tripodi, M., Tozzi, E. & Barillari, U. (2005), Disturbo semantico-pragmatico del linguaggio ed emisfero destro, *XXXIX Congresso Nazionale S.I.F.E.L. (Società Italiana Foniatria e Logopedia)*, Bagni di Tivoli (Rm), 13-16 aprile.

Angelillo, N. (2007), *Development and evaluation of semantic-pragmatic competences in normal hearing and deaf subjects*, Tesi di Dottorato in: "Scienze Biomediche, Endocrinologiche e Neurofisiologiche", curriculum "Elettrofisiologia e Biologia Cellulare dell'Apparato Uditivo", Università degli Studi di Ferrara, Relatore Prof. A. Martini.

Angelillo, N., Vitiello, A., Di Costanzo, B., Barillari, M.R. & Barillari, U. (2007), Abilità metacognitive e competenze semantico-pragmatiche nel ritardo mentale lieve, *Acta Phoniatria Latina*, 29, 85-91.

Angelillo, N., Di Costanzo, B., Angelillo, M., Vitiello, A., Barillari, M.R. & Barillari U. (in press), Valutazione delle competenze extralinguistiche in bambini già sottoposti a terapia logopedica per disturbo del linguaggio, *Acta Phoniatria Latina*.

Bara, B.G., Bosco, F.M. & Bucciarelli, M. (1999), Developmental pragmatics in normal and abnormal children, *Brain and Language*, 68, 507-528.

Baum, S., Pell, M., Leonard, C. & Gordon, J. (1997), The ability of right-and left-hemisphere damaged individuals to produce and interpret prosodic cues marking phrasal boundaries, *Language and Speech*, 40, 313-330.

Boss, B.J. (1996), Pragmatics: right brain communication, *Axone*, 17, 81-85.

- Brownell, H. & Stringfellow, A. (1999), Making requests: illustrations of how right-hemisphere brain damage can affect discourse production, *Brain and Language*, 68, 442-465.
- Campbell, T.F. & Shriberg, L.D. (1982), Associations among pragmatic functions, linguistic stress, and natural phonological processes in speech-delayed children, *Journal of Speech and Hearing Research*, 25, 547-553.
- Carlomagno, S., Santoro, A., Menditti, A., Pandolci, M. & Marini, A. (2005), Referential communication in Alzheimer's type dementia, *Cortex*, 41, 520-534.
- de Villiers, P. (2004), Assessing pragmatic skills in elicited production, *Seminars in Speech and Language*, 25, 57-71.
- Di Ruocco, G. (A.A. 2004-05), *La valutazione e la riabilitazione delle competenze semantico-pragmatiche nel bambino ipoacusico con Impianto Cocleare*, Tesi di Laurea in Logopedia, Seconda Università degli Studi di Napoli, Relatore Dott. N. Angelillo.
- Fasano, V. (A.A. 2006-07), *Deficit a carico delle competenze semantico-pragmatiche: proposta di inquadramento etiologico e nostro protocollo di trattamento logopedico*, Tesi di Laurea in Logopedia, Seconda Università degli Studi di Napoli, Relatore Dott. N. Angelillo.
- Happé, F. (1993), Communicative competence and theory of mind in autism: A test of relevance theory, *Cognition*, 48, 101-119.
- Juncos-Rabadan, O. (1992), The processing of negative sentences in fluent aphasics: semantic and pragmatic aspects, *Brain and Language*, 43, 96-106.
- Kaplan, J.A., Brownell, H.H., Jacobs, J.R. & Gardner, H. (1990), The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks, *Brain and Language*, 38, 315-333.
- Kasher, A., Batori, G., Soroker, N., Graves, D. & Zaidel, E. (1999), Effects of right- and left-hemisphere damage on understanding conversational implicatures, *Brain and Language*, 68, 566-590.
- McDonald, S. & Pearce, S. (1995), The 'dice' game: a new test of pragmatic language skills after closed-head injury, *Brain Injury*, 9, 255-271.
- McDonald, S. (1999), Exploring the process of inference generation in sarcasm: a review of normal and clinical studies, *Brain and Language*, 68, 486-506.
- Monfort, M. (2004), Intervention in children with pragmatic language and communication disorders, *Revue Neurologique*, 38, S85-87.
- Petranja, G. (A.A. 2003-04), *Correlazione tra Disturbo Semantico-Pragmatico ed Emisfero Destro*, Tesi di Laurea in Logopedia, Seconda Università degli Studi di Napoli, Relatore Dott. N. Angelillo.
- Prutting, C.A. & Kirchner, D.M. (1987), A clinical appraisal of the pragmatic aspects of language, *Journal of Speech and Hearing Disorders*, 52, 105-119.
- Shields, J. (1991), Semantic-pragmatic disorder: a right hemisphere syndrome?, *The British Journal of Disorders of Communication*, 26, 383-392.

Siegal, M., Carrington, J. & Radel, M. (1996), Theory of mind and pragmatic understanding following right hemisphere damage, *Brain and Language*, 53, 40-50.

Vitiello, A. (A.A. 2004-05), *Nuova batteria di test per la valutazione delle competenze Semantico-Pragmatiche*, Tesi di Laurea in Logopedia, Seconda Università degli Studi di Napoli, Relatore Dott. N. Angelillo.

Vitiello, A., Angelillo, N., Di Rocco, G., Tozzi, E., Angelillo, M. & Barillari U. (2007), La riabilitazione delle competenze semantico-pragmatiche, *XLI Congresso Nazionale S.I.F.E.L. (Società Italiana Foniatria e Logopedia)*, Cervia Milano Marittima, 21-24 novembre.

Zaidel, E., Kasher, A., Soroker, N., Batori, G., Giora, R. & Graves, D. (2000), Hemispheric contributions to pragmatics, *Brain and Cognition*, 43, 438-443.

Zanini, S., Bryan, K., De Luca, G. & Bava, A. (2005), Italian Right Hemisphere Language Battery: the normative study, *Neurological Sciences*, 26, 13-25.

## IL SISTEMA MDVP: SOGLIE DI NORMALITÀ A CONFRONTO

Antonio Mendicino, Vincenzo Galatà, Maria Tucci, Rosita Lio, Luciano Romito  
Laboratorio di Fonetica – Dip.to di Linguistica – Università della Calabria  
*antonio.mendicino@unical.it, vgalata@libero.it, tucci.maria@libero.it, lio.rosita@libero.it,*  
*luciano.romito@unical.it*

### 1. SOMMARIO

Il *Multi-Dimensional Voice Program* (MDVP) è un algoritmo che da un'unica emissione fonica permette l'estrazione di 33 diversi parametri, 22 dei quali sono relativi alle perturbazioni a medio e lungo termine della voce, e la loro rappresentazione su di un grafico radiale che Ferrero *et al.* (1994) hanno proposto di chiamare *vocaligramma*.

L'MDVP è stato valutato da Ferrero-Lanni (1994), Ferrero-Vagges (1994), Ferrero *et al.* (1995) e De Colle (2001), sia sul piano delle perturbazioni *estrinseche* o *metodologiche* (dipendenti essenzialmente dalle modalità di acquisizione del segnale, ma che riguardano anche il tipo di vocale e il numero di ripetizioni), sia riguardo alla validità delle soglie normative ai fini della discriminazione normale / patologico.

Il confronto tra i modelli di *variabilità estrinseca individuale* definiti in Mendicino (2004, 2005), le soglie di normalità di *default* del sistema MDVP e quelle di De Colle (*idem*) ha mostrato una maggiore affidabilità delle soglie normative proposte dallo stesso De Colle.

La difficoltà di standardizzazione di tutti i parametri d'analisi ha evidenziato la necessità che ciascun Laboratorio disponga di una propria normativa adattando i limiti di soglia proposti dal sistema.

In Mendicino (2004) si è fatto riferimento ad uno studio tuttora in corso, relativo alla creazione di una banca dati di voci *normofoniche* e alla possibile definizione di una normativa adeguatamente "robusta" – e, quindi, standardizzata – fruibile non soltanto dai singoli laboratori ma su scala più ampia.

La redazione di una normativa che possa essere utile a questi scopi è pertanto l'obiettivo principale del presente lavoro.

### 2. INTRODUZIONE

Il *Multi-Dimensional Voice Program* (MDVP) è un sistema che estrae da un'unica emissione fonica 33 diversi parametri, 22 dei quali si riferiscono alle perturbazioni a medio e lungo termine della voce e vengono rielaborati dal sistema sotto forma di un grafico radiale che Ferrero *et al.* (1994) hanno proposto di chiamare *vocaligramma*.

I valori di soglia per ciascun parametro sono stati originariamente calcolati su emissioni sostenute della «vocale 'a'» da parte di 15 soggetti normali (7 maschi, 8 femmine) e di 53 pazienti (25 maschi, 28 femmine) affetti da diverse patologie laringee, dal cancro alla laringite cronica (Deliyski 1993).

La valutazione del *Multi Dimensional Voice Program* (MDVP) è stata successivamente effettuata da Ferrero-Lanni (1994), Ferrero-Vagges (1994), Ferrero *et al.* (1995) e De Colle (2001) che hanno concentrato la loro attenzione in massima parte sulla valutazione delle soglie di normalità di *default* del sistema – utile ai fini della discriminazione normale / patologico –, nonché sulla rilevanza delle perturbazioni *estrinseche* o *metodologiche* che

dipendono dalle modalità di acquisizione del segnale e dall'ambiente in cui essa viene effettuata. Ferrero-Vagges (1994: 292) includono, tra le perturbazioni *estrinseche*, anche le «modalità di produzione verbale (parlato fluente o emissione di una vocale sostenuta e in questo caso, tipo di vocale e numero di ripetizioni)».

La particolare costituzione del campione di voci da cui sono state ricavate le soglie di normalità di *default* del sistema MDVP ha indotto questi autori ad aumentare opportunamente il numero dei campioni di voce esaminati. I limiti di normalità del sistema sono stati, quindi, riformulati da De Colle (2001: 131) sulla base dei risultati relativi all'emissione di una singola vocale [a], sostenuta per circa 3 secondi, da parte di 100 soggetti normofonici e di altrettanti soggetti disfonici.

Tenendo conto della perturbazione *estrinseca* di ciascun parametro, in Mendicino (2004, 2005) si è proceduto alla valutazione della "robustezza" dei limiti di normalità proposti da De Colle in alternativa a quelli di *default* del sistema MDVP. Sulla base di 100 emissioni della vocale [a], prodotte da due soggetti, uno di sesso maschile e uno di sesso femminile, si sono definiti due modelli di *variabilità estrinseca individuale* che sono stati messi a confronto con i limiti di normalità di Deliyski (1993) e di De Colle (2001).

Il confronto tra i modelli di *variabilità estrinseca individuale* definiti in Mendicino (2004, 2005), le soglie di normalità di *default* del sistema MDVP e quelle di De Colle ha mostrato una complessiva maggiore "robustezza" dei limiti normativi proposti da De Colle, che continuano così a costituire un «punto di riferimento senz'altro più affidabile rispetto a quelle di *default*» (Mendicino, 2005: 475).

Un altro aspetto connesso all'utilizzo di questa metodica riguarda la difficoltà di standardizzazione di tutti i parametri d'analisi che ha indotto De Colle a ribadire la necessità che ciascun Laboratorio disponga di una propria normativa adattando i limiti di soglia proposti dal sistema (De Colle, 2001:130).

Il presente studio vuole essere la continuazione di quanto intrapreso in Mendicino (2004) in riferimento alla creazione di una banca dati di voci *normofoniche* e alla possibile definizione di una normativa adeguatamente "robusta" – e, quindi, standardizzata – fruibile su scala più ampia dei singoli laboratori.

### 3. MATERIALI E METODI

Dei 100 soggetti contattati per la presente indagine, solo 43 sono riusciti a portare a termine la sessione completa di 100 produzioni sostenute della vocale [a]. Nella maggior parte dei casi restanti (36 su 100) i soggetti sono stati scartati dopo le prime dieci produzioni per una evidente difficoltà degli stessi ad avere un certo controllo sulla costanza delle produzioni; nei rimanenti casi (21 su 100) sono stati gli stessi soggetti a voler rinunciare dopo le prime produzioni. La difficoltà riscontrata nel reperimento dei soggetti, nonché ovvie ragioni di tempo, ci hanno indotto a non insistere oltre nella ricerca di ulteriori individui. Ciononostante riteniamo che il totale dei soggetti effettivamente registrati sia adeguatamente rappresentativo.

Ciascuna emissione vocalica è stata digitalizzata direttamente su computer tramite una scheda audio *Sound Blaster mod. Live Platinum 5.1* ed un microfono direzionale dinamico mod. *Sennheiser e835* ad essa collegato, tenuto a non più di 5 cm di distanza dalle labbra del locutore e inclinato di 45° rispetto ad esso (per la procedura seguita si veda De Colle 2001).

Tutte le acquisizioni sono state effettuate in camera silente, onde evitare la presenza di perturbazioni estrinseche da rumore ambientale (cfr. Di Nicola *et al.* 2000, Ricci Maccarini,

et al, 2002), mantenendo costanti la frequenza di campionamento e la quantizzazione, a 44,1 KHz e 16 bit rispettivamente.

Per i 43 soggetti, normofonici, sono state acquisite 100 emissioni della vocale [a], tutte di durata 4÷6 s e quanto più possibile costanti riguardo ad altezza e intensità, osservando un'adequata pausa tra una ripetizione e ogni qualvolta richiesto da ciascun soggetto.

Tutte le emissioni vocaliche sono state successivamente riportate ad una durata di 3 s – eliminando opportunamente la parte iniziale e finale delle stesse – e sottoposte, quindi, ad analisi mediante l'MDVP. A questo proposito ci preme evidenziare un aspetto a nostro avviso rilevante: durante tutte le fasi dell'analisi abbiamo potuto notare come, in moltissimi casi, millesimali scostamenti nella selezione temporale potessero causare il passaggio da valori del tutto normali/anormali a valori ampiamente esterni/interni alla norma per lo stesso parametro, senza che ciò fosse giustificato da evidenti indicazioni sulla forma d'onda visualizzata. Riportiamo di seguito a titolo esemplificativo alcune schermate relative ad uno dei casi esaminati.

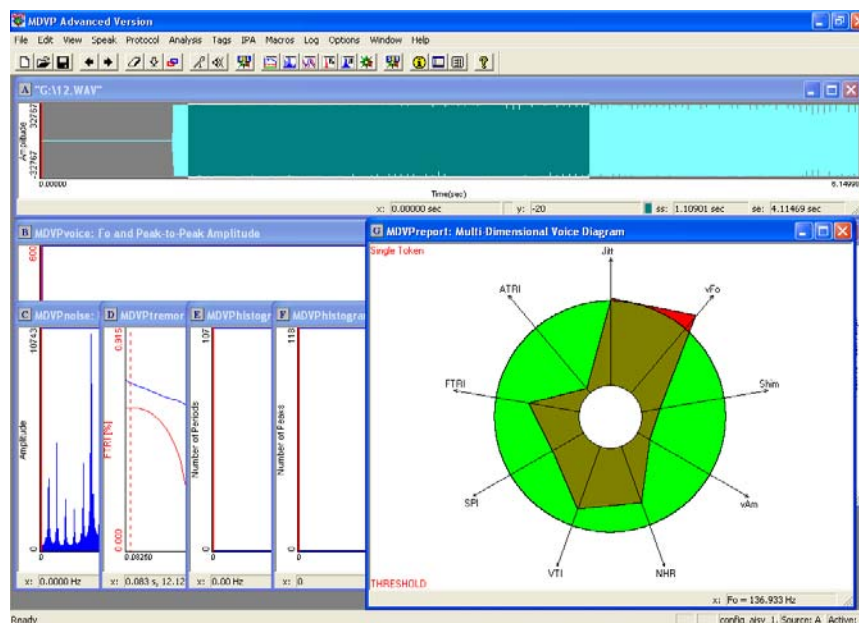
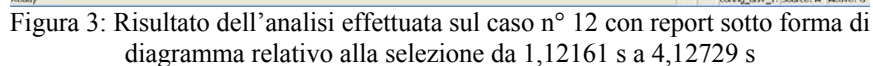
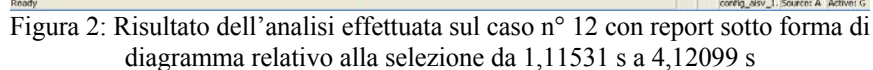


Figura 1: Risultato dell'analisi effettuata sul caso n° 12 con report sotto forma di diagramma relativo alla selezione da 1,10901 s a 4,11469 s



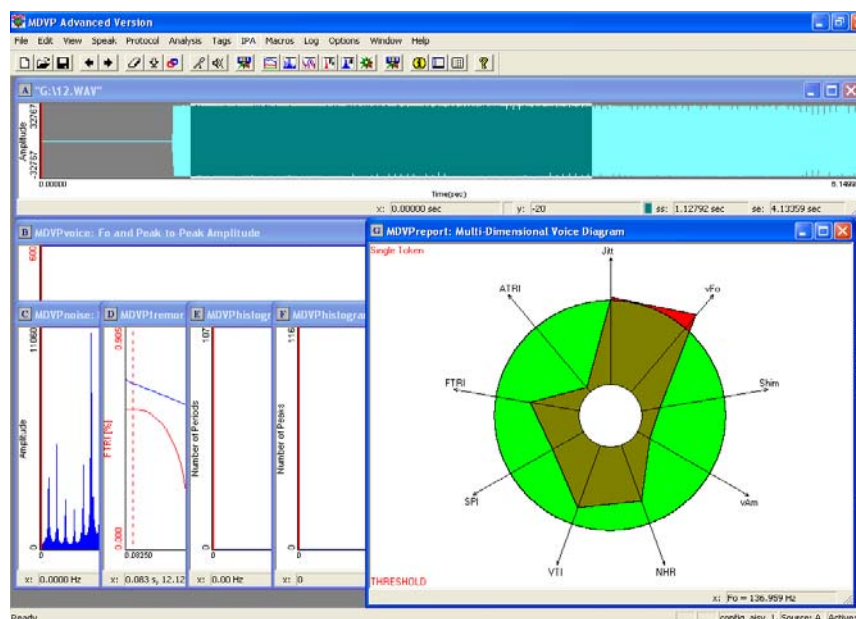


Figura 4: Risultato dell'analisi effettuata sul caso n° 12 con report sotto forma di diagramma relativo alla selezione da 1,12792 s a 4,13359 s

Come si può notare, nelle Figure 2 e 3 i valori di Jitt e  $vF_0$  sono molto al di sopra dei limiti normativi di default del sistema. Nelle Figure 1 e 4 si nota invece un notevole abbassamento dei medesimi valori. Ci si può rendere conto della problematicità di questo aspetto quando si prendono in considerazione i margini temporali della selezione sottoposta ad analisi che riportiamo nella Tabella seguente:

| Analisi | Riferimento | Inizio selezione | Fine selezione |
|---------|-------------|------------------|----------------|
| 1       | Figura 1    | 1,10901 s        | 4,11469 s      |
| 2       | Figura 2    | 1,11531 s        | 4,12099 s      |
| 3       | Figura 3    | 1,12161 s        | 4,12729 s      |
| 4       | Figura 4    | 1,12792 s        | 4,13359 s      |

Tabella 1: Riepilogo inizio e fine selezione per le analisi riportate nelle Figure 1-4

Come si può notare, infatti, dai valori di inizio e fine selezione riportati in Tabella 1, nelle 4 analisi consecutive si ha uno spostamento in avanti della selezione di soli 0,006 secondi. Si tratta di un carattere aleatorio del sistema MDVP da tenere in debito conto e che corrobora ancor più la scelta di effettuare le nostre analisi su più di una ripetizione.

Per il giudizio di normofonicità ci siamo basati, come nei precedenti lavori, sulla valutazione delle componenti di rumore presenti nello spettrogramma di una vocale [a]



comprendente un intervallo di 4KHz, per una frequenza di campionamento di 11.025Hz (vedi Figura 5 a titolo esemplificativo)<sup>1</sup>.

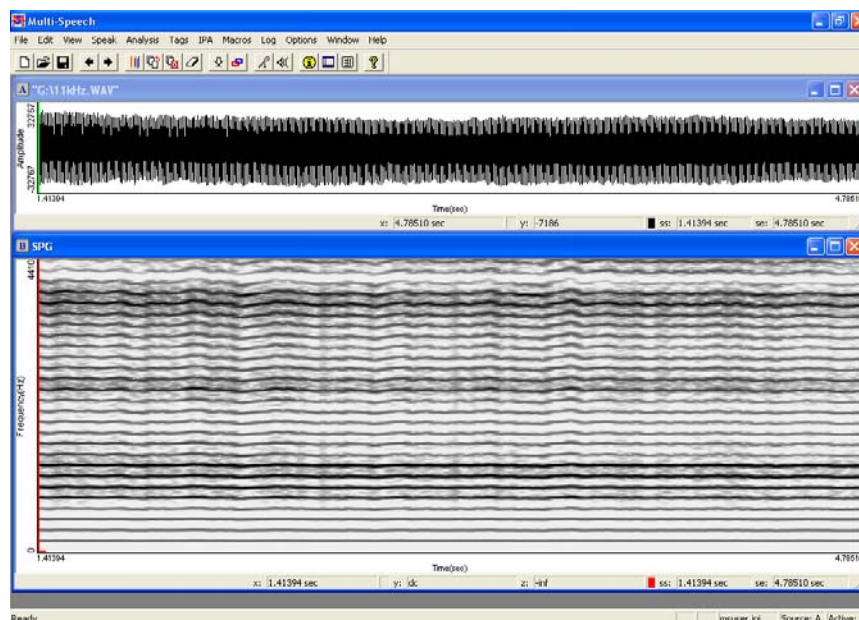


Figura 5: Oscillogramma e spettrogramma di una vocale [a] della durata di 3 s, comprendente un intervallo di 4KHz, ottenuto con una frequenza di campionamento pari a 11.025 Hz e una finestra di analisi di 1024 punti (15,77 Hz). Si nota l'assenza di diplofonia e un bilancio energetico spettrale nettamente a favore delle componenti armoniche in tutto l'intervallo di frequenza.

Sui dati così estrapolati è stata condotta un'analisi statistica per definire delle possibili soglie di normalità da mettere successivamente a confronto con quelle proposte da De Colle (vedi § 5.2).

<sup>1</sup> Secondo Ricci Maccarini *et al.* (2002: 168-69) i campioni di voce che non presentano caratteristiche di Tipo I, II, III e IV, così come loro stessi li hanno definiti, o diplofonia, sono classificati come normali.

#### 4. RISULTATI

In Tabella 2 sono riportate le medie ottenute per ciascun soggetto e per ciascun parametro<sup>2</sup>, nonché la loro media e deviazione standard (DS)<sup>3</sup>.

| Soggetto | Jitt  | vF <sub>0</sub> | Shim  | vAm    | NHR   | VTI   | SPI    | FTRI         | ATRI          |
|----------|-------|-----------------|-------|--------|-------|-------|--------|--------------|---------------|
| 1        | 0,494 | 0,725           | 2,886 | 6,148  | 0,138 | 0,052 | 7,006  | 0,211 (n=70) | 2,031 (n=52)  |
| 2        | 0,444 | 0,731           | 3,805 | 7,485  | 0,133 | 0,041 | 15,690 | 0,237 (n=94) | 2,639 (n=87)  |
| 3        | 0,566 | 0,897           | 2,423 | 4,896  | 0,145 | 0,061 | 3,844  | 0,244 (n=71) | 1,583 (n=75)  |
| 4        | 0,899 | 1,079           | 3,365 | 7,354  | 0,139 | 0,054 | 12,514 | 0,289 (n=96) | 2,500 (n=65)  |
| 5        | 0,387 | 0,619           | 1,364 | 3,742  | 0,127 | 0,043 | 8,934  | 0,154 (n=70) | 0,941 (n=57)  |
| 6        | 0,607 | 0,903           | 0,597 | 1,621  | 0,131 | 0,048 | 5,525  | 0,252 (n=78) | 0,424 (n=82)  |
| 7        | 0,337 | 1,214           | 1,076 | 2,663  | 0,136 | 0,047 | 6,412  | 0,350 (n=64) | 0,807 (n=66)  |
| 8        | 0,464 | 0,947           | 3,561 | 9,008  | 0,139 | 0,056 | 7,698  | 0,263 (n=77) | 2,628 (n=65)  |
| 9        | 0,697 | 1,012           | 1,740 | 4,883  | 0,131 | 0,051 | 8,366  | 0,255 (n=82) | 1,288 (n=55)  |
| 10       | 0,457 | 1,004           | 1,175 | 4,810  | 0,133 | 0,044 | 8,603  | 0,290 (n=73) | 1,697 (n=62)  |
| 11       | 0,383 | 0,669           | 0,746 | 2,255  | 0,125 | 0,055 | 3,548  | 0,162 (n=88) | 0,553 (n=72)  |
| 12       | 1,294 | 1,497           | 1,637 | 4,065  | 0,118 | 0,036 | 18,651 | 0,336 (n=89) | 1,130 (n=77)  |
| 13       | 0,331 | 0,691           | 0,767 | 1,457  | 0,124 | 0,051 | 3,765  | 0,180 (n=86) | 0,426 (n=75)  |
| 14       | 0,741 | 0,972           | 2,943 | 6,052  | 0,132 | 0,052 | 13,774 | 0,193 (n=62) | 1,906 (n=64)  |
| 15       | 0,401 | 0,871           | 1,055 | 4,773  | 0,121 | 0,035 | 13,120 | 0,172 (n=57) | 1,288 (n=31)  |
| 16       | 0,498 | 0,687           | 2,550 | 5,255  | 0,123 | 0,036 | 10,330 | 0,165 (n=71) | 1,427 (n=54)  |
| 17       | 0,341 | 0,672           | 0,911 | 0,924  | 0,121 | 0,043 | 4,303  | 0,134 (n=53) | 0,226 (n=98)  |
| 18       | 1,301 | 1,318           | 2,634 | 4,620  | 0,138 | 0,049 | 7,614  | 0,267 (n=91) | 1,572 (n=76)  |
| 19       | 0,532 | 0,946           | 1,829 | 9,620  | 0,108 | 0,035 | 14,764 | 0,238 (n=68) | 2,633 (n=32)  |
| 20       | 0,466 | 1,003           | 2,777 | 7,322  | 0,133 | 0,049 | 6,617  | 0,280 (n=65) | 2,326 (n=46)  |
| 21       | 0,617 | 0,903           | 0,966 | 3,975  | 0,128 | 0,042 | 7,668  | 0,244 (n=88) | 1,145 (n=58)  |
| 22       | 0,490 | 1,039           | 1,191 | 3,895  | 0,129 | 0,050 | 3,957  | 0,302 (n=76) | 1,045 (n=73)  |
| 23       | 0,430 | 0,759           | 0,887 | 1,377  | 0,125 | 0,042 | 6,827  | 0,204 (n=88) | 0,466 (n=95)  |
| 24       | 0,884 | 1,055           | 2,875 | 5,725  | 0,123 | 0,041 | 12,136 | 0,271 (n=83) | 1,986 (n=77)  |
| 25       | 0,333 | 0,848           | 1,822 | 9,046  | 0,136 | 0,056 | 4,980  | 0,298 (n=92) | 2,447 (n=42)  |
| 26       | 0,667 | 0,859           | 2,401 | 3,551  | 0,118 | 0,054 | 7,726  | 0,203 (n=77) | 1,037 (n=81)  |
| 27       | 0,469 | 0,789           | 2,959 | 5,145  | 0,128 | 0,054 | 6,668  | 0,214 (n=91) | 1,783 (n=90)  |
| 28       | 1,045 | 1,490           | 2,107 | 5,906  | 0,130 | 0,053 | 10,631 | 0,493 (n=94) | 2,455 (n=95)  |
| 29       | 1,115 | 1,373           | 2,069 | 4,219  | 0,141 | 0,060 | 5,368  | 0,295 (n=80) | 1,546 (n=62)  |
| 30       | 0,660 | 0,833           | 2,862 | 7,536  | 0,126 | 0,049 | 7,148  | 0,175 (n=82) | 1,932 (n=53)  |
| 31       | 0,568 | 0,819           | 1,352 | 2,151  | 0,139 | 0,061 | 3,295  | 0,188 (n=88) | 0,580 (n=86)  |
| 32       | 0,700 | 1,107           | 1,043 | 2,365  | 0,116 | 0,043 | 7,091  | 0,308 (n=71) | 0,438 (n=66)  |
| 33       | 0,335 | 0,643           | 1,713 | 5,403  | 0,133 | 0,042 | 10,985 | 0,181 (n=71) | 1,743 (n=40)  |
| 34       | 1,358 | 1,526           | 2,307 | 5,234  | 0,130 | 0,049 | 13,174 | 0,330 (n=82) | 1,419 (n=66)  |
| 35       | 0,995 | 1,047           | 1,226 | 7,031  | 0,121 | 0,032 | 23,914 | 0,278 (n=82) | 2,253 (n=27)  |
| 36       | 0,453 | 0,671           | 2,695 | 6,644  | 0,130 | 0,047 | 8,190  | 0,157 (n=77) | 1,854 (n=43)  |
| 37       | 0,534 | 0,725           | 2,798 | 5,472  | 0,139 | 0,055 | 7,426  | 0,184 (n=90) | 2,161 (n=83)  |
| 38       | 1,052 | 1,148           | 4,214 | 7,747  | 0,143 | 0,056 | 4,898  | 0,273 (n=76) | 2,165 (n=64)  |
| 39       | 0,445 | 1,087           | 2,241 | 6,767  | 0,123 | 0,053 | 4,889  | 0,334 (n=87) | 2,235 (n=74)  |
| 40       | 0,487 | 0,909           | 1,676 | 5,379  | 0,139 | 0,044 | 10,216 | 0,222 (n=22) | 1,268 (n=38)  |
| 41       | 0,675 | 1,032           | 2,156 | 5,910  | 0,131 | 0,045 | 10,782 | 0,241 (n=67) | 1,561 (n=47)  |
| 42       | 0,629 | 0,858           | 3,614 | 7,383  | 0,137 | 0,056 | 8,259  | 0,246 (n=88) | 2,538 (n=71)  |
| 43       | 1,114 | 1,241           | 8,516 | 10,188 | 0,148 | 0,060 | 16,607 | 0,262 (n=91) | 3,586 (n=100) |
| Media    | 0,644 | 0,959           | 2,222 | 5,279  | 0,131 | 0,048 | 8,928  | 0,246        | 1,620         |
| DS       | 0,288 | 0,239           | 1,353 | 2,282  | 0,008 | 0,008 | 4,438  | 0,068        | 0,769         |

Tabella 2: Riepilogo medie ottenute per ciascun soggetto e per ciascun parametro

<sup>2</sup> Vengono segnalati tra parentesi solo quei casi in cui il numero di occorrenze è minore di 100.

<sup>3</sup> I dati relativi a ciascun soggetto sono disponibili facendone esplicita richiesta agli autori.

Nella Figura 6 sono riportate le distribuzioni di frequenza per ciascuna delle nove variabili considerate (Jitt,  $vF_0$ , Shim,  $vAm$ , NHR, VTI, SPI, FTRI, ATRI)<sup>4</sup> relative ai dati in Tabella 2, al fine di cogliere anche visivamente la dispersione delle singole variabili. Come si può notare, per alcune di esse si registra una buona approssimazione alla *normale* (Shim,  $vAm$ , NHR, SPI, FTRI).

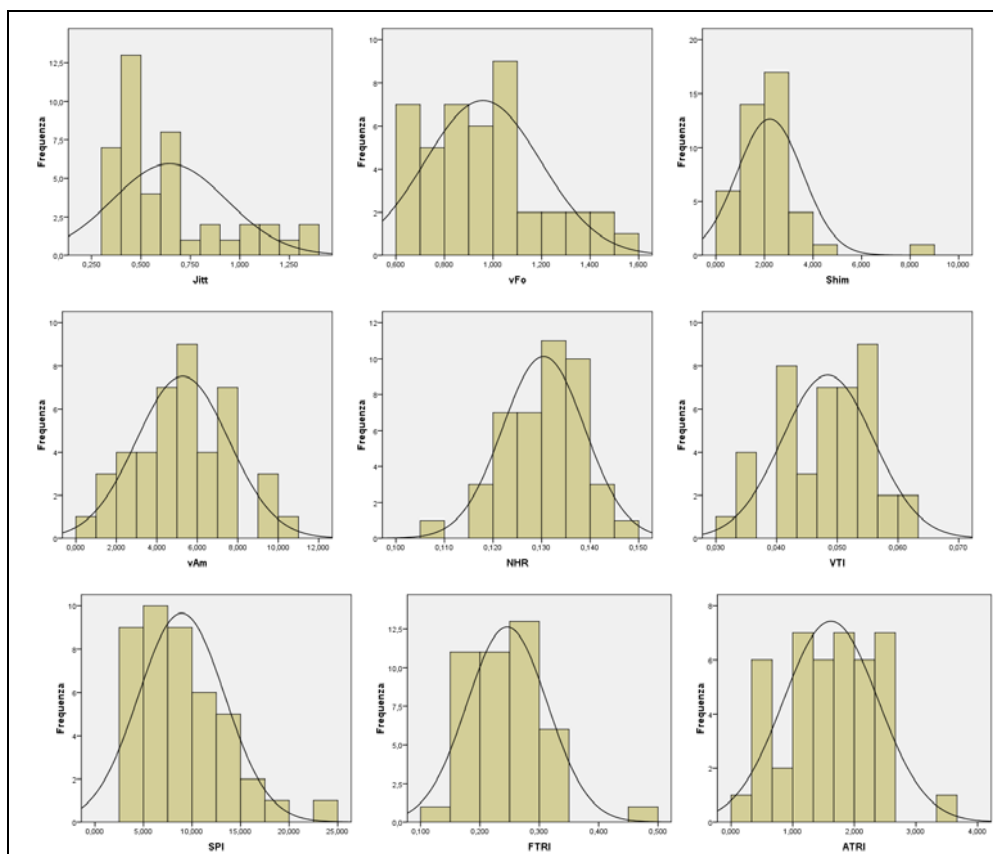


Figura 6: Distribuzioni di frequenza per ciascuna delle nove variabili considerate (Jitt,  $vF_0$ , Shim,  $vAm$ , NHR, VTI, SPI, FTRI, ATRI) relative ai dati in Tabella 2

<sup>4</sup> Per il significato delle sigle qui riportate si rimanda, oltre che al manuale dell'MDVP, a Ferrero *et al.* (1995) e De Colle (2001).

## 5. SOGLIE DI NORMALITÀ

### 5.1. Valori di riferimento

Per la definizione di una possibile soglia di normalità per ciascun parametro, nel prosieguo della discussione faremo riferimento ai valori medi di ciascun parametro, riportati in Tabella 2, e alla media delle deviazioni standard riportate in Tabella 3:

| Soggetto        | Jitt         | vFo          | Shim         | vAm          | NHR          | VTI          | SPI          | FTRI         | ATRI         |
|-----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1               | 0,331        | 0,286        | 0,872        | 2,154        | 0,011        | 0,012        | 3,864        | 0,115        | 1,107        |
| 2               | 0,214        | 0,213        | 0,743        | 1,664        | 0,017        | 0,008        | 3,922        | 0,129        | 1,327        |
| 3               | 0,230        | 0,169        | 1,459        | 2,919        | 0,010        | 0,010        | 0,917        | 0,130        | 1,307        |
| 4               | 0,278        | 0,218        | 0,788        | 2,073        | 0,014        | 0,011        | 2,631        | 0,143        | 1,218        |
| 5               | 0,143        | 0,153        | 0,657        | 2,022        | 0,010        | 0,011        | 2,660        | 0,073        | 0,761        |
| 6               | 0,387        | 0,310        | 0,292        | 1,638        | 0,009        | 0,012        | 1,737        | 0,147        | 0,596        |
| 7               | 0,283        | 0,397        | 0,516        | 2,382        | 0,010        | 0,009        | 1,704        | 0,226        | 0,973        |
| 8               | 0,204        | 0,214        | 0,983        | 2,092        | 0,015        | 0,017        | 3,730        | 0,130        | 1,304        |
| 9               | 0,256        | 0,222        | 1,001        | 2,819        | 0,012        | 0,011        | 2,115        | 0,161        | 1,116        |
| 10              | 0,203        | 0,193        | 0,639        | 2,517        | 0,010        | 0,011        | 2,572        | 0,174        | 1,292        |
| 11              | 0,213        | 0,210        | 0,279        | 1,888        | 0,007        | 0,010        | 0,561        | 0,081        | 0,600        |
| 12              | 0,526        | 0,329        | 1,530        | 3,974        | 0,020        | 0,010        | 8,283        | 0,147        | 1,496        |
| 13              | 0,173        | 0,209        | 0,195        | 1,129        | 0,009        | 0,011        | 0,794        | 0,103        | 0,428        |
| 14              | 0,274        | 0,209        | 0,850        | 2,231        | 0,011        | 0,010        | 2,828        | 0,117        | 1,031        |
| 15              | 0,201        | 0,207        | 0,481        | 2,218        | 0,007        | 0,012        | 3,388        | 0,100        | 1,027        |
| 16              | 0,198        | 0,201        | 0,577        | 1,515        | 0,025        | 0,010        | 2,706        | 0,086        | 0,851        |
| 17              | 0,093        | 0,172        | 0,169        | 0,149        | 0,010        | 0,007        | 0,696        | 0,104        | 0,106        |
| 18              | 0,634        | 0,449        | 1,537        | 2,572        | 0,012        | 0,011        | 2,690        | 0,110        | 1,141        |
| 19              | 0,265        | 0,236        | 0,641        | 2,860        | 0,027        | 0,012        | 5,253        | 0,146        | 1,488        |
| 20              | 0,195        | 0,287        | 0,750        | 2,013        | 0,009        | 0,010        | 1,964        | 0,154        | 1,182        |
| 21              | 0,207        | 0,160        | 0,538        | 2,206        | 0,011        | 0,011        | 3,054        | 0,139        | 1,027        |
| 22              | 0,206        | 0,207        | 0,734        | 2,727        | 0,010        | 0,010        | 0,738        | 0,173        | 1,099        |
| 23              | 0,163        | 0,180        | 0,416        | 1,134        | 0,009        | 0,008        | 1,676        | 0,123        | 0,450        |
| 24              | 0,316        | 0,217        | 0,620        | 1,699        | 0,019        | 0,010        | 3,285        | 0,109        | 0,856        |
| 25              | 0,160        | 0,205        | 0,366        | 2,376        | 0,010        | 0,011        | 1,666        | 0,142        | 1,153        |
| 26              | 0,343        | 0,239        | 1,493        | 2,513        | 0,025        | 0,011        | 1,894        | 0,110        | 1,099        |
| 27              | 0,233        | 0,179        | 0,564        | 1,251        | 0,006        | 0,008        | 1,052        | 0,104        | 0,816        |
| 28              | 0,447        | 0,448        | 1,342        | 3,705        | 0,015        | 0,014        | 3,488        | 0,218        | 1,670        |
| 29              | 0,730        | 0,613        | 0,999        | 1,897        | 0,010        | 0,014        | 1,442        | 0,185        | 0,919        |
| 30              | 0,301        | 0,256        | 0,587        | 1,889        | 0,017        | 0,011        | 2,182        | 0,101        | 1,054        |
| 31              | 0,435        | 0,362        | 1,157        | 1,953        | 0,017        | 0,016        | 1,037        | 0,104        | 0,514        |
| 32              | 0,474        | 0,382        | 0,831        | 2,468        | 0,014        | 0,010        | 2,911        | 0,143        | 0,494        |
| 33              | 0,100        | 0,133        | 0,632        | 1,358        | 0,012        | 0,011        | 3,326        | 0,096        | 0,770        |
| 34              | 0,562        | 0,431        | 1,659        | 3,709        | 0,009        | 0,012        | 2,661        | 0,220        | 1,298        |
| 35              | 0,355        | 0,301        | 0,629        | 2,545        | 0,028        | 0,011        | 6,897        | 0,155        | 1,287        |
| 36              | 0,182        | 0,158        | 0,533        | 1,409        | 0,013        | 0,009        | 2,493        | 0,078        | 0,942        |
| 37              | 0,211        | 0,169        | 0,944        | 2,177        | 0,009        | 0,011        | 2,223        | 0,089        | 1,108        |
| 38              | 0,667        | 0,445        | 0,631        | 1,720        | 0,013        | 0,011        | 1,630        | 0,130        | 1,017        |
| 39              | 0,254        | 0,255        | 0,513        | 1,659        | 0,008        | 0,010        | 0,920        | 0,155        | 1,138        |
| 40              | 0,266        | 0,212        | 0,652        | 2,537        | 0,016        | 0,010        | 2,581        | 0,158        | 0,923        |
| 41              | 0,216        | 0,207        | 0,413        | 1,517        | 0,012        | 0,009        | 2,891        | 0,156        | 0,937        |
| 42              | 0,284        | 0,219        | 0,574        | 1,668        | 0,011        | 0,011        | 2,813        | 0,114        | 1,074        |
| 43              | 0,388        | 0,518        | 2,014        | 2,412        | 0,030        | 0,012        | 2,712        | 0,128        | 1,658        |
| <b>Media DS</b> | <b>0,298</b> | <b>0,265</b> | <b>0,786</b> | <b>2,125</b> | <b>0,013</b> | <b>0,011</b> | <b>2,572</b> | <b>0,133</b> | <b>1,015</b> |

Tabella 3: Riepilogo deviazioni standard per ciascun soggetto e per ciascun parametro e media delle deviazioni standard

La scelta di utilizzare la media delle deviazioni standard ci sembra più opportuna ai nostri fini, perché indicativa di quella che possiamo considerare essere la dispersione media di ciascuna variabile tra i soggetti. Nella Tabella seguente forniamo pertanto le medie di

ciascun parametro e i valori medi delle loro deviazioni standard su cui si baserà la nostra discussione sulle soglie di normalità.

|                 | Jitt  | vFo   | Shim  | vAm   | NHR   | VTI   | SPI   | FTRI  | ATRI  |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| <b>Media</b>    | 0,644 | 0,959 | 2,222 | 5,279 | 0,131 | 0,048 | 8,928 | 0,246 | 1,620 |
| <b>Media DS</b> | 0,298 | 0,265 | 0,786 | 2,125 | 0,013 | 0,011 | 2,572 | 0,133 | 1,015 |

Tabella 4: Riepilogo valori medi di ciascun parametro con i rispettivi valori medi delle deviazioni standard

## 5.2. Soglie di normalità a confronto

In Tabella 5 sono riportate, per ciascuna variabile, la media e la deviazione standard di riferimento per il presente lavoro; questi valori sono messi a confronto con quelli di De Colle (2001: 118 e 131) che già in altra sede (Mendicino, 2004, 2005) abbiamo considerato essere più robusti dei limiti di *default* che, per questo, non vengono qui riportati.

| Parametro | Media | DS    | Media De Colle | DS De Colle |
|-----------|-------|-------|----------------|-------------|
| Jitt - %  | 0,644 | 0,298 | 0,464          | 0,227       |
| vF0 - %   | 0,959 | 0,265 | 0,815          | 0,201       |
| Shim - %  | 2,222 | 0,786 | 2,241          | 0,698       |
| vAm - %   | 5,279 | 2,125 | 7,441          | 1,970       |
| NHR       | 0,131 | 0,013 | 0,1255         | 0,016       |
| VTI       | 0,048 | 0,011 | 0,0414         | 0,011       |
| SPI       | 8,928 | 2,572 | 11,3019        | 4,181       |
| FTRI - %  | 0,246 | 0,133 | 0,343          | 0,142       |
| ATRI - %  | 1,620 | 1,015 | 2,601          | 1,185       |

Tabella 5: Valori di media e deviazione standard di riferimento per il presente lavoro a confronto con quelli proposti da De Colle (2001)

Nella Tabella 6, invece, sono messi a confronto, per ciascun parametro, i limiti superiori costituiti dalla media + 1DS e dalla media + 2DS, al di sotto dei quali, per una distribuzione *normale*, cadono, rispettivamente, l'84,13% e il 97,72% dei dati osservati.

| Parametro | Media + 1DS | Media + 1DS De Colle | Media + 2DS | Media + 2DS De Colle |
|-----------|-------------|----------------------|-------------|----------------------|
| Jitt - %  | 0,942       | 0,69                 | 1,240       | 0,92                 |
| vF0 - %   | 1,223       | 1,02                 | 1,488       | 1,22                 |
| Shim - %  | 3,008       | 2,94                 | 3,794       | 3,64                 |
| vAm - %   | 7,404       | 9,41                 | 9,528       | 11,38                |
| NHR       | 0,144       | 0,142                | 0,158       | 0,158                |
| VTI       | 0,059       | 0,052                | 0,070       | 0,063                |
| SPI       | 11,500      | 15,483               | 14,072      | 19,663               |
| FTRI - %  | 0,379       | 0,48                 | 0,511       | 0,63                 |
| ATRI - %  | 2,636       | 3,79                 | 3,651       | 4,97                 |

Tabella 6: Limiti superiori di media + 1DS e di media + 2DS calcolati nel presente lavoro messi a confronto con quelli di De Colle (2001)

Come è noto, ogniquale si vogliano stabilire, per determinati parametri, limiti o soglie di normalità oltre i quali considerare una persona sana o malata, è estremamente importante stimare, a partire dai dati osservati, le probabilità di falsa positività/riconoscimento e di falsa negatività/esclusione. Una volta stabiliti i limiti di soglia, occorre, cioè, calcolare quanti soggetti, tra quelli sani, possono erroneamente essere classificati come malati (*falsi positivi*) e quanti, tra quelli malati, possono erroneamente essere considerati come sani (*falsi negativi*), tenuto conto che le distribuzioni di frequenza

per i due diversi campioni di soggetti, normofonici e patologici, possono sovrapporsi in qualche punto<sup>5</sup>. In modo del tutto analogo, possono poi essere calcolate le percentuali di identificazioni corrette, sia sul lato della popolazione malata (*sensibilità*) che su quello della popolazione sana (*specificità*). Prima di passare al calcolo di queste percentuali di errore, ci preme, però, sottolineare che, con riferimento ai *falsi negativi* e alla *sensibilità*, essendo il nostro campione relativo ad una popolazione di soli soggetti sani, baseremo la nostra analisi su dati presenti in letteratura, in particolare su quelli di De Colle (2001: 122), che riportiamo nella Tabella 7 seguente:

| Parametro | Media<br>De Colle | DS<br>De Colle |
|-----------|-------------------|----------------|
| Jitt - %  | 1,976             | 1,744          |
| vF0 - %   | 5,025             | 9,333          |
| Shim - %  | 6,018             | 5,112          |
| vAm - %   | 17,959            | 10,705         |
| NHR       | 0,1989            | 0,146          |
| VTI       | 0,0584            | 0,073          |
| SPI       | 19,8044           | 12,811         |
| FTRI - %  | 1,200             | 2,596          |
| ATRI - %  | 6,857             | 5,687          |

Tabella 7: Riepilogo medie e deviazioni standard di soggetti patologici utilizzati per la presente analisi

Proprio per questo motivo, mentre per i *falsi positivi* e la *specificità* i dati presentati in tabella corrispondono a percentuali calcolate su dati direttamente osservati, per la *sensibilità* e i *falsi negativi* vengono riportate soltanto delle mere percentuali di probabilità, non essendo a noi disponibili i singoli valori di ogni soggetto che nel lavoro di De Colle (op. cit.) non vengono forniti. Queste percentuali di probabilità per i soggetti patologici sono stati desunte dopo aver calcolato gli *intervalli di confidenza* secondo la formula:

$$\text{Intervallo di confidenza} = \frac{\text{media De Colle} - (\text{media} + 1\text{DS sperimentali})}{\text{DS De Colle}}$$

in modo da poter esprimere la distanza dei valori di De Colle dalle nostre soglie di normalità nei termini di multipli della deviazione standard. Nel caso del parametro Jitt., ad es., si ha:

$$(1,976-0,942) / 1,744 = 0,59289 \text{ DS}$$

A partire da questo risultato, facendo riferimento ad una tabella della distribuzione normale<sup>6</sup>, è possibile determinare la probabilità che i dati osservati cadano al di sotto o al di sopra del limite calcolato (28% e 72% rispettivamente).

Le percentuali relative alla *specificità* e ai *falsi positivi* che si basano, quindi, sui dati effettivamente osservati sono riportati in Tabella 8, mentre le percentuali di probabilità relative alla *sensibilità* e ai *falsi negativi* sono elencate nella Tabella 9. Nelle stesse tabelle sono riportati anche, per confronto, i dati di De Colle:

<sup>5</sup> Si veda a tal proposito l'approfondita discussione in De Colle (2001: 122-132).

<sup>6</sup> Cfr. Butler C., *Statistics in Linguistics*, Basil Blackwell, Oxford-New York, 1985, p. 171.

| Parametro           | Specificità/falsi positivi |                         |             |                         |
|---------------------|----------------------------|-------------------------|-------------|-------------------------|
|                     | Media + 1DS                | Media + 1DS<br>De Colle | Media + 2DS | Media + 2DS<br>De Colle |
| Jitt - %            | 81/19                      | 82/18                   | 93/7        | 93/7                    |
| vF <sub>0</sub> - % | 86/14                      | 82/18                   | 93/7        | 96/4                    |
| Shim - %            | 86/14                      | 87/13                   | 93/7        | 97/3                    |
| vAm - %             | 84/16                      | 82/18                   | 95/5        | 97/3                    |
| NHR                 | 95/5                       | 92/8                    | 100/0       | 100/0                   |
| VTI                 | 91/9                       | 82/18                   | 100/0       | 98/2                    |
| SPI                 | 77/23                      | 83/17                   | 88/12       | 100/0                   |
| FTRI - %            | 98/2                       | 86/14                   | 100/0       | 96/4                    |
| ATRI - %            | 95/5                       | 82/18                   | 100/0       | 96/4                    |

Tabella 8: Percentuali relative a *specificità/falsi positivi* sui dati effettivamente osservati

| Parametro           | Sensibilità/falsi negativi |                         |             |                         |
|---------------------|----------------------------|-------------------------|-------------|-------------------------|
|                     | Media + 1DS                | Media + 1DS<br>De Colle | Media + 2DS | Media + 2DS<br>De Colle |
| Jitt - %            | 72/28                      | 84/16                   | 66/34       | 75/25                   |
| vF <sub>0</sub> - % | 66/34                      | 88/12                   | 65/35       | 78/22                   |
| Shim - %            | 72/28                      | 81/19                   | 67/33       | 72/28                   |
| vAm - %             | 84/16                      | 84/16                   | 79/21       | 72/28                   |
| NHR                 | 64/36                      | 64/36                   | 61/39       | 36/64                   |
| VTI                 | 50/50                      | 30/70                   | 44/56       | 15/85                   |
| SPI                 | 74/26                      | 59/41                   | 67/33       | 32/68                   |
| FTRI - %            | 62/38                      | 57/43                   | 61/39       | 48/52                   |
| ATRI - %            | 77/23                      | 71/29                   | 71/29       | 54/46                   |

Tabella 9: Percentuali di probabilità relative a *sensibilità/falsi negativi*

Passando ora al confronto tra i nostri dati e quelli di De Colle osserviamo che, con riferimento ai dati riportati in Tabella 8, emerge quanto segue:

- si registra un sostanziale bilanciamento con i dati di De Colle riguardo ai parametri Jitt, Shim, vAm e NHR;
- per quanto attiene al parametro SPI, la percentuale di falsi positivi da noi calcolata è sensibilmente superiore al dato registrato da De Colle (23 *falsi positivi* vs. 17 per 1DS e 12 *falsi positivi* vs. 0 per 2DS);
- per il parametro vF<sub>0</sub>, invece, il rapporto cambia di segno a seconda che si faccia riferimento ad 1 o 2 deviazioni standard, con la percentuale di *falsi positivi* che, nel nostro caso, è leggermente inferiore a quella di De Colle per 1DS (14 vs. 18) mentre è di poco superiore ad essa per 2DS (7 vs. 4);
- con riferimento ai parametri VTI, FTRI e ATRI le percentuali di falsi positivi calcolate con i nostri limiti di soglia sono nettamente inferiori a quelli di De Colle, soprattutto per il limite costituito dalla media + 1DS.

In merito ai dati riguardanti la sensibilità e i falsi negativi, riportati in Tabella 9, osserviamo quanto segue:

- con riferimento ai parametri Jitt, vF<sub>0</sub> e Shim, tre dei parametri ritenuti ad alto potere diagnostico da De Colle (op. cit.: 132), notiamo un notevole peggioramento del potere discriminante delle nostre soglie, siano esse costituite dalla media + 1DS o + 2DS, con la probabilità di falsi negativi da noi calcolata nettamente superiore ai dati effettivamente osservati da De Colle;
- con riferimento ai parametri vAm, NHR, VTI, SPI, FTRI, ATRI si rileva, invece, una sostanziale riduzione delle probabilità calcolate tenendo conto dei nostri limiti di soglia rispetto ai dati realmente osservati nel lavoro a confronto.

Considerando ora i nostri dati, se mettiamo a confronto le percentuali di probabilità relative alla sensibilità/falsi negativi con le percentuali calcolate sui dati osservati

riguardanti la specificità/falsi positivi, emerge un dato non trascurabile. Come si può notare dalla Tabella 10, infatti, vi è un evidente scostamento tra le percentuali di probabilità della sensibilità/falsi negativi e le percentuali relative alla specificità/falsi positivi.

| Parametro           | Sensibilità/falsi negativi |             | Specificità/falsi positivi |             |
|---------------------|----------------------------|-------------|----------------------------|-------------|
|                     | Media + 1DS                | Media + 2DS | Media + 1DS                | Media + 2DS |
| Jitt - %            | 72/28                      | 66/34       | 81/19                      | 93/7        |
| vF <sub>0</sub> - % | 66/34                      | 65/35       | 86/14                      | 93/7        |
| Shim - %            | 72/28                      | 67/33       | 86/14                      | 93/7        |
| vAm - %             | 84/16                      | 79/21       | 84/16                      | 95/5        |
| NHR                 | 64/36                      | 61/39       | 95/5                       | 100/0       |
| VTI                 | 50/50                      | 44/56       | 91/9                       | 100/0       |
| SPI                 | 74/26                      | 67/33       | 77/23                      | 88/12       |
| FTRI - %            | 62/38                      | 61/39       | 98/2                       | 100/0       |
| ATRI - %            | 77/23                      | 71/29       | 95/5                       | 100/0       |

Tabella 10: Confronto tra percentuali di probabilità relative a sensibilità/falsi negativi e percentuali calcolate sui dati osservati riguardanti specificità/falsi positivi

Se si prendono in considerazione i dati relativi alla Media + 1DS<sup>7</sup> è possibile notare che:

- lo scostamento è abbastanza contenuto per due dei parametri (Jitt e Shim) considerati a più alto potere diagnostico da De Colle;
- la differenza tra le due percentuali risulta perfettamente bilanciata per il parametro vAm e lievemente sbilanciata verso i falsi negativi per il parametro SPI;
- per i parametri vF<sub>0</sub> e ATRI risulta invece una consistente differenza tra le due percentuali con il numero dei falsi negativi che supera ampiamente quello dei falsi positivi;
- per NHR, VTI ed FTRI il divario è, invece, piuttosto ampio.

## 6. CONCLUSIONI

A fronte della disamina dei dati sopra riportati restano sicuramente confermati i risultati e le considerazioni di De Colle per quel che riguarda alcuni parametri, quali ad esempio VTI, che rimane quello più problematico ai fini della discriminazione tra una popolazione sana ed una malata con una percentuale di probabilità di falsi negativi che si mantiene, anche nel nostro caso, assai elevata (50% per 1DS e 56% per 2DS vs. 70% per 1DS De Colle e 85% per 2DS De Colle).

Fatta eccezione, poi, per SPI (tra i falsi positivi) e Jitt, vF<sub>0</sub> e Shim (tra i falsi negativi), per tutti gli altri parametri si registra un netto miglioramento del potere discriminante dei nostri limiti di soglia rispetto a quelli in De Colle (2001).

Nonostante Jitt, vF<sub>0</sub> e Shim, assieme a vAm, sembrano costituire i parametri «più “robusti” nella classificazione diagnostica»<sup>8</sup>, riteniamo che l’esito confortante dei dati qui ottenuti, sia riguardo ai falsi positivi che ai falsi negativi, confermi e segni la correttezza della strada da noi intrapresa, che ribadisce la necessità di caratterizzare una voce a partire da più di una riproduzione.

Se i dati qui presentati non risolvono definitivamente il problema della standardizzazione dei parametri di questo sistema di analisi, riteniamo che non vi possa essere altra strada percorribile se non quella qui intrapresa.

<sup>7</sup> Si vedano le analoghe considerazioni in De Colle (op. cit.: 131-132).

<sup>8</sup> *Ibidem*.



## BIBLIOGRAFIA

- Butler C. (1985), *Statistics in Linguistics*, Basil Blackwell, Oxford-New York.
- De Colle W. (2001), *Voce & Computer, Analisi Acustica Digitale del Segnale Verbale (Il sistema CSL-MDVP)*, Ed. Omega, Torino.
- Deliyski D. D. (1993), Acoustic Model and Evaluation of Pathological Voice Production, *Proceedings of the 3-rd Conference on Speech Communication and Technology, EUROSPEECH '93*, Berlin, Germany, Vol. 3, pp. 1969-1972.
- Di Nicola V, Fiorella ML, Luperto P, Staffieri A, Fiorella R. (2000), La valutazione obiettiva della disfonia. Possibilità e limiti, *Acta ORL Italica*, Vol. 21, pp. 10-21.
- Ferrero F. E., Lanni R. (1994), Valutazione del sistema CSL-MDVP con campioni di voce normale e patologica, digitale e analogica, microfonica e elettroglottografica, *Quaderni del Centro di Studio per le Ricerche in Fonetica del CNR*, Padova, Vol. 13, pp. 257-289.
- Ferrero F. E., Lanni R., De Colle W. (1995), Primi risultati di uno studio per la validazione del sistema MDVP come strumento per una caratterizzazione multiparametrica della voce, *Acta Phoniatria Latina*, vol. 17:3, pp. 161-180.
- Ferrero F. E., Vaggies K., Di Ottavio R., Sfakianu T. (1994), Caratterizzazione multiparametrica della voce: il vocaligramma, *Atti del XXVIII Congresso Nazionale della Società Italiana di Foniatria e Logopedia*, Montesilvano (PE), 7-9 Aprile 1994, pp. 88-89.
- Ferrero F. E., Vaggies K. (1994) Perturbazioni a breve e lungo termine del segnale verbale: soglie di normalità da voce naturale e sintetica, *Atti del XXII Convegno Nazionale dell'Associazione Italiana di Acustica*, a cura di B. Perrone (Lecce 13-15 Aprile 1994), pp. 473-478.
- Kay Elemetrics Corp. (1999), *Multi-Dimensional Voice Program (MDVP), Model 5105. Software Instruction Manual*, Kay Elemetrics Corp., Lincoln Park-NJ (USA).
- Mendicino A. (2004), Il sistema MDVP: *variabilità estrinseca individuale* e soglie di normalità, *Acta Phoniatria Latina*, La Garangola, Padova, vol. 26, fasc. 3-4, pp. 354-370.
- Mendicino A. (2005), Il sistema MDVP: *variabilità estrinseche individuale* e soglie di normalità a confronto, *Acta Phoniatria Latina*, La Garangola, Padova, vol. 27, fasc. 3, pp. 467-476.
- Ricci Maccarini A., De Colle W., Lucchini E., Casolino D. (2002), L'esame spettroacustico della voce, *Atti del LXXXIX Congresso Nazionale della Società Italiana di Otorinolaringologia e Chirurgia Cervico-Facciale*, San Benedetto del Tronto, 22-25 Maggio 2002, pp. 150-182.

# **FONETICA FORENSE**

# **VERSO UN FORMATO STANDARD NELLE INTERCETTAZIONI E UNA PROPOSTA PER L'ARCHIVIAZIONE E LA CONSERVAZIONE DELLE REGISTRAZIONI**

Luciano Romito, Maria Tucci  
Laboratorio di Fonetica, Università della Calabria  
[luciano.romito@unical.it](mailto:luciano.romito@unical.it); [tucci.maria@libero.it](mailto:tucci.maria@libero.it)

## **1. SOMMARIO**

Le intercettazioni o registrazioni a scopo forense, sono un mezzo di ricerca della prova ai fini dell'individuazione di un reato. Per sua natura quindi è una registrazione di un parlato spontaneo e naturale, da qui il grande interesse da parte di linguisti, dialettologi e sociolinguisti. Tale interesse è stimolato anche dal fatto che in Italia le intercettazioni non riguardano solo ed esclusivamente la criminalità organizzata di 'basso livello' dove il linguista troverebbe un registro basso di italiano e più verosimilmente l'uso di un dialetto molto marcato, ma al contrario le persone intercettate risultano essere: ministri, principi, primari, direttori di banca, industriali, giornalisti, docenti, magistrati, avvocati, preti, poliziotti, calciatori, arbitri, vallette, impiegati ecc. presentando di fatto tutta la stratificazione sociale della popolazione italiana e tutti i registri di italiano e dialetto possibili. Tutto ciò senza avere la necessità di costruire liste di parole o frasi o, ancora più importante, di ricercare un parlato semispontaneo attraverso strategie complesse e costose. Tanto materiale spontaneo quindi, differenziato per produzione, età, sesso, variabili diafasiche e distrattive con l'unico problema legato alla qualità della registrazione e degli strumenti utilizzati. Grazie a questo interesse da non esperti legali, negli anni, abbiamo però notato una grande discrepanza e superficialità nelle tecniche di acquisizione, di conservazione, di catalogazione e archiviazione e nell'uso del materiale registrato.

La proposta di ideare un progetto di digitalizzazione del materiale sonoro intercettato si pone due importanti obiettivi il primo di tipo culturale nel tentativo del recupero e della corretta conservazione di tutto questo materiale per di aprire nuove prospettive alla ricerca e il secondo è quello di proporre un servizio innovativo e multimediale che possa, in un certo senso, garantire maggiore sicurezza e chiarezza al sistema delle intercettazioni in Italia.

## **2. LE INTERCETTAZIONI**

Fin dall'inizio delle intercettazioni su larga scala e solo fino a pochi anni fa, le intercettazioni venivano effettuate con l'ausilio di un registratore marca UHER modello RT 2000 o RT 4000. Salvo alcune rarissime eccezioni, il formato era unico per tutte le Procure Italiane. Le registrazioni avvenivano su supporto analogico ovvero su una bobina normalmente di marca BASF (cfr. Romito 2000). L'unica variabile riscontrabile era la velocità di scorrimento del nastro 2,38 cm/sec per le intercettazioni di tipo telefonico (su rete fissa) e 4,75 cm/sec per le intercettazioni ambientali.

Nell'era del digitale, il miglioramento sarebbe dovuto essere quasi scontato, invece si rileva la nascita e l'uso contemporaneo di una infinità di formati digitali (MP3, Wav, formati proprietari e criptati molti dei quali compressi ecc.) con supporti differenti (cassette analogiche, microcassette digitali DDS, dischi ottici, dischi ORB, ecc.), frequenze di campionamento differenti con oscillazioni che variano da 8000 a 44100 Hz con numero di bit diversi e registrazioni sia stereo che mono. Anche i tipi di intercettazione sono molto

diversi tra loro: le intercettazioni telefoniche possono essere su rete fissa o mobile, su ponte radio, su centrale digitale o meccanica, le reti possono essere GSM, UMTS o VOIP. Anche l'operatore è molto diverso, infatti con l'avvento del digitale e la lentezza delle istituzioni nell'adeguarsi al cambiamento è divenuta prassi comune delegare al privato qualunque tipo di operazione a partire dal noleggio della microspia e della apparecchiatura per l'intercettazione alla sistemazione e installazione della stessa, fino al trattamento analitico della registrazione attraverso consulenza tecnica di trascrizione, comparazione, filtraggio ripulitura ecc.

I registratori analogici (RT2000 e 4000) sono stati soppiantati dai registratori digitali RT6000 (con supporto analogico: cassetta DDS), RT8000 (con supporto digitale: CD) e RT10000 (senza alcun supporto; la registrazione avviene tramite server e memorizzata direttamente su Hard Disk). Tutto ciò è avvenuto in un tempo ridottissimo e, ancor più grave, in momenti differenti da Procura a Procura. Si hanno, quindi, contemporaneamente intercettazioni analogiche in una parte del paese e digitali in un'altra, formati e supporti completamente differenti tra loro ecc.

Anche l'informatizzazione degli uffici preposti è arrivata in momenti differenti nel paese, così non è affatto raro trovare bobine e CD ROM catalogati e conservati in ugual modo o archiviazioni che oscillano dal *libro rubrica* al software fatto in casa nato dalla volontà dell'operatore di gestire le registrazioni in maniera quanto meno più moderna.

Eppure molte sentenze della Cassazione (cfr in seguito) riportano che, la **prova** in un processo, non deve essere intesa la trascrizione, la trasposizione su carta o l'analisi della registrazione sonora, ma bensì la **bobina originale**, sulla quale è avvenuta la registrazione stessa. È proprio per questo motivo che le bobine una volta registrate e munite oltre che di brogliaccio cartaceo anche di *striscetta* sulla quale ogni conversazione veniva identificata attraverso l'ora, il giorno il mese e l'anno, il numero di telefono in entrata e in uscita, il numero progressivo e il numero di giri all'inizio e alla fine, veniva sigillata con ceralacca e il plico veniva firmato da un operatore responsabile.

Molte delle indagini dipendono dalla qualità degli strumenti utilizzati per l'intercettazione, per la conservazione dei supporti e dalle disponibilità economiche di ogni singola Procura. Il ricorso ai privati è aumentato esponenzialmente e spesso grossi investimenti economici non corrispondono ad alta qualità. E' eclatante il caso presentato in una puntata della trasmissione televisiva "Chi l'ha visto" in onda su Rai 3 (12 febbraio 2007) nella quale si fa riferimento ad una inchiesta della Procura della Repubblica presso il Tribunale di Reggio Calabria denominata "Gioco D'azzardo". L'inchiesta basata (anche) su intercettazioni telefoniche e ambientali, ha portato all'arresto di 16 importanti personaggi di Messina, tra cui imprenditori, poliziotti, magistrati e anche un ex sottosegretario al tesoro. In particolare, ciò che in queste sede interessa è che una tra le intercettazioni più importanti effettuata dalla DIA nel 2001 è avvenuta tramite un microregistratore analogico per appunti su una micro-cassetta a velocità LP. Su tale registrazione hanno lavorato e sono intervenuti decine di periti a livello nazionale. Come dire, a fronte di enormi investimenti e spinte verso una sempre crescente digitalizzazione e informatizzazione dei processi, la prova più importante è stata fornita grazie ad un piccolo registratore analogico portatile dal costo di una decina di euro scarsi.

Lo scopo di questo lavoro quindi è quello di evidenziare la necessità di una formato unico e standard per le registrazioni sonore in ambito forense in tutte le procure italiane; la necessità di una creazione di un protocollo per l'archiviazione, la conservazione e la consultazione dei flussi sonori intercettati; l'utilizzazione di protocolli metodologici Standard come AES ad esempio; l'identificazione della registrazione originale e quindi

della **prova** in un processo, nonché la necessità di accertare la validità giuridica della registrazione digitale oggi; la creazione di un documento sonoro informatico; la riservatezza e la sicurezza di tale *documento*; lo scambio di dati con formati simili e confrontabili su tutto il territorio nazionale con un notevole abbattimento dei costi e dei tempi.

### **3. DALL'ANALOGICO AL DIGITALE: IL CASO DELLE INTERCETTAZIONI**

Il processo di digitalizzazione è legato a uno dei più importanti e significativi sviluppi tecnologici del nostro tempo. E' noto che ciò che avviene nella conversione di un documento da analogico a digitale comporta la suddivisione in unità discrete di qualcosa che in realtà è continuo. Il procedimento è molto semplice infatti in fase di digitalizzazione qualsiasi carattere viene tradotto in una sequenza numerica composta da otto cifre (1 bit). Il bit come unità di informazione rappresenta una vera e propria rivoluzione tecnologica che inizia con l'invenzione dei primi computer e raggiunge la sua massima diffusione con la rete Internet. La potenzialità del digitale consiste nel far convivere codici e linguaggi differenti sulla stessa macchina mentre nella rappresentazione di tipo analogica informazioni diverse devono essere archiviate su supporti differenti e non possono essere decodificate dallo stesso dispositivo. Questo è quanto succedeva nel campo delle intercettazioni quando queste venivano effettuate utilizzando un registratore analogico<sup>1</sup>. L'intercettazione avveniva in parallelo su due differenti registratori e le registrazioni venivano segnalate come originale PG e copia AG di fatto un secondo originale registrato contemporaneamente su un identico registratore (cfr §§ succ.). Attraverso un filtro posto in entrata, nella banda di frequenza intorno a 2138 Hz, venivano registrati contemporaneamente sul nastro e attraverso una piccola stampante collegata al registratore su una striscia cartacea (tipo scontrino fiscale) tutte le informazioni relative all'operazione in corso: numero progressivo della registrazione, anno, mese giorno e ora di inizio della registrazione, numero telefonico composto, ora, minuto e secondo relativo alla fine della registrazione, numeri di giri della bobina e infine numero della pista utilizzata sul nastro per la registrazione.

Le stesse indicazioni venivano riportate su un brogliaccio o verbale cartaceo ad opera di un operatore. Ogni turno di servizio aveva quindi un verbale ed un responsabile. Alla fine della intercettazione, le bobine PG ed AG venivano sigillate con ceralacca e il reperto veniva firmato da un operatore-responsabile. Il formato, il supporto e le procedure, salvo rarissime eccezioni, erano uniche in tutta Italia. Il reperto veniva in seguito consegnato all'Ufficio Reperti dove avveniva la catalogazione e la conservazione.

Il consulente che abbia necessità di consultare la registrazione, dietro autorizzazione di un giudice, si reca presso l'ufficio reperti e preleva la copia AG della registrazione, lasciando l'originale PG presso gli uffici della Procura.

È capitato in alcuni casi che il tempo abbia logorato o rovinato parte della bobina registrata, l'usura per esempio può aver portato allo strappo o alla rottura in alcuni casi del nastro o del supporto. In tali casi è stato necessario effettuare prima un restauro del supporto, recuperare le informazioni presenti e non ancora danneggiate sul nastro e confrontarle con quelle presenti nella bobina originale PG. Proprio come potrebbe accadere per un vecchio libro come una cinquecentina ad esempio.

---

<sup>1</sup> Marca UHER modello RT 2000 o RT 4000.



Figura 1: Anche se solo in parte il documento può ancora essere letto e ciò anche dopo 600 anni dalla sua produzione e una non ottima conservazione.

Nell'era del digitale la situazione si è capovolta completamente i registratori sono stati sostituiti da modelli differenti che utilizzano formati e supporti differenti. Si oscilla da una cassetta DDS analogica ma registrata in digitale che contiene fino a 54 ore di registrazione (al contrario di un massimo di 5 ore di una bobina registrata a velocità bassa) RT6000<sup>2</sup> fino alla registrazione tramite server su Hard Disk. In questo ultimo caso alla fine dell'intercettazione viene creata una copia della registrazione su un CD o su un DVD. Il registratore durante la fase di registrazione crea un File TXT con tutte le informazioni relative alla registrazione: anno mese giorno e ora, inizio e durata della conversazione, cella utilizzata, progressivo della registrazione ecc. si ha anche un campo note che può essere riempito dall'operatore con un riassunto della conversazione o con appunti relativi all'indagine. Il seguito della procedura è rimasta inalterata. Il CD viene confezionato e chiuso con ceralacca:



Figura 2: Supporto digitale

Questo viene debitamente firmato e consegnato all'ufficio reperti. Come si può leggere dalla figura risulta essere una copia, in quanto l'originale è presente sull'Hard Disk che per

<sup>2</sup> Il Ministero di Grazia e Giustizia precisa che per quelle Procure che non superano i 200 bersagli annui non è conveniente evolvere verso il sistema informatizzato, potendosi conseguire già apprezzabili i risultati dei sistemi tipo RT 6000 con apparati di proprietà.

ovvi motivi non può certo essere archiviato. Nonostante quindi la grande qualità, almeno in linea teorica, del digitale non si ha alcuna certezza riguardo la correttezza della registrazione, della conservazione, del formato e soprattutto non sappiamo se le macchine costruite nei prossimi cinque o dieci anni saranno ancora capaci di leggere i supporti attuali. In meno di venti anni infatti questi si sono evoluti passando da schede forate (ormai inutili ed illeggibili) a memorie statiche o flash.

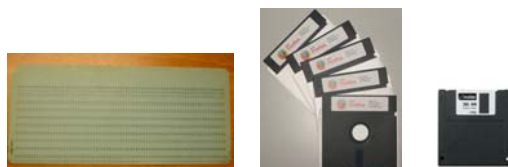


Figura 3: Alcuni supporti digitali ormai in disuso

Ulteriori argomenti di discussione riguardano la durata ed il restauro, al contrario del nastro magnetico o del vinile, la durata del supporto digitale (CD o DVD) non supera i dieci anni di vita e, ancora più drammatico, al contrario di una pagina scritta nel 500, un piccolo danneggiamento del supporto digitale rende tutto il supporto illeggibile con la conseguente perdita di tutti i dati contenuti nel CD o nel DVD.

La rapida evoluzione verso i sistemi digitali ha determinato una proliferazione di formati differenti soprattutto nell'ambito dei files audio. Riteniamo che le risorse del digitale dovrebbero essere sfruttate per uniformare e rendere trasparente l'intero intervento relativo alle intercettazioni. In questo momento il sistema *analogico* risulta, al contrario, essere più "fedele" rispetto all'attuale sistema digitale.

#### 4. ORIGINALE E COPIA NEL SISTEMA ANALOGICO E IN QUELLO DIGITALE

Nel paragrafo precedente abbiamo ripercorso per grandi linee come in pochissimo tempo la procedura legata alle intercettazioni sia cambiata evolvendo verso il sistema informatizzato. Abbiamo anche accennato al concetto di supporto e registrazione originale e in copia.

La corte di Cassazione Penale Sez. V con una Nota del 11/03/2002, n° 9633 sancisce che <<[...] la prova è costituita dalla bobina. [...] La trasposizione su carta del contenuto delle registrazioni rappresenta solo un'operazione di secondo grado>>

Nel sistema analogico l'intercettazione avveniva *in parallelo* su due differenti registratori e le registrazioni venivano segnalate come *originale PG* e *copia AG*. Sul termine *copia* molte pagine sono state scritte, molte delle quali anche inutilmente. La *copia AG* di fatto non è mai stata una vera copia ma bensì un *secondo* originale registrato in parallelo su un *secondo* registratore. L'unica differenza tra le due registrazioni o meglio tra i due originali sta nel fatto che la registrazione AG avveniva di continuo su un'unica pista senza alcuna interruzione mentre invece la registrazione PG (o di lavoro) veniva ascoltata anche durante la fase della intercettazione (durante i momenti in cui nessuna registrazione era in atto) quindi poteva accadere che tutte e quattro le piste del nastro venissero utilizzate.

Altra condizione per noi importante è che la copia di una registrazione analogica non ha mai la stessa qualità dell'originale e comparando le due registrazioni è sempre possibile identificare l'originale.

L'intercettazione in *era digitale*, avviene direttamente sulla memoria di massa di un calcolatore (hard disk). Alla fine dell'intercettazione, o anche prima, l'addetto o l'operatore trasferisce i file su supporti digitali (CD-ROM/ DVD o altro). A questo punto si pone un quesito, almeno sotto il profilo formale: quale è la registrazione originale?

È noto che la *copia* informatica di un file audio ha caratteristiche identiche, in termini acustici e di qualità del segnale, e senza adeguati sistemi di protezione non è assolutamente possibile distinguere la *copia* dall'*originale*. Spesso però, ci si trova a lavorare su supporti la cui origine rimane ignota, ed è difficile, senza l'apporto di verbali o di indicazioni riguardo le caratteristiche della strumentazione utilizzata in fase di registrazione o di copiatura, stabilire se si tratti di un originale o di una copia. La maggior parte delle volte la *copia-originale* è un semplice CD ROM contenuto in una bustina in plastica o in una busta da lettere senza alcun riferimento con una scritta effettuata con un pennarello indelebile del numero di RIT (Registro di intercettazione) o di Procedimento Penale (cfr figura 2). Nulla viene riportato riguardo al formato, all'attrezzatura utilizzata, alle metodiche di riversaggio o di copiatura. L'unica informazione è la traccia informatica dell'avvenuta registrazione.

L'utilizzo delle tecnologie digitali pone il problema di come rendere inconfutabile la *prova* della registrazione. Mentre per il procedimento analogico le bobine e le relative striscette venivano prodotte in duplice copia (doppio originale), venivano inserite in un plico firmato dall'operatore responsabile e sigillato con ceralacca, e al consulente o al perito veniva consegnata solo uno dei due originali, nel mondo digitale la registrazione avviene su hard disk e le copie che vengono prodotte non hanno alcun sigillo o sistema di protezione o certificato di autenticità. Quindi nel caso della registrazione analogica il problema della *copia* è un problema tecnico, mentre nel caso del digitale il problema della *copia* diventa un problema di **validità giuridica**<sup>3</sup>. Quanto detto ha portato la Cassazione a assumere alcune posizioni come la seguente ad esempio<sup>4</sup>:

---

<sup>3</sup> E' necessario introdurre a questo punto il concetto di impronta. Dato un documento di lunghezza arbitraria, un algoritmo crittografico di Hash produce una stringa di lunghezza fissa detta anche *impronta*. Detta impronta ha le seguenti caratteristiche: resistenza alle pre-immagini, è impossibile ricostruire il documento a partire dall'impronta; resistenza alle collisioni, l'impronta è unica per ogni documento; resistenza alle correlazioni, una modifica in un documento genera una modifica nell'impronta.

<sup>4</sup> Articolo apparso sulla Gazzetta del Sud del 27 giugno 2008.



## Garanzia sul riscontro originale delle trascrizioni della polizia

### Cassazione: "server " nelle Procure per memorizzare le registrazioni

ROMA Sulle intercettazioni – andando anche nel solco della previsione del ddl del governo che in maniera ancora più netta prevede centrali uniche di ascolto presso le Corti di Appello – la Cassazione ha messo a punto un piccolo assestamento «garantista» che, comunque, non mette in discussione i processi in corso in quanto – già adesso – in tutte le Procure c'è una «memoria» centrale che custodisce il materiale captato.

Gli ermellini delle Sezioni Unite penali di Piazza Cavour hanno stabilito che affinché le intercettazioni autorizzate dai pubblici ministeri siano utilizzabili in dibattimento, devono essere memorizzate dal «server» della Procura che non può solo fare da «ponte» per la trasmissione del segnale verso i centri d'ascolto esterni (quelli collocati negli uffici di polizia giudiziaria). Ma deve essere in grado di fornire, agli avvocati che lo richiedono, il riscontro originale alle copie dei dischetti contenenti le trascrizioni realizzate nei comandi di polizia e carabinieri.

Insomma l'indagato, se non si fida delle forze dell'ordine, deve poter comparare le intercettazioni che lo incastrano con la registrazione originale rimasta impressa nella memoria della centrale di registrazione dislocata in Procura.

Ad ogni modo, fanno presenti fonti della stessa magistratura, già adesso in ogni Procura c'è il server fisso e quindi gli avvocati che sospettano «taroccamenti», possono confrontare le registrazioni della polizia giudiziaria con quelle del disco rigido del pm.

A sollevare il caso è stato il ricorso di un indagato di Comacchio (Ferrara), al quale era stata imposta la misura cautelare dell'obbligo di dimora nel comune di residenza col divieto di uscire nelle ore notturne. Il legale dell'uomo sosteneva che i resoconti delle intercettazioni, in base alle quali era stata emessa la misura cautelare, erano inutilizzabili in quanto fatte nel Comando dei carabinieri di Comacchio e non in Procura.

I supremi giudici – in sostanza – gli hanno risposto che se ha simili dubbi può benissimo bussare in Procura e consultare il server fisso alla ricerca di imprecisioni nelle trascrizioni delle intercettazioni che «danneggiano» il suo cliente.

Ovviamente tutto ciò diventa molto costoso sia dal punto di vista delle attrezzature ma anche e soprattutto sotto il profilo della manutenzione.

Tutto ciò accade, nonostante lo stesso Ministero per le Riforme e le Innovazioni nella Pubblica Amministrazione abbia pubblicato la *“Proposta di regole tecniche in materia di formazione e conservazione di documenti informatici”*. La proposta è particolarmente interessante per quanto riguarda le intercettazioni soprattutto per gli articoli 5 e 7 del decreto. Nel primo relativo alla **Riproduzione di documenti informatici** si legge:

“Il processo di riproduzione di documenti informatici ai fini della conservazione avviene mediante memorizzazione su diverso e adeguato supporto fisico e termina con l'apposizione sull'insieme dei documenti o su una evidenza informatica contenente una o più **impronte** dei documenti o di insiemi di essi del riferimento temporale e della **firma digitale** da parte del responsabile della conservazione che attesta il corretto svolgimento del processo”

e nel secondo relativo ai **Formati per la conservazione**:

“Il formato di conservazione dei documenti informatici assicura la conservazione del documento e delle sue caratteristiche nel rispetto della normativa vigente. Il formato di conservazione è un **formato standard** aperto, compreso tra quelli riconosciuti dagli organismi nazionali e internazionali preposti alla relativa normazione”.

Se, dunque, il processo di realizzazione e riproduzione dei documenti informatici frutto di intercettazioni fosse così gestito dagli operatori del settore anche nel caso del digitale non parleremmo più di originale e di copia in quanto tutte le copie sarebbero degli originali. Il contenuto del documento, della firma digitale e di tutte le informazioni di sicurezza, sono riportate all'interno del documento stesso, per cui effettuando una copia informatica, di

fatto viene riprodotto integralmente il documento e non esiste la possibilità di distinguere la copia dall'originale, mentre è sempre possibile verificare l'integrità del documento e la validità della firma digitale. Apporre la firma digitale sia sui dati di partenza che su quelli copiati garantisce il servizio di intercettazione nel possedere sempre una registrazione in originale; bisogna ricordare, infatti, che il flusso di registrazioni su Hard Disk è temporaneo e dopo un certo periodo di tempo i dati acquisiti sul server centrale vengono cancellati lasciando quindi esistere l'*originale* solo nella versione *copia*. La tecnologia digitale, con tutte le procedure di sicurezza che prevede, permette di garantire autenticità a qualsiasi tipo di documento.

## 5. LA STRUTTURA DEL DATABASE

Da uno studio condotto dall'istituto di legge criminale internazionale Max Planck in Italia si effettuano più intercettazioni che in tutti gli altri paesi d'Europa. Diventa quindi un problema la catalogazione e la conservazione di tutti questi documenti sonori o supporti.

Oggi nella maggior parte dei casi il materiale audio finisce in scatoloni e in ambienti che non rispettano le misure necessarie per la conservazione di tali supporti. I fattori che influenzano maggiormente l'alterabilità dei supporti audio sono la polvere, le impurità, l'umidità, la temperatura e i campi magnetici ma anche l'integrità del supporto è di fondamentale importanza infatti una minima deformazione può essere il fattore determinante per la perdita di informazione. Il tutto viene complicato, come detto già in precedenza, dal fatto che i supporti da archiviare e catalogare sono sia digitali che analogici e i supporti sono di infinita diversità.

Riportiamo di seguito solo alcuni degli standard AES:

- AES7\_2000 (r2005) standard AES per la preservazione la restaurazione delle registrazioni audio. Metodo per la misurazione del flusso registrato delle registrazioni sonore magnetiche a lunghezza d'onda media (Revisione dell'AES7-1982);
- AES22\_1997 (r2003) raccomandazioni pratiche dell'AES per la preservazione e la restaurazione dell'audio - stoccaggio e manutenzione – stoccaggio di nastri magnetici basati su poliestere;
- AES22\_1996 (r2002) raccomandazioni pratiche dell'AES per scopi legali. Trattamento del materiale audio registrato espressamente per essere sottoposto a esame;
- AES43\_2000 (r2005) standard AES per scopi legali. Criteri per l'autenticazione di registrazioni analogiche audio su nastro.

E' da questa constatazione dei fatti che nasce la necessità di pensare e progettare un database *dinamico* che dia oltre alla possibilità di archiviare il documento informatico frutto di intercettazione ma anche di associare a esso alcuni metadati riguardanti le informazioni legate in qualche modo alla registrazione, includendo quindi tutti i dati concernenti le caratteristiche acustiche del segnale sonoro, il tipo di supporto sul quale il segnale è stato acquisito o semplicemente riversato e anche le indicazioni connesse ai software per la loro creazione e per la lettura.

I metadati dovrebbero costituire in qualche modo il *curriculum o la storia della registrazione* dalla loro creazione a tutte le possibili valutazioni che nel caso di perizia fonica la registrazione viene sottoposta. L'archiviazione potrebbe essere organizzata in moduli fissi contenenti i dati ed in moduli dinamici contenenti i metadati. Questi ultimi conterrebbero informazioni analitiche e dettagliate sul file come il tipo di registrazione, se si tratta di registrazione ambientale o telefonica, il canale utilizzato, il numero di

procedimento penale, l'ufficio richiedente (si può verificare il caso in cui la registrazione sonora sia stata già oggetto di valutazione in altro procedimento), la trascrizione dialettale e la sua eventuale traduzione/interpretazione, i commenti e le note di operatori e consulenti, le operazioni di filtraggio con il relativo file sonoro associato ecc.

Una corretta archiviazione delle registrazioni e anche una corretta gestione dei dati potrebbe rappresentare un punto di partenza per la creazione di un corpus di voci anonime e voci note ai fini del riconoscimento del parlatore nel caso di perizia di comparazione per la costituzione di una comunità linguistica geograficamente dettagliata, l'omologazione di alcuni metodi e procedure di analisi quali i filtraggi, le statistiche utilizzate ecc. Il modello del DataBase potrebbe essere il seguente:

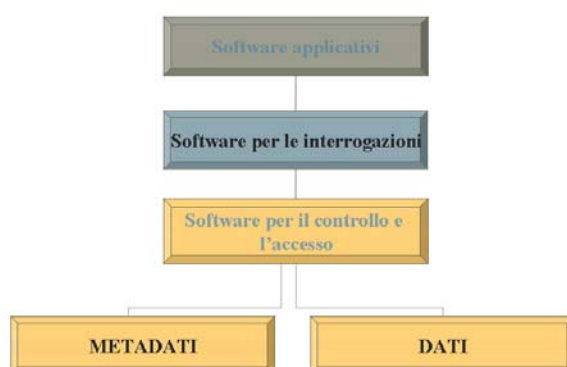


Figura 4: modello del Database

I dati sarebbero costituiti dalla registrazione sonora originale:

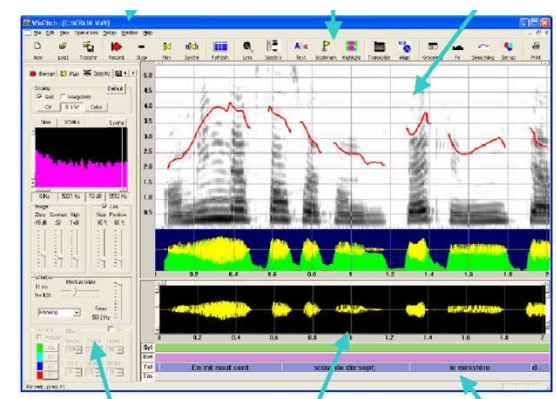


Figura 5

mentre invece i metadati avrebbero una sezione ovviamente fissa, certificata e immutabile e una sezione dinamica e aggiornabile.

## ***METADATI***



Figura 6: Schema del modulo dei Metadati

## **6. CONCLUSIONI**

Il codice di procedura penale precisa che le intercettazioni di conversazioni o comunicazioni (artt. 226-271) rientrano tra i mezzi di ricerca della prova ai fini dell'accertamento della verità e la perizia (artt. 220-223) rappresenta il mezzo di prova. Di conseguenza le intercettazioni come mezzo di ricerca della prova e la validità giuridica che assume la perizia diventano di cruciale importanza in un processo e *i risultati raggiunti sono pienamente utilizzabili dal giudice ai fini della decisione.*

Questo lavoro nasce dall'osservazione di tutto il processo di captazione in ambito forense: dalla installazione della microspia, alla registrazione e intercettazione, dalla copia all'archiviazione e alla catalogazione sia dei materiali che dei supporti.

I problemi riscontrati in questo campo, da punto di vista meramente scientifico ovviamente sono tanti, alcuni di questi potrebbero essere risolti solo attuando delle piccole procedure e rispettando dei protocolli già noti ed esistenti.

Riteniamo che qualsiasi documento, al di là del suo valore caratterizzante o del supporto sul quale è registrato, acquisisce valore in relazione alla sua utilizzabilità e fruibilità. La forza dei DataBase sta nel fatto che riescono a contenere informazioni di ogni genere (testi, grafica, filmati, suoni, ecc.) e di renderli immediatamente disponibili su dispositivi di basso costo e di grande diffusione. Un Database in questo settore così delicato potrebbe essere utile non solo per il recupero del materiale audio intercettato o per la conservazione e la veloce consultazione di una grande mole di dati archiviati ma anche per la loro relazione in ambito di indagine.

In questo preciso momento storico in cui si parla molto di Giustizia e di spese relative all'intercettazione riteniamo che non sia stata effettuata una corretta analisi del attuale situazione.

Il problema e il grande dispendio economico, a nostro avviso, non è da ricercare *nello strumento*, ma nel suo utilizzo. Un gruppo di studio che affrontasse seriamente il problema del digitale e dell'informatizzazione in ambito forense con la messa appunto di reti protette e di formati unici in modo da sfruttare le potenzialità del digitale come per esempio lo scambio veloce di dati, l'archiviazione con più chiavi ecc renderebbe tutto più facile ed economico. Ad oggi invece paghiamo lo scotto relativo alle spese del digitale senza sfruttarne le potenzialità. Tale gruppo di lavoro scoprirebbe che molte delle voci di spesa

riguardano il noleggio di differenti attrezzature. Apparati strumentali di proprietà dello Stato eviterebbero il giogo dei privati, personale dipendente preparato attraverso percorsi formativi strutturati per l'occorrenza, migliorerebbero le acquisizioni e quindi i risultati accorciando i tempi delle indagini e riducendo i costi.

## **BIBLIOGRAFIA**

Bellucci P., (2002), *A onor del vero Fondamenti di linguistica giudiziaria*, UTET.

Paoloni A., Zavattaro D., (2007), *Intercettazioni telefoniche e ambientali*, Centro Scientifico Editore.

Romito L., (2000) *Manuale di Fonetica articolatoria, acustica e forense*, Università degli Studi della Calabria: centro editoriale e Librario.

Romito L., (2003), Passato Presente e Futuro nelle Analisi di Speaker Recognition. In *Voce Canto Parlato*, Zamboni A. (a cura di), Padova: Unipress, pp. 237-246.

Romito L., Galata' V., (2004), Towards a protocol in speaker recognition analysis. in *Forensic Science International*, pp. 105-113.

Romito L., (2004), La misura dell'intelligibilità e il rapporto segnale-rumore. *Atti del convegno "AISV (Associazione Italiana di Scienze della Voce)"*, Padova, 2004.

Romito L., (2005), La competenza linguistica nelle trascrizioni Forensi: l'intelligibilità, l'oggettività e il rapporto segnale/rumore. *Detective And Crime*.

## SPECIFICHE LINGUISTICHE DEL DATABASE UTILIZZATO PER LO SPEAKER RECOGNITION IN S.M.A.R.T.<sup>1</sup>

Luciano Romito, Tommaso Bove, Stefano Delfino, Carla Rossi, Giovanna Jona Lasinio  
Università della Calabria – Laboratorio di Fonetica  
*luciano.romito@unical.it*

### 1. INTRODUZIONE

All'interno dei Programmi europei OISIN I, OISIN II e AGIS 2005<sup>2</sup>, è stato sia sviluppato un progetto di ottimizzazione di un metodo per il riconoscimento del parlatore sia in ambito forense sia, di investigazione preventiva, dal titolo SMART (Statistical Methods Applied to the Recognition of the Talker).

Il progetto, che volge ormai al termine, aveva come obiettivi: il miglioramento delle tecniche di Speaker Recognition; l'implementazione e l'elaborazione statistica dei dati vocali; e soprattutto la costruzione di un corpus etichettato e archiviato (db d'ora in poi), che offre la possibilità di selezionare e organizzare sottocorpora sulla base di variabili linguistiche, anagrafiche e tecniche quali il dialetto, l'inventario fonologico, l'età del parlante, lo zona geografica di provenienza, il canale di registrazione ecc. Tali sottocorpus potrebbero essere utilizzati come comunità linguistica di confronto (Reference Population) per una precisa comparazione fonica.

I metodi statistici utilizzati all'interno del progetto per la comparazione, sono già stati presentati in lavori precedenti come: Brutti P et al. 2002, Bove T. et al. 2002, Bove T. et al. 2003, Bove T. et al. 2004, Bove T. 2006.



Figura 1: Pagina iniziale del software SMART

<sup>1</sup> Lo S.M.A.R.T. (Statistical Methods Applied to the Recognition of the Talker) è un progetto di ricerca finanziato dalla Comunità Europea – Direzione Generale Giustizia, Libertà e Sicurezza.

<sup>2</sup> Programma OISIN II (Rif. JAI/2002/OIS/035).

## 2. LA COMPARAZIONE FONICA

Nel caso di una comparazione fonica, l'esperto, opera su una voce anonima, una voce nota e deve disporre di una comunità linguistica di riferimento per poter stimare il rapporto di verosimiglianza e gli errori di falsa identificazione e di mancato riconoscimento<sup>3</sup>. Al momento attuale l'unico sistema che utilizza un database di voci di riferimento viene utilizzato dall'Arma dei Carabinieri e prodotto dalla FUB (Fondazione Ugo Bordonì)<sup>4</sup>. Tale metodo utilizza un database unico per la popolazione italiana, è costruito su 4 vocali e 4 Frequenze Formantiche. Resta implicito che i segmenti scelti siano ritenuti ugualmente distribuiti su tutto il territorio.

In questa sede la nostra attenzione, è rivolta alla struttura interna del db, in particolare alle categorie linguistiche scelte e alla etichettatura di ogni singolo file. Le motivazioni di tali scelte nascono dalla convinzione che non esista un italiano standard, che la maggior parte delle intercettazioni contenga voci dialettali e che i dialetti italiani sono molto differenti tra loro per segmenti vocalici e inventari fonologici utilizzati (la letteratura a riguardo è talmente vasta da risultare inutile qualunque citazione).

In definitiva crediamo che considerare gli *italiani* come appartenenti ad una unica comunità linguistica, sia un grande errore.

## 3. IL DATABASE

Il db è allocato in un server centrale e prevede l'aggiunta, la modifica o l'eliminazione dei dati da parte di client autorizzati distribuiti sul territorio italiano e europeo<sup>5</sup>. Prevede, anche, la possibilità di interrogazioni attraverso un motore di ricerca e di analisi su tutto il materiale esistente o su una sua parte. Ovviamente tutti i processi, come anche la sola consultazione sono regolati da un sistema di autorizzazioni e privilegi assegnati ad ogni singolo utilizzatore (o esperto).

Le operazioni che potranno essere effettuate avranno due finalità. Innanzitutto, dopo un'analisi linguistica sulle voci da comparare, si estrapolerà ed identificherà il sottocorpus appartenente alla stessa comunità linguistica di parlatori ed in seguito si impiegheranno gli algoritmi per il riconoscimento del parlatore e si effettuerà la comparazione fonica.

Le specifiche linguistiche e non-linguistiche date alle etichette del db, possono essere differenziate in diverse tipologie: informazioni riguardo il supporto; informazioni riguardo la registrazione del file e le sue caratteristiche acustiche principali; le impressioni acustico-percettive sul segnale archiviato; le informazioni anagrafiche e geografiche relative al parlatore e infine le informazioni che entrano nel merito del contenuto della registrazione quindi la lingua o dialetto, l'inventario fonologico di detta lingua/dialetto, il modo di fonazione, informazioni su alcune variabili linguistico-fonologiche ecc. Sono presenti anche alcune parole chiave che riguardano le variabili diafasiche (saggio, telefonata estorsiva, ecc) o l'argomento della conversazione (droga, armi, rivendicazioni ecc). Tale scelta può risultare molto importante perché, nel caso della variabile diafasica è ovvio che

---

<sup>3</sup> Per le metodologie adottate in Italia in ambito di Speaker Recognition, si veda Romito 2006, 2007a e 2007b.

<sup>4</sup> Ovviamente è possibile che alcuni privati abbiano costruito in proprio un database da utilizzare in casi di Speaker recognition anche se l'indagine di Romito 2006 rivela tendenze esattamente opposte.

<sup>5</sup> Almeno per quanto riguarda i partner europei del progetto.

lo stato psicologico del parlatore durante una estorsione non è lo stesso di quello presente durante un saggio o una telefonata amorosa, quindi è indubbio che al variare della situazione il suo parlato risulterà differente (cfr. Romito et al. 1997 e 1998). Questo tipo di archiviazione permette di estrapolare un sottogruppo, dal db, costituito di solo voci caratterizzate dall'essere di qualità omogenea (ad esempio telefonica) e di variabile, ad esempio, estorsiva. In quest'ultimo caso, dal punto di vista sia scientifico che di indagine, potrebbe risultare interessante comparare la produzione linguistica di diversi parlatori sullo stesso argomento, per esempio per individuare gerghi o codici particolari (si pensi all'importanza di tale accertamento in relazione alle rivendicazioni).

La scheda di archiviazione comprende alcune informazioni storiche sul file come il numero del fascicolo o del protocollo, la data dell'analisi, l'ufficio richiedente, l'autorità giudiziaria, il numero di procedimento penale, il tipo di reato, la data di registrazione e la data di inizio lavori.

Informazioni sul canale come Telefono, Cellulare GPS, Cellulare UMTS, VOIP, registrazione ambientale, ambientale in auto, ambientale durante un colloquio in carcere ed è inoltre possibile aggiungere nuovi canali anche se non immediatamente previsti dagli attuali menu.

Informazioni sul supporto come Analogico e Digitale. Nel caso di supporto analogico verrà specificato il tipo di supporto: *cassetta*, *microcassetta* o *bobina*; nel caso di supporto digitale verrà segnalato se il supporto sarà un *CD*, un *DVD* o altro. Se si tratta di un nastro allora dovrà essere indicato se si tratta di *cassetta normal*, *Chrome*, *metal* e se lo stato del supporto è *eccellente*, *buono* ecc. Bisogna indicare la velocità di registrazione, se questa è analogica, la frequenza di campionamento se questa è digitale. Verrà segnalato il numero di bit e se la registrazione è *stereo* o *mono*.

Altre informazioni che non riguardano il supporto ma la registrazione sono di tipo percettivo o analitico. Nel primo caso la registrazione sarà segnalata come *distorta*, *saturata* nel secondo invece verrà segnalato il tipo di fonazione come voce *mormorata*, *bisbigliata* ecc. o il tipo di lingua utilizzata e nel caso di italiano, il dialetto utilizzato. In questo caso viene segnalato anche se tutta la registrazione è in dialetto, in italiano o se la produzione è mistilingue. All'interno delle analisi linguistiche viene segnalato il numero di segmenti vocalici presenti nella lingua o nel dialetto e la distribuzione di detti segmenti nello spazio vocalico (utilizzo di tratti come arrotondamento (F3) o nasalizzazione (FN) o durata (ms)).

Le informazioni anagrafiche sul singolo parlatore invece, richiedono il nome, l'età, il grado di scolarizzazione, provincia di provenienza, la provincia di residenza, ecc.

#### 4. INTERVENTI POSSIBILI SUL DATABASE

Il database come già descritto prevede l'inserimento di dati fissi e di dati dinamici. È possibile inserire un file wav contenente una registrazione completa (per esempio una conversazione telefonica tra due interlocutori) o una singola voce presente in una registrazione più ampia (per esempio solo la voce appartenente all'interlocutore intercettato o sotto controllo). Il file viene etichettato e catalogato attraverso informazioni fisse (utenza telefonica, giorno mese ed anno nonché ora durata della registrazione, ecc) e informazioni dinamiche come nuove trascrizioni, nuove operazioni di filtraggio e di ripulitura, nuove misure acustiche effettuate in seguito ad un accertamento di riconoscimento del parlatore con algoritmi diversi o altri accertamenti attraverso analisi dialettali, fonetiche e fonologiche, morfologiche o sintattiche e stilistiche, ecc.



È possibile modificare o manipolare le **informazioni dinamiche** presenti nel db. Aggiungendo informazioni e analisi su più livelli, misure acustiche o relative alla durata, parametri come *articulation rate*, *speech rate*, o *fluency* o anche parametri che al momento attuale non vengono considerati come *andamento storico della Frequenza Fondamentale o delle Frequenze Formantiche* ecc. Ovviamente per effettuare qualunque modifica l'operatore<sup>6</sup> deve essere effettuata dall'amministratore del db.

## 5. IL RISULTATO DELLA COMPARAZIONE

Una volta che la comunità linguistica è stata identificata e la stessa viene *trasferita* sul *client* (in modo da non appesantire le operazioni del server), è possibile effettuare la comparazione fonica.

Usando l'approccio del massimo likelihood, viene stimata la soglia di confidenza cioè la peculiare caratteristica dello speaker analizzato, l'errore A-priori di *Falsa Reiezione* e di *Falsa Accettazione*.

Se la distanza tra la voce nota e la voce anonima è maggiore della soglia stimata allora le due voci non appartengono allo stesso parlatore, viceversa se la distanza è minore allora le due voci appartengono allo stesso parlatore o quanto meno sono compatibili. In entrambi i casi vengono presentati anche i valori di errore di Falsa Identificazione e di Falsa Accettazione.

I risultati dell'analisi comprendono una completa valutazione prevedendo un grafico rappresentante la distanza intra e inter parlatore con la soglia massima di likelihood (LR); un *report* contenente due tipi di errori a-priori e tutte le informazioni legate all'analisi; un *grafico* rappresentante la distanza intra e inter speaker all'interno della attuale distanza tra voce nota e voce anonima; un *report* contenente due tipi di errori a-posteriori e tutte le informazioni relative alla comparazione tra voce anonima e voce nota (per approfondimenti sul metodo parametrico si vedano i lavori già citati di Brutti P et al. 2002, Bove T. et al. 2002, Bove T. et al. 2003, Bove T. et al. 2004, Bove T. 2006).

## 6. CONCLUSIONI

Il progetto SMART non vuole essere solo un metodo per la comparazione di voci note e anonime, ma anche un software di supporto per l'esperto nella complessa operazione di comparazione. L'individuazione della comunità linguistica della voce nota e della voce anonima, l'identificazione di caratteristiche fonetico-fonologiche simili, l'identificazione di una sottocomunità ristretta di voci prodotte da locutori non solo appartenenti alla stessa comunità linguistica ma anche con caratteristiche simili quali l'età, il canale di registrazione e addirittura la tipologia diafasica rendono più agevole e sicuramente più preciso il risultato. Il database può anche essere consultato per studiare gerghi, codici ristretti o variabili stilistiche particolari delle sole estorsioni o delle sole rivendicazioni, ecc.

La stessa creazione del database risulta essere utile sia agli esperti sia agli operatori, infatti è da stimolo e da volano per procedure di standardizzazione sui metodi di misura,

---

<sup>6</sup> Per operatore, in questo caso si intende anche un dialettologo che effettua analisi e misure su un corpus di parlato spontaneo telefonico o ambientale di una certa zona geografica. I suoi risultati scientifici potranno essere utilizzati da un esperto perito in fase di indagine.

sugli algoritmi da utilizzare, sulle statistiche decisionali e sicuramente sulla compartecipazione di diversi metodi all'interno dello stesso ambiente informatico.

Allo scopo sono già iniziati lavori di analisi per studiare fenomeni quali la variabilità interparlatore o intercomunità rispetto a quella intra parlatore ecc. si veda Romito Lio 2008.

## BIBLIOGRAFIA

Brutti P, Fabi F. Jona Lasinio G. (2002), Una proposta di meta-analisi basata sulla combinazione di classificatori per il problema del riconoscimento del parlatore, *Statistica*, 3 pp. 455:473.

Bove T., Brutti P., Fabi F., Jona Lasinio G., Giua P.E., Forte A. e Rossi C., (2002), Tecnologie Informatiche nella Promozione della Lingua Italiana, 25-26 giugno 2002, Progetto di ricerca S.M.A.R.T. 2, Conferenza: *T.I.P.I.*, pp. 121-124.

Bove T., Brutti P., Fabi F., Jona Lasinio G., Giua P.E., Forte A. e Rossi C. (2003), Three approaches to the speaker identification problem for forensic use, in *atti del Convegno Cladag 22-24 settembre 2003* (relazione invitata).

Bove T. Jona Lasinio G., Rossi C.. (2004), The speaker recognition problem, XLII Riunione Scientifica della Società Italiana di *Statistica*, pp. 429:440, (relazione invitata)

Bove T., (2006), SMART II: Analisi della voce Conferenza T.A.L.–in *Trattamento Automatico Linguaggio – Sessione 3<sup>a</sup>: Intelligence* 9-10 marzo 2006.

Trumper J., Romito L., Maddalon M., Mendicino A., Belluscio G. M. G., (1993), Stime manuali: un esperimento. Atti del convegno *Teoria e Sperimentazione: Parametri, tratti e segmento*, Calabria, 28-29 novembre, 1991, Esagrafica Roma:Roma, Vol. XIX, pp. 61-79.

Trumper J., Romito L., Mendicino A., Li J., (1994), Stime vocaliche problematiche. Atti del convegno *Giornate di Studio del Gruppo di Fonetica Sperimentale*, Padova, 1992, Vol. XX, pp. 49-59.

Romito L., Maddalon M., Trumper J., (1996), La parametrizzazione nei test di riconoscimento. Atti del convegno *VIIe Giornate di Studio del Gruppo di Fonetica Sperimentale (G.F.S.)*, Roma, 1996, pp. 87-93.

Romito L., Maddalon M., Trumper J., (1996), Atteggiamento della Magistratura nei confronti delle perizie foniche. Atti del convegno *VIIe Giornate di Studio del Gruppo di Fonetica Sperimentale (G.F.S.)*, Roma, 1996, Roma, pp. 34-45.

Romito L., *Manuale di Fonetica articolatoria, acustica e forense*, (2000), Università degli Studi della Calabria: centro editoriale e Librario.

Romito L., (2003), Passato Presente e Futuro nelle Analisi di Speaker Recognition. In *Voce Canto Parlato*, Zamboni A. (a cura di), Padova: Unipress, pp. 237-246.

Romito L., Galata' V., (2004), Towards a protocol in speaker recognition analysis. *Forensic Science International*, pp. 105-113.

Romito L., (2004), La misura dell'intelligibilità e il rapporto segnale-rumore. *Atti del convegno "AISV (Associazione Italiana di Scienze della Voce)"*, Padova, 2004.

Romito L., (2005), La competenza linguistica nelle trascrizioni Forensi: l'intelligibilità, l'oggettività e il rapporto segnale/rumore. *Detective And Crime*.

Romito L., Blefari M., (2005), Towards a new parameter in Recognition. *Speech Language and the Law*.

Romito L., Galata' V., Lio R., (2006), Fluency Articulation and Speech Rate as new parameters in the Speaker Recognition. Atti del convegno "*III Congreso de Fonética Experimental*", Santiago de Compostela, 26-24 ottobre, 2005, Xunta de Galicia:Santiago de Compostela, pp. 537-549.

Romito L., Galata' V., (2007), Speaker Recognition: Stato dell'arte in Italia. Atti del 3° Convegno Nazionale AISV, "*Scienze Vocali e del Linguaggio Metodologie di Valutazione e Risorse Linguistiche*", Trento, 29/11-1/12, 2006, EDK Editore SRL:RN, Vol. 3.

Romito L., Galata' V., (2008), Speaker Recognition in Italy: evaluation of methods used in forensic cases. *Atti del IV Convegno CFE*, Granada - Spagna, 11-14/02, 2008.

Romito L. , Tucci M. , (2008), Verso un formato standard nelle intercettazioni e una proposta per l'archiviazione e la conservazione delle registrazioni. Atti del IV convegno AISV - *La fonetica sperimentale. Metodo e applicazioni*, Cosenza, 3-5/12/2007, 2007, EDK s.r.l.:RN, Vol. IV.

Romito L., Scullari V., (2008), Un protocollo delle procedure di restauro all'interno dell'Archivio sonoro calabrese. *Atti del IV convegno AISV*, Università della Calabria, 3-5/12, 2007, EDK Editore SRL:RN, Vol. 4.

Romito L., Lio R., (2008), Stabilità dei parametri nello Speaker Recognition: la variabilità intra e interparlatore. *Atti del IV convegno AISV* (Associazione Italiana Scienze della voce), Unical, Campus di Arcavacata di Rende (CS), 3-5/12, 2007, EDK Editore s.r.l.:RN, Vol. IV.

G. Cavarretta L. Romito, M. Tucci, (2008), Verso un formato standard nelle intercettazioni: archiviazione, conservazione, consultazione e validità giuridica della registrazione sonora, in *Atti del Convegno Internazionale Ass.I.Term*, Università degli studi della Calabria 6-8/6/2008.

## 7. APPENDICE

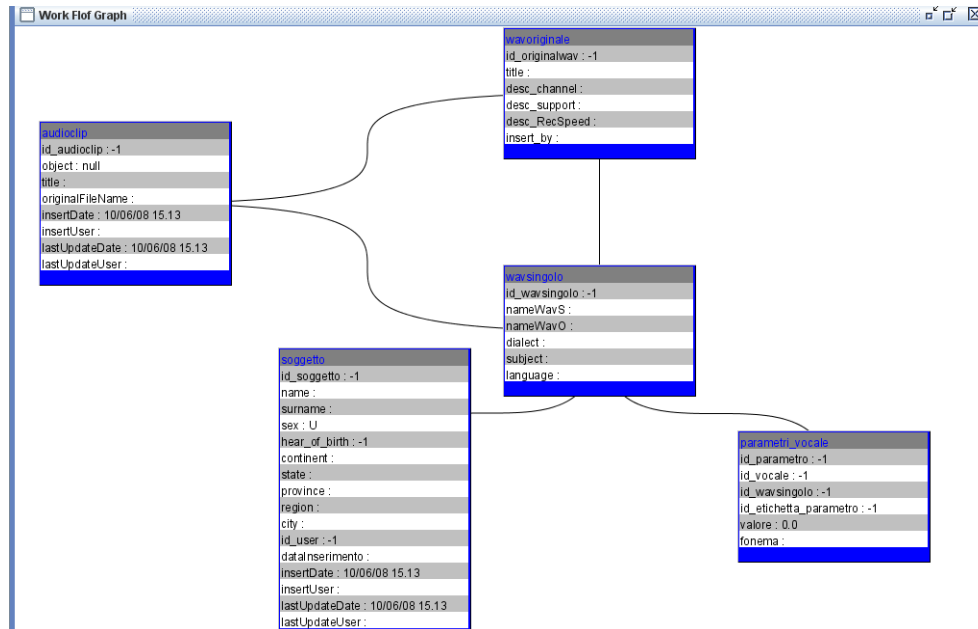


Figura 2: Diagramma di flusso del software SMART

The 'Original Wav' form is used for entering data related to the support and file. It includes the following fields and controls:

- ID**: A text input field with the value '-1'.
- Title**: A text input field.
- Support**: A dropdown menu.
- Vel. Record**: A dropdown menu.
- Type**: A dropdown menu.
- Frq. Rec.**: A dropdown menu.
- Support State**: A dropdown menu.
- Record State**: A dropdown menu.
- Chanel**: A dropdown menu.
- Record Date**: A text input field.
- AudioClip**: A checkbox.
- Search into DataBase**: A button.
- Save**, **Delete**, and **Exit**: Buttons at the bottom.

RecordGui linking to Record by: smart3revb.client.gui.WavOriginaleGui,0,0,0x0,invalid,layout=j

Figura 3: Maschera per l'inserimento dei dati riguardanti il supporto e il file.

RecordGui linking to Record by: smart3revb.client.gui.WavSingoloGui[0,0,0x0,invalid,layout=java.awt.B...

Figura 4: Maschera per l'inserimento dei dati riguardanti la registrazione.

RecordGui linking to Record by: smart3revb.client.gui.SoggettoGui[0,0,0x0,invalid,layout=java.awt.Bor...

Figura 5: Maschera per l'inserimento dei dati anagrafici del locutore.

WIZARD - STORE FILE FORMANTS INTO SMART DATABASE

SELECT VOWELS-

- ☐ Vowel: a
- ☐ Vowel: e
- ☐ Vowel: i
- ☐ Vowel: o
- ☐ Vowel: u
- ☐ Vowel: Ä ɤ
- ☐ Vowel: Ä ʰ
- ☐ Vowel: Ä ʱ
- ☐ Vowel: Ä ʷ
- ☐ Vowel: w

SINGLE WAV IDENTIFIER-

Single Wav Identifier: -1

Search ID

INSPECT SPEAKER

NEXT >> CANCEL

Figura 6: Maschera per l'inserimento dei dati acustici

SPEAKER RECOGNITION ALGORITHM - Select Speaker Set

SELECT VOWELS-

- ☐ Vowel: a
- ☐ Vowel: e
- ☐ Vowel: i
- ☐ Vowel: o
- ☐ Vowel: u
- ☐ Vowel: Ä ɤ
- ☐ Vowel: Ä ʰ
- ☐ Vowel: Ä ʱ
- ☐ Vowel: Ä ʷ
- ☐ Vowel: w

SELECT LANGUAGE-

Language: Italian

Phonation:

SELECT DIALECTS-

- NULL or not present
- calabrese
- milanese
- napoletano
- romano

SELECT SEX-

☒ MALE

☐ FEMALE

☐ UNKNOWN

SELECT AGE-

☐ Don't care age

Minimum age: 20

Maximum age: 60

SELECT KEYWORDS-

- ATTENTATO
- BOMBA
- CAPODANNO
- DECAPITAZIONE
- DETONAZIONE
- ESPLODERE
- ESPLOSIONE
- FUCILE
- GRANATA
- MISSILE

Additional options

NEXT >> CANCEL

Figura 7: Maschera per l'inserimento e per la ricerca di parole chiave nelle trascrizioni allegate alle registrazioni