Studi AISV 2

LA FONETICA NELL'APPRENDIMENTO DELLE LINGUE

PHONETICS AND LANGUAGE LEARNING

a cura di
Renata Savy e Iolanda Alfano

AISV
Associazione Italiana Scienze della Voce

LA FONETICA NELL'APPRENDIMENTO DELLE LINGUE

PHONETICS AND LANGUAGE LEARNING

a cura di
Renata Savy e Iolanda Alfano

Milano 2016

Studi AISV è una collana di volumi collettanei e monografie dedicati alla dimensione sonora del linguaggio e alle diverse interfacce con le altre componenti della grammatica e col discorso. La collana, programmaticamente interdisciplinare, è aperta a molteplici punti di vista e argomenti sul linguaggio: dall'attenzione per la struttura sonora alla variazione sociofonetica e al mutamento storico, dai disturbi della parola alle basi cognitive e neurobiologiche delle rappresentazione fonologiche, fino alle applicazioni tecnologiche.

I testi sono sottoposti a processi di revisione anonima fra pari che ne assicurano la conformità ai più alti livelli qualitativi del settore.

I volumi sono pubblicati nel sito dell'Associazione Italiana di Scienze della Voce con accesso libero a tutti gli interessati.

Curatore/Editor
Cinzia Avesani (CNR-ISTC)

Curatori Associati/Associate Editors
Franco Cutugno (Università di Napoli), Barbara Gili Fivela (Università di Lecce), Daniel Recasens (Università di Barcellona), Mario Vayra (Università di Bologna).

Comitato Scientifico/Scientific Committee
Giuliano Bocci (Università di Ginevra), Silvia Calamai (Università di Siena), Mariapaola D'Imperio (Università di Aix-en-Provence), Giovanna Marotta (Università di Pisa), Renata Savy (Università di Salerno), Stephan Schmid (Università di Zurigo), Carlo Semenza (Università di Padova), Claudio Zmarich (CNR-ISTC).

© 2016 AISV - Associazione Italiana Scienze della Voce
c/o LUSI Lab - Dip. di Scienze Fisiche
Complesso Universitario di Monte S. Angelo
via Cynthia snc
80135 Napoli

Edizione realizzata da
Officinaventuno
Via Doberdò, 21
20126 Milano - Italy
email: info@officinaventuno.com
sito: www.officinaventuno.com

ISBN edizione digitale: 978-88-97657-16-3

Indice

DOI: 10.17469/O2102AISV000000

RENATA SAVY

Prefazione

Il volume è una raccolta di contributi orientati a vari aspetti fonetici e fonologici dell'apprendimento linguistico, che presentano i risultati di ricerche su tratti articolatori, acustici e percettivi di suoni non nativi, sia in italiano L2, sia nelle produzioni in diverse lingue straniere da parte di nativi italofoni. Nel piano complessivo della raccolta, emergono almeno due prospettive di indagine particolarmente interessanti: l'attenzione rivolta alle caratteristiche prosodico-intonative delle produzioni degli apprendenti e il *focus* di buona parte dei lavori su sperimentazione e ricadute della speculazione teorica in campo glottodidattico.

L'idea del volume nasce infatti sulla spinta dell'interessante dibattito sviluppatosi in seno al XII Convegno Nazionale dell'AISV (Associazione Italiana di Scienze della Voce), che si è svolto presso l'Università degli Studi di Salerno (Campus di Fisciano) nel gennaio del 2016, dal titolo *La fonetica sperimentale nell'insegnamento e nell'apprendimento delle lingue straniere*. Il Convegno proponeva una riflessione sia sugli aspetti più propriamente 'didattici' (strumenti, metodi, protocolli utilizzati nell'insegnamento delle lingue seconde e straniere, con specifico riferimento alla fonetica), sia sui meccanismi e i processi di sviluppo delle competenze fonetiche e 'metafonetiche' da parte degli apprendenti di lingua straniera. Su queste premesse, l'incontro salernitano ha coinvolto un pubblico vivace ed eterogeneo, composto da un lato da fonetisti italiani e stranieri[1], dall'altro da docenti impegnati in prima persona nella didattica di lingue straniere.

Particolarmente rilevante è stato, in quell'occasione, il confronto proposto da una Tavola Rotonda che ha visto partecipi alcuni rappresentanti e direttori dei Centri Linguistici degli atenei italiani[2]. L'incontro aveva lo scopo di sensibilizzare le istituzioni competenti e gli erogatori dei servizi di supporto linguistico allo svilup-

[1] Ci preme ricordare le preziose letture plenarie di Joaquim Llisterri Boix (Universitat Autònoma de Barcelona) *De la fonética a la enseñanza de la pronunciación* e di Mariapaola D'Imperio (Aix Marseille Université, Laboratoire Parole et Langage) *Acquisition of L2 prosodic features and individual differences: the role of L1 use and musical training.*

[2] Approfittiamo di questa sede per ringraziare della loro cortese collaborazione: Anna De Meo (Direttrice del Centro Interdipartimentale dei servizi Linguistici e Audiovisivi dell'Università degli Studi di Napoli L'Orientale) che ha introdotto e moderato il dibattito, Carmen Argondizzo (Presidente dell'Associazione Italiana dei Centri Linguistici Universitari e del Centro Linguistico di Ateneo dell'Università della Calabria), Giuliano Bernini (Direttore del Centro di Competenza delle Lingue dell'Università di Bergamo), Elisabetta Bonvino (Direttrice del Centro Linguistico di Ateneo dell'Università Roma Tre), Elisabetta Jezek (Presidente del Centro Linguistico di Ateneo dell'Università di Pavia) e Marie Berthe Vittoz (Direttrice del Centro Linguistico di Ateneo dell'Università di Torino).

po di metodi e tecniche della didattica fonetica. Il dibattito, se da un lato ha messo in luce le difficoltà e anche alcune carenze nell'inclusione degli aspetti fonetici nei programmi d'insegnamento, dall'altro ha evidenziato la necessità e il desiderio da parte della comunità scientifica di lavorare alla costruzione di modelli e metodologie che siano in grado di rispondere adeguatamente alle esigenze didattiche.

Il libro, articolato in tre parti, include, quindi, una selezione dei saggi presentati in quella sede e sottoposti a *blind peer review*, nei quali trovano posto spunti di riflessione critica sui metodi e gli strumenti didattici.

La prima parte, dedicata a *Aspetti fonetici e prosodici nell'apprendimento di L1 e di L2*, presenta due distinte sezioni. La prima, dedicata agli aspetti fonetici segmentali, si apre con il contributo di Romito, Frontera e Ciardullo che investiga la realizzazione di alcune opposizioni fonematiche del sistema vocalico (vocali in posizione post-tonica finale) e consonantico (opposizione /b v/ e /ʦ ʣ/) dell'italiano da parte di apprendenti nativi galiziani, in rapporto a due variabili sociolinguistiche, tradizionalmente considerate come fattori in gioco nella letteratura su *Second Language Acquisition*: la percentuale d'uso della L1 (in questo caso, galiziano *vs* castigliano) e il grado di esperienza e familiarità dell'apprendente con la L2, unitamente a specifiche competenze metafonetiche. I risultati di due esperimenti, uno di percezione e uno di produzione, discussi con riferimento al quadro dei principali modelli sull'acquisizione di categorie fonologiche in LS, mostrano un quadro complesso in cui percezione e produzione non sembrano viaggiare di pari passo nel processo di apprendimento e in cui, soprattutto, la competenza metalinguistica gioca un ruolo fondamentale come fattore di accelerazione dello sviluppo acquisizionale dell'apprendente. Il lavoro pertanto suggerisce, nella conclusione, "la necessità e l'utilità di integrare i percorsi di insegnamento e apprendimento di una L2/LS con attività rivolte alla pratica fonetica e fonologica" (p. 40).

Nel saggio successivo, Schmid e Pedrazzini presentano un esperimento didattico condotto su due gruppi distinti di apprendenti italofoni di tedesco in classi di scuola secondaria della Svizzera italiana: un gruppo sottoposto a sessioni di training di pronuncia/fonetica, un secondo gruppo di controllo senza alcun addestramento preventivo. Lo studio si focalizza sull'acquisizione e realizzazione corretta del tratto del VOT e del %Voice (percentuale di sonorità) delle consonanti tedesche. La ricerca affronta esplicitamente le questioni dell'insegnamento della pronuncia di una seconda lingua e degli approcci metodologici impiegati (dal metodo di ispirazione 'comportamentista', basato su esercizi ripetitivi, ai vari utilizzi del 'metodo fonetico', centrati sullo studio esplicito delle differenze tra L1 ed L2). I risultati della sperimentazione, nel mettere in evidenza il *transfer* da L1 ad L2 nei due parametri acustici considerati, mostrano anche che non esiste apparentemente un incremento significativo delle *performance* dei soggetti sottoposti ad addestramento fonetico: gli autori, tuttavia, avanzano per quest'aspetto un'interessante ipotesi di differenza intersoggettiva di sensibilità al *training* metalinguistico e, soprattutto, discutono i limiti di un metodo didattico tradizionale articolato in 'presentazione-pratica-pro-

duzione' e implementato su parlato letto, condiviso da buona parte della ricerca sull'insegnamento della 'pronuncia'.

Il contributo di d'Apolito e Gili Fivela indaga le realizzazioni di apprendenti italofoni in francese L2 di nessi di sibilanti a confine di parola (es. *Il dit ta**sse chi**noise rapidement*) che presentano difficoltà coarticolatorie e in cui, di norma, è ipotizzabile l'intervento di due strategie 'riparative' distinte: quella dell'inserimento di vocale epentetica, specificata da un proprio *target* articolatorio, e quella del fenomeno noto come *gestural mistiming*, che equivale alla realizzazione di una vocale intrusiva transizionale, priva cioè di stabilità articolatoria (*targetless*). Anche in questo caso, le strategie sono riconducibili ad una pratica glottodidattica diffusa che suggerisce la produzione di uno *schwa* per la pronuncia di tali nessi. Nella ricerca vengono riportati dati acustici e articolatori delle suddette sequenze in due contesti prosodici e a velocità d'elocuzione diverse; i risultati delle misurazioni inducono gli autori a considerare il fenomeno come un aspetto di *transfer* pervasivo dalla L1 dei parlanti con chiare caratteristiche di *gestural mistiming*, cioè variabilità coarticolatoria legata alla scarsa coordinazione tra le due consonanti.

Nel successivo lavoro, Sisinni, Gili Fivela e Grimaldi conducono uno studio elettrofisiologico su parlanti nativi italiani che mira a stabilire se e in che misura gli stessi sono in grado di discriminare il tratto di durata in suoni (vocalici) non nativi. L'esperimento fa uso del metodo della *Mismatch Negativity* (MMN) e si basa sull'elaborazione di dati elettroencefalografici (EEG) registrati durante le sessioni di compiti percettivi di vocali in stimoli pseudo-linguistici. Le conclusioni cui arrivano gli autori sembrano dimostrare l'esistenza di strategie diverse da parte dei parlanti di lingue in cui la quantità vocalica non è un tratto del sistema fonologico rispetto ai nativi di lingue cosiddette *quantity system*: i soggetti italofoni, infatti, elaborano i valori formantici e la durata di suoni non nativi come eventi separati e strettamente sequenziali, mentre nei soggetti nativi è ipotizzata l'esistenza di due meccanismi neurali indipendenti e paralleli.

In una seconda sezione di questa prima parte del volume, è ospitato un nutrito gruppo di saggi su aspetti dello sviluppo prosodico nell'apprendimento di lingue straniere (o seconde), un tema rimasto per diverso tempo marginale nella letteratura fonetica italiana (a parte rare eccezioni), ma che di recente sta appassionando molti ricercatori. I saggi di seguito illustrati spaziano dall'analisi dei diversi gradi di approssimazione alla lingua *target* nelle realizzazioni prosodiche degli stadi di "interlingua", all'indagine sulle interfacce dell'intonazione con strategie pragmatiche, intenzioni comunicative e stati emotivi. Trovano posto, infine, alcune analisi dello sviluppo di tratti prosodici di bambini nelle fasi di apprendimento della lingua materna.

Vigliano, Pellegrino e Pettorino presentano una ricerca che si prefigge lo scopo "di estendere le indagini sperimentali sull'efficacia pedagogica della tecnica dell'auto-imitazione per il miglioramento della competenza prosodica" (p. 103), effettuata su soggetti nativi giapponesi alle prese con produzioni in italiano L2. Nel lavoro vengono raccolti e manipolati dati intonativi in L2 legati a tre diverse fun-

zioni comunicative (concessione, comando e richiesta). L'attenzione fondamentale del lavoro si concentra sulla sperimentazione metodologica, articolata in cinque fasi progressive: pre-esercitazione, manipolazione del segnale acustico, auto-imitazione, test percettivo e analisi acustica. I risultati ottenuti vengono discussi, quindi, a supporto della validità delle tecniche didattiche utilizzate (in particolare l'esercitazione prosodica) che incrementano positivamente l'efficacia comunicativa in L2. Di particolare rilievo è, nella premessa, la segnalazione di una serie di risorse e tecnologie web a disposizione dei docenti per la sperimentazione di sistemi di insegnamento della pronuncia assistiti dal computer o *Computer Assisted Pronunciation Teaching* (CAPT).

Su alcuni aspetti intonativi di 'interlingua' si sofferma il saggio di Orrico, Cataldo, Savy e Barone, frutto di un progetto di monitoraggio delle competenze prosodiche di apprendenti italofoni di Inglese-Lingua Straniera (*English Foreign Language*) nel contesto del curriculum di studi universitario. Il contributo indaga i profili intonativi di domande polari prodotte in lingua straniera da cinque gruppi di soggetti a diversi livelli di competenza linguistica, testando e verificando l'ipotesi che in mancanza di un esplicito percorso di acquisizione di competenze 'metaprosodiche', l'avanzamento sul piano grammaticale (in senso lato e omnicomprensivo) non corrisponda ad un parallelo miglioramento del piano fonetico e intonativo; su quest'ultimo si manifestano quindi in maniera preponderante fenomeni di *transfer*, fossilizzazione dell'*habitus* nativo e di 'deriva' prosodica, vale a dire un allontanamento dalla propria L1 non indirizzato al modello della lingua *target*. Anche in questo caso, gli autori suggeriscono la necessità di un percorso guidato e dell'inclusione dell'insegnamento prosodico nei programmi didattici.

Il contributo di Gamal propone un confronto prosodico di strutture sospensive in arabo (-cairota) L1 e italiano L2 di arabofoni, in conversazioni semi-spontanee, elicitate con lo stesso metodo anche in italiano L1. L'analisi, condotta attraverso il metodo ToBI su differenti profili non finali, correlati ad altrettante modalità pragmatiche sospensive, riguarda i toni di confine, lo *slope* e la struttura accentuale. I risultati del confronto, sebbene preliminari, suggeriscono un'interpretazione in direzione di fenomeni di '*transfer* positivo': le affinità tra L1 e lingua target (ad esempio nel contorno fortemente ascendente sul sintagma intonativo pre-finale) favoriscono le performance dell'apprendente, che tuttavia si discostano dal modello del sistema di arrivo su altri piani (la tipologia degli accenti intonativi e l'entità dello *slope*). Differenze ancora più significative vengono osservate preliminarmente sul piano ritmico, ma in ogni caso viene sottolineata una distanza tra L1 ed L2 che limita l'ipotesi di trasferimento tout court dei profili della lingua materna.

Sorianello e De Marco indagano la realizzazione prosodica di tre emozioni primarie (collera, gioia, tristezza) in diverse L1 (arabo tunisino, russo e spagnolo) e in italiano nativo e non nativo. Nella formulazione delle autrici, "l'ipotesi di partenza è che le differenze nell'espressione delle emozioni da parte dei soggetti coinvolti siano riconducibili alla loro L1 e cultura di origine. In tal senso il ruolo del transfer agirebbe come filtro sulla struttura prosodica dei parlanti" (p. 163). L'analisi dei parame-

tri acustici considerati (velocità d'eloquio, intensità media, valori puntuali di f0 ed escursione melodica dell'intero enunciato) e la discussione dei risultati sperimentali vengono così messe in relazione a fattori culturali, in particolare al modello di Hall (Hall, Hall, 1990) che distingue tra culture ad alto o basso contesto comunicativo. Il quadro finale che ne risulta è molto eterogeneo e dimostra che le produzioni delle emozioni in L2, piuttosto che da fenomeni di *transfer* dalla propria L1, sono caratterizzate da un altissimo livello di confusione, incertezza e incongruenza, correlate alla difficoltà da parte dell'apprendente di cimentarsi in un compito paralinguistico complesso come quello dell'espressione emotiva.

Il lavoro di Zanchi, D'Imperio, Zampini e Fasolo si incentra su caratteristiche dello sviluppo prosodico in L1, mettendo a confronto l'intonazione delle narrazioni di bambini in età prescolare e adulti italiani. Il discorso narrativo è stato elicitato da un *Narrative Competence Task* ed è stata condotta un'analisi autosegmentale-metrica, attraverso il paradigma ToBI, dei profili di enunciati assertivi, prendendo in considerazione: *pitch accent nucleare*, *boundary tones* e numero di *breaks* 3 e 4 degli enunciati. Dai risultati emerge come, già a 3 anni di età, i bambini siano in grado di utilizzare correttamente, in modo simile agli adulti, il *pitch accent* nucleare tipico dell'enunciato *broad focus* in italiano, mentre non dominano l'utilizzo di un tono di confine di tipo *continuation rise*, usato di norma, nelle narrazioni, per fornire informazioni all'interlocutore circa la necessità di interpretare il significato dell'enunciato in collegamento col successivo. Gli autori interpretano tale risultato come prova di un *gap* tra bambino e adulto non tanto sul piano prosodico-intonativo, quanto sul piano di un sistema cognitivo globale ancora in evoluzione che impedisce al bambino la corretta valutazione delle presupposizioni dell'interlocutore e della somma delle conoscenze condivise o meno.

La prosodia di una varietà di italiano L1 di soggetti adulti è oggetto di indagine del contributo sperimentale di Badan e Crocco, che presentano una descrizione intonativa molto dettagliata degli accenti melodici e dei toni di confine caratteristici delle *domande wh-* (neutre e eco) nell'italiano parlato in Veneto (provincia di Padova), ancora poco descritto in letteratura. I risultati di questo lavoro mostrano che, alle diverse strutture sintattiche e proprietà interpretative della *domanda wh-* neutra e di quella eco corrispondono strutture prosodiche chiaramente differenziate, sul piano della sequenza degli accenti nucleari e del *range* melodico (espanso nelle domande eco).

A concludere la prima parte del volume, il saggio di Olivucci, Pasqualetto, Vayra e Zmarich investiga in prospettiva fonetico-acustica lo sviluppo dell'accento lessicale in italiano nel bambino in età prescolare, in uno studio longitudinale su un periodo di sei mesi. Sul piano descrittivo e sperimentale, la letteratura mostra una notevole discrepanza tra le numerose ricerche condotte su soggetti adulti e gli studi assai scarsi sui soggetti in età evolutiva; rare eccezioni riguardano quasi esclusivamente la lingua inglese. I risultati di questo studio, relativi a bambini italiani dai 21 ai 27 mesi, mostrano come la capacità di produrre sillabe toniche e sillabe atone distinte fra loro emerga in età molto precoce, sia per quanto riguarda il tratto di du-

rata, sia, in misura meno netta, per quanto riguarda intensità e qualità timbrica. In quest'ultimo caso, le differenze riscontrate nella ricerca con le produzioni di soggetti adulti, sono spiegabili in termini di difficoltà, da parte dei bambini, a controllare una coordinazione gestuale complessa necessaria a raggiungere il target articolatorio, piuttosto che chiamare in causa una imperfetta distinzione fonologica.

La seconda parte del volume, intitolata *Strumenti e tecnologie per l'apprendimento e la didattica delle lingue*, raccoglie contributi descrittivi di una serie di tools e strumenti per l'apprendimento e per la didattica fonetica multimodale, di cui si propongono i risultati di sperimentazioni pilota o si descrivono le funzioni e le utilità metodologiche.

Il contributo di Cole e Cosi descrive la struttura e i componenti di *Mindstar Books*, una compagine di tutorial intelligenti per attività di apprendimento multimodale di contenuti scientifici. Il *toolkit* consiste di moduli orientati a diverse attività successive: apprendimento di un vocabolario concettuale di base, attività di *listening comprehension* assistita da un agente virtuale, esercitazioni di lettura a voce alta con autovalutazione e correzione delle sequenze. Il tutorial è già arrivato al pieno sviluppo per la lingua inglese e quella spagnola ed è in fase di prototipo e sperimentazione per l'italiano.

Lidia Calabrò presenta e discute *PhonetIC(T)s*, un metodo didattico-fonetico multimodale che integra movimenti corporei, apprendimento cooperativo e utilizzo di applicazioni *mobile* (*Apps*) gratuitamente scaricabili dai principali *markets* digitali; queste ultime invitano gli apprendenti a cimentarsi con una L2 attraverso il gioco e la ricerca in rete e a prendere consapevolezza dei propri livelli di competenza, fornendo contemporaneamente supporto al docente nella preparazione dei test di valutazione. Nel contributo al volume, l'utilità del metodo viene presentata attraverso una sperimentazione di insegnamento fonetico di italiano LS su studenti cinesi e viene testata tramite questionari di soddisfazione.

Un'altra sperimentazione in aula (ancora su apprendenti sinofoni di italiano) è quella riportata da Mairano e Calabrò, relativa all'utilizzo del tool *Minimal Pair Finder* (MPF) come ausilio all'apprendimento di contrasti fonologici: un 'classico' dell'insegnamento fonetico, il riconoscimento di 'coppie minime' di parole, viene rimodulato in un protocollo didattico articolato in tre tipi di attività successive che utilizzano uno strumento di facile implementazione da parte del docente e di semplice uso da parte del discente.

I successivi due saggi presentano, invece, due corpora di parlato sviluppati all'interno di progetti europei: un corpus di parlato infantile *Childit2* (riedizione del precedente Childit) progettato per l'implementazione di sistemi di riconoscimento automatico, nel contributo di Cosi, Paci, Sommavilla e Tesser; e il corpus DIRHA-ENGLISH una raccolta di registrazioni multi-microfoniche di parlato reale e in ambiente simulato, utilizzato nell'ambito della domotica, descritto da Ravanelli, Cristoforetti, Gretter, Pellin, Sosi e Omologo.

Infine, chiudono la sezione due lavori ad alto contenuto tecnologico e sperimentale. Nel primo Tesser, Sommavilla, Paci e Cosi discutono la realizzazione e

l'impiego di un metodo di creazione automatica di un test d'intellegibilità per la valutazione di un sistema di sintesi vocale da testo (TTS – *Text-to-speech synthesis*); nel secondo Manca, De Nunzio e Grimaldi propongono i risultati sperimentali di un metodo di classificazione di suoni linguistici basato su tecniche elettro-encefalografiche (EEG), messo a punto come base per lo sviluppo di un sistema di *Silent Speech Interfaces* (SSI).

Arriviamo, quindi, alla terza e ultima parte del volume dedicata a *Questioni di fonetica articolatoria, acustica e percettiva*, i cui contributi, sebbene non direttamente indirizzati agli aspetti dell'apprendimento linguistico, fanno luce su aspetti specifici e ancora poco studiati, in prospettiva socio-fonetica e anche clinico-fonetica.

Nel saggio a firma di Ciaurelli, Namasivayam, Tisato, van Lieshout e Zmarich si riportano dati articolatori raccolti tramite tecnica di articulografia elettromagnetica (EMA) relativi ai gesti vocali coinvolti nella produzione dei *glides* /j, w/ e [i̯]-[u̯]. dell'italiano, con l'obiettivo di contribuire ad una più approfondita e completa comprensione della natura di tali segmenti consonantici e della loro differenza contrastiva con le rispettive vocali omorgamiche.

Nodari fornisce, invece, un'analisi acustica molto dettagliata di alcuni parametri legati alla coarticolazione di consonante occlusiva aspirata e vocale successiva. Il dataset del lavoro è costituito da registrazioni di parlato della varietà di italiano calabrese (Lamezia Terme), nella quale tradizionalmente sono annoverate consonanti aspirate; l'originalità della ricerca consiste nel metodo di analisi che affianca alle classiche misurazioni del VOT, l'osservazione di altri due parametri, indici della modalità di fonazione della vocale (il quoziente di apertura o *Open Quotient* e lo *Spectral Tilt*).

Il contributo di Iraci, Grimaldi e Gili Fivela affronta su base sperimentale un tema fonetico ad alto impatto clinico: la realizzazione di contrasti fonologici nelle produzioni di pazienti affetti da Morbo di Parkinson. La patologia disartrica di questi pazienti limita, infatti, la loro capacità di controllo dei gesti articolatori, ma sembra, dai risultati riportati in questo studio (e supportati da precedente letteratura clinico-medica) che non abbia alcun effetto sulla categorizzazione fonologica; infatti, i soggetti affetti da Disartria Ipocinetica sono in grado di rendere tali distinzioni fonologiche attraverso strategie compensatorie. Il focus della ricerca è un esperimento di valutazione dell'intellegibilità e discriminazione di alcune opposizioni (legate al tratto di lunghezza consonantica) da parte di un gruppo di soggetti, con lo scopo di integrare tale metodologia nella procedura di diagnosi clinica delle disartrie.

Parallelamente, nel saggio di Zmarich, Bernardini, Lenoci, Natarelli e Pisciotta si propone l'uso della tecnica del *Disfluency Profile* come metodo diagnostico precoce e predittivo del rischio di persistenza di balbuzie in neonati e bambini entro i tre anni di età. I risultati di una sperimentazione in atto mostrano diversi vantaggi del metodo proposto (semplicità d'uso, riduzione dell'arbitrarietà diagnostica e maggiore accuratezza) rispetto ad altri comunemente utilizzati in ambito clinico.

Chiude il volume un lavoro di ambito squisitamente sociofonetico di Avesani, Vayra e Longo. Gli autori pongono la questione della percezione di allungamenti vocalici prosodici nella varietà regionale di parlato bolognese e ipotizzano l'esistenza di un fenomeno di *transfer* (per via lessicale) dalla corrispettiva varietà dialettale, in cui la quantità vocalica costituisce tratto fonologicamente pertinente: la ricerca, condotta attraverso tre diversi esperimenti di produzione e analisi delle durate vocaliche in posizioni sillabiche e prosodiche distinte, non fornisce risultati definitivi e lascia un alone persistente di mistero e strade aperte all'investigazione del fenomeno.

L'insieme dei contributi qui sopra brevemente illustrati nelle loro linee fondamentali rende, a nostro parere, la pubblicazione di questo volume un prezioso contributo alla riflessione scientifica della comunità dei fonetisti e dei linguisti italiani, fornendo un buon esempio della possibilità di coniugare la ricerca su temi altamente specialistici ad aspetti applicativi di impatto sociale, come la didattica linguistica e la diagnosi clinica. Il merito di questo valore va tutto agli autori per i loro sforzi e la loro collaborazione.

Le curatrici desiderano, infine, ringraziare singolarmente i membri del Comitato Scientifico del XII Convegno AISV del 2016, per la loro pazienza e generosa disponibilità alla revisione dei lavori:

Cinzia Avesani (ISTC-CNR, Padova), Elisabetta Bonvino (Università di Roma Tre), Maria Grazia Busà (Università di Padova), Silvia Calamai (Università di Siena), Chiara Celata (Scuola Normale Superiore di Pisa), Piero Cosi (ISTC-CNR, Padova), Lidia Costamagna (Università per Stranieri di Perugia), Claudia Crocco (Università di Ghent, Belgio), Franco Cutugno (Università di Napoli "Federico II"), Amedeo De Dominicis (Università La Tuscia), Anna De Meo (Università di Napoli L'Orientale), Mariapaola D'Imperio (Aix Marseille Université, Laboratoire Parole et Langage, CNRS, Francia), Vincenzo Galatà (Libera Università di Bolzano), Barbara Gili Fivela (Università di Lecce), Mirko Grimaldi (Università di Lecce), Joaquim Llisterri Boix (Universitat Autonoma de Barcelona, Spagna), Pietro Maturi (Università di Napoli "Federico II"), Antonio Origlia (Università di Napoli "Federico II"), Elisa Pellegrino (Università di Napoli L'Orientale), Massimo Pettorino (Università di Napoli L'Orientale), Antonio Romano (Università di Torino), Luciano Romito (Università della Calabria), Pier Luigi Salza (Socio onorario AISV, ex dirigente Loquendo), Carlo Schirru (Università di Sassari), Stefan Schmid (Università di Zurigo, Svizzera), Patrizia Sorianello (Università di Bari), Lorenzo Spreafico (Libera Università di Bolzano), Fabio Tamburini (Università di Bologna), Fabio Tesser (ISTC-CNR, Padova), Mario Vayra (Università di Bologna), Alessandro Vietti (Libera Università di Bolzano), Marilisa Vitale (Università di Napoli L'Orientale), Miriam Voghera (Università di Salerno), Claudio Zmarich (ISTC-CNR, Padova)

DOI: 10.17469/O2102AISV000001

GIULIANO MION

Un ricordo personale di Andrea Paoloni (1947-2015)

After an education in engineering, Andrea Paoloni specialized in phorensic phonetics and investigated several fields of this science like, in particular, speaker recognition, speaker characterization, phorensic transcription and intelligibility of the signals. The few lines proposed here intend to recall some aspects of the personality of our scholar who passed away in October 2015.

L'ultima volta che incontrai Andrea Paoloni fu nel suo studio della Fondazione Ugo Bordoni (FUB), in via del Policlinico a Roma, verso la fine di maggio 2015 quando, orgogliosamente, mi mostrò l'imponente Spagnolo (2015), un volume di oltre mille pagine che aveva da poco ricevuto sulla scrivania e che conteneva il lungo capitolo *La voce* da lui redatto.

I nostri rispettivi impegni professionali avevano fatto sì che dal nostro incontro precedente fosse trascorso molto tempo e, come spesso avviene dopo periodi di lontananza prolungata, nel momento in cui ci si rivede è possibile cogliere i mutamenti sfumati che il tempo attribuisce inesorabile a ciascuno. E fu così che ebbi la vaga impressione di percepire un lievissimo tremolio nella sua voce (che chissà come avrebbe saputo classificare scientificamente lui), ma mai avrei potuto pensare che di lì a qualche mese (ottobre 2015) Andrea Paoloni ci avrebbe improvvisamente lasciati.

Di formazione ingegneristica, Andrea Paoloni ha sempre avuto un rapporto significativo con "gli umanisti" (così amava definirli) perché, dopo la laurea in Ingegneria Elettronica alla Sapienza di Roma nel 1973, ben presto sarebbe approdato a ricerche sul linguaggio e, in particolare, sulla fonetica. In Fondazione Bordoni sin dal 1974, lì si occuperà di analisi del segnale vocale, riconoscimento del parlato, caratterizzazione e riconoscimento del parlante. Nel 1976, sempre in Fondazione, avrebbe poi ottenuto la qualifica di ricercatore senior e quindi di responsabile dei progetti sul "Trattamento Automatico della Lingua". Avrà da quel momento in poi collaborazioni continue con numerose istituzioni di ricerca, come l'Università La Sapienza, la Tuscia di Viterbo, Roma Tre, il CNR, e altre ancora.

Il nostro primo incontro risaliva forse all'inizio del 2006, o forse anche prima, verso la fine del 2005, in un convegno di linguistica che si teneva a Roma. Un incontro che ebbe un seguito, qualche giorno dopo, nel suo studio in Fondazione che, all'epoca, aveva sede in via Baldassarre Castiglione, quartiere Montagnola. Un incontro senza dubbio singolare per entrambi: lui abituato a rapportarsi con linguisti e ingegneri, chi scrive queste righe proveniente invece da una formazione perlopiù

orientalistica e, in particolare, semitistico-arabistica. Desideroso di trovare un modo di far interagire la fonetica acustica alla dialettologia araba, il mio interesse di base, concepii una ricerca dottorale che sarebbe confluita in una tesi sul riconoscimento del parlante in arabo marocchino della quale Andrea Paoloni fu cotutore.

In quegli anni, infatti, insegnava "Fisica acustica" e "Fisiologia vocale e auditiva" presso la Facoltà di Conservazione dei Beni Culturali dell'Università della Tuscia di Viterbo e, soprattutto, era componente del Collegio del Dottorato in "Linguistica storica e Storia linguistica italiana" della Sapienza di Roma, nel quale rimase – a quanto mi risulta – fino all'ultimo. Il dottorato era articolato in diversi *curricula* e vedeva il suo contributo sostanziale in quello di "Fonetica acustica" al quale, fra l'altro, la Fondazione Bordoni destinava, per quasi ogni ciclo dottorale, una borsa per progetti di ricerca in fonetica sperimentale e trattamento automatico del linguaggio.

Da quel momento in poi, ciascuno dei due ebbe modo di scambiare con l'altro le proprie competenze: lui insegnava all'arabista i segreti della fonetica forense, l'arabista insegnava all'ingegnere in cosa consistesse la lingua araba. Lui, in particolare, avrebbe continuato a confessarmi nel tempo come una realtà sociolinguistica complessa come quella diglottica araba gli rimanesse sempre sfuggente ma al contempo affascinante. E quando, alla fine dell'anno 2007, gli feci dono di un mio profilo descrittivo dell'arabo appena pubblicato, mi confidò in seguito di aver trascorso le feste natalizie accompagnato dalla sua "piacevole lettura", che per me non poteva rappresentare migliore complimento.

Andrea Paoloni è stato fra gli ideatori di IDEM (Voice Identification Method), un sistema semiautomatico realizzato dalla Fondazione Bordoni volto all'analisi del segnale vocale e al riconoscimento del parlante. Da qualcuno, infatti, il nostro ingegnere è stato definito come patriarca della fonetica forense italiana, una definizione postuma che non sono in grado di valutare se il suo carattere riservato, a tratti apparentemente schivo, gli avrebbe fatto o meno apprezzare ma che – occorre riconoscere – al contempo non si discosta dalla realtà.

Ha prestato la sua collaborazione professionale in centinaia di casi giudiziari in cui si rendesse necessaria una consulenza di fonetica forense, in particolare per l'analisi di nastri e registrazioni, la trascrizione di segnali rumorosi e il riconoscimento del parlante. Quando, in un caso particolarmente difficile, si disponeva della registrazione (spesso un'intercettazione telefonica o ambientale) di una voce anonima *X* e occorreva verificare se questa fosse attribuibile o meno all'indiziato *Y*, il suo era uno dei nomi più quotati e rappresentava una garanzia di successo. In Italia, non c'è stato caso giudiziario "scottante", anche tra quelli che hanno avuto un risalto mediatico imponente, che non abbia visto la sua presenza in qualità di consulente: dalle stragi di mafia ai sequestri di persona, dal traffico di droga all'omicidio politico, dalle inchieste sulla corruzione fino alla cronaca nera destinata a occupare per giorni i rotocalchi televisivi.

Durante e dopo il periodo di formazione dottorale, ho avuto la fortuna di affiancarlo in diversi casi di analisi forense riconducibili a numerose regioni d'Italia: Calabria, Campania, Puglia, Sardegna, Sicilia, e così via, in una geografia forense

che andava di pari passo con la geolinguistica dell'italofonia, in cui vocalismi e consonantismi erano di volta in volta vivisezionati con i criteri rigorosi della fonetica sperimentale.

L'applicabilità alla lingua araba (o, meglio, alle varietà dell'arabo) dei sistemi di riconoscimento che la Fondazione Bordoni aveva ideato per l'italiano era uno degli obiettivi che ci eravamo ripromessi più di una volta di perseguire.

Fu così, ad esempio, che collaborai alla preparazione di un suo esperimento registrando un breve enunciato in arabo giordano che ha poi utilizzato per uno studio sull'intelligibilità dei segnali rumorosi, un altro dei campi d'indagine che Andrea Paoloni ha praticato maggiormente, proprio per via dei risvolti applicativi legati alla trascrizione dei segnali rumorosi in contesto forense. Quello studio sarebbe confluito in una sua relazione al III convegno dell'Associazione Italiana Scienze della Voce, tenutosi a Trento nel 2006 e pubblicata in seguito nei relativi atti (Paoloni, 2008). Di quell'esperimento ridemmo spesso insieme: una brevissima quanto innocente frase in arabo, opportunamente coperta di rumore rosa, filtrata in banda telefonica e con riverbero, era stata trascritta da un gruppo di ascolto secondo le interpretazioni più svariate e fantasiose, tutte peraltro in lingua italiana e senza la benché minima ombra di dubbio che potesse trattarsi di una qualsivoglia lingua straniera.

Numerosi erano i suoi progetti scientifici, come del resto numerose le sue pubblicazioni scientifiche. Fra queste, va senz'altro segnalata Paoloni, Zavattaro (2007), una monografia scritta a quattro mani con Davide Zavattaro, ufficiale dell'Arma dei Carabinieri, che inaugurava una collana editoriale dedicata alle scienze forensi.

Sempre elegantissimo nell'aspetto e nei modi, per via della sua precisione meticolosa e della sua ferrea fiducia nella scienza (era anche lettore abituale della rivista edita dal Cicap – Comitato Italiano per il Controllo delle Affermazioni sul Paranormale), agli occhi di "un umanista" Andrea Paoloni al primo impatto non poteva non sembrare "un ingegnere" (definizione che finirà per stridere, apparentemente, con quella di "umanisti" che amava tanto adoperare). Eppure, chi aveva l'opportunità di conoscerlo in maniera più che superficiale finiva poi per scoprirne un *link* inatteso con le scienze umane: bibliofilo convinto, grande lettore e ottimo conoscitore delle letterature classiche.

Sempre lieto di ritrovare coloro che, a vario titolo, l'avevano conosciuto durante la propria formazione scientifica (i suoi "allievi"), era ben contento di scoprirne gli sviluppi professionali.

Personalmente, feci in tempo a scrivergli un breve messaggio in cui gli comunicavo alcune felici novità di natura personale. A quel messaggio tuttavia non seguì risposta e deduco – a posteriori – perché già preso da problemi molto seri.

Nella sua amata Fondazione Bordoni si recò fino all'ultimo, fino a qualche giorno prima della scomparsa.

Riferimenti bibliografici

Paoloni, A. (2008). Limiti della trascrizione giudiziaria. In Giordani, V., Bruseghini, V. & Cosi, P. (Eds.), *AISV 2006. Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche*. Torriana: EDK.

Paoloni, A., Zavattaro, D. (2007). *Intercettazioni telefoniche e ambientali. Metodi, limiti e sviluppo nella trascrizione e verbalizzazione*. Torino: Centro Scientifico Editore.

Spagnolo, R. (Ed.) (2015). *Acustica. Fondamenti e applicazioni*. Torino: Utet.

PARTE I

ASPETTI FONETICI E PROSODICI NELL'APPRENDIMENTO DI L1 E DI L2

Aspetti fonetici

DOI: 10.17469/O2102AISV000002

LUCIANO ROMITO, MANUELA FRONTERA, MARIA ASSUNTA CIARDULLO

Studio acustico-percettivo di contrasti fonemici in Italiano L2/LS

The aim of this paper is to show how Galician learners of Italian L2/FL, perceive and produce specific Italian segmental oppositions by taking into account similarities and discrepancies between the two linguistic systems (Italian vs. Galician) and the role played by two sociolinguistic variables such as L1 use (Galician vs. Castilian) and the learner's amount of L2 experience. The oppositions examined here deal with vowels (final unstressed /e – i/, /o – u/, an opposition that in Galician is reduced to /e – o/) and consonants (/b – v/, /ʦ – ʣ/, both absent in Galician). The goals are: (a) to verify the extent to which Italian vowel and consonant oppositions are identified by Galician learners of Italian L2/FL; (b) to analyse discrepancies in the way the oppositions are produced by Italian native speakers and learners of Italian L2/FL. Despite good percentages of correct perceptual identifications, speech analyses results show a considerable impact of L1 specific processes on L2 production.

Key words: Italian L2/LS, Phonetics, Phonology, perception, production.

Introduzione

L'acquisizione del sistema fonetico e fonologico di una lingua seconda/straniera in età adulta è ad oggi considerato il processo più difficoltoso per un apprendente. Sebbene un apprendente adulto possa raggiungere facilmente e in breve tempo un'ottima competenza linguistica su tutti i livelli, è altamente improbabile (seppur possibile) che la sua pronuncia diventi indistinguibile rispetto a quella dei parlanti nativi. La *ri-sintonizzazione* del sistema nativo su nuove categorie e nuovi suoni è un processo che richiede particolare sforzo[1], e il repertorio della lingua materna interferisce in modo inevitabile assurgendo al ruolo di base di partenza, controllo e confronto tanto nei processi di percezione che di produzione linguistica. Nei successivi paragrafi si cercherà *in primis* di offrire un quadro sintetico dei principali modelli e teorie sulla percezione linguistica in L2, alla luce del dibattito riguardante il rapporto diretto fra produzione e percezione; si propone in seguito una rassegna sulle variabili linguistiche e socioculturali maggiormente indagate nella ricerca sull'acquisizione fonologica in L2, a cui segue un breve stato dell'arte degli studi più

[1] Pur non facendo espressamente riferimento a fattori legati all'acquisizione linguistica, bensì alla permeabilità dei livelli di un sistema linguistico, è indicativo chiamare in causa la *scala di potere del parlante* formulata da Romaine (1984), nella quale la fonetica si inserisce all'ultimo gradino dei livelli linguistici di cui il parlante ha maggiore controllo e maggiore consapevolezza (pragmatica>semantica >sintassi>fonologia>fonetica).

recenti relativi alla percezione e la produzione linguistica di suoni dell'italiano L2/LS. Seguono i materiali e i metodi impiegati nell'esperimento qui proposto, l'analisi dei dati e i risultati ottenuti.

1. *Percezione e produzione dei suoni di una lingua straniera*

La relazione esistente tra percezione e produzione di suoni in una lingua seconda o straniera è, ed è stato, un fattore profondamente dibattuto e indagato sia sul piano prettamente fonologico e fonetico, sia sul versante della linguistica acquisizionale e della psicolinguistica.

Ciò che si è cercato *in primis* di comprendere, è quale dei due processi abbia maggiore influenza sull'altro: considerando fattori strettamente fisici, l'attività motoria legata all'articolazione del suono, come ogni altra attività anatomica, è soggetta al trascorrere del tempo e risente di un'alterazione nella plasticità correlata al normale processo di maturazione neurale, ragion per cui apprendenti tardivi (adolescenti e adulti) di una L2 faticherebbero tanto a raggiungere una pronuncia vicina a quella nativa rispetto ad apprendenti precoci; l'ipotesi dell'esistenza di un *periodo critico* nella fase di produzione, prettamente correlata a cause fisiologiche, spiegherebbe da sé la subordinazione dei processi percettivi e di decodifica del segnale a quelli strettamente legati alla produzione[2].

D'altro canto, è stato ampiamente dimostrato come i limiti nella percezione e produzione di suoni in una L2 sorgano dall'esperienza pregressa legata alla L1, più che da una perdita di plasticità, e siano frutto di errori di attribuzione dei suoni target a quelli noti (Iverson, Kuhl, 1995; Kuhl, 2000; Flege, 2003). Certamente, la L1 di riferimento esercita la funzione di filtro fonologico sulla percezione dei suoni in lingua straniera, così che una percezione fortemente plasmata dal sistema di provenienza può indurre a fraintendimenti e a una produzione inappropriata o imprecisa (Trubeckoj, 1939); ma è anche vero che non tutti i suoni che vengono percepiti e *categorizzati* correttamente, sono poi prodotti in modo altrettanto corretto (Llisterri, 1995). Questo può dipendere da molteplici fattori, sia interni che esterni al parlatore, primo fra tutti dal peso che i diversi correlati acustici svolgono nella discriminazione degli stessi suoni a livello interlinguistico (Lisker, Abramson, 1970; Bohn, Flege, 1990; Llisterri, 1995), e che vengono calibrati in modo differente anche nella fase di produzione del suono (Iverson, Kuhl, Akahane-Yamadac, Dieschd, Tohkurae, Kettermannf & Siebert, 2003; Escudero, 2009). Tali differenze evidenti dal punto di vista fonetico, non possono prescindere da caratteristiche intrinseche a ciascuna delle lingue considerate, dalla loro vicinanza in termini tipologici, ai livelli di marcatezza dei relativi sistemi[3] (Major, 2001; Flege, 2003; Best, Tyler, 2007). La percezione linguistica, soprattutto in età adulta, è un processo altamente linguo-

[2] Per un resoconto, su questa corrente di studi, si rimanda a Flege (2003).

[3] Nel suo *Ontogeny Phylogeny Model* (OPM), Major affronta la relazione esistente fra fenomeni di transfer, universali linguistici e somiglianza fra L1 ed L2 nel processo di acquisizione linguistica: la somiglianza fra lingue e la marcatezza sono i fattori che più rallentano il percorso di acquisizione, ma

specifico (Kuhl, 2000; Escudero, 2009), di conseguenza la chiave interpretativa fornita dalla L1 di provenienza può aiutare o svantaggiare, in termini di somiglianza/discrepanza rispetto alla lingua *target*, nel processo di percezione linguistica e categorizzazione dei suoni L2. È necessario innanzi tutto tener conto di una dicotomia imprescindibile, quella tra suoni 'nuovi' e suoni 'simili'[4]. I primi, non riconducibili a nessuna categoria della L1 e molto distanti/divergenti da esse, vengono notati più facilmente e acquisiti con più rapidità, per cui attivano la creazione di un nuovo spazio categoriale, adibito a inglobare suoni non nativi di una lingua seconda[5]; tale differenziazione può avvenire sulla base di dati acustici-fonetici (Flege, 1987; 1995) o attraverso l'individuazione di parametri stabili[6] legati ai gesti articolatori impiegati nella realizzazione del suono (Best, 1995; Best, Hallé, Bohn & Faber, 2003; Best, Tyler, 2007). Le complicazioni riguardano l'atteggiamento percettivo assunto rispetto a suoni considerati 'simili' fra L1 e L2, o condivisi dai due sistemi, per i quali è possibile, per lo meno inizialmente, una totale assimilazione alle categorie fonologiche e/o fonetiche note[7] (Flege, 1987; Best, Tyler, 2007; Vayra, Avesani, Best & Bohn, 2012), quindi un approccio relativamente semplice, o si ammette l'insorgere di difficoltà nella differenziazione, che causerà nell'apprendente un riadattamento della mappatura percettiva di L1 (Escudero, 2009), quindi una fusione tra categorie simili (*merger hypothesis*, Flege, 1987). In altre parole, maggiore è la distanza tra suono L2 e *prototipo* della categoria L1 costruita per quel suono, maggiore la sensibilità percettiva (Lacerda, 1995; Iverson, Kuhl, 1995); parallelamente, la sensibilità si affievolisce con l'approssimazione a suoni prototipici della L1, così come tra opposizioni della L2 con realizzazioni tra esse molto prossime (Major, 2001; Best, Tyler, 2007). In fase di produzione, questo può chiaramente condurre ad una sostituzione sistematica dei suoni di L2 con quelli di L1 più simili a cui sono stati assimilati, o che vengono realizzati nello stesso spazio fonologico e una realizzazione più o meno accurata di suoni correttamente identificati come nuovi.

In sintesi, i suddetti modelli presuppongono che, sebbene i livelli di accuratezza nella produzione (quindi nella pronuncia) non potranno mai raggiungere completamente quelli dei parlanti nativi della stessa L2, i processi di categorizzazione, as-

mentre il transfer opera prevalentemente su fenomeni simili, sulla marcatezza agiscono con più forza i principi relativi agli universali linguistici.

[4] Così come li differenzia Flege nel proprio modello di riferimento, SLM (*Speech Learning Model*).

[5] Questo principio è riconducibile alla *Similarity Differential Rate Hypothesis* (SDRH), che scardina uno dei capisaldi legati alle teorie sull'Analisi Contrastiva secondo cui, sulla base di un confronto fra sistemi fonologici, suoni simili fra L1 ed L2/LS verrebbero acquisiti più rapidamente per effetto di *transfer positivo*.

[6] Quelli che Best, nel modello percettivo PAM (*Perceptual Assimilation Model*), definisce *invariants*.

[7] È il fenomeno che Flege (1987; 1995) identifica come *equivalence classification* e che, se non superato mediante un progressivo *attunement* alla nuova lingua, è causa di un blocco nella categorizzazione e di una fase di stallo nel percorso di miglioramento produttivo e di perdita dell'accento straniero (finché suoni della L2 verranno assimilati a categorie equivalenti della L1, saranno sistematicamente sostituiti a questi e prodotti come tali; si veda anche la discussione fornita al riguardo da Vayra et al. 2012).

similazione o attivazione percettiva siano alla base della competenza linguistica di ciascun apprendente, quindi indispensabili e propedeutici alla fase di produzione.

1.1 Variabili correlate ai processi di percezione e produzione di suoni in L2/LS

Chiaramente, l'analisi linguistica di tali fenomeni non può prescindere dal considerare l'effetto di fattori sociali, culturali ed emotivi nel contatto fra lingue. Piske, Mackay & Flege (2001) identificano alcune variabili preminentemente influenti nelle fasi di acquisizione[8] di una L2/LS, che possono influire, di conseguenza, sul grado di accento straniero prodotto e percepito:

- *età* di primo contatto con la lingua straniera;
- *durata della permanenza*, inteso come il periodo di permanenza nel paese di cui la L2 o LS sia la lingua ufficiale;
- *sesso*;
- *istruzione formale*, inteso come il periodo di studio dedicato all'apprendimento della lingua seconda o straniera di riferimento;
- *motivazione*, valutata come la necessità di acquisire una buona pronuncia con fine integrativo e/o strumentale e professionale, nonché come il desiderio e la volontà intrinseche di imparare la nuova lingua;
- *attitudine*, intesa come l'atteggiamento e la predisposizione individuale all'acquisizione/apprendimento di una lingua straniera;
- *uso della lingua,* ovvero la valutazione in termini quantitativi e qualitativi dell'uso della lingua materna rispetto alla L2 o LS, soprattutto in contesto estero.

Le variabili esposte possono essere adattate tanto in un contesto di acquisizione/apprendimento di una lingua seconda, come di una lingua straniera; tuttavia alcune di esse appaiono maggiormente calzanti in alcuni casi. La *durata della permanenza* ad esempio fa espressamente riferimento al periodo di tempo trascorso all'estero e al contatto con la nuova lingua, consentendo di differenziare apprendenti non esperti (con periodi di permanenza inferiori ai 6 mesi) da apprendenti esperti (con permanenza dai 6-12 mesi in poi); la categoria di apprendenti non esperti può includere parlanti che stiano apprendendo la lingua target come lingua straniera e che abbiano trascorso brevi periodi di tempo (per motivi di studio o di lavoro) all'estero, pur senza implicare un trasferimento permanente, che renda la nuova lingua una vera e propria lingua seconda[9]. Lo stesso vale per la variabile *uso della lingua*, la quale nel caso di apprendenti guidati che utilizzano la lingua straniera in contesto formale (accademico o scolastico), può rappresentare un indice di variabilità poco significativo. È chiaro che fra queste variabili alcune hanno un peso specifico maggiore di altre, *in primis* l'età di prima esposizione alla L2 che di per sé può rappresentare

[8] Naturalmente, si tratta di fattori influenti su tutti i livelli linguistici nel processo generale di acquisizione.

[9] Classificazione tratta da Best, Tyler (2007).

un fattore significativo e determinante; le altre acquisiscono più valore agendo in concomitanza a fattori più forti, qualificandoli di nuove sfumature.

A questi si aggiungano ulteriori fattori, sulla base di quanto esposto in precedenza (§ 1.), ovvero:

– la vicinanza/distanza in termini strutturali fra la L1 e la L2;
– le differenze e somiglianze tra gli inventari fonetici e fonologici delle due lingue;
– quindi, la condivisione o meno di determinate categorie di suoni.

Molti studi sulle lingue seconde tendono a prendere in considerazione ulteriori variabili sia 'esterne', come l'*input* (in termini di qualità e quantità degli stimoli offerti ai discenti), i livelli di *interazione* e *socializzazione* (chiusura vs. coesione, congruenza culturale vs. integrazione), sia 'interne' o affettive, ovvero *ansietà* degli apprendenti (apprensione comunicativa, ansietà sociale, autostima) e *personalità* (inibizione, estroversione, empatia)[10].

Nel presente studio verrà considerata un'ulteriore variabile, ovvero l'esperienza di alcuni soggetti in studi linguistici e fonetici, che chiaramente può alterare il comportamento percettivo e produttivo degli informatori coinvolti, favorendo la differenziazione di suoni nuovi in opposizioni anche molto simili, e un atteggiamento di riflessione *propriocettiva* al momento della produzione.

1.2 Percezione e produzione di suoni dell'italiano L2/LS: stato dell'arte

L'interesse nello sperimentare il comportamento percettivo e produttivo, a livello segmentale, di apprendenti di lingua italiana è giustificato dalla presenza di relativamente poche ricerche specifiche in questo ambito. Esiste, ad esempio, una copiosissima bibliografia riguardante lo studio percettivo e produttivo di segmenti vocalici e consonantici della lingua inglese (si vedano a titolo esemplificativo Aoyama, Flege, 2011; Flege, Schmidt, 1995), che considerano le variabili citate sinora, su apprendenti di madrelingua diversa, incluso l'italiano (Flege, MacKay & Meador, 1999; Flege, Mackay, 2004), ma, al confronto, pochi studi recenti di questo tipo in cui la lingua *target* sia appunto l'italiano.

I maggiori interessi di ricerca (per lo meno a conoscenza degli autori) sono ad esempio indirizzati verso aspetti prosodico/intonativi dell'italiano come lingua seconda – percezione dell'accento lessicale, come in Alfano, Llisterri & Savy (2007); aspetti prosodici e ritmici nell'acquisizione dell'intonazione italiana in De Meo, Pettorino (2012); valutazione dell'accento straniero su base prosodico-intonativa, attraverso compiti di imitazione e autoimitazione svolti da apprendenti di madrelingua cinese (De Meo, Vitale & Pellegrino, 2016) e giapponese (Pellegrino, Vigliano, 2015); le produzioni di apprendenti sinofoni sono analizzate a livello ritmico-prosodico, con relativa valutazione del grado di accento straniero percepito da nativi italiani, anche in Pettorino, De Meo, Pellegrino, Salvati & Vitale (2011); Pellegrino (2012) presenta ancora un'analisi segmentale – in termini di

[10] Si vedano Pallotti (2003) e Larsen-Freeman, Long (2014).

durate vocaliche e sillabiche – e prosodica di soggetti sinofoni; da un punto di vista strettamente segmentale, Costamagna (2007) propone uno studio longitudinale sull'acquisizione delle quattro affricate dell'italiano L2 appreso da studenti brasiliani; Celata e Costamagna (2012) analizzano il *timing* di consonanti geminate prodotte da estoni apprendenti di italiano L2; il confronto tra *timing* di consonanti geminate e scempie è trattato ancora da Kabak, Reckziegel & Braun (2011), nella produzione di parlanti di madrelingua tedesca, apprendenti di italiano come lingua seconda; Pape e Jesus (2014) conducono invece uno studio percettivo e produttivo sulla desonorizzazione delle occlusive velari nel portoghese e nell'italiano. Mori (2007) presenta un'analisi approfondita dell'interlingua di marocchini apprendenti di italiano, concentrandosi sulla variazione nella produzione di segmenti vocalici e consonantici e la valutazione della salienza percettiva delle marche consonantiche che contribuiscono all'identificazione dell'accento straniero.

1.2.1 Obiettivi del presente lavoro

L'esperimento proposto ha voluto testare il comportamento percettivo e produttivo di studenti di madrelingua galiziana apprendenti di italiano come lingua straniera. Sono state analizzate opposizioni vocaliche e consonantiche, di suoni dell'italiano sia nativi che non nativi, simili e nuovi.

Tra le vocali, si è scelto di esaminare l'opposizione fra le anteriori alte e semi-alte /i e/ e le corrispondenti posteriori /u o/ tutte in contesto atono finale di parole e non-parole[11] piane bisillabiche e trisillabiche.

Si è deciso di concentrarsi sulle differenze riscontrate nel sistema vocalico atono, poiché quello galiziano, rispetto all'italiano, presenta notevoli peculiarità (si confronti la Tabella 1). Le vocali atone in posizione post-tonica finale si riducono a tre /e a o/, neutralizzando l'opposizione tra vocali alte e medio-alte sia in anteriorità che in posteriorità.

Tabella 1 - *Sistemi vocalici del galiziano e dell'italiano*

SISTEMA VOCALICO TONICO		SISTEMA VOCALICO ATONO	
GALIZIANO	ITALIANO	GALIZIANO	ITALIANO
/i, e, ɛ, a, ɔ, o, u/	/i, e, ɛ, a, ɔ, o, u/	Posizione Pretonica /i, e, ɛ, a, ɔ, o, u/	/i, e, a, o, u/
–	–	Posizione postonica centrale /i, e, a, o, u/	/i, e, a, o, u/
–	–	Posizione postonica finale /e, a, o/	/i, e, a, o, u/

[11] La scelta delle non-parole mira a far sì che gli informatori facciano affidamento esclusivamente su informazioni acustico/fonetiche e non lessicali (Mora, 2008).

Ulteriore aspetto tipico del galiziano è che /e o/ in posizione postonica finale vengono realizzare rispettivamente come [ɪ] e [ʊ] (Martínez Celdrán, 2002; Regueira, 2007) da cui l'interesse nell'indagare la produzione degli stessi suoni in italiano LS.

Le due opposizioni consonantiche considerate nell'esperimento sono invece /b v/ e /ʦ ʤ/. La prima coppia in opposizione include un fonema condiviso dai due sistemi di riferimento (galiziano e italiano), /b/, sebbene in galiziano esso presenti due possibili realizzazioni allofoniche assenti nella lingua italiana (cfr. Tabella 2), le quali realizzano indistintamente i grafemi *b* e *v*. L'occlusiva bilabiale sonora viene analizzata in opposizione al suono fricativo /v/, assente nel sistema galiziano.

Tabella 2 - *Varianti galiziane del fonema /b/*

/b/	[b]	PRINCIPIO ASSOLUTO	*vago* [ˈbaɣo]
		POST-NASALE	*un vaso* [ˈum ˈbaso]
		POSIZIONE IMPLOSIVA[12]	*obstruir* [obˈstrwir]
	[β̞]	ALTRI CONTESTI	*aba* [ˈaβ̞a]
		(INTERVOCALICA, POST-LIQUIDA,	*el vaso* [ˈelˈβ̞aso]
		DOPO [s])	*esvarar* [ezβ̞aˈrar]

L'opposizione consonantica tra le affricate alveolari dell'italiano comprende due suoni assenti dal sistema galiziano. Si è scelto di prendere in esame questi suoni come esempio lampante di opposizione in lingua seconda o straniera di suoni estremamente simili e *vicini* fra essi, entrambi *nuovi* e fortemente marcati. Si tratta inoltre di un'opposizione dell'italiano non prevedibile[13].

Anche le opposizioni consonantiche riguardano parole e non-parole; i segmenti oggetto di indagine sono sempre in attacco di sillaba tonica aperta con struttura CV, e inseriti in ossitoni, parossitoni e proparossitoni foneticamente bilanciati. L'opposizione /b v/ è esaminata sempre in contesto intervocalico.

Le parole contenenti i segmenti in opposizione sono tutte inserite in frasi cornice ("Dico X rapidamente/con calma").

1.2.2 Ipotesi

Le ipotesi legate al comportamento percettivo e produttivo degli informatori galiziani si fondano sulle teorie e i modelli proposti in *incipit*. È da premettere che, trattandosi di un confronto tra sistemi linguistici molto vicini (entrambi di derivazione latina), il processo di acquisizione fonologica è per certi versi avvantaggiato, date inoltre le numerose categorie condivise tra galiziano e italiano e la presenza di

[12] Dei suoni occlusivi in coda sillabica.

[13] In italiano non esiste una regola fonologica che indichi la distribuzione delle affricate alveolari. Escludendo le coppie minime esistenti, la distribuzione dei due suoni è, come anticipato, imprevedibile. Sono riconducibili tuttavia a varietà regionali per cui, ad esempio, al sud la sorda /ʦ/ sonorizza in contesto intervocalico (*polizia* [poliˈ ʤiːa]) o post-laterale (*alzo* [ˈalʤo]), mentre in inizio di parola è frequente la realizzazione sonora nelle varietà del nord, sorda in quelle del sud ([ˈʤiːo] vs. [ˈʦiːo]) (Mioni, 2001; Gili Fivela, 2010).

molti suoni simili. È prevedibile che suoni simili vengano assimilati alle categorie equivalenti della L1 e che questo si rifletta in una produzione impropria; nel caso in oggetto, ciò potrebbe verificarsi tanto nelle opposizioni vocaliche come in quelle consonantiche[14].

Le vocali atone finali medio-alte, seppure identificate correttamente, potrebbero essere realizzate con la centralizzazione tipica del galiziano.

Il caso delle consonanti si complica: il suono fricativo /v/, ad esempio, è per il galiziano un suono marcato, corrispettivo sonoro della fricativa labiodentale sorda /f/, condivisa dai due sistemi. Allo stesso tempo però, il galiziano presenta la variante approssimante [β̞] del fonema occlusivo bilabiale sonoro, identificata anche come suono *fricativo approssimante*[15], di realizzazione molto prossima alla fricativa; ciò può compromettere la capacità di identificare e riprodurre tale suono tenendo conto, in aggiunta, del fatto che il galiziano non discrimina le grafie *b* e *v*, bensì, come suddetto, le realizza negli stessi suoni. È prevedibile quindi che i suoni oggetto di questa prima opposizione consonantica vengano identificati con difficoltà (soprattutto il suono fricativo) e che, in contesto intervocalico in fase di produzione, tanto il suono occlusivo come il fricativo vengano sostituiti dalla approssimante/fricativa approssimante, ossia dalla variante fonetica applicata dal galiziano in quello specifico contesto.

I suoni affricati in opposizione sono i suoni della lingua italiana più marcati rispetto al galiziano (e al castigliano), quindi potenzialmente di più semplice acquisizione. Essendo discriminati dal solo tratto della sonorità, la loro identificazione potrà risultare complessa al pari della produzione, escludendo con buona probabilità i casi di soggetti maggiormente esperti in linguistica e/o fonetica, casi che nella analisi saranno trattati singolarmente.

2. *Metodi*

2.1 Esperimento 1: percezione

2.1.1 Soggetti

I test percettivi sono somministrati a: 1) 23 madrelingua galiziani, apprendenti di italiano L2 (12 con istruzione formale di 1-2 anni, 11 con istruzione formale di 3-4 anni), tutti studenti dell'Università di Santiago de Compostela (Galizia, Spagna) di età compresa fra 18 e 28 anni; 2) un gruppo di controllo costituito da 7 madrelingua italiani fra i 25 e i 43 anni di età.

[14] Per una analisi preliminare dei risultati relativi alla percezione e produzione dei suoni vocalici, si vedano Frontera (*in stampa*) e Romito, Frontera (2015).

[15] In Navarro, Quilis Merín (2012), i suoni [ß] e [ɣ] vengono presentati come fricative approssimanti dello spagnolo e a livello acustico identificate come approssimanti o fricative a bassa risonanza.

2.1.2 Stimoli

Le prove percettive sono state espletate mediante due test di *identificazione*[16], uno inerente le opposizioni vocaliche, uno quelle consonantiche. Le parole *target*, contenenti i segmenti in opposizione indagati, sono state estrapolate dalle registrazioni effettuate da 3 parlanti madrelingua italiani (si rimanda a § 2.2.2 e § 2.2.1). Le parole sono state isolate tramite il software di elaborazione del segnale Sound Forge 7. Ciascun esperimento è stato costruito e somministrato attraverso il software Folerpa[17] (Fernández Rei, 2014) e dettagliato selezionando 9 parole *target* per ciascun segmento oggetto di indagine, ciascuna ripetuta per 5 volte (45 stimoli per segmento, 90 stimoli per opposizione vocalica e consonantica, 180 stimoli totali suddivisi e randomizzati in 4 blocchi, sia per l'esperimento condotto su opposizioni vocaliche che per quello inerente le opposizioni consonantiche). Ogni stimolo sonoro è ripetuto due volte con intervallo di 1 secondo. Il compito degli informatori è quello di selezionare la parola associata allo stimolo proposto, scegliendo tra due opzioni; vengono calcolati i tempi di risposta. I test sono stati somministrati *on-line*, mediante l'invio degli URL di riferimento per posta elettronica, e preceduti da una nota introduttiva in cui si spiega ed esemplifica lo svolgimento delle prove[18]. È stato vivamente consigliato di eseguire il test mediante l'uso di auricolari o cuffie e in un ambiente silenzioso.

2.1.3 Analisi

I dati, collezionati dallo stesso software, sono stati elaborati tramite Folerpa ed Excel, le analisi statistiche eseguite mediante il software SPSS. Le prime analisi condotte sugli esiti dei test percettivi sono di tipo descrittivo e fanno riferimento alle percentuali di identificazioni corrette ed errate per ciascuna categoria, comparate ai dati ottenuti dal gruppo di controllo. Le analisi statistiche preliminari, condotte tramite test chi-quadrato di Pearson, mirano a verificare la correlazione di quattro

[16] La metodologia prevalentemente associata alla creazione e somministrazione di prove percettive consiste, generalmente, in compiti di identificazione e discriminazione categoriale. In alcune discussioni incentrate su questioni prettamente metodologiche (Repp, 1984; Mora, 2008) si sostiene che le prove siano entrambe funzionali alla verifica della classificazione categoriale tra stimoli, ma mentre l'identificazione risponde ad una prova di categorizzazione dello stimolo, inserito in una specifica categoria – *labeling* –, la discriminazione è maggiormente utile per giudicare l'accuratezza nella distinzione tra stimoli. Nel caso in oggetto si è scelto di utilizzare test di identificazione tanto per le opposizioni vocaliche quanto per quelle consonantiche, allo scopo di fornire una predizione del livello di difficoltà percettiva dei segmenti in questione. Nelle future fasi di sperimentazione, i test di identificazione verranno integrati con test di discriminazione categoriale ABX, soprattutto nell'opposizione fra affricate sorde e sonore, per le quali, trattandosi di variazioni intra-fonemiche, un test di discriminazione risulta più appropriato a valutare l'accuratezza con cui sono distinte le due varianti.

[17] http://ilg.usc.es/FOLERPA.

[18] Nella sezione della presentazione introduttiva riguardante l'opposizione fra affricate, i segmenti sono stati identificati come *'z' dura* (sorda) e *'z' dolce* (sonora) e accompagnati da un file audio rappresentativo.

specifiche variabili con i dati ottenuti dai *task* di identificazione. Le variabili considerate sono state:

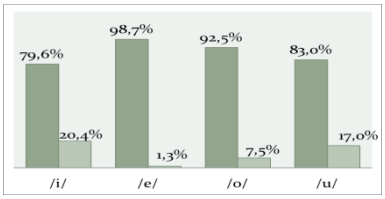
– L1 degli informatori – nonostante gli studenti siano tutti di provenienza galiziana, si è chiesto loro di specificare quale fosse la lingua maggiormente utilizzata in contesto quotidiano, tra varietà galiziana e castigliana –;
– periodo di istruzione formale ricevuta in italiano L2/LS (da 1 a 4 anni);
– sesso dei locutori registrati in fase di preparazione al test – si è voluto testare la possibile incidenza di genere (voce maschile *versus* voce femminile) nella percezione degli stimoli sonori proposti –;
– tempi di risposta – per verificare la presenza di una correlazione diretta fra tempi e qualità delle risposte –.

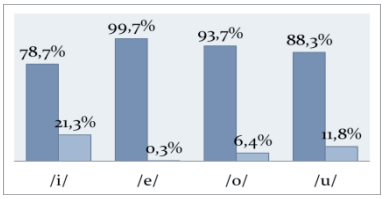
2.1.4 Risultati

2.1.4.1 Opposizioni vocaliche

Nonostante gli ottimi risultati sul riconoscimento di tutte le vocali in esame, le maggiori difficoltà si riscontrano proprio sulle vocali alte /i/ e /u/. Tale tendenza si rileva, seppure in maniera ridotta, anche nel gruppo di controllo italiano, confermando la propensione alla centralizzazione, intrinseca nelle atone finali italiane, a discapito dei suoni posizionati agli estremi. Ciò può pertanto compromettere anche la categorizzazione prettamente percettiva (cfr. Grafico 1).

Grafico 1 - *Risultati in percentuale dei test di identificazione percettiva di vocali atone dell'italiano, svolti da studenti galiziani (a sinistra) e dal gruppo di controllo di madrelingua italiani (a destra)*





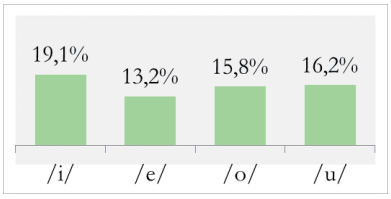
I dati statistici non rilevano valori che possano essere considerati significativi per nessuna delle quattro variabili considerate. Si presentano in aggiunta le percentuali di tempi di risposta superiori alla media, in relazione alle quattro categorie vocaliche presentate nei test (Grafico 2)[19]. Da queste si conferma quanto evinto dalle analisi descrittive, vale a dire maggiore difficoltà nell'identificazione delle vocali alte, in modo particolare le anteriori /i/, tanto negli apprendenti come nel gruppo di controllo.

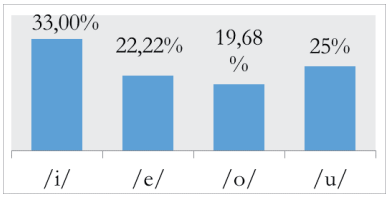
[19] La media è calcolata sulla durata in millisecondi dei tempi di risposta totali, ottenuti da tutti i partecipanti nelle risposte agli stimoli vocalici. Il valore della deviazione standard, che supera in alcuni casi in modo evidente lo stesso valore della media, rivela la poca omogeneità nel comportamento degli ascoltatori. Il Grafico 2 riporta invece le percentuali delle singole occorrenze, superiori al valore della media.

Tabella 3 - *Medie e deviazioni standard dei tempi di risposta ottenuti nelle identificazioni vocaliche*

TEMPI DI RISPOSTA - VOCALI				
	GALIZIANI		GRUPPO DI CONTROLLO	
	MEDIA (MS)	DEV.ST.	MEDIA (MS)	DEV.ST.
[i]	923	1423	1286	1981
[e]	1268	10021	1059	2080
[u]	1031	3209	914	1230
[o]	972	3774	1021	3096

Grafico 2 - *Opposizioni vocaliche: tempi di risposta superiori alla media, ottenuti dal gruppo di informatori galiziani (a sinistra) e dal gruppo di controllo di madrelingua italiani (a destra)*



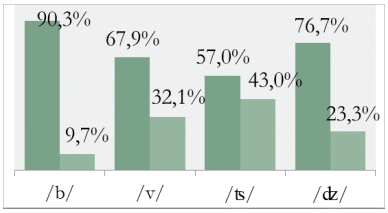


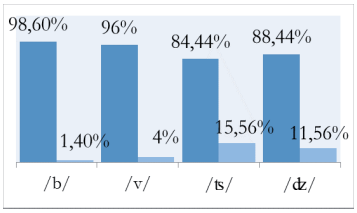
2.1.4.2 Opposizioni consonantiche

Le ipotesi relative alla categorizzazione di suoni simili e/o condivisi trovano in parte riscontro nei dati ottenuti (cfr. Grafico 3). Le occlusive bilabiali sonore, condivise da entrambi i sistemi di riferimento (si veda § 1.2.1.1), rappresentano la categoria identificata correttamente nella quasi totalità dei casi (90,32% di identificazioni corrette). I modelli percettivi di riferimento suggeriscono che sia possibile identificare suoni condivisi con le corrispettive categorie del sistema nativo, ma presuppongono anche che le differenze fra sistemi siano le prime ad essere interiorizzate e categorizzate; il fono occlusivo prodotto in italiano in contesto intervocalico non corrisponde alla variante galiziana realizzata nello stesso contesto, il che può aver favorito l'identificazione del suono, facendo in parte affidamento anche su informazioni acustiche, avallate dal supporto grafico delle etichette proposte. Al contrario, la fricativa sonora dell'italiano, prodotta in contesto intervocalico, confonde più spesso gli informatori (32,06% di errori), i quali presumibilmente riconducono tale suono alla realizzazione fricativa approssimante di *b* e *v* intervocaliche del sistema nativo. Se si considera la seconda opposizione indagata, tra affricate alveolari sorda e sonora dell'italiano, l'identificazione della corretta variante risulta un task complesso, come previsto a causa dell'estrema vicinanza tra i suoni, in modo maggiore nel caso della sorda (43,02% di errori contro il 23,34% commessi per la sonora); anche il gruppo di controllo in questo caso non sempre identifica correttamente la differenza fra i segmenti indagati: vale la pena ribadire che si tratta di un'opposizione

molto debole, resistente solo in poche coppie minime, spesso frutto di regole fonologiche e variabili locali e che di norma le realizzazioni allofoniche, come in questo caso, vengono prodotte inconsciamente dai parlanti. Per gli informatori galiziani è possibile che tale opposizione venga assimilata a un'unica categoria i cui allofoni risultano particolarmente complesse da individuare.

Grafico 3 - *Risultati in percentuale dei test di identificazione percettiva di consonanti dell'italiano, svolti da studenti galiziani (in verde, a sinistra) e dal gruppo di controllo di madrelingua italiani (in blu, a destra)*



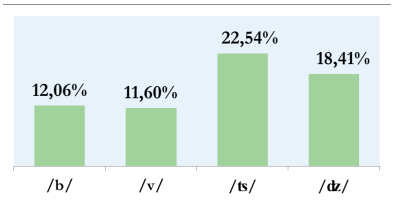


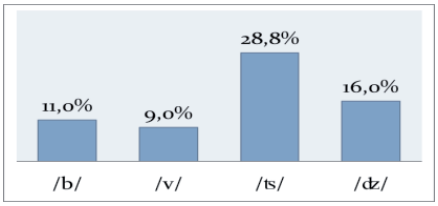
I tempi di risposta associati alle identificazioni di consonanti confermano una maggiore attenzione verso l'ascolto dei suoni affricati (Grafico 4), tanto nel gruppo di studenti, come nel gruppo di controllo (Tabella 4).

Tabella 4 - *Medie e deviazioni standard dei tempi di risposta ottenuti nelle identificazioni consonantiche*

TEMPI DI RISPOSTA - CONSONANTI				
	GALIZIANI		GRUPPO DI CONTROLLO	
	MEDIA (MS)	DEV.ST	MEDIA (MS)	DEV.ST
[b]	1051	4079	1152	2744
[v]	1094	3520	810	1116
[ts]	1756	8514	1625	2996
[dz]	1166	3595	1956	7768

Grafico 4 - *Opposizioni consonantiche: tempi di risposta superiori alla media, ottenuti dal gruppo di informatori galiziani (a sinistra) e dal gruppo di controllo di madrelingua italiani (a destra)*





Le prime analisi statistiche restituiscono valori significativi unicamente in relazione alla lingua materna degli informatori: coloro i quali utilizzano in prevalenza la lingua castigliana rispetto alla galiziana sembrano rispondere più correttamente alle prove di identificazione consonantica. Per valutare l'intensità dell'evidenza del test chi-quadro, le analisi vengono ampliate con un test di regressione multinomiale, il quale conferma la significatività della variabile L1 (dato p=1, il rapporto di verosimiglianza è =.267).

2.2 Esperimento 2: produzione

2.2.1 Soggetti

Le prove di produzione sono state realizzate da 3 madrelingua italiani (di varietà linguistica centro-settentrionale), 2 donne e 1 uomo, di età compresa fra i 40 e i 51 anni e 8 studenti galiziani (dell'Università di Santiago de Compostela, Galizia, Spagna), 6 donne e 2 uomini di età compresa fra 18 e 25 anni, con istruzione formale in lingua italiana da 1 a 4 anni e periodi di permanenza in Italia variabili, da 0 a 24 mesi. I soggetti verranno in seguito suddivisi, ai fini delle analisi, in due gruppi: il primo comprende 4 soggetti (3 donne e 1 uomo) con istruzione formale in lingua italiana di 1-2 anni; fra questi, due soggetti hanno seguito almeno un corso universitario di fonetica e fonologia, due sono invece iscritti a un corso di dottorato in studi linguistici. Il secondo gruppo è costituito da 4 soggetti (3 donne e 1 uomo) con istruzione formale in lingua italiana di 3-4 anni; tutti hanno seguito almeno un corso universitario di fonetica e fonologia.

2.2.2 Stimoli

Il compito dei soggetti è stato quello di leggere ad alta voce, per due volte consecutive, 30 frasi per ciascuna opposizione, contenenti 15 coppie minime di parole *target* (per un totale di 240 frasi). Le registrazioni sono state effettuate in ambiente insonorizzato, tramite registratore digitale Tascam DR-100MK2 e microfono esterno Sennheiser ME 3-ew, acquisite con frequenza di campionamento a 44.100 Hz, 16 bit, mono.

2.2.3 Analisi

Per le analisi ci si è avvalsi dei *software* di elaborazione del segnale PRAAT (Boersma, Weenink, 2014) e Sound Forge 7. Per una maggiore consistenza del campione analizzabile, le misure acustiche delle vocali sono state eseguite su voci femminili appartenenti al gruppo con esposizione alla lingua italiana di 1-2 anni. I file sonori di ciascun parlante sono stati etichettati manualmente tramite il software PRAAT; i valori formantici (calcolati nel punto medio a partire dal secondo *pulse* dopo l'attacco di sonorità della vocale) e le durate di ciascuna vocale sono stati estratti automaticamente[20]. Le analisi sui segmenti consonantici sono state condotte tramite ascolto e osservazione delle immagini sonografiche ottenute su PRAAT. Sono state quindi calcolate le percentuali di occorrenze corrette e di errori produttivi, e differenziati

[20] Tramite script, appositamente costruito dagli autori del presente lavoro.

per livelli di istruzione formale in italiano LS di tutti i soggetti facenti parte del campione (categorie 1-2 anni e 3-4 anni, uomini e donne)[21].

2.2.4 Risultati

2.2.4.1 Vocali

I valori ottenuti dal gruppo di controllo di madrelingua evidenziano una forte dispersione, soprattutto delle vocali anteriori, dovuta a fenomeni di centralizzazione tipici del vocalismo atono dell'italiano, e minore differenziazione in quelle posteriori (cfr. Tabella 5)[22]. Comparando i dati di riferimento del galiziano presenti in letteratura (Regueira, 2007) con i dati ottenuti dal gruppo di controllo di voci italiane e quelli delle studentesse galiziane emerge un dato inatteso: la produzione di vocali medie atone dell'italiano L2 ben si distanzia dai valori di riferimento della L1, in modo particolare nelle vocali anteriori, le quali subiscono una notevole anteriorizzazione (cfr. valori medi della vocale italiana atona [i] prodotta da italiani e da galiziani – F_2 2003 Hz *versus* F_2 2425 Hz –), probabilmente dovuta a iperarticolazione dei suoni da discriminare. Anche le vocali posteriori vengono distinte, seppure sia in anteriorità che in posteriorità le vocali medie L2 risultino innalzate rispetto a quelle del gruppo di controllo, riflettendo quindi un minore controllo nella produzione in L2 e un maggiore influsso della lingua materna (valori di F_1 [e] Ita_L1 = 586 Hz *vs.* [e] Ita_L2 = 538 Hz; valori di F_1 [o] Ita_L1 = 511 Hz *vs.* [o] Ita_L2 = 459 Hz). La polarizzazione delle vocali anteriori in italiano L2 è direttamente correlata a valori in durata maggiori (cfr. Grafico 5 e Tabella 5).

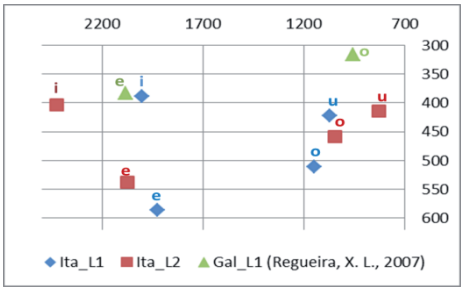
Tabella 5 - *F1, F2 e durate in ms di vocali anteriori [i e] e posteriori [u o] prodotte da madrelingua italiani (Ita_L1) e galiziani (Ita_L2); valori di riferimento di F1 e F2 delle vocali medie [e o] del galiziano (Gal_L1)*

	Ita_L1						**Ita_L2**					**Gal_L1**	
	F1	**d.s.**	**F2**	**d.s.**	**(ms)**		**F1**	**d.s.**	**F2**	**d.s.**	**(ms)**	**F1**	**F2**
[e]	586	66,5	1927	306,1	61.78	**[e]**	538	50,5	2074	329,8	69.88	383	2083
[i]	389	41,3	2003	504,6	62.40	**[i]**	404	44,7	2425	441,8	75.16	–	–
[o]	511	80,9	1147	100,5	52.72	**[o]**	459	42,6	1045	69,1	45.26	316	955
[u]	422	39,2	1072	86,9	46.15	**[u]**	415	25,3	828	60	43.73	–	–

[21] Oltre a un'analisi sonografica, gli autori si sono avvalsi di un ausilio percettivo.

[22] Si è deciso per varie ragioni di non trattare dati normalizzati: le analisi sono state condotte, come anticipato, su voci esclusivamente femminili, includendo vocali atone finali, delle quali si è voluta mettere in evidenza soprattutto la variazione nella dispersione delle aree di esistenza. Inoltre, i valori acustici di riferimento del galiziano, presenti in letteratura, sono espressi già in medie in Hz e non sarebbero stati trattabili ai fini della normalizzazione.

Grafico 5 - *Medie dei valori formantici di vocali dell'italiano prodotte da madrelingua italiani (Ita_L1) e galiziani (Ita_L2) e valori di riferimento delle vocali medie del galiziano (Gal_L1)*



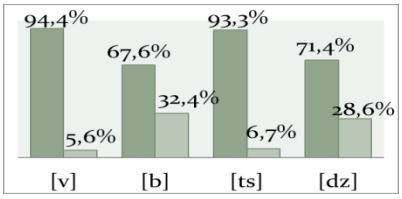
2.2.4.2 Consonanti

Considerando la prima opposizione consonantica presentata, il segmento [v] è prodotto appropriatamente in quasi tutte le occorrenze (94,4% per il gruppo di apprendenti meno esperti, 97,3% per i secondi[23]); il suono occlusivo in contesto intervocalico è, al contrario, il segmento su cui l'interferenza fonologica della lingua materna esercita più pressione, causando in molti casi la realizzazione approssimante del suono occlusivo (32,4 % di produzioni errate nel gruppo Ita_1-2, ben 77,3% per gli apprendenti più esperti del gruppo Ita_3-4). Nell'opposizione fra suoni affricati, l'alveolare sorda [ʦ] viene realizzata correttamente nella quasi totalità delle occorrenze (93,3% per il gruppo Ita_1-2, 100% per il secondo gruppo): i dati confermano in parte le previsioni teoriche e dei modelli di riferimento relative alla produzione di suoni nuovi, non nativi; la produzione del suono affricato sonoro, al contrario, rappresenta un compito particolarmente difficoltoso per gli apprendenti. Nella scala di marcatezza fonologica i suoni affricati sono fra i più marcati, seguiti unicamente dalle vibranti; fra suoni sordi e sonori i secondi rappresentano i più marcati fra i due e, rispetto al luogo di articolazione, le affricate meno marcate sono quelle palatali, più marcate invece quelle alveolari. Queste premesse possono motivare il relativo ed elevato numero di errori compiuti nella produzione del segmento [ʣ], quasi sempre realizzato come sordo, che tuttavia è maggiore nei soggetti afferenti al gruppo più esperto Ita_3-4 (71,4% di produzioni corrette dell'affricata alveolare sonora per il gruppo Ita_1-2, versus il 24% del gruppo di apprendenti più esperti; cfr. Grafico 6). Questa dissonanza, riguardante anche l'opposizione consonantica /b v/, ha spinto ad indagare più a fondo e nel dettaglio i profili anagrafici dei singoli soggetti, per poter spiegare la ragione per cui fossero proprio i soggetti con minore istruzione formale in italiano L2 a produrre il minor numero di errori. È emerso che alcuni dei soggetti con un solo anno di corso di lingua italiana siano dei dottorandi in studi linguistici, con esperienza metafonetica particolarmente rilevante rispetto agli altri informatori. Differenziando quindi i dati del primo gruppo in due ulteriori sotto-

[23] Da questo momento in poi, identificati rispettivamente come gruppo *Ita_1-2* e gruppo *Ita_3-4*.

gruppi di soggetti, con maggiore o minore esperienza in linguistica/fonetica[24], i dati si conformano ai livelli di competenza dei soggetti esaminati, unicamente nelle produzioni relative ai suoni affricati: il *transfer* fonologico (/b/→[β̞]/V_V) persiste anche tra gli informatori con maggiore esperienza in ambito fonetico (cfr. Grafico *7*).

Grafico 6 - *Percentuali di produzioni corrette (barre scure) e di errori (barre chiare) in opposizioni consonantiche dell'italiano realizzate da apprendenti con 1-2 anni (a sinistra) e 3-4 anni (a destra) di istruzione formale di italiano L2*



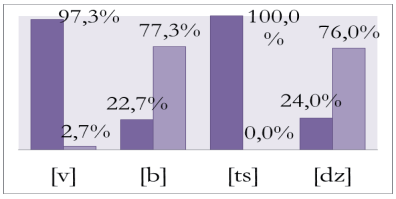
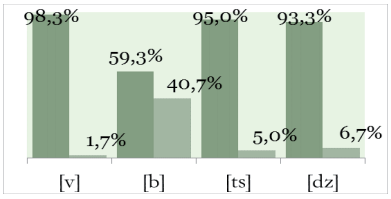


Grafico 7 - *Percentuali di produzioni corrette (barre scure) e di errori (barre chiare) in opposizioni consonantiche dell'italiano realizzate da apprendenti con 1-2 anni di istruzione formale di italiano L2, studenti (a sinistra) e dottorandi in studi linguistici (a destra)*



Si evidenzia inoltre il caso di un singolo soggetto appartenente al gruppo Ita_3-4 con *durata della permanenza* in Italia di 2 anni, le cui produzioni evidenziano una capacità molto più avanzata rispetto ai soggetti dello stesso gruppo, come si evince dal seguente dettaglio delle percentuali (Grafico 8):

Grafico 8 - *Percentuali di produzioni corrette e di errori in opposizioni consonantiche dell'italiano realizzate da un apprendente con 3-4 anni di istruzione formale di italiano L2 e durata della permanenza in Italia di 2 anni (a sinistra) e studenti dello stesso gruppo Ita_3-4 con permanenza inferiore (a destra)*

[24] Non si esclude una minima esperienza in linguistica/fonetica per gli altri soggetti del gruppo, in quanto tutti studenti afferenti ai corsi di laurea in Filología Gallega o Románica, i quali includono discipline linguistiche, oltre a uno o più corsi di fonetica e fonologia.

I livelli di istruzione formale in LS non risultano in ogni caso particolarmente determinanti, a giudicare dal persistere delle difficoltà produttive di suoni sia nuovi che simili anche negli apprendenti più esperti. Le variabili legate al *training* fonologico e alla permanenza all'estero in contesto linguistico L2, si confermano al contrario più incisive (cfr. Piske et al., 2001).

3. *Discussione e conclusioni*

I processi di percezione e produzione linguistica di suoni di una L2/LS sono correlati e variano per una contingenza di fattori, spesso interdipendenti. Tuttavia, non sempre si riscontra una corrispondenza diretta tra i due e per capirne le dinamiche è necessario ancora indagare a fondo. Nel caso preso in esame ad esempio, non sembra sempre esserci un rapporto di dipendenza diretta fra percezione e produzione[25]. Analizzando in modo trasversale i dati ottenuti dalle prove di produzione e percezione per ciascuna delle opposizioni in esame, emerge *in primis* che il filtro fonologico della lingua materna agisce con maggiore pressione sul versante produttivo: anche laddove le prove di identificazione diano degli ottimi risultati, la realizzazione del suono simile o condiviso è sempre influenzata dai parametri relativi alla produzione del suono corrispondente in L1, in modo preponderante se condizionati da processi linguo-specifici[26], si veda il caso dell'opposizione vocalica tra posteriori alte e medio-alte, la relativa centralizzazione della /o/ e la sovrapposizione delle aree di esistenza in un luogo prossimo alla realizzazione della stessa semi-alta in galiziano; è quanto accade ancora nella produzione della /b/ intervocalica, identificata percettivamente in modo corretto con percentuali elevatissime di successo, ma realizzata come approssimante/fricativa approssimante anche dai parlanti più esperti e con più cognizione metafonetica[27]. Si è notato inoltre quanto possa contribuire in casi come questi il supporto dell'informazione grafica, quindi di un grafema di riferimento per disambiguare suoni simili, come avviene ancora nella percezione della coppia di suoni /b v/. I suoni non condivisi fra sistemi (il riferimento è alla coppia di affricate) denotano una sorta di complementarietà nei dati degli informatori: quanto è percepito con più chiarezza nelle prove di identificazione (le alveolari sonore) risulta più difficile da riproporre in fase produttiva, laddove le alveolari sorde (identificate con minore successo rispetto alla controparte sonora) vengono prodotte in modo appropriato nella quasi totalità dei casi e a tutti i livelli di competenza. Le peculiarità legate alla fonologia e la complessità articolatoria delle affricate, le rende dei suoni meritevoli di ulteriori indagini nei processi di acquisizione da parte di apprendenti di italiano L2/LS; sarà interessante valutare le differenze tra dati percettivi e di produzione ottenuti da informatori afferenti a

[25] Comportamento che pare emergere anche in nuovi recenti studi (si cita, a titolo esemplificativo, Cheng, Zhang, 2015).

[26] Flege (1995), Major (2001).

[27] Tali risultati inducono a confermare che i processi di categorizzazione percettiva siano più propriamente basati sul riconoscimento di parametri acustico/fonetici (come sostenuto nei modelli PAM e SLM), che esulano dalla classificazione prettamente fonologica; quest'ultima interviene, al contrario e con maggiore peso nei processi di produzione, laddove categorie di L2, simili a quelle corrispondenti della lingua materna, vengono spesso sostituite da quelle native, assorbendo, laddove presenti, le possibili variazioni determinate dall'azione di processi fonologici linguo-specifici (come sostenuto da Flege, 1987; 1995 e Major, nel suo modello OPM, 2001).

lingue anche tipologicamente distanti dall'italiano, per cercare di fornire una visione più globale e ulteriori informazioni sull'acquisizione di tali suoni. I test percettivi somministrati saranno inoltre integrati con prove di discriminazione, per indagare maggiormente a livello infra-categoriale fra suoni in opposizione nuovi ma fra loro molto simili. Altro elemento da considerare è l'apporto al processo di acquisizione fornito dalle variabili studiate. Tra quelle proposte nello studio in oggetto, sono risultate preminenti il periodo di permanenza all'estero (quindi di contatto con un contesto madrelingua L2)[28] e il *training* fonologico implicitamente fornito a studenti di linguistica (quindi fonetica e fonologia)[29], ribadendo la necessità e l'utilità di integrare i percorsi di insegnamento e apprendimento di una L2/LS con attività rivolte alla pratica fonetica e fonologica. Tralasciando fattori come l'età o l'uso della lingua, ampiamente dibattuti e verificati come determinanti nel percorso di acquisizione fonologica (quindi di 'perdita' dell'accento straniero), è ulteriore intenzione degli autori quella di effettuare nuovi studi che, coinvolgendo parlanti di L1 molto distanti dall'italiano, quindi soprattutto soggetti immigrati, indaghino sull'influenza dei fattori motivazionali e sull'attitudine di questi ultimi verso la lingua italiana, in termini di prestigio e inclinazione alla socialità, i quali sembrano potenzialmente determinanti nell'acquisizione di una produzione in L2 non accentata (Bongaerts, Van Summeren, Planken & Schils, 1997; Moyer, 1999; 2007[30]). Infine, mentre lo studio proposto è stato rivolto prevalentemente a studenti di italiano come lingua straniera in contesto estero (L1), gli studi futuri saranno focalizzati su apprendenti autonomi di italiano come lingua seconda e momentaneamente o permanentemente residenti in territorio nazionale.

Riferimenti bibliografici

Alfano, J., 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*, Saarbrücken, Germany, 6-10 August 2007, 1793-1796.

Aoyama, K., Flege, J. (2011). Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers. In *Journal of the Phonetic Society of Japan*, 15 (3), 5-13.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, MD: York Press, 171-204.

Best, C.T., Hallé, P.A., Bohn, O.-S. & Faber, A. (2003). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners native language. In Solé, M.J., Recasens, D. & Romero J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona: Causal Productions, 2889-2892.

Best, C., Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O.-S., Munro, M. (Eds.), *Language*

[28] Si confrontino le discussioni in merito proposte da Piske, Mackay & Flege (2001).

[29] Come dimostrato anche, ad esempio, in Kuhl (2000).

[30] Gli stessi autori evidenziano come parlanti di lingue di valore sociale, il cui prestigio è universalmente riconosciuto, non siano motivati a lavorare sulla propria pronuncia; ne sono esempio lampante la maggior parte dei parlanti anglofoni.

Experience in Second language Speech Learning. In honor of James Emil Flege. Amsterdam: John Benjamins, 13-34.

Boersma, P., Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.59. http://www.praat.org/ Accessed 20.11.14.

Bohn, O.-S., Flege, J. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. In *Applied Psycholinguistics*, 11, 303-328.

Bongaerts, T., van Summeren, C., Planken, B. & Schils, E. (1997). Age and Ultimate Attainment in the Pronunciation of a Foreign Language. In *Studies in Second Language Acquisition*, 19, 447-465.

Celata, C., Costamagna, L. (2012). Geminate timing in the speech of Estonian L2 learners of Italian. In De Meo, A., Pettorino, M. (Eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition. The case of Italian*. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Cheng, B., Zhang, Y. (2015). Syllable Structure Universals and Native Language Interference in Second Language Perception and Production: Positional Asymmetry and Perceptual Links to Accentedness. In *Frontiers in Psychology*. http://dx.doi.org/10.3389/fpsyg.2015.01801/ Accessed 18.06.16.

Costamagna, L. (2007). The Acquisition of Italian L2 Affricates: The Case of a Brazilian Learner. In *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianopolis: Federal University of Santa Catarina, 137-148.

De Meo, A., Pettorino, M. (Eds.) (2012). *Prosodic and Rhythmic Aspects of L2 Acquisition: The Case of Italian*. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

De Meo, A., Vitale, M. & Pellegrino, E. (2016). Tecnologia della voce e miglioramento della pronuncia in una L2: imitazione e autoimitazione a confronto. Uno studio su sinofoni apprendenti di italiano L2. In Bianchi, F., Leone, P. (Eds.), *Linguaggio e apprendimento. Metodi e strumenti tecnologici*, 6-13.

Escudero, P. (2009). Linguistic Perception of "similar" L2 sounds. In Boersma, P., Silke, H. (Eds.), *Phonology in Perception*. Berlin: Mouton de Gruyter.

Fernández Rei, E. (coord.) (2014). FOLERPA: Ferramenta On-Line para ExpeRimentación PerceptivA. Santiago de Compostela: Instituto da Lingua Galega. http://ilg.usc.es/FOLERPA.

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

Flege, J. (1995). Second language speech learning: theory, findings, and problems. In strange, W. (Ed.), *Speech perception and language experience: issues in cross-language research*. Baltimore. MD: York Press, 233-277.

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

Flege, J., Schmidt, A. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. In *Phonetica*, 52, 90-111.

Flege, J., Mackay, I. & Meador, D. (1999). Native Italian speakers' production and perception of English vowels. In *Journal of the Acoustical Society of America*, 106, 2973-2987.

Flege, J., Mackay, I. (2004). Perceiving vowels in a second language. In *Studies in Second Language Acquisition*, 26, 1-34.

Frontera, M. (in press). Hablantes gallegos frente a sonidos italianos: percepción y producción de oposiciones vocálicas y consonánticas. In *Working papers in Spanish in Society*, 7th International Conference of Hispanic Linguistics (5th Biennial Meeting of the International Association for the Study of Spanish in Society [SiS]), Heriot-Watt University, Edinburgh, 28-29 May 2015.

Gili Fivela, B. (2010). *Definizione di affricate*. Enciclopedia Treccani online.

Iverson, P., Kuhl, P.K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. In *Journal of the Acoustical Society of America*, 97, 553-562.

Iverson, P., Kuhl, P.K., Akahane-Yamadac, R., Dieschd, E., Tohkurae, Y., Kettermannf, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. In *Cognition*, 87, B47-B57.

Kabak, B., Reckziegel, T. & Braun, B. (2011). Timing of second language geminates and singletons. In *Proceedings of the 17th International Congress of the Phonetic Sciences*, 994-97.

Kuhl, P.K. (2000). A new view of language acquisition. In *Proceedings of the National Academy of Sciences USA*, 97, 11850-11857.

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In Elenius, K., Branderud, P. (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm: KTH and Stockholm University, 2, 140-147.

Larsen-Freeman, D., Long, M.H. (2014). *An Introduction to Second Language Acquisition Research*. London and New York: Longman.

Lisker, L., Abramson, A.S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 1967, 563-567.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *ICPhS 1995. Proceedings of the 13th International Congress of Phonetic Sciences*, August 13-19, 1995, Stockholm, Sweden, 4, 92-99.

Major, R.C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Martínez Celdrán, E. (2002). *Introducción á fonética. O son na comunicación humana*. Vigo: Editorial Galaxia.

Mioni, A.M. (2001). *Elementi di fonetica*. Padova: Unipress.

Mora, J.C. (2008). Methodological issues in assessing L2 perceptual phonological competence. In *Proceedings of the PTLC 2007 Phonetics Teaching and Learning Conference,* London: Dept. of Phonetics and Linguistics, University College London, 1-5.

Mori, L. (2007). *Fonetica dell'italiano L2: un'indagine sperimentale sulla variazione nell'interlingua dei marocchini*. Roma: Carocci Editore.

Moyer, A. (1999). Ultimate attainment in L2 phonology. The critical factors of age, motivation and instruction. In *Studies in Second Language Acquisition*, 21, 81-108.

Moyer, A. (2007). Do language attitudes determine accent? A study of bilinguals in the USA. In *Journal of Multilingual and Multicultural Development*, 28, 502-518.

Navarro, A.H., Quilis merín, M. (2012). *La voz del lenguaje: fonética y fonología del español*. Valencia: Tirant Humanidades.

Pallotti, G. (2003). *La seconda lingua*. Milano: Bompiani.

Pape, D., Jesus, L.M.T. (2014). Production and perception of velar stop (de)voicing in European Portuguese and Italian. In *EURASIP Journal on Audio, Speech, and Music Processing*, 6, 1-10.

Pellegrino, E. (2012). The perception of foreign accented speech. Segmental and suprasegmental features affecting degree of foreign accent in Italian L2. In Mello, H., Pettorino, M. & Raso, T. (Eds.), *Proceedings of the VIIth GSCP International Conference - Speech and Corpora*. Firenze: Firenze University Press, 261-267.

Pellegrino, E., Vigliano, D. (2015). Self imitation in prosody training: A study on Japanese learners of Italian. In Steidl, S., Batliner, A. & Jokisch, O. (Eds.), *Workshop on Speech and Language Technology in Education*, September 4-5, 2015, Leipzig, 53-57.

Pettorino, M., De Meo, A., Pellegrino, E., Salvati, L. & Vitale, M. (2011). Accento straniero e credibilità del messaggio: un'analisi acustico-percettiva. In Gili Fivela, B., Stella, A., Garrapa, L. & Grimaldi, M. (Eds.), *Contesto comunicativo e variabilità nella produzione e percezione della lingua*, Atti del 7° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV 2011). Roma: Bulzoni editore.

Piske, T., Mackay, I.R.A. & Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: a review. In *Journal of Phonetics*, 29(2), 191-215.

Regueira, X.L. (2007). Vocais finais en Galego e en portugués: un estudio acústico. In González Fernández, H., Lama López, M.X. (Eds.), *Actas VII Congreso Internacional de Estudos Galegos. Mulleres en Galicia. Galicia e os outros pobos da Península*. Sada: Ediciós do Castro.

Repp, B. (1984). Categorical Perception: Issues, Methods, Findings. In *Speech and Language*, 10, 243-335.

Romaine, S. (1984). The status of sociological models and categories in explaining language variation. In *Linguistische Berichte*, 90, 25-38.

Romito, L., Frontera, M. (2015). Perception and production of Italian L2 sounds. In *Proceedings of the 6th ISEL Conference on Experimental Linguistics*, 26-27 June 2015, Athens, Greece, 70-73.

Trubeckoj, N.S. (1939). *Grundzüge der Phonologie*; Mazzuoli Porru, G. (Ed.) (1971). *Fondamenti di fonologia*. Torino: Einaudi.

Vayra, M., Avesani, C., Best, C.T. & Bohn, O.-S. (2012). Non solo dettagli fonetici, non solo categorie fonologiche: L'interazione tra fonetica e fonologia nella percezione di suoni non-nativi. In *Studi e Saggi Linguistici*, 50 (2), 119-146.

DOI: 10.17469/O2102AISV000003

STEPHAN SCHMID, GIULIA PEDRAZZINI

La pronuncia delle occlusive nel tedesco L2 di apprendenti italofoni: un esperimento didattico

The present contribution investigates the pronunciation of German plosives by 15-year-old students from the Italian part of Switzerland. In particular, VOT and %Voice (the percentage of duration by which the signal of 'voiced plosives' is periodic) are analysed. In a classroom experiment, 10 students received detailed instructions about the phonetic differences between German and Italian plosives and were recorded twice in a reading task (prior and after the instruction). A control group of 10 students without explicit pronunciation training was recorded twice as well. No statistically significant effects of the explicit pronunciation training were found, at least for the group as a whole; only four students showed clearly higher VOT values of the voiceless German plosives in the second recording. Implications for pronunciation teaching and further research are discussed.

Key words: voice onset time; German as a second language; Italian; pronunciation teaching.

Introduzione

Il presente studio intende dare un contributo alla discussione sull'insegnamento della pronuncia in una lingua straniera riportando i risultati di un esperimento didattico. Nella parte empirica si esamina la produzione delle consonanti occlusive del tedesco da parte di due gruppi di apprendenti italofoni, misurando per le occlusive sorde il tempo dell'attacco della sonorità e per le occlusive sonore la percentuale della durata in cui il segnale è periodico. Com'è noto, la realizzazione fonetica del contrasto tra occlusive omorganiche può dare adito a fenomeni di interferenza della L1 sulla L2 (e viceversa), in particolare quando tale contrasto viene implementato diversamente nelle due lingue in questione. Nel nostro caso, infatti, la distinzione tra occlusive 'sorde' e 'sonore' si deve in tedesco essenzialmente all'aspirazione delle sorde, mentre in italiano il contrasto viene veicolato soprattutto attraverso l'attività glottidale nelle sonore (v. 2.1).

Al nostro esperimento, che prende spunto da un precedente studio pilota (cfr. 3.1), hanno partecipato due classi di studenti liceali che sono state registrati due volte mentre eseguivano un compito di lettura. Entrambi i gruppi sono stati sensibilizzati alla resa delle occlusive tedesche tramite una lezione di pronuncia, consistente in una spiegazione esplicita dei dettagli fonetici e in una fase di esercitazione; tuttavia, nel primo gruppo questa attività didattica è avvenuta prima della seconda registrazione, mentre il secondo gruppo è stato istruito solo dopo la seconda lettura.

Questo contributo è strutturato come segue. Nel prossimo paragrafo esporremo alcune riflessioni generali sullo stato della ricerca nel campo dell'acquisizione e dell'insegnamento della pronuncia in una seconda lingua (1). Successivamente delineeremo un breve schizzo di fonetica contrastiva delle occlusive in italiano e in tedesco (2.1) che servirà alla formulazione di ipotesi concrete per la nostra ricerca (2.2). In 3.1 riporteremo brevemente i risultati di una precedente ricerca pilota, dopodiché illustreremo la metodologia adottata per la raccolta e l'analisi dei dati (3.2-3.3) e lo svolgimento dell'unità didattica elaborata per questo esperimento (3.4). Presenteremo quindi i risultati ottenuti dall'analisi acustica (*Voice Onset Time* e *%Voice*) delle occlusive prodotte dagli allievi sia in tedesco che in italiano nella prima e nella seconda registrazione (4). Concluderemo con una breve discussione delle implicazioni dei nostri risultati per l'insegnamento della pronuncia e per la ricerca futura in questo ambito (5).

1. *La pronuncia nell'acquisizione e nell'insegnamento di una seconda lingua*

1.1 Il ruolo della pronuncia negli studi sull'acquisizione di lingue seconde e nell'insegnamento delle lingue straniere

La pronuncia costituisce una delle prime difficoltà che l'apprendente di una lingua straniera deve affrontare nel suo percorso di acquisizione. Viceversa, per i parlanti nativi il cosiddetto 'accento straniero' viene spesso considerato un fenomeno evidente e difficilmente superabile (il che può condizionare a sua volta in modo negativo gli atteggiamenti e la motivazione degli stessi parlanti nativi quando imparano una lingua straniera). Oggigiorno, a questa centralità della dimensione fonetico-fonologica nell'acquisizione si contrappone però una certa marginalità tanto nella ricerca scientifica quanto nella prassi didattica. In manuali, riviste e congressi dedicati alla *Second Language Acquisition* (SLA) prevalgono di gran lunga gli studi dedicati a fenomeni morfosintattici e pragmatici, laddove il *Second Language Speech* viene studiato prevalentemente da una comunità di specialisti, che attira comunque un numero via via crescente di ricercatori[1].

Sul versante dell'insegnamento delle lingue straniere si nota che i libri di testo comunemente adottati sono incentrati più che altro sui diversi atti linguistici e su determinate strutture grammaticali, mentre si suppone che gli aspetti sonori della lingua bersaglio vengano acquisiti in modo più o meno automatico e inconscio. In sostanza, l'opinione comune viene sintetizzata nell'affermazione secondo la quale un grado anche notevole di accento straniero è accettabile finché l'intelligibilità degli enunciati in L2 non viene compromessa (Saville-Troike, 2006: 143).

[1] V. ad esempio il convegno *New Sounds* che si svolge ogni tre anni e la rivista *Journal of Second Language Pronunciation* lanciata nel 2015; per una rassegna sommaria di alcuni modelli teorici v. ad esempio Schmid (2012: 633-637) e Schmid, Wachter (2015: 203-204).

Varie cause hanno portato a questa situazione, ma se si volge uno sguardo alla storia degli studi sull'acquisizione delle lingue seconde, si scopre che il componente fonologico assumeva un ruolo centrale nei modelli imperanti dopo la seconda guerra mondiale. Ad esempio, la classica 'ipotesi dell'analisi contrastiva' (Lado, 1957: 2) assumeva che – una volta individuate le principali differenze strutturali tra L1 e L2 – si potesse migliorare la pronuncia degli apprendenti mediante esercizi di natura ripetitiva. Nella prassi glottodidattica, tale approccio – di chiara ispirazione comportamentista – trovava la sua emanazione tecnologica più evidente nei 'laboratori di lingua' istituiti negli edifici scolastici, dove i discendenti si esercitavano con cuffie, microfoni e cassette. Negli anni più vicini a noi, con il superamento della visione comportamentista dell'acquisizione nel linguaggio non solo sono stati abbandonati i laboratori di lingua e i rispettivi metodi didattici, ma al contempo è stato ridimensionato anche il ruolo della pronuncia nell'insegnamento delle lingue straniere.

Risulta invece interessante notare che il cosiddetto 'metodo fonetico' (Canepari, 1979: 9) non condivideva affatto le presupposizioni del comportamentismo, ma insisteva piuttosto sull'insegnamento esplicito delle differenze tra L1 e L2, con l'ausilio di nozioni di fonetica articolatoria (cfr. 1.2). Appare chiaro che questo approccio presti a sua volta il fianco a critiche da parte di chi insiste sulla fondamentale differenza tra il sapere metalinguistico da un lato e la vera e propria competenza linguistica dall'altro, mettendo l'accento sulla scarsa interazione tra i due tipi di conoscenza e sulla conseguente efficacia limitata dell'insegnamento esplicito (v. Gass, Selinker, 2008: 368-394 e Chini, 2005: 111-119 per un riassunto dello *status quaestionis*). Non sarà dunque un caso che nella didattica delle lingue straniere la cosiddetta 'svolta comunicativa' – di cui si trova l'espressione più autorevole nel 'Quadro comune europeo di riferimento per la conoscenza delle lingue' (QCER, v. Consiglio d'Europa, 2010) – abbia notevolmente favorito l'applicazione di metodi impliciti nella glottodidattica.

Pur non potendo andare a fondo di questo problema, possiamo comunque sintetizzare il 'principio dell'intelligibilità' (cfr. Thomson, Derwing, 2015: 327) nell'affermazione che, negli approcci odierni, la soglia di intervento didattico è focalizzata sul livello fonemico, laddove differenze allofoniche o di 'dettaglio fonetico fine' non attirano l'attenzione degli insegnanti. Sembra quindi che la glottodidattica non faccia altro che reiterare i processi cognitivi degli stessi apprendenti, i quali acquisiscono più facilmente contrasti tra suoni nuovi che non differenze tra suoni simili, com'è stato messo in evidenza a più riprese dalle ricerche empiriche sull'acquisizione della pronuncia di una lingua seconda (a partire da Flege, 1987).

Ebbene, un esempio classico di questa casistica risiede nei diversi modi in cui le lingue – e di conseguenza gli apprendenti – implementano foneticamente il contrasto fonologico tra occlusive dello stesso luogo di articolazione attraverso

il Voice Onset Time (VOT), ovvero proprio il fenomeno indagato in questa sede (v. 2.1)[2].

1.2 Studi sull'insegnamento della pronuncia delle lingue straniere

In un importante contributo apparso recentemente nella rivista *Applied Linguistics*, Thomson e Derwing (2015) riassumono i caratteri principali di 75 studi sull'efficacia dell'insegnamento della pronuncia[3]. Nella loro rassegna sullo stato dell'arte, i due autori affermano innanzitutto che buona parte della ricerca sull'insegnamento della pronuncia sia spesso priva di un fondamento teorico: "Almost all the studies we examined lacked an overt theoretical stance" (Thomson, Derwing, 2015: 334).

Per quanto riguarda le metodologie adottate nelle varie ricerche esaminate, si rileva tra l'altro che il 79% degli studi si basava su giudizi di ascoltatori (di contro al 21% che adopera invece delle analisi acustiche), che il 73% esaminava il parlato letto e che il 60% degli esperimenti includeva un gruppo di controllo (Thomson, Derwing, 2015: 331). Nel 52% degli studi i fenomeni esaminati erano di tipo segmentale e nel 18% di tipo soprasegmentale; il restante 30% combinava aspetti segmentali e soprasegmentali (Thomson, Derwing, 2015: 330).

Per quanto riguarda poi i metodi di insegnamento della pronuncia (Thomson, Derwing, 2015: 330), il 61% degli studi focalizzava l'insegnamento in classe, mentre il 39% esaminava l'esercitazione con supporto informatico (*Computer Assisted Pronunciation Training*, CAPT). L'insegnamento in classe sembra seguire prevalentemente un approccio di tipo PPP (*Presentation, Practice Production*)[4].

Inserendo dunque la nostra ricerca nel quadro della ricerca internazionale constatiamo che essa mette al centro un fenomeno segmentale (il contrasto tra occlusive sorde e sonore) nel parlato letto, adoperando delle misure acustiche (3.3). Come vedremo di seguito, l'esperimento coinvolge un gruppo di controllo (3.2) e l'intervento didattico segue essenzialmente il protocollo 'presentazione, pratica, produzione' (3.4). Prima di presentare il metodo e i dati raccolti, occorre però descrivere brevemente la natura fonetica delle occlusive in tedesco e in italiano (2.1), il che ci permetterà di formulare due ipotesi specifiche sulla pronuncia del tedesco da parte di apprendenti italofoni (2.2).

---

[2] Per una rassegna parziale dei numerosi studi sul VOT in varie situazioni di contatto linguistico v. ad esempio Laeufer (1997: 331-340) e Chang (2012: 252).

[3] Ringraziamo un revisore anonimo del nostro abstract per averci segnalato questo titolo prima del convegno AISV di Salerno.

[4] Notiamo tra parentesi che questo tipo di procedura didattica era già stato raccomandato nel tradizionale 'metodo fonetico' di Canepari (1979: 7, 9), con un'enfasi aggiunta sulle abilità percettive: "Per riuscire a pronunciar bene una lingua straniera si deve esercitare l'orecchio a riconoscere suoni nuovi [...] ma non si può fare a meno d'esercizi sistematici [...]"; "Il metodo fonetico consiste nel rendersi pienamente conto delle possibilità articolatorie dell'apparato fonatorio [...]".

2. *Schizzo di fonetica contrastiva delle occlusive in tedesco e in italiano*

2.1 Il *Voice Onset Time* (VOT)

Com'è noto, con il termine 'tempo di attacco della sonorità' (ingl. *Voice Onset Time*, VOT)[5] ci si riferisce al lasso di tempo che trascorre nell'articolazione di una consonante occlusiva tra il rilascio dell'occlusione e l'inizio della vibrazione delle pliche vocali. Il VOT può quindi essere positivo, se le pliche vocali iniziano a vibrare dopo il rilascio, oppure esso può essere negativo (ingl. *lead* VOT) se le pliche vibrano già durante la fase di occlusione; il VOT positivo può inoltre essere breve (*short-lag*) o più lungo (*long-lag*). Benché il VOT costituisca ovviamente un continuum (Cho, Ladefoged, 1999: 223), questo parametro permette comunque di suddividere grosso modo le lingue del mondo in due grandi classi (cfr. Lisker, Abramson, 1964; Beckman, Jessen & Ringen, 2013), ovvero le *true voice languages* e le *aspirating languages*.

La differenza tra i due tipi fonetici può essere esemplificata con le due lingue prese in esame in questo studio: l'italiano (una *true voice language*) mostra un VOT positivo piuttosto contenuto per le sorde, mentre le sonore mostrano un VOT chiaramente negativo. In tedesco, invece, le occlusive /b d ɡ/ vengono realizzate come sonore soltanto nel contesto intervocalico (infatti si parla anche di *passive voicing*: v. Beckman, Jessen & Ringen, 2013: 259), mostrando invece in posizione iniziale spesso un VOT lievemente positivo; in questo contesto, il contrasto tra le due categorie fonologiche viene invece accentuato tramite un VOT fortemente positivo delle sorde /p t k/ che vengono realizzate come aspirate [p^h t^h k^h] (Reetz, 1999: 143-148).

2.2 Ipotesi

In base all'analisi contrastiva tra italiano L1 e tedesco L2 possiamo quindi formulare due ipotesi specifiche per la nostra indagine:

- Ipotesi 1: gli apprendenti italofoni del tedesco tenderanno a pronunciare le occlusive sorde in posizione prevocalica secondo il modello della loro L1, con un VOT breve ovvero senza aspirazione.
- Ipotesi 2: gli apprendenti italofoni del tedesco tenderanno a pronunciare le occlusive 'sonore' secondo il modello della loro L1, con un VOT negativo ovvero con vibrazione delle pliche vocaliche.

Assumiamo inoltre che questi due fenomeni contribuiscano in modo decisivo a creare l'impressione di un 'accento italiano' in tedesco; non a caso, essi si annoverano tra i sette tratti elencati da Maturi (2006: 137) per caratterizzare la pronuncia italiana del tedesco.

[5] Scegliamo qui la convenzione terminologica, diffusa negli studi di fonetica in Italia, di usare per il concetto di 'tempo di attacco della sonorità' l'acronimo inglese VOT, in sintonia con la prassi internazionale (cfr. Albano Leoni, Maturi, 2011: 65, 169; Schmid, 1999: 61, 218).

Al fine di verificare queste due ipotesi sono stati registrati due gruppi di allievi di un liceo nella Svizzera italiana, a due riprese. Entrambi i gruppi hanno ricevuto una lezione di fonetica con spiegazioni esplicite sulla pronuncia e una fase di esercitazione (v. 3.4 per una descrizione dettagliata della lezione), ma nel primo gruppo la lezione è stata tenuta prima della seconda registrazione, mentre nel secondo gruppo la lezione ha avuto luogo soltanto in seguito. Ben consapevoli delle discussioni controverse sull'efficacia dell'insegnamento esplicito della pronuncia, aggiungiamo comunque alle due ipotesi fonetiche precedenti una terza ipotesi riguardo all'intervento didattico:

– Ipotesi 3: l'insegnamento esplicito e l'esercitazione della pronuncia avrà un effetto positivo sulle realizzazioni delle occlusive del primo gruppo, per cui ci aspettiamo di riscontrare nella seconda registrazione un avvicinamento della pronuncia alla norma della lingua bersaglio. Al contrario, il secondo gruppo (che non riceve un'istruzione esplicita tra la prima e la seconda registrazione) non mostrerà un cambiamento nella pronuncia delle occlusive.

3. *Ricerca empirica e esperimento didattico*

3.1 Lo studio pilota

Va premesso che a favore delle due ipotesi formulate nel paragrafo precedente depongono anche i risultati di uno studio pilota condotto dalla seconda autrice del presente contributo (Pedrazzini, 2015).

Nello studio pilota, cinque studentesse dell'età di 17-18 anni, e che avevano seguito delle lezioni di tedesco durante sei anni, hanno letto 23 parole tedesche contenenti le occlusive /p t k/ e /b d ɡ/ in posizione iniziale e interna di parola; le stesse parlanti hanno letto inoltre 23 parole italiane contenenti le stesse occlusive negli stessi contesti. Sono state misurate le fasi di chiusura e di rilascio (VOT) delle occlusive nonché la fase di transizione dalla vocale precedente (*Voice Offset Time*). I risultati mostrano un sostanziale transfer del modello di pronuncia italiano alle parole tedesche: da un lato si è osservato un VOT solo lievemente positivo delle occlusive sorde (20 ms nel caso di /p/ iniziale e 50 ms nel caso di /k/ intervocalico), mentre dall'altro lato i parlanti hanno mostrato un VOT fortemente negativo delle occlusive sonore (da -29 a -69 ms).

3.2 La raccolta dei dati: il campione e il *corpus*

Incoraggiati dai risultati dello studio pilota ci siamo quindi proposti di verificare le prime due ipotesi formulate in 2.2. Per la presente indagine è stato scelto un campione più cospicuo, consistente di 20 alunni del Liceo cantonale di Bellinzona (Canton Ticino, Svizzera).

Gli allievi appartengono a due prime classi del liceo (denominate in modo arbitrario 1H e 1L), per cui presentano lo stesso tipo di retroterra biografico e scolastico. Di ciascuna classe sono stati registrati 10 soggetti, 5 di sesso femminile e 5 di sesso

maschile. Al momento della registrazione (novembre 2015), l'età media dei ragazzi era di 15 anni e mezzo in ambedue le classi; nella precedente scuola media tutti gli studenti avevano studiato il tedesco per tre anni (con tre ore di insegnamento settimanali). Il loro livello di padronanza del tedesco può essere caratterizzato come A2-B1, secondo la classificazione proposta dal 'Quadro comune europeo di riferimento per la conoscenza delle lingue' (QCER, v. Consiglio d'Europa, 2010).

Per le registrazioni sono state scelte 12 parole bersaglio in ciascuna lingua (tedesco e italiano) che contenessero le 6 occlusive /p t k/ e /b d g/ in posizione sia iniziale sia interna di parola. La tabella 1 presenta le 24 parole del *corpus* nelle due lingue secondo il tipo di consonante e di contesto:

Tabella 1 - *Parole bersaglio*[6]

	Occlusive	*Tedesco*		*Italiano*	
Contesto		**#_**	**V_V**	**#_**	**V_V**
Labiali	p	***p**acken*	*Ka**pp**e*	***p**ane*	*a**p**e*
	b	***b**aden*	*A**b**end*	***b**anco*	*fia**b**e*
Coronali	t	***T**age*	*ha**tt**e*	***t**anto*	*fa**t**e*
	d	***d**anke*	*scha**d**e*	***d**anno*	*ca**d**e*
Dorsali	k	***k**annte*	*ba**ck**en*	***c**assa*	*po**ch**e*
	g	***G**abel*	*Fra**g**e*	***g**allo*	*pa**gh**e*

Al fine di comparare al meglio le due lingue, sono state scelte parole tedesche e italiane bisillabiche con una struttura fonotattica 'C[a](C).C[e], che avessero cioè la prima vocale tonica /a/ e la seconda vocale atona /e/. Di conseguenza, la differenza tra le consonanti nella prima e la terza colonna da un lato e quelle nella seconda e la quarta colonna dall'altro lato non è solo fonotattica (posizione iniziale o interna di parola), ma anche prosodica (sillaba tonica o atona).

Com'è noto, in tedesco /e/ atona viene ridotta a schwa in posizione finale di parola, mentre può essere cancellata del tutto quando segue una /n/ tautosillabica che di conseguenza assume la posizione di nucleo della sillaba atona (cfr. Kohler, 1999: 87). A rigore, le due parole *Abend* 'sera' e *backen* 'infornare' non presentano le occlusive bersaglio in posizione intervocalica come le altre parole; il contesto fonotattico contiene comunque una sonorante (inoltre, molti allievi non hanno ridotto la vocale atona come previsto dalla norma del tedesco standard, seguendo invece una specie di *spelling pronunciation*). Ai nostri fini era più importante scegliere delle

[6] Notiamo tra parentesi che il fonema tedesco /k/ viene scritto con due allografi <k> e <ck> (v. *kannte* 'conosceva' e *backen* 'infornare'); anche le doppie consonanti ortografiche (ad esempio in *Kappe* 'berretto' e *hatte* 'aveva') sono meri allografi che stanno ad indicare la brevità fonologica della vocale precedente.

parole tedesche di cui i ragazzi conoscevano il significato; anche per l'italiano abbiamo cercato delle parole del lessico comune (per questo motivo la scelta di *poche* che ha una vocale tonica diversa da /a/).

Per il compito di lettura, le parole erano inserite in una posizione prosodica prominente (ma non prepausale) in due frasi cornice, rispettivamente *Ho detto _ due volte* e *Ich habe _ gesagt*. L'ordine dei quattro blocchi nella lettura era: prima 12 frasi in italiano e 12 frasi in tedesco, poi la seconda ripetizione di queste 24 frasi.

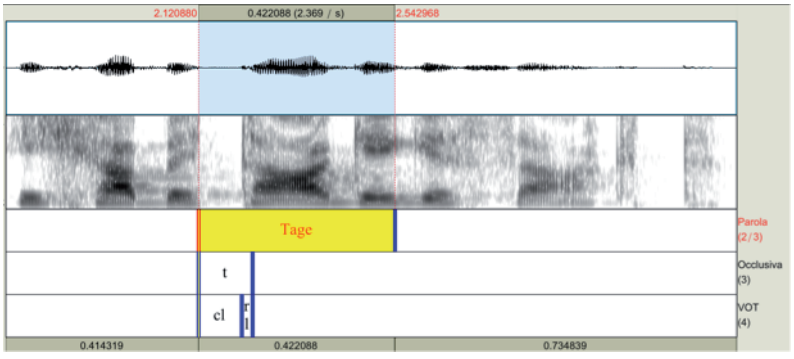
Le due sessioni di registrazione sono state effettuate durante le lezioni di tedesco e sono avvenute esattamente a una settimana di distanza tra di loro. Per la registrazione, le frasi sono state proiettate sullo schermo di un computer portatile; la voce dei ragazzi è stata registrata con il programma *SpeechRecorder* (Drachsler, Jänsch, 2004), in modo da ottenere un singolo file sonoro per ogni frase letta. Il computer era collegato ad un'interfaccia audio USB Pre 2 e a un microfono Sennheiser MKE 2 P-C[7]. Le sessioni di lettura hanno avuto luogo in una piccola stanza dell'edificio scolastico, la quale tuttavia non era dotata di particolari attrezzature fonoassorbenti. Al fine di ridurre eventuali interferenze acustiche, il microfono è stato fissato mediante un collarino a una distanza di ca. 5 cm davanti alla bocca dei locutori. I file sonori sono stati salvati con una frequenza di campionamento di 44.100 Hz e una quantizzazione di 16 bit.

3.3 Procedura di analisi

In totale il *corpus* da analizzare ammonta a 1920 occorrenze di occlusive: 20 locutori x 6 consonanti bersaglio (/p t k/ e /b d g/) x 2 contesti fonotattici e prosodici (posizione iniziale e interna di parola ovvero sillaba tonica e atona) x 2 lingue (italiano e tedesco) x 2 sessioni di registrazione x 2 ripetizioni per ogni sessione di registrazione.

I file sonori sono stati segmentati ed etichettati manualmente dai due autori mediante dei TextGrid nel programma *Praat* (Boersma, Weenink, 2015). La procedura di segmentazione e annotazione viene illustrata nella Figura 1:

Figura 1 - *Esempio di segmentazione e annotazione mediante un TextGrid di Praat*



[7] Dati tecnici: direttività omnidirezionale, gamma di frequenza di 20–20.000 Hz, ±23dB, e coefficiente di trasmissione a vuoto di 10 mV/Pa, ±2,5 dB.

L'esempio riporta la prima ripetizione della frase cornice *Ich habe Tage gesagt* ('ho detto giorni'), letta dalla parlante AMB. Nel primo livello viene trascritta ortograficamente la parola messa in evidenza dalla frase cornice, nel secondo livello viene indicato il segmento bersaglio e nel terzo livello è visibile la durata delle fasi di occlusione (cl) e di rilascio (rl) della consonante [t].

Di tutte le realizzazioni di /p t k/ e /b d g/ sono quindi state misurate la durata delle fasi di occlusione e di rilascio nonché la durata totale del segmento. Di tutte le durate è stata calcolata la percentuale di periodicità (*%Voice*) nel segnale mediante l'apposita funzione nel *Voice Report* di *Praat*. Le misurazioni e i calcoli sono stati automatizzati tramite uno script programmato da Dieter Studer-Joho. Per l'analisi ci siamo invece concentrati su due parametri, ovvero per [b d g] sulla percentuale di periodicità della durata dell'intero segmento (*%Voice*) e per [p t k] sulla durata della fase di rilascio (rl) ovvero sul VOT.

3.4 L'insegnamento della pronuncia: la 'lezione di fonetica'

La 'lezione di fonetica' è durata 45 minuti ed è stata impartita dalla seconda autrice con l'ausilio di una presentazione *Powerpoint*. Dopo un'introduzione generale sulla fonetica come disciplina scientifica, è stato spiegato il funzionamento delle occlusive sorde e sonore. Gli allievi hanno poi avuto un'esperienza di propriopercezione, in cui sentivano la vibrazione delle pliche vocali toccando con due dita la laringe. Inoltre hanno avuto accesso a una rappresentazione visiva dei vari tipi consonante occlusiva, nella fattispecie di spettrogrammi segmentati. In particolare è stata illustrata non solo la differenza tra italiano e tedesco, ma per la L2 sono state comparate anche le produzioni di alcuni allievi con la pronuncia standard prodotta dal primo autore, rendendo gli studenti attenti sulle differenze.

In sintonia con il succitato modello PPP (*Presentation Practice Production*), dopo questa fase di sensibilizzazione è seguita una fase di esercitazione in cui gli allievi lavoravano in coppia, leggendo ad alta voce le frasi cornice dell'esperimento. Inoltre essi avevano come compito a casa la lettura delle stesse frasi (prima della seconda registrazione).

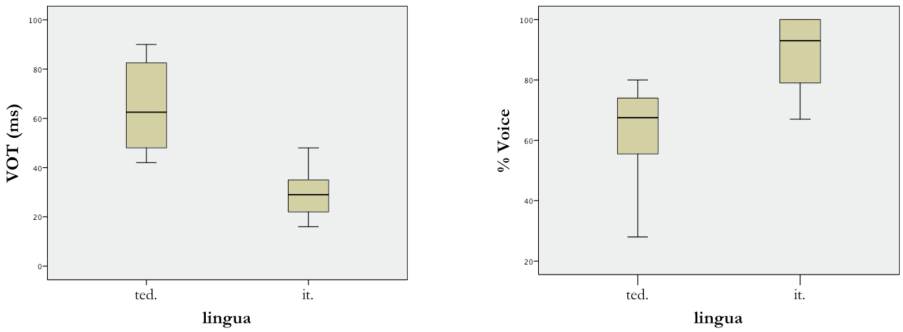
4. *Risultati*

Nella presentazione dei risultati esaminiamo innanzitutto la pronuncia dell'insegnante, che può essere considerato il modello principale di *input* per gli allievi. Successivamente si illustrano i dati ricavati dalla registrazione della prima settimana, comparando i valori *%Voice* e VOT per le due classi e nelle due lingue indagate. Segue infine un confronto degli stessi valori fra la prima e la seconda registrazione, per ciascuna delle due classi, al fine di verificare se la lezione di fonetica impartita nel frattempo nella classe 1L abbia avuto un effetto sulle realizzazioni delle occlusive da parte di questi allievi.

4.1 La natura dell'*input*: la pronuncia dell'insegnante

Un primo aspetto da verificare riguarda la natura dell'*input* al quale gli allievi sono esposti. Evidentemente le fonti di tedesco parlato con cui i giovani ticinesi possono entrare in contatto sono di vario genere (insegnanti precedenti, turisti germanofoni, mass media, ecc.). Tuttavia è probabile che la pronuncia della loro attuale insegnante (la seconda autrice) costituisca in qualche modo il modello di pronuncia che loro cercano di imitare. Vediamo dunque come l'insegnante realizza le occlusive delle due lingue nelle stesse frasi lette dagli allievi:

Figura 2 - *VOT di [p t k] (a sinistra) e %Voice di [b d g] (a destra) nelle due lingue pronunciate dall'insegnante*



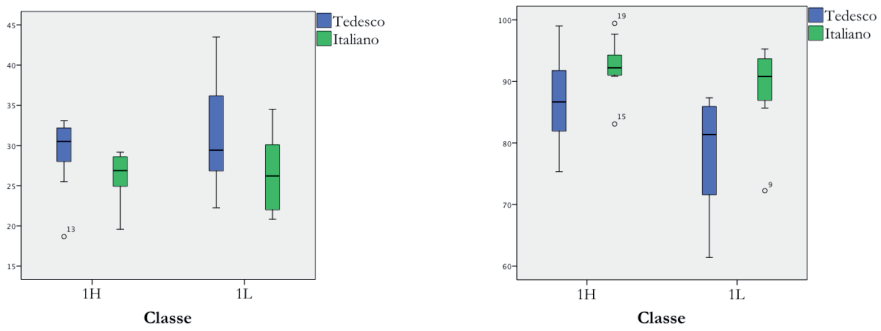
Come si vede dai boxplot riportati nella parte destra del grafico, l'insegnante realizza i fonemi /b d g/ in maniera piuttosto diversa nelle due lingue, pronunciandole in modo molto più sonoro in italiano (con un *%Voice* al di sopra dell'80%) che non in tedesco (il *%Voice* sta sotto l'80%). In effetti, un t-test non accoppiato rivela una differenza significativa tra le due lingue: ($t(22)= -4.641$, $p<0.001$). Altrettanto netta è la differenza del VOT nella pronuncia delle occlusive sorde che viene riportata nella parte sinistra del grafico: qui le durate della fase di rilascio sono sensibilmente più lunghe in tedesco (sopra i 40 ms) che non in italiano (sotto i 40 ms). Di nuovo, il t-test non accoppiato indica una differenza altamente significativa tra le due lingue: ($t(22)=6.301$, $p<0.001$).

In altre parole, possiamo considerare la pronuncia dell'insegnante come molto vicina alla norma standard della lingua bersaglio; un eventuale 'accento italiano' nel tedesco degli allievi non sarà quindi da imputare all'*input* al quale essi sono principalmente esposti.

4.2 La pronuncia delle occlusive nelle registrazioni della prima settimana: confronto tra le due classi (1H e 1L) e le due lingue (italiano e tedesco)

Vediamo innanzitutto come gli allievi pronunciavano le occlusive sorde e sonore del tedesco e dell'italiano nella prima sessione di registrazione, quando nessuna delle due classi aveva avuto un insegnamento esplicito sulla pronuncia.

Figura 3 - *VOT di [p t k] (a sinistra) e %Voice di [b d g] (a destra) nelle registrazioni della prima settimana*



Considerando innanzitutto i boxplot delle occlusive sorde (a sinistra), notiamo un VOT un po' più alto in tedesco e una dispersione leggermente maggiore dei valori (per tutte e due le lingue) nella classe 1L. Un'analisi della varianza mostra solo un lieve effetto per il fattore 'lingua' ($F(1, 18)=13{,}94$, $p<0.05$), ma nessun effetto per il fattore 'classe'. Nei boxplot delle occlusive sonore (a destra) osserviamo invece una dispersione maggiore del *%Voice* in tedesco per ambedue le classi e una percentuale di periodicità nettamente superiore in italiano che non in tedesco. In questo caso, l'analisi della varianza rileva un chiaro effetto per il fattore 'lingua' ($F(1, 18)=30{,}51$, $p<0.001$), mentre non vi è nessun effetto per il fattore 'classe'.

I dati mostrano perciò che non vi è una differenza sensibile tra i 10 allievi della classe 1H e i 10 allievi della classe 1L. Questa situazione di partenza si presta quindi in modo ottimale per il disegno metodologico del nostro esperimento: ricordiamo che l'efficacia dell'insegnamento verrà testata nella classe 1L, mentre la classe 1H servirà come gruppo di controllo.

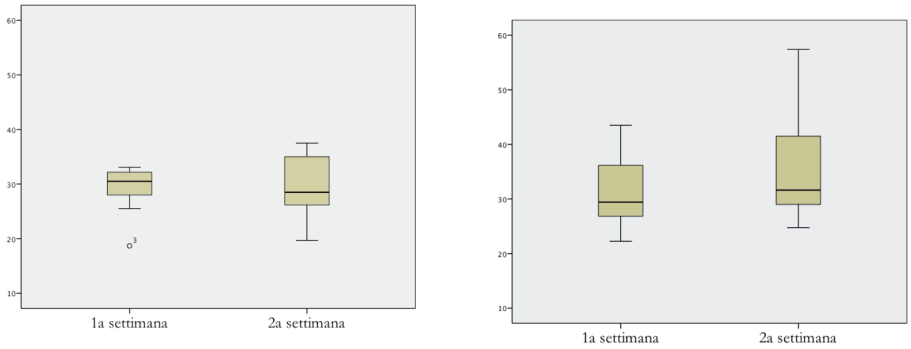
Si osserva che nella Figura 3 i boxplot del tedesco occupano una posizione leggermente diversa rispetto a quelli dell'italiano, il che sta a indicare un tenue avvicinamento verso la lingua bersaglio. Tuttavia le mediane del VOT dei ragazzi si aggirano ancora attorno ai 30 ms, laddove quella della professoressa supera i 60 ms (cfr. Fig. 2). Una differenza analoga tra allievi e insegnante – benché meno evidente – si riscontra anche per la sonorità di [b d g] in tedesco: la mediana del *%Voice* dell'insegnante (cfr. Fig. 2) è già relativamente alta (ca. 70%), ma quella degli allievi si avvicina ancora di più ai valori dell'italiano (più dell'80% in ambedue le classi).

4.3 Confronto del VOT e di *%Voice* nella pronuncia del tedesco tra la prima e la seconda registrazione nelle due classi

Avendo constatato nel sottoparagrafo precedente che gli allievi delle due classi pronunciavano le occlusive del tedesco in modo simile nella prima registrazione, ci accingiamo ora a verificare se l'insegnamento esplicito della pronuncia abbia avuto un

effetto nella classe 1L. I risultati relativi al VOT delle occlusive sorde del tedesco vengono illustrati nella Figura 4:

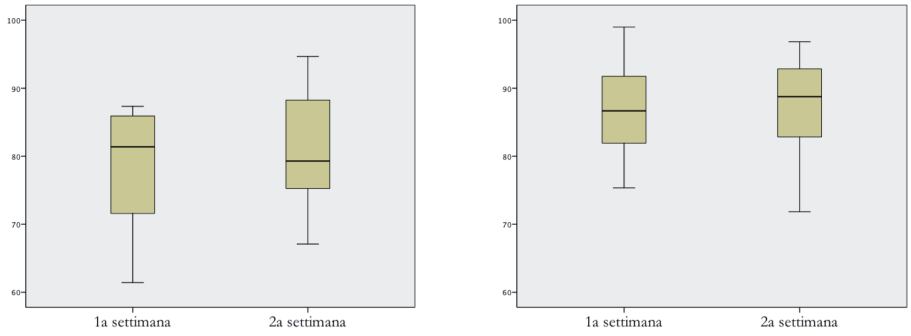
Figura 4 - *VOT (ms) di [p t k] del tedesco nella classe 1L (con insegnamento esplicito, a sinistra) e della classe 1H (senza insegnamento esplicito, a destra)*



Analizzando la pronuncia della classe 1L (con istruzione esplicita, grafico a sinistra), in base alla mediana possiamo persino constatare un abbassamento delle durate del VOT sotto i 30 ms (cioè un 'peggioramento' della pronuncia), ma un t-test accoppiato non mostra nessuna differenza significativa tra le due settimane: t(9)=-1.57, p=0.15; si nota comunque un aumento della dispersione dei valori nella seconda settimana. Anche la classe 1H (senza istruzione esplicita, grafico a destra) mostra un lieve aumento della dispersione dei valori nella seconda settimana, e si nota addirittura un lievissimo aumento del VOT, ma anche in questa classe il t-test accoppiato non rivela nessuna differenza significativa tra le due letture: t(9)= -0.31, p=0.76.

Vediamo ora se l'istruzione esplicita ha avuto un effetto sulla pronuncia delle occlusive 'sonore' del tedesco:

Figura 5 - *%Voice di [b d g] nel tedesco della classe 1L (con insegnamento esplicito, a sinistra) e della classe 1H (senza insegnamento esplicito, a destra)*



A differenza del VOT, nel caso delle occlusive sonore la dispersione dei valori di *%Voice* non sembra essere aumentata in nessuna delle due classi. Anche lo sposta-

mento delle mediane non è notevole, nonostante il leggero calo nella classe 1L (grafico a sinistro) e l'altrettanto leggero aumento nella classe 1H (grafico a destra). In effetti, il t-test accoppiato non fornisce risultati significativi né per la classe 1L che ha ricevuto un'istruzione esplicita prima della seconda lettura ($t(9)=-1.27$, $p=0.24$) né per la classe 1H che ha ricevuto un'istruzione esplicita soltanto dopo la seconda lettura ($t(9)=-0.07$, $p=0.95$).

5. *Discussione e conclusioni*

Volgendoci a qualche riflessione sommaria sui risultati esposti nel paragrafo precedente, notiamo innanzitutto come le misurazioni acustiche effettuate sulle registrazioni della prima settimana (4.2) hanno in effetti evidenziato una certa interferenza del modello italiano sulla pronuncia delle occlusive del tedesco. Ciò è più evidente nel caso delle occlusive sorde, dove gli allievi hanno prodotto un VOT nettamente inferiore rispetto a quello dell'insegnante; ma anche nel caso delle occlusive sonore i valori di *%Voice* sono molto simili nella lettura in tedesco e in italiano. In generale sono quindi state confermate le prime due ipotesi formulate all'inizio della nostra ricerca, dato che gli studenti hanno realizzato delle occlusive sorde prevalentemente non aspirate e delle occlusive sonore con un VOT negativo (2.2).

Nonostante questa evidente interferenza della L1 sulla pronuncia in L2, i boxplot della Figura 3 (relativi alle registrazioni della prima settimana) mostrano anche che gli studenti realizzano le occlusive del tedesco in modo leggermente diverso rispetto a quelle dell'italiano. Infatti, nella loro pronuncia del tedesco la mediana del VOT è più alta nel caso delle occlusive sorde, così come la mediana del *%Voice* è più bassa nel caso delle occlusive sonore; inoltre troviamo una maggiore dispersione dei valori nella L2, il che rispecchia una maggiore variabilità generale delle interlingue. Possiamo comunque interpretare questi indizi acustici come primo passo verso un avvicinamento alla norma della lingua bersaglio.

Per quanto riguarda invece la terza ipotesi sull'efficacia dell'insegnamento esplicito della pronuncia (2.2), i risultati esposti (4.3) non sembrano evidenziare un netto miglioramento nella classe 1L che ha ricevuto tale tipo di istruzione rispetto alla classe 1H che non l'ha avuto, e questa affermazione vale sia per l'attacco del tempo di sonorità nelle occlusive sorde sia per la percentuale di periodicità nelle occlusive sonore.

Nonostante questo risultato generale, corroborato dai test di statistica inferenziale calcolati in base ai valori medi delle due classi, la notevole dispersione dei dati ci induce però a confrontare le misurazioni anche tra i singoli parlanti, almeno nella classe 1L che ha ricevuto un'istruzione esplicita prima della seconda registrazione. La Tabella 2 riporta perciò le medie per tutti gli studenti di questa classe, calcolando anche la differenza tra i valori della prima e della seconda settimana.

Tabella 2 - *Medie del VOT (in ms) nelle occlusive sorde prodotte dai singoli allievi della classe 1L che ha ricevuto un insegnamento esplicito della pronuncia*

Allievo	*1a settimana*	*2a settimana*	*Differenza*
AL	36	32	-5
BD	22	38	16
CT	27	25	-2
GC	30	32	2
ME	29	29	0
MF	29	44	15
MN	25	29	4
RD	33	42	9
RON	40	31	–10
TL	44	57	14
Media	31	36	4

Certo, nella media delle medie di tutti gli allievi lo scarto non è rilevante (4 o 5 millisecondi a seconda della procedura di arrotondamento), e ben tre allievi (AL, CT e RON) hanno prodotto persino un VOT più breve nella seconda settimana. D'altro canto vanno però valorizzate le medie evidenziate in rosso (degli allievi BD, MF, TL e RD), le quali mostrano un notevole aumento del VOT tra la prima e la seconda registrazione. I dati della Tabella 2 mostrano che 4 soggetti su 10 realizzano effettivamente le occlusive sorde con un VOT più lungo nella seconda registrazione rispetto alla prima. Pur rappresentando la minoranza dei casi, questo tipo di evoluzione può essere visto come prova del fatto che, a livello individuale, l'esperienza didattica può sortire degli effetti positivi.

Se quindi a prima vista i nostri risultati ci inducono a rispondere negativamente alla domanda di ricerca principale (vale a dire "no, l'insegnamento fonetico esplicito non serve a migliorare la pronuncia"), uno sguardo alle realizzazioni dei singoli allievi ci permette di rilevare notevoli differenze individuali che potrebbero essere determinate da diversi stili di apprendimento di una lingua straniera (cfr. Chini, 2005: 64-65). Non tutti gli apprendenti sembrano essere ugualmente pronti a costruire un'interfaccia tra il sapere metalinguistico e la competenza fonetica nella L2; nel nostro esperimento, solo una minoranza – quattro allievi su dieci – si è dimostrata sensibile a questo tipo di insegnamento. A nostro avviso sembra comunque prematuro bandire del tutto l'istruzione esplicita della pronuncia dall'aula di lingua straniera: forse un sano pluralismo negli approcci didattici (ad esempio con l'aggiunta di un metodo di tipo verbo-tonale) permetterebbe di raggiungere anche discendenti con capacità di apprendimento eterogenee.

Avviandoci a qualche breve osservazione conclusiva, occorre evidenziare i limiti che il nostro esperimento condivide – purtroppo – con buona parte della ricerca sull'insegnamento della pronuncia (cfr. Thomson, Derwing, 2015). Tra gli aspetti problematici vanno menzionati l'impiego di un solo metodo didattico (presentazione pratica produzione) e l'analisi di parlato letto, a pochi giorni di distanza dalla

lezione di fonetica. Sarebbe quindi importante non solo testare l'efficacia di una metodologia didattica alternativa, ma analizzare anche la produzione degli allievi nel parlato spontaneo e a una distanza temporale maggiore. È evidente che occorrono ulteriori ricerche con metodi più affinati sull'insegnamento della pronuncia nelle L2.

Ringraziamenti

Il nostro primo ringraziamento va ai venti allievi delle classi 1H e 1L che hanno collaborato con entusiasmo all'esperimento di lettura. Teniamo a ringraziare il Direttore del Liceo Cantonale di Bellinzona, il professor Omar Gianora, per la sua disponibilità e per l'atteggiamento estremamente positivo nei confronti di questo studio svolto con gli allievi.

Siamo poi debitori nei confronti di due colleghi di lavoro per il valevole sostegno, senza il quale il presente lavoro non avrebbe potuto essere realizzato: grazie a Dieter Studer-Joho per aver programmato lo script di *Praat* che ci ha permesso di automatizzare l'analisi acustica, e un grazie di cuore anche a Sandra Schwab per la preziosa consulenza nel trattamento statistico dei dati. Ringraziamo infine tre revisori anonimi per le osservazioni critiche e i numerosi suggerimenti che ci hanno permesso di migliorare il testo e di chiarire alcuni punti; naturalmente, la responsabilità di eventuali debolezze rimaste in questo lavoro ricade unicamente sui due autori.

Riferimenti bibliografici

Albano Leoni, F., Maturi, P. (2011). *Manuale di fonetica.* Terza edizione. Roma: Carocci.

Beckman, J., Jessen, M. & Ringen, C. (2013). Empirical Evidence for Laryngeal Features: German vs. True Voice Languages. In *Journal of Linguistics*, 49, 259-284.

Boersma, P., Weenink, D. (2015). Praat: doing phonetics by computer. Versione 5.4.09. http://www.praat.org/Accessed 15.06.15.

Canepari, L. (1979). *Introduzione alla fonetica.* Torino: Einaudi.

Chang, C.B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. In *Journal of Phonetics*, 40, 249-268.

Chini, M. (2005). *Che cos'è la linguistica acquisizionale.* Roma: Carocci.

Cho, T., Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. In *Journal of Phonetics*, 27, 207-229.

Consiglio d'Europa (2010). *Quadro comune europeo di riferimento per le lingue: apprendimento insegnamento valutazione.* Terza ristampa. Oxford: La Nuova Italia.

Drachsler, C., Jänsch, K. (2004). SpeechRecorder. http://www.bas.unimuenchen.de/Bas/software/speechrecorder/Accessed 15.06.15.

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

Gass, S., Selinker, S. (2008). *Second Language Acquisition. An Introductory Course* (Third Edition). London: Routledge.

Kohler, K. (1999). German. In *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 86-89.

Lado, R. (1957). *Linguistics Across Cultures.* Ann Arbor: University of Michigan Press.

Lauefer, C. (1997). Towards a typology of bilingual phonological systems. In James, A., Leather, J. (Eds.), *Second language speech. Structure and process.* Berlin: de Gruyter, 325-342.

Lisker, L., Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustic Measurements. In *Word*, 20, 527-565.

Maturi, P. (2006). *I suoni delle lingue, i suoni dell'italiano*. Bologna: il Mulino.

Pedrazzini, G. (2015). Eine kontrastive Analyse der VOT im Italienischen und Deutschen bei italienischen Muttersprachlern aus dem Tessin. Manoscritto, Università di Zurigo.

Reetz, H. (1999). *Artikulatorische und akustische Phonetik.* Trier: Wissenschaftlicher Verlag.

Saville-Troike, M. (2006). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.

Schmid, S. (1999). *Fonetica e fonologia dell'italiano.* Torino: Paravia.

Schmid, S. (2012). The pronunciation of voiced obstruents in L2 French: a preliminary study of Swiss German learners. In *Poznan Studies in Contemporary Linguistics*, 48(4), 627-659.

Schmid, S., Wachter, S. (2015). Le ostruenti sonore nella pronuncia dell'italiano di apprendenti svizzero-tedeschi. In *Studi AISV*, 1, 203-217.

Thomson, R.I., Derwing, T.M. (2015). The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. In *Applied Linguistics*, 36(3), 326-344.

DOI: 10.17469/O2102AISV000004

SONIA D'APOLITO, BARBARA GILI FIVELA

Targetless schwa in francese L2: primi risultati in area italofona

This paper focuses on how non-native consonant clusters are produced by Italian learners of French L2 depending on speech rate, prosodic conditions and sequence type. We compare the productions by three Italian learners and one native speaker, paying particular attention to epenthetic vowels, which are possibly inserted between the two consonants. We examine the acoustic and articulatory characteristics of these contexts to observe how consonant gestures are coordinated and if the insertion of a vowel corresponds to a full vowel insertion (i.e. there is an articulatory target) or to a gestural mistiming (i.e. there is no articulatory specific vowel target). Results show that the inserted vowel has a high variability in acoustic quality and duration; moreover, on the articulatory level, no target has been identified for the tongue dorsum gesture and different patterns are actually realized because of the influence of speech rate, prosodic conditions and sequence type. Our results suggest that vowel insertion in Italian speakers production seems to be due to the failure in reaching a correct coordination between the two consonantal gestures (gestural mistiming).

Key words: inserzione vocalica, vocale epentetica e intrusiva, coordinazione intragestuale, dati articolatori EMA, francese L2.

Introduzione

L'inserzione vocalica può avvenire in base a due meccanismi articolatori: i) inserzione di un gesto vocalico articolatorio – vocale epentetica; ii) un gesto di transizione tra due gesti consonantici che produce un suono vocalico – vocale intrusiva. Secondo Hall (2003; 2006; 2011), le vocali intrusive mostrano una qualità variabile, poiché possono avere le caratteristiche di una vocale schwa, oppure essere più simili alla vocale adiacente o essere influenzate dal luogo di articolazione delle consonanti adiacenti. Peraltro, Levin (1987) afferma che una vocale intrusiva può anche esibire le qualità di una vocale non presente nel sistema linguistico, quindi, una vocale schwa può essere realizzata tra due consonanti benché non appartenga al sistema linguistico in esame. Inoltre, la durata della vocale è variabile ed è fortemente influenzata dalla velocità di eloquio, al punto che a velocità sostenuta una vocale intrusiva non si realizza. La sua funzione è quella di favorire la percezione dei segmenti di un nesso consonantico. Al contrario, una vocale epentetica ha una qualità fissa o caratteristiche simili alla vocale adiacente, il suo inserimento non è influenzato dalla velocità di eloquio e ha la funzione è di riparare una struttura marcata.

Per quanto riguarda la durata acustica, una vocale epentetica ha una durata più lunga di una vocale intrusiva. Ad esempio, nella lingua Salish si realizzano entrambe le vocali,

ossia degli schwa epentetici, fonologicamente presenti sia dal punto di vista acustico che percettivo, e degli schwa intrusivi che si presentano come elementi di transizione. Uno studio acustico ha dimostrato che le vocali epentetiche hanno effettivamente una durata maggiore rispetto a quelle intrusive (Shanin, Kimary & Black, 2004). Visto che la durata di una vocale epentetica è maggiore, ci si aspetta quindi che anche la durata di una sequenza sia maggiore in caso di inserimento di una vocale epentetica. Davidson, Roon (2008) in effetti hanno osservato nessi consonantici in posizione iniziale nella lingua russa e i loro risultati confermano che sequenze #CəC, con inserimento di uno schwa fonologico, hanno una durata maggiore rispetto alle sequenze con uno schwa intrusivo. La durata delle sequenze è stata anche osservata confrontando lo schwa intrusivo con schwa lessicalizzati (con un target sottostante), in produzioni da parte di parlanti americani di sequenze non native della lingua polacca (es./sC-, səC- zC-/; Davidson, 2005) e della lingua ceca (es. /C_1əC_2/, in cui C_1 è una fricativa combinata con tutte le possibili consonanti; Davidson, 2006). I risultati mostrano che le sequenze con uno schwa intrusivo hanno una durata minore rispetto alle sequenze con schwa lessicalizzato. Inoltre, lo schwa intrusivo presenta valori di F1 e F2 più bassi rispetto allo schwa lessicalizzato, per via della coarticolazione con la vocale successiva. Davidson interpreta questi risultati affermando che uno schwa intrusivo si realizza attraverso un *gestural mistiming*, cioè i parlanti falliscono nel coordinare in modo appropriato, di fatto in termini di sovrapposizione, i gesti consonantici della sequenza. A supporto di questa interpretazione ci sono anche gli studi con dati ricavati mediante *Ultrasound Tongue Imaging* (Davidson, 2005; Davidson, Stone, 2003), in cui è stata osservata la conformazione della lingua durante la produzione di [zᵊC], con schwa intrusivo, di [səC], con schwa lessicalizzato e di [sC]. I risultati mostrano che la configurazione della lingua durante la produzione di [zᵊC] è più simile a [sC] rispetto a [səC] poiché [s] seguita da [ə] ha una posizione di partenza più bassa nel cavo orale essendo coarticolata con la vocale [ə] che ha una posizione del corpo della lingua più bassa. Al contrario, nelle sequenze [sC] e [zᵊC], la fricativa alveolare si coarticola direttamente con la consonante successiva. Ad esempio, [s] seguita da [k] assume una posizione più alta poiché il dorso della lingua durante [k] è innalzato, sia nella sequenza [sk] che [zᵊk] con schwa intrusivo (Davidson, 2005). I risultati dimostrano che i parlanti non necessariamente riparano una sequenza marcata con uno schwa epentetico, ma falliscono nel coordinare i gesti consonantici, per cui tra le due consonanti si crea uno spazio in cui si realizza uno schwa di transizione. I risultati dello studio precedente, indicano inoltre che uno schwa intrusivo non è specificato da un proprio target articolatorio dal momento che le due consonanti si coarticolano direttamente, al contrario di quanto accade nel caso in cui lo schwa sia una vocale vera e propria. Infatti, come abbiamo detto precedentemente, una vocale epentetica si realizza con un gesto articolatorio con *target* vocalico, mentre una vocale intrusiva si realizza grazie a un gesto di transizione tra due gesti consonantici. Infine, Browman e Goldstein (1992) hanno osservato il movimento della lingua durante la realizzazione di una vocale schwa in posizione mediana all'interno della pseudo-parola [pipəpapə] in americano. La posizione della lingua durante la produzione dello schwa mediano non è una semplice interpolazione tra consonanti e vocali adiacenti. Questo fa pensare che

in inglese non si possa escludere la presenza di un target per lo schwa. In altre lingue, invece, come l'arabo marocchino, la vocale schwa sembra derivare da gesti consonantici non sufficientemente sovrapposti (Gafos, 2002) e non dalla presenza di un bersaglio articolatorio di tipo vocalico. Per quanto riguarda la lingua francese, Rialland (1986) distingue due tipi di schwa individuati in: 1) un morfema interno che ha un nucleo fonologico e 2) un morfema finale che deriva dal rilascio acustico della consonante. Non ci risulta però che siano disponibili dati articolatori relativi alla presenza o meno di bersagli per le vocali.

In base alla Fonologia articolatoria (Browman, Goldstein, 1986; 1987; 1989; 1992; 2007; Browman, 1995), la presenza di un *target* per la vocale e la scarsa sovrapposizione di gesti consonantici corrispondono alla presenza, rispettivamente, di vocali specificate e non specificate: una vocale non specificata non ha *target* articolatorio (*targetless*), mentre una vocale specificata ha un proprio bersaglio. Pertanto, nella produzione di sequenze consonantiche non-native potrebbe realizzarsi questo tipo di meccanismo, cioè i parlanti, per riparare un nesso marcato (Eckman, 2008), potrebbero anche realizzare un tipo di coordinazione caratterizzato da un basso livello di sovrapposizione gestuale tra i gesti consonantici che fa sì che si crei uno spazio vocalico aperto sufficiente per l'inserzione vocalica (Davidson, 2005).

1. *Obiettivi e Ipotesi*

L'obiettivo di questo studio acustico-articolatorio è quello di osservare le sequenze di sibilanti (alveolare-postalveolare - AP - e postalveolare-alveolare - PA) a confine di parola, realizzate mediante l'inserimento di una vocale d'appoggio da parte di apprendenti italiani di francese L2 avanzato. In italiano, le sequenze di consonanti al confine di parola compaiono in pochi casi, principalmente in preposizioni, prestiti stranieri, o nei verbi in la vocale finale sia stata troncata (Muliačić, 1973; Farnetani, Busà, 2004). In francese, invece, le sequenze consonantiche a confine di parola sono molto frequenti e quelle di sibilanti possono anche essere dominio di assimilazioni di luogo (Niebuhr, Lancia & Meunier, 2008). Di conseguenza, gli apprendenti avranno difficoltà nel realizzare queste sequenze e una delle possibili strategie per riparare i nessi potrà essere quella di inserire una vocale d'appoggio. Peraltro, in molte varietà di francese, tra cui il francese del Sud della Francia, è previsto uno schwa e nella didattica del francese in Italia molti insegnanti producono questo schwa. Di conseguenza, la maggior parte degli italofoni apprendenti francese è convinta che lo schwa sia standard. Del resto alcuni materiali didattici trascrivono tuttora parole (es. *tasse*) con uno schwa finale. In particolare, in questo studio si vuole osservare dal punto di vista articolatorio l'*output* acustico corrispondente a una eventuale vocale d'appoggio per capire se esso sia dovuto all'inserimento di una vocale specificata da un proprio *target* articolatorio – vocale epentetica con *target* – oppure sia *targetless* – vocale intrusiva dovuta a un *gestural mistiming*. L'ipotesi è che l'inserzione vocalica sia dovuta ad un'epentesi come strategia generale adottata dagli apprendenti in base alle caratteristiche fonetiche e fonologiche della L1. Tuttavia, dal punto di vista articolatorio, non si esclude un processo di *gestural*

mistiming dovuto alla difficoltà da parte degli apprendenti di coordinare in modo appropriato i due gesti consonantici. Inoltre, la produzione delle sequenze sarà osservata al variare della struttura prosodica e della velocità di eloquio. Circa l'influenza della condizione prosodica, la presenza di un confine prosodico interferisce con il processo coarticolatorio (Byrd, Choi, 2006), mentre una velocità di eloquio sostenuta facilita la coproduzione di segmenti successivi (Byrd, Tan, 1996). Ci si aspetta, quindi, che l'inserzione vocalica, intrusiva e epentetica, possa essere favorita in caso di confine intonativo mentre potrebbe ridursi, se non addirittura scomparire, a velocità sostenuta. Infine, circa il tipo di sequenza ci si aspetta l'inserimento di una vocale d'appoggio indipendentemente dall'ordine del luogo di articolazione.

2. *Metodo*

Le sequenze oggetto di indagine, che riguardano le fricative alveolari e post-alveolari /sʃ - ʃs - sʒ - ʒs - zʃ - zʒ/, sono state studiate all'interno del contesto vocalico /a_i/, al confine di parole inserite all'interno di una frase cornice. Gli stimoli sono stati osservati variando la struttura prosodica, poiché, come già specificato, la presenza di un confine può interferire con il processo coarticolatorio riducendone la sovrapposizione (Byrd, Choi, 2006), e la velocità di eloquio, poiché una velocità sostenuta facilita la coarticolazione (Byrd, Tan, 1996). Per quanto riguarda la struttura prosodica, sono stati creati due contesti, inserendo gli stimoli all'interno di una frase con: 1) confine di sintagma fonologico (es. *Il dit t**asse chi**noise rapidement* – Dice tazza cinese rapidamente) e con 2) confine intonativo (es. *D'abord il a dit t**asse. Chi**noise l'a dit après*" – Prima ha detto tazza. Cinese l'ha detto dopo). A velocità sostenuta di eloquio e in caso di confine fonologico è possibile attendersi una ristrutturazione, mentre nessuna ristrutturazione è prevista quando sia presente un confine intonativo (cfr. fonologia prosodica, Nespor, Vogel, 1986).

In una fase preliminare, la naturalezza di tutte le frasi è stata verificata da 4 parlanti nativi (due studenti Erasmus provenienti da Nantes di età di 24 e 25 anni e due lettori di francese della facoltà di lingue dell'Università del Salento). Successivamente, tre studentesse (PI1, PI2 e PI3) della facoltà di Lingue e Letterature straniere dell'Università del Salento apprendenti di francese avanzato L2[1] (PI1, PI2, PI3) e una studentessa Erasmus di Nantes (PF4)[2] hanno letto per 7 volte il corpus, una volta a velocità normale e una volta a velocità sostenuta (per un totale di 168 frasi a parlante= 6 sequenze x due

[1] Al momento della raccolta dati, PI2 svolgeva un dottorato in linguistica francese presso l'Università del Salento, mentre PI3 aveva perseguito l'abilitazione SSIS in francese e collaborava presso la stessa Università. Il soggetto PI1 frequentava il primo anno di specialistica presso l'Università del Salento. Tutti e tre i soggetti avevano effettuato un soggiorno all'estero di sei mesi per Erasmus. Il parlante PI1 è stato a Nantes, il parlante PI2 è stato a Reims ritornando più volte in Francia per ricerche e PI3 si è recata a Saarbrücken (in Germania, ma vicino al confine francese), dove ha praticato comunque la lingua francese.

[2] Nelle varietà di francese dell'area settentrionale lo schwa inserito non ha una rappresentazione sottostante per cui o è assente o mostra una grande variabilità (Andreassen, Durand & Racine, 2016; Lyche, 2016), al contrario di quanto si riscontra per le varietà di francese dell'area meridionale, in cui lo schwa è mantenuto (Andreassen et al., 2016). Il nostro parlante nativo PF4 proviene dall'area settentrionale per cui lo schwa non ha una rappresentazione specificata. Ai fini del nostro studio questo aspetto

contesti prosodici x 2 velocità di eloquio x 7 ripetizioni). I materiali acustici e articolatori sono stati raccolti presso il CRIL (Università del Salento). I dati articolatori sono stati registrati utilizzando l'articulografo AG-500, posizionando i sensori sulle parti di interesse: asse sagittale della lingua (4); labbro inferiore e superiore (2); incisivi superiori e inferiori (2) e dietro le orecchie (2). Il segnale audio è stato registrato simultaneamente con una scheda audio Edirol a 44.1kHz. I dati sono stati analizzati dal punto di vista percettivo, acustico e articolatorio utilizzando scripts di PRAAT (Boersma, Weenink, 2010) e MatLab (Sigona, Stella, Grimaldi & Gili Fivela, 2015).

2.1 Analisi percettiva

Un'analisi percettiva è stata effettuata dai due autori per verificare, in modo autonomo, se il confine prosodico atteso fosse stato realizzato e la realizzazione delle sequenze e fenomeni (v. presenza di vocale epentetica).

2.2 Analisi acustica e misurazioni

L'etichettatura acustica ha riguardato i segmenti della sequenza $V_1C_1\#C_2V_2$, (ad es. /as/#/ʃi/) includendo una possibile vocale d'appoggio (V0) e/o una pausa (P) al confine tra le due consonanti. Le misurazioni hanno riguardato: i) durata di ciascun segmento; ii) durata della sequenza; iii) F1 e F2 per /a/, /i/ e l'eventuale vocale d'appoggio V0, calcolate nella porzione stabile della vocale, cioè nella porzione in cui i correlati acustici non variavano (si assumeva dunque che la conformazione linguale avesse raggiunto il *target* e l'influenza del contesto adiacente fosse minima). Inoltre sono state calcolate: iv) velocità di eloquio e di articolazione, ossia il rapporto tra la durata della sequenza e il numero di sillabe prodotte (includendo le pause per calcolare la velocità di eloquio ed escludendole per ottenere la velocità di articolazione).

2.3 Analisi articolatoria e misurazioni

L'etichettatura articolatoria ha riguardato il gesto di chiusura e apertura, sia per quanto riguarda i dati di posizione che per i picchi di velocità. Sono state considerate le traiettorie relative ai:

- movimenti della punta della lingua (*Tongue Tip*, TT), asse verticale (asse z) e orizzontale (asse x), per entrambe le fricative (alveolare e postalveolare), essendo consonanti coronali;
- movimenti del dorso della lingua (*Tongue Dorsum*, TD), asse verticale (asse z) e orizzontale (asse x), per il passaggio vocalico V-to-V ([a]- [i]);
- movimenti del labbro inferiore (*Lower Lip*, LL), lungo orizzontale (asse x), per la protrusione della fricative postalveolare, sebbene anche le fricative alveolari possano presentare un proprio gesto di protrusione (Perkell, 1986 in Kühnert, Nolan, 1999; Engwall, 2000).

---

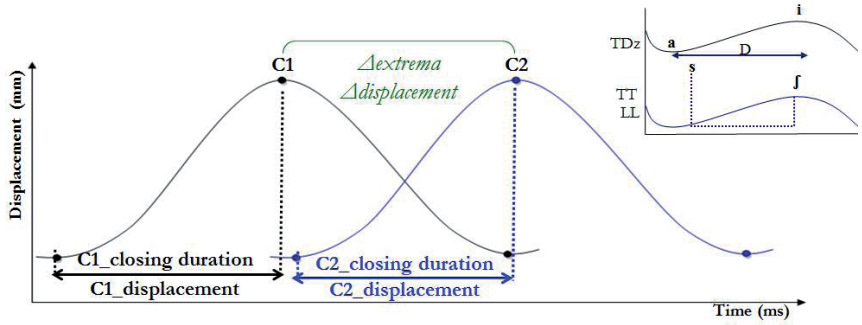
risulta particolarmente utile, poiché ci permette di confrontare e capire meglio le realizzazione degli apprendenti anche sulla base della variabilità fonetica nelle produzione del parlante nativo.

Durante la fase di etichettatura si è osservato principalmente il dorso della lingua per individuare i *target* relativi alle vocali, essendo l'articolatore maggiormente coinvolto per la loro realizzazione. Per le fricative sono state osservate, invece, le traiettorie della punta della lingua e del labbro inferiore.

Sono state calcolate successivamente le seguenti misurazioni:

– durata (ms) e ampiezza (mm) del gesto di chiusura per ciascuna fricativa (Figura 1: *C1/C2 closing_duration; C1/C2 closing_displacement*) (Byrd, Kaun, Narayanan & Saltzman, 2000);
– durata intervallo (ms) e ampiezza (mm) tra il *target* della seconda consonante (C2) e il target della prima consonante (C1) (Figura 1: *Δextrema; Δdisplacement*) (Byrd et al., 2000);
– fase relativa (ms): il rapporto tra l'intervallo tra C2-C1 e la durata del passaggio vocalico [a]-[i] (Tiede, Shattuck-Hufnagel, Johnson, Ghosh, Mattheis, Zandipur & Perkell, 2007) (Figura 1: riquadro in alto a destra).

Figura 1 - *Schema delle misurazioni articolatorie*



Le analisi statistiche sono state effettuate in SPSS. Le variabili in esame non presentano una distribuzione normale per cui si è scelto di procedere con i test non parametrici Kruskal-Wallis e Mann-Whitney, effettuando le analisi separatamente per soggetto. In questo lavoro, saranno presentate solo le analisi statistiche relative alla differenza di durata tra i *target* delle due fricative (*Δextrema*) al fine di osservare come i due gesti si coordinano in base alla presenza o meno di una vocale d'appoggio. Questa misurazione è rappresentata nella figura in basso. Tale misurazione è stata calcolata relativamente allo stesso articolatore e asse quando erano presenti entrambi i *target* consonantici (ad esempio, entrambi realizzati grazie alla punta della lingua TT o al labbro inferiore LL), in altri casi si è proceduto incrociando le informazioni o all'interno dello stesso articolatore guardando assi diversi (ad esempio punta della lingua asse verticale TTz e punta della lingua asse orizzontale TTx) o tra due articolatori diversi (ad esempio punta della lingua e labbro inferiore).

Figura 2 - *Rappresentazione grafica delle misurazioni articolatorie*

Prima di passare ai risultati dei dati articolatori, presentiamo i risultati relativi alla frequenza con la quale i parlanti inseriscono una vocale d'appoggio, e ai risultati sulla durata e qualità di tale vocale.

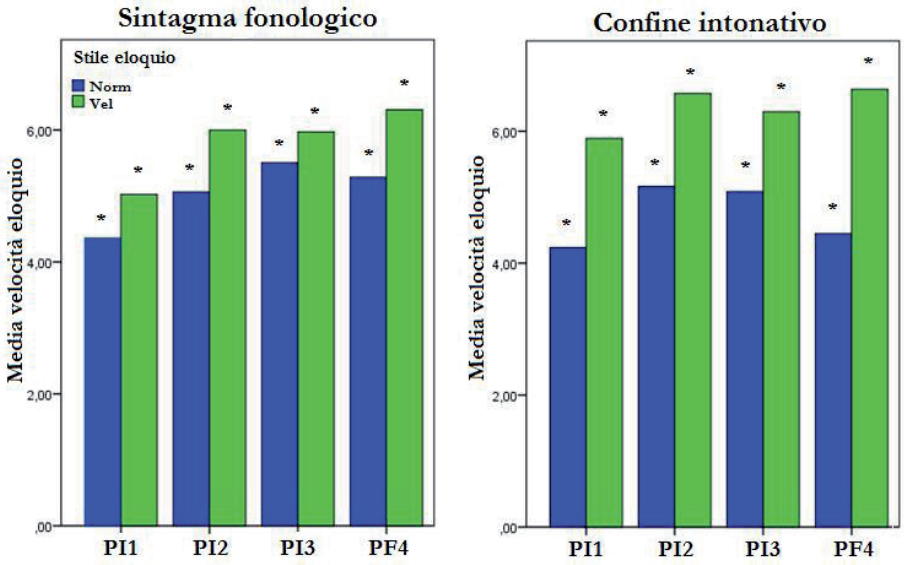
3. *Risultati*

3.1 Risultati acustici

3.1.1 Velocità di eloquio

Tutti i parlanti hanno effettivamente variato lo stile di eloquio Il test di Wilcoxon, infatti, indica una differenza statistica per tutti i parlanti, sia per le produzioni con sintagma fonologico [PI1: Z=–5.256 p=,000; PI2: Z=–5.287 p=,000; PI3: Z=–3.589 p=,000; PF4: Z=–5.309 p=,000; PF5: Z=–5.483 p=,000] che con confine intonativo [PI1: Z=–5.232 p=,000; PI2: Z=–5.086 p=,000; PI3: Z=–5.579 p=,000; PF4: Z=–5,656 p=,000; PF5: Z=–5.590 p=,000].

Figura 3 - *Velocità di eloquio in caso di sintagma fonologico (a sinistra) e di confine intonativo (a destra) per tutti i parlanti*



La velocità di articolazione è stata calcolata per confrontare le produzioni degli apprendenti con quella del nativo. Una serie test di Mann-Whitney, effettuando

confronti incrociati (velocità normale degli apprendenti *vs* velocità normale del francofono; velocità sostenuta degli apprendenti *vs* velocità normale e sostenuta del francofono), mostra una differenza significativa (p=,000).

3.1.2 Frequenza dell'inserimento di vocale

A velocità normale di eloquio, tutti i parlanti inseriscono una vocale tra le due consonanti, per entrambe le condizioni prosodiche e per le due sequenze. Le tabelle riportate di seguito mostrano le frequenze relative all'inserimento di una vocale per gli apprendenti (Tabella 1) e per il parlante nativo (Tabella 2). Come si può osservare, tra gli apprendenti solo il parlante PI3 realizza alcuni casi in cui non inserisce alcuna vocale, per la sequenza AP in assenza di confine prosodico (5 casi su 30). PF4, invece, in presenza di confine intonativo realizza sempre una pausa che può essere preceduta o meno da una vocale.

Tabella 1 – *Frequenze relative all'inserimento di una vocale d'appoggio da parte degli apprendenti italofoni a velocità normale di eloquio e nelle due condizioni prosodiche*

Velocità normale di eloquio – Apprendenti italofoni							
Sintagma fonologico							
Alveolare-postalveolare				*Postalveolare-alveolare*			
P	Schwa	Tot	%	P	Schwa	Tot	%
PI1	28	28	100	PI1	14	14	100
PI2	29	29	100	PI2	14	14	100
PI3	25	30	83,3	PI3	14	14	100
Confine intonativo							
PI1	28	28	100	PI1	13	13	100
PI2	27	27	100	PI2	14	14	100
PI3	26	26	100	PI3	14	14	100

Tabella 2 - *Frequenze relative all'inserimento di una vocale d'appoggio da parte del parlante nativo a velocità normale di eloquio e nelle due condizioni prosodiche*

Velocità normale di eloquio – parlante nativo							
Sintagma fonologico							
Alveolare-postalveolare				*Postalveolare-alveolare*			
P	Schwa	Tot	%	P	Schwa	Tot	%
PF4	28	28	100	PF4	14	14	100
Confine intonativo							
PF4	16	28	57,1	PF4	5	14	35,7

A velocità sostenuta, tra gli apprendenti si distingue PI1 che, come mostra la Tabella 3, per la sequenza AP in presenza di confine intonativo realizza una vocale all'interno del nesso solo in 3 casi. In tutti gli altri contesti, PI1 realizza assimilazioni di luogo. I parlanti PI2 e PI3, al contrario, continuano a inserire una vocale soprattutto in presenza di confine intonativo. Come mostra la Tabella 4, invece, PF4 produce una vocale solo per la sequenza AP in caso di un confine intonativo, in tutti gli altri casi realizza assimilazioni di luogo[3].

Tabella 3 - *Frequenze relative all'inserimento di una vocale d'appoggio da parte degli apprendenti italofoni a velocità sostenuta di eloquio e nelle due condizioni prosodiche*

Velocità sostenuta di eloquio – Apprendenti italofoni							
Sintagma fonologico							
Alveolare-postalveolare				*Postalveolare-alveolare*			
Spk	Schwa	Tot	%	Spk	Schwa	Tot	%
PI1	0	26	–	PI1	0	12	–
PI2	7	30	23,3	PI2	8	17	47,1
PI3	9	29	31,0	PI3	8	14	57,1
Confine intonativo							
PI1	3	24	12,5	PI1	0	11	–
PI2	12	24	50	PI2	7	11	63,6
PI3	16	27	59,3	PI3	11	14	78,6

Tabella 4 - *Frequenze relative all'inserimento di una vocale d'appoggio da parte del parlante nativo a velocità sostenuta di eloquio e nelle due condizioni prosodiche*

Velocità sostenuta di eloquio – parlante nativo							
Sintagma fonologico							
Alveolare-postalveolare				*Postalveolare-alveolare*			
Spk	Schwa	Tot	%	Spk	Schwa	Tot	%
PF4	0	28	-	PF4	0	13	-
Confine intonativo							
PF4	12	20	42,9	PF4	0	13	-

[3] Per i risultati sulla realizzazione delle assimilazioni di luogo da parte di questi soggetti si rimanda ai lavori di d'Apolito (2012) e d'Apolito, Gili Fivela (2013).

3.1.3 Durata del segmento

A velocità normale di eloquio e in assenza di confine prosodico, la durata del segmento vocalico si colloca con maggiore frequenza tra i 50 ms e i 110 ms, mentre in presenza di confine intonativo si ha una maggiore variabilità per cui l'intervallo di durata oscilla tra i 70-150 ms per PI1 e PI2 e tra 50-110 ms per PI3 e PF4. A velocità sostenuta di eloquio, per PI2 e PI3 la vocale ha una durata tra i 30 ms e gli 80 ms in entrambe le condizioni prosodiche. Per PI1, nei pochi casi in cui inserisce una vocale, si osserva una durata variabile tra 20-100 ms; per il parlante francofono (PF4), il range è, invece, tra 50-140 ms.

Figura 4 - *Durata del segmento vocalico (ms) a velocità normale (a sinistra) e sostenuta (a destra) di eloquio, nelle due condizioni prosodiche per tutti i parlanti*

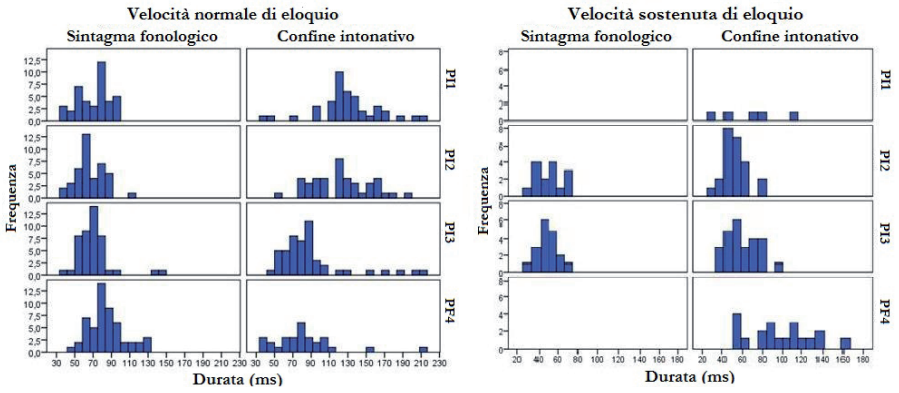
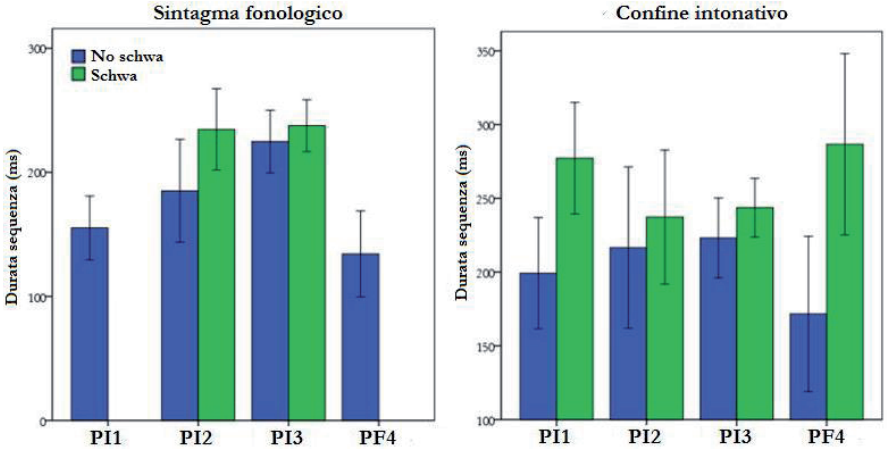


Tabella 5 - *Valori medi (in ms) e deviazioni standard relativi alla durata del segmento vocalico V0 in caso di sintagma fonologico (a sinistra) e di confine intonativo (a destra) per tutti i parlanti a velocità normale (sopra) e sostenuta (sotto) di eloquio*

Velocità normale			
Sintagma fonologico		*Confine intonativo*	
Spk	Media (dev. st.)	Spk	Media (dev. st.)
PI1	66.83 (22.39)	PI1	130.46 (35.11)
PI2	68.12 (21.43)	PI2	121.75 (32.24)
PI3	64.45 (25.24)	PI3	91.35 (42.29)
PF4	82.25 (21.90)	PF4	42.52 (48.47)
Velocità sostenuta			
PI1	0	PI1	5.48 (17.64)
PI2	20.61 (26.83)	PI2	28.62 (26.85)
PI3	20.20 (24.75)	PI3	37.54 (31.22)
PF4	0	PF4	49.74 (53.34)

Per le produzioni a velocità di eloquio sostenuta è stato anche osservato se la presenza o meno di una vocale d'appoggio contribuisse alla maggiore durata della sequenza. Come si può osservare dal grafico in Figura 5, nel caso in cui venga realizzata una vocale tra le due fricative, la durata della sequenza è maggiore rispetto alla sequenza senza vocale. Un T-test a campioni indipendenti indica che in assenza di confine prosodico la significatività è raggiunta solo per PI2 ($t_{(39)}$=-3,680; p=,001). In presenza di confine intonativo, la differenza è significativa per PI1 ($t_{(40)}$=-4,688; p=,000), per PI3 ($t_{(40)}$=-2,825; p=,000) e per PF4 ($t_{(40)}$=-6,513; p=,000).

Figura 5 - *Durata della sequenza (ms) a velocità sostenuta di eloquio in base alla presenza (verde) e assenza (blu) del segmento vocalico*



3.1.4 Formanti F1 e F2

Al fine di osservare le caratteristiche qualitative del segmento vocalico realizzato tra le due fricative, ne sono state calcolate le prime due formanti. I grafici scatterplot riportati in basso rappresentano le aree di esistenza delle vocali realizzate all'interno della sequenza *target*, cioè [a] (in blu), [i] (in verde) e V0 (in rosso), a velocità di eloquio normale (Figura 6) e sostenuta (Figura 7) e per entrambe le condizioni prosodiche. Come si può osservare, per gli apprendenti italofoni l'area di esistenza della vocale d'appoggio è molto variabile e, almeno per alcuni parlanti, sembra assuma caratteristiche diverse a seconda del contesto prosodico nel quale viene realizzata: per il francofono è qualitativamente più circoscritta. Inoltre, la vocale realizzata dagli apprendenti ha le caratteristiche di una vocale medio-alta o alta, più anteriore di [a] (bassa F1 e alta F2), mentre per il francofono mostra caratteristiche maggiormente simili a una vocale centralizzata.

Figura 6 - *F1 e F2 delle vocali realizzate a velocità normale e per entrambe le condizioni prosodiche (sint = sintagma fonologico; inton = confine intonativo):/a/ in blu; /i/ in verde e V0 in rosso*

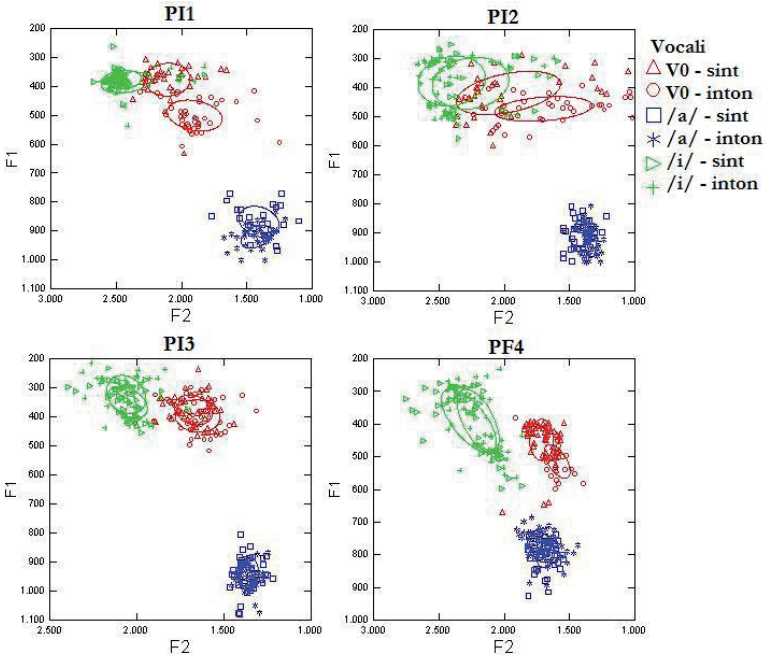
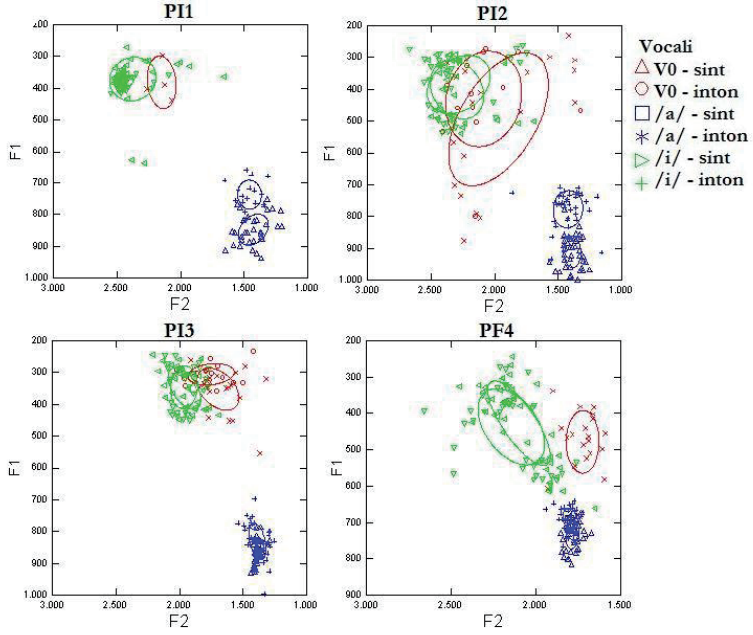


Figura 7 - *F1 e F2 delle vocali realizzate a velocità sostenuta e per entrambe le condizioni prosodiche (sint = sintagma fonologico; inton = confine intonativo):/a/ in blu; /i/ in verde e V0 in rosso*



Quindi, riassumendo, dal punto di vista acustico, a velocità normale di eloquio, tutti i parlanti realizzano un segmento vocalico tra le due consonanti, mentre a velocità sostenuta il numero di vocali diminuisce notevolmente e, di fatto, solo due parlanti italofoni su tre (PI2 e PI3) continuano a inserire una vocale all'interno del nesso, con una frequenza maggiore in caso di confine intonativo. L'altro parlante italofono (PI1) riduce invece drasticamente il numero di casi di inserzione vocalica, realizzando perlopiù assimilazioni di luogo e comportandosi, da questo punto di vista, in modo simile al parlante francofono. La durata del segmento è variabile per tutti parlanti e si attesta tra 50-150 ms e, a velocità sostenuta di eloquio, l'inserzione della vocale contribuisce ad una maggiore durata della sequenza. Per gli apprendenti anche la qualità della vocale, in termini di F1 e F2, è molto variabile e si presenta come una vocale più chiusa e anteriore rispetto a un canonico schwa. Per il parlante nativo, invece, la vocale mostra meno variabilità e presenta le caratteristiche di una tipica vocale schwa. Queste caratteristiche fanno pensare, rispetto a quanto descritto in letteratura, che la vocale inserita dagli apprendenti italofoni sia una vocale intrusiva, benché si possa comunque considerare che svolga la funzione di una vocale epentetica, ossia favorisca la produzione di un nesso marcato.

3.2 Risultati articolatori

Un risultato importante riguarda il fatto che sul dorso della lingua sono stati sempre individuati, per tutti i parlanti, i *target* relativi alle due vocali estreme della sequenza, cioè [a] e [i], mentre non è stato individuato alcun *target* per la vocale intermedia. Anche nel caso sia udibile una vocale, quindi, tra i bersagli articolatori dei due gesti consonantici si realizza solo un gesto di apertura.

A velocità normale di eloquio, il gesto di apertura si realizza in modo sistematico per entrambe le sequenze e in entrambe le condizioni prosodiche. L'apprendente PI3 realizza alcuni casi (5) senza un vero gesto di apertura tra le due fricative e senza alcun *output* acustico per la sequenza AP (/zʃ/e /zʒ/) in assenza di confine prosodico. Considerando il *Δextrema*, cioè la differenza tra i due target consonantici, il test di Kruskal-Wallis indica una differenza significativa su TTz [$\chi 2(1,29)=5,602$; p=,018] e LL-TTz [$\chi 2(1,29)=6,770$; p=,009]. I valori medi indicano una durata maggiore per i casi in cui sia stata inserita una vocale (165ms) rispetto ai casi senza *output* acustico e gesto di apertura (125ms). In caso di un confine intonativo, il francofono si distingue dagli italofoni, poiché inserisce sempre una pausa che può essere preceduta o meno da schwa. Tuttavia, il test di Kruskal-Wallis non indica alcuna significatività nel confronto delle misure articolatorie effettuate circa le sue produzioni, per cui l'intervallo tra i bersagli delle due fricative non cambia se interviene un solo elemento (pausa) o due elementi (pausa + schwa).

A velocità sostenuta di eloquio si ha una maggiore variabilità. In assenza di confine prosodico, PI1 e PF4 non inseriscono alcuna vocale e realizzano sequenze senza un gesto di apertura intermedio. Gli apprendenti PI2 e PI3, invece, realizzano i nessi con due o tre *pattern* articolatori: 1) gesto di apertura tra C_1 e C_2 con inserimento

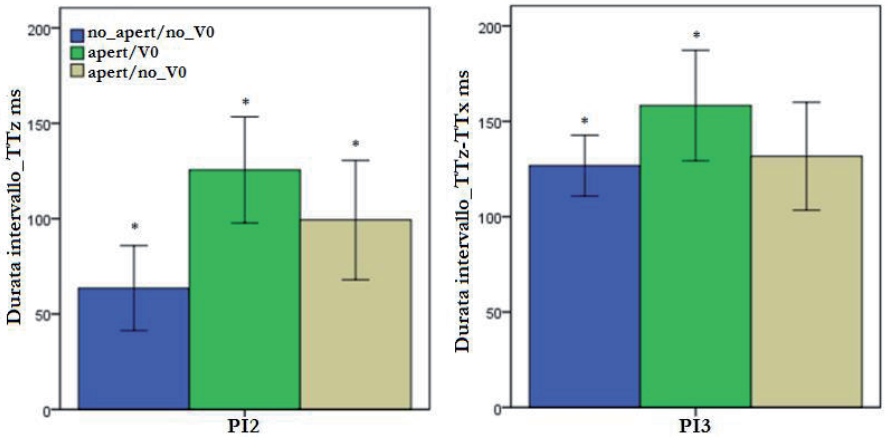
di vocale udibile (apertura/V0); 2) gesto di apertura senza *output* acustico (apertura/no V0); 3) nessun gesto di apertura e nessun *output* acustico (no apertura/no V0). Nella Tabella 6 sono riportate le frequenze di realizzazione dei tre *pattern* articolatori per entrambe le sequenze. Come si può osservare, la sequenza AP viene realizzata con tre *pattern*, mentre la sequenza PA viene realizzata con due *pattern* (casi descritti in 1 e 2).

Tabella 6 - *Frequenze relative alla realizzazione dei tre pattern articolatori a velocità sostenuta e in assenza di confine prosodico da parte di tutti i parlanti*

Sintagma fonologico									
Alveolare-postalveolare					*Postalveolare-alveolare*				
P	Casi	N	Tot	%	P	Casi	N	Tot	%
PI1	No apert/no V0	26	26	100	PI1	No apert/no V0	12	12	100
PI2	Apert/V0	7	30	23,3	PI2	Apert/V0	8	17	47,1
	Apert/no V0	14		46,7		Apert/no V0	9		52,9
	No apert/no V0	9		30		No apert/no V0	0		0
PI3	Apert/V0	9	29	31,0	PI3	Apert/V0	8	14	57,1
	Apert/no V0	5		17,2		Apert/no V0	6		42,9
	No apert/no V0	15		51,8		No apert/no V0	0		0
PF4	No apert/no V0	28	28	100	PF4	No apert/no V0	13	13	100

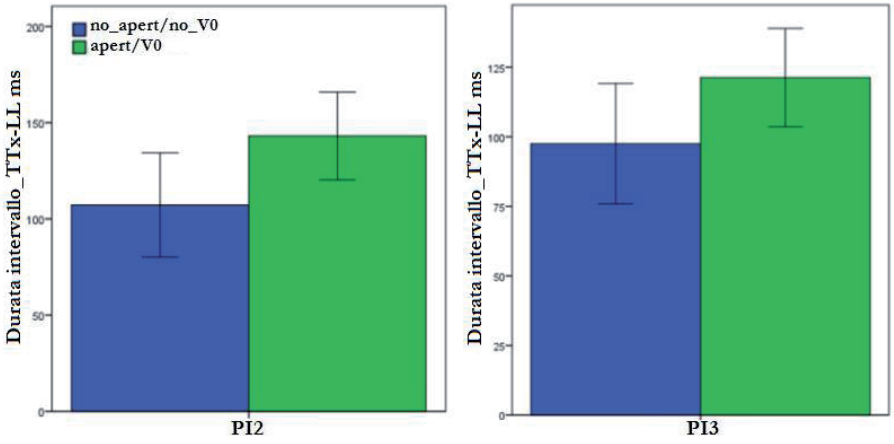
Per la sequenza AP, il test indica una differenza significativa su TTz per PI2 [χ2(2,30)=12,932; p=,002] e su TTz-TTx per PI3 [χ2(2,29)=7,609; p=,022]. Il test di Mann-Whitney mostra che per PI2 i casi no_apertura/no_V0 si distinguono dai casi apertura/V0 [Z=-3,228; p=,000] e dai casi apertura/no_V0 [Z=-2,521; p=,011]. L'intervallo medio per i casi no_apertura/no_V0 è di 63,54 ms (dev. st. 22,26), mentre per i casi apertura/V0 è di 125,54 ms (dev. st. 27,38) e per i casi apertura/no_V0 è di 99,28 ms (dev. st. 31,31). Per i PI3 il test di Mann-Whitney [Z=–2,772; p=,004] mostra una differenza significativa solo tra i casi apertura/V0 (158,26 ms, dev. st. 28,99) e i casi no_apertura/no_V0, (126,75 ms, dev. st. 15,98). Il *pattern* apertura/V0 mostra l'intervallo maggiore, a cui segue il *pattern* apertura/no_V0 e infine il *pattern* no_apertura/no_V0 con l'intervallo più breve.

Figura 8 - *Grafico a barre per il Δextrema per la sequenza alveolare-postalveolare in assenza di confine prosodico per PI2 su TTz (a sinistra) e PI3 su TTz-TTx (a destra)*



Per la sequenza PA, il test di Kruskal-Wallis è significativo su TTx-LL per PI2 [$\chi2(1,17)=5,346$; p=,021] e PI3 [$\chi2(1,14)=4,067$; p=,044]. In particolare, l'intervallo è maggiore per i casi apertura/V0 (PI2: 143,12 ms, dev. st. 22.,82; PI3: 121,25 ms, dev. st. 17,67) rispetto ai casi apertura/no_V0 (PI2: 107,25 ms, dev. st. 27,09; PI3: 97,50 ms, dev. st. 21,62).

Figura 9 - *Grafico a barre per il Δextrema per la sequenza postalveolare-alveolare in assenza di confine prosodico per PI2 (a destra) e PI3 s (a sinistra) su TTx-LL*



Per quanto riguarda i risultati relativi alle sequenze in presenza di confine intonativo, come si può osservare dalla Tabella 7 si realizzano tre *pattern* per la sequenza AP e due *pattern* per la sequenza PA, come nel contesto precedente. Il madrelingua PF4 realizza la sequenza inserendo uno schwa o assimilando per il luogo di articolazione. Il parlante PI1, comportandosi in modo simile, tende a non inserire una vocale tra

le due fricative realizzando, quindi, la sequenza con o senza gesto di apertura. PI2 e PI3, invece, realizzano sempre il gesto di apertura accompagnato o meno da *output* acustico.

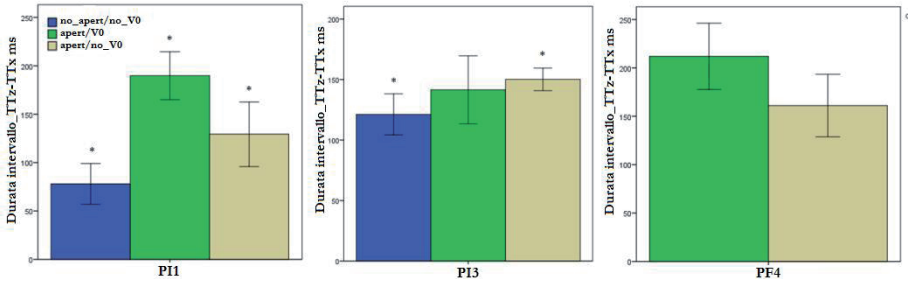
Tabella 7 - *Frequenze relative alla realizzazione dei tre pattern articolatori a velocità sostenuta e in presenza di confine intonativo da parte di tutti i parlanti*

Confine intonativo									
Alveolare-postalveolare					*Postalveolare-alveolare*				
P	Casi	N	Tot	%	P	Casi	N	Tot	%
PI1	Apert/V0	4	24	16,66	PI1	Apert/V0	0	11	0
	Apert/no V0	12		50		Apert/no V0	5		45,5
	No apert/no V0	8		33.34		No apert/no V0	6		54,5
PI2	Apert/V0	13	24	54,16	PI2	Apert/V0	7	11	63,6
	Apert/no V0	8		33,34		Apert/no V0	4		36,4
	No apert/no V0	3		12,5		No apert/no V0	0		0
PI3	Apert/V0	15	27	55,56	PI3	Apert/V0	11	14	78,6
	Apert/no V0	4		14,82		Apert/no V0	3		21,4
	No apert/no V0	8		29,62		No apert/no V0	0		0
PF4	Apert/V0	13	28	46.42	PF4	Apert/V0	7	14	50
	Apert/no V0	7		25		Apert/no V0	0		0
	No apert/no V0	8		28,58		No apert/no V0	7		50

Per la sequenza AP, il test di Kruskal-Wallis sui valori del *Δextrema* raggiunge la significatività su TTz per PI1 [$\chi2(2,24)=18,184$; p=,000] e PI3 [$\chi2(2,27)=10,268$; p=,006]. Il test di Mann-Whitney indica che per PI1 i casi no_apertura/no_V0 si distinguono dai casi apertura/V0 [Z=−2,717; p=,004] e dai casi apertura/no_V0 [Z=−3,626, p=,000] poiché i casi senza gesto di apertura e senza *output* acustico corrispondono all'intervallo più breve (71,69 ms; dev. st.15,39) rispetto ai casi apertura/no_V0 (120,55 ms; dev. st. 28,84) e apertura/V0 (186,05 ms; dev. st. 22,56). Inoltre, si riscontra anche una differenza significativa tra i casi apertura/V0 e i casi apertura/no_V0 [Z=−2,668, p=,008] con valori medi di 186,05 ms (dev. st. 22,56) e 120,55 ms (28,84) rispettivamente. Per il parlante PI3, il test indica una differenza tra no_apertura/no_V0 e apertura/V0 (Z=−3,099; p=,001). I casi con inserimento di vocale corrispondono invece ad un intervallo maggiore (150,55 ms; dev. st. 23,92) rispetto ai casi no_apertura/no_V0 (107,46 ms, dev. st. 25,44). I casi con apertura intermedia e senza *output* acustico hanno una durata di circa 145ms (dev. st. 13,91), quindi simile a quella dei casi in cui è stato effettivamente inserita una vocale. Per il parlante francofono, l'analisi ha riguardato solo i casi in cui fosse presente il gesto di apertura, con o senza output acustico: i casi con *output* acustico, hanno un intervallo maggiore (200,64 ms; dev. st. 30,70) rispetto ai casi con ge-

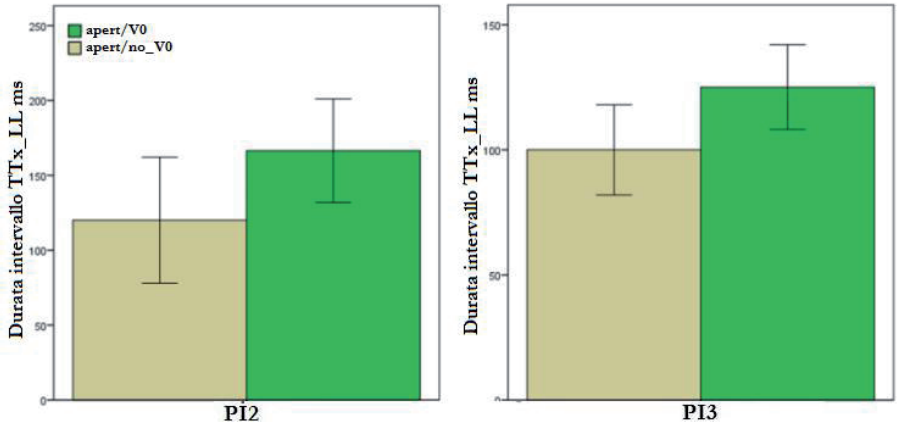
sto di apertura e senza che si percepisca una vocale (142,74 ms; dev. st. 36,77) PF4 [$\chi 2(1,20)=8,391$; p=,004.

Figura 10 - *Grafico a barre per il Δextrema per la sequenza alveolare-postalveolare in presenza di confine intonativo su TTz-TTx per PI2 (a sinistra), PI3 (centro), PF4 (a destra)*



Infine, come abbiamo già detto, gli apprendenti PI2 e PI3 realizzano la sequenza PA sempre con un gesto di apertura che può avere o meno un *output* acustico. Per PI2 questo corrisponde ad una differenza significativa in termini di *Δextrema* su TTz [$\chi 2(1,11)=6,036$; p=,014], mentre per PI3 su TTx [$\chi 2(1,14)=6,303$; p=,012]. Per entrambi i parlanti, l'intervallo è maggiore nel caso in cui ci sia *output* acustico rispetto ai casi in cui sia stato realizzato solo un gesto di apertura (rispettivamente per PI2: 149,28 ms dev. st. 7,86 vs 124,08 ms; dev. st. 14,99; per PI3 167,77 ms dev. st. 17,60 vs. 133,33 dev. st. 5,77). Anche PI1 realizza la sequenza con o senza gesto di apertura, ma questo non corrisponde ad una differenza significativa delle misure da noi considerate. Il francofono PF4 inserisce una vocale o assimila per il luogo di articolazione. Trattandosi di *pattern* articolatori corrispondenti a strategie che riguardano segmenti diversi non si è ritenuto opportuno procedere con un'analisi statistica.

Figura 11 - *Grafico a barre per il Δextrema per la sequenza postalveolare-alveolare in presenza di confine intonativo su TTx-LL per PI2 (a sinistra) e PI3 (a destra)*



Riassumendo, quindi, a velocità normale di eloquio, il *pattern* articolatorio realizzato da tutti i parlanti corrisponde a un gesto di apertura tra le due fricative, senza un *target* specifico per la vocale inserita benché il segmento vocalico risulti sul piano acustico e percettivo. A velocità sostenuta si ha una maggiore variabilità. Si individuano tre tipi di realizzazione: 1) gesto di apertura tra C_1 e C_2 con inserimento di vocale (apertura/V0); 2) gesto di apertura senza *output* acustico (apertura/no V0); 3) nessun gesto di apertura e nessun *output* acustico (no apertura/no V0). La frequenza delle tre realizzazioni varia in base al parlante e al tipo di sequenza. Generalmente, si riscontra una differenza significativa nella misura del *Δextrema* tra i casi in cui si ha il gesto di apertura con *output* acustico, con l'intervallo più lungo, e i casi senza gesto di aperura e senza *output* acustico, con intervallo più breve. Quando è stato realizzato un gesto di apertura senza segmento vocalico, invece, la durata è intermedia e differisce dai casi con gesto di apertura e *output* vocalico; si hanno, invece, poche evidenze rispetto ai casi senza gesto di apertura e senza *output* acustico.

4. *Discussione e conclusioni*

In questo studio si è osservata la realizzazione di sequenze di sibilanti, alveolare-postalveolare e postalveolare-alveolare, da parte di apprendenti italofoni di francese L2 avanzato, focalizzando l'attenzione, dal punto di vista acustico e articolatorio, sulla vocale d'appoggio che viene realizzata al fine di riparare nessi consonantici non nativi. Nell'articolo, è stato sottolineato che l'inserzione vocalica può realizzarsi o come epentesi vocalica, cioè una vocale che presenta un proprio *target* articolatorio, oppure come vocale intrusiva, senza alcun *target* articolatorio dovuta ad un gesto di transizione tra le due consonanti.

In questa sede, l'inserzione della vocale è stata studiata prendendo in considerazione due fattori che possono interferire con la realizzazione dei nessi consonantici e possono quindi avere un effetto sulla realizzazione della vocale d'appoggio: l'influenza della velocità di eloquio e della presenza di confini prosodici diversi. In generale, l'inserzione vocalica si riscontra in modo sistematico a velocità normale di eloquio, in entrambe le condizioni prosodiche, per entrambe le sequenze e per tutti i parlanti. Le sole eccezioni si hanno per un apprendente, PI3, che realizza pochi casi senza una vocale di appoggio e per PF4 che, in presenza di confine intonativo, realizza sempre una pausa preceduta o meno dal segmento vocalico. A velocità sostenuta di eloquio, il segmento vocalico viene inserito soprattutto da due apprendenti (PI2 e PI3). L'apprendente PI1 tende a non inserire una vocale a velocità sostenuta, di fatto avvicinandosi al comportamento del francofono, che realizza una vocale solo in caso di confine intonativo e per la sequenza alveolare-postalveolare.

Dal punto di vista acustico, sono state osservate la durata e i valori della prima e seconda formante. La durata del segmento vocalico è variabile per tutti i parlanti e incide sulla durata maggiore dell'intera sequenza. Per gli italofoni la vocale inserita

non mostra uno spazio fonetico ben definito, né centralizzato, tale da far pensare a un canonico schwa, poiché il segmento inserito risulta essere una vocale abbastanza chiusa e più anteriore di [a] (sembra influenzata dalla vocale successiva [i]). Per il francofono, invece, il segmento vocalico mostra uno spazio fonetico definito con le caratteristiche centralizzate di una vocale schwa. Il segmento inserito, quindi, nel presentare durata e qualità variabile pare influenzato dallo stile di eloquio per cui, secondo quanto descritto in letteratura, tale vocale potrebbe essere considerata una vocale intrusiva. Tuttavia, tenuto conto del fatto che per gli apprendenti italofoni l'inserzione vocalica è supposta avere la funzione di riparare il nesso marcato nella loro L1, è verosimile pensare comunque che la funzione svolta sia quella di una vocale epentetica. D'altronde, bisogna considerare che l'elemento vocalico si inserisce in posizione di coda in una sillaba non accentata, posizione che già di per sé può favorire un'alta variabilità in durata e qualità poiché fortemente influenzata dal contesto adiacente. Poiché dal punto di vista acustico è parso piuttosto difficile poter distinguere chiaramente tra vocale intrusiva e epentetica, si sono osservate anche le produzioni delle sequenze dal punto di vista articolatorio.

Osservando la traiettoria del dorso della lingua, che dà indicazioni sui gesti vocalici, non si è riscontrato alcun *target* specifico per la vocale inserita, per nessun parlante, mentre sono chiaramente visibili i *target* delle due vocali della sequenza [a] e [i]. È stato quindi calcolato l'intervallo tra i due *target* consonantici (differenza tra il *target* di C_2 e di C_1 - *Δextrema*), al fine di osservare la coordinazione tra le due fricative in base alla presenza o meno di una vocale inserita. A velocità normale di eloquio, il pattern articolatorio tipico che si è riscontrato per tutti i parlanti, per le due sequenze e per entrambe le condizioni prosodiche, è un gesto di apertura tra i due gesti consonantici. Si distingue il parlante francofono che in presenza di confine intonativo e per entrambe le sequenze realizza una pausa che può essere preceduta o meno dallo schwa. Dal punto di vista articolatorio, la coordinazione temporale tra le sue due fricative non sembra subire modificazioni poiché l'intervallo tra i due target consonantici non varia in base all'inserimento di due elementi (pausa e vocale) o di un solo elemento (vocale). Si è osservato un effetto di compensazione sulla durata della fricativa in sillaba finale (C_1) poiché si osserva una durata minore nel caso di inserimento di pausa e vocale e una durata maggiore nel caso di inserimento solo di pausa. A velocità sostenuta di eloquio, si riscontrano tre tipi di realizzazione: 1) gesto di apertura tra C_1 e C_2 con inserimento di vocale; 2) gesto di apertura senza *output* acustico; 3) nessun gesto di apertura e nessun *output* acustico. I tre tipi di realizzazione si riscontrano nelle produzioni di due apprendenti su tre (PI2 e PI3) per la sequenza alveolare-postalveolare in entrambe le condizioni prosodiche, mentre per il terzo italofono (PI1) e per il parlante francese nativo solo in presenza di confine intonativo. In ogni caso. la sequenza postalveolare-alveolare viene realizzata con due dei tre *pattern* articolatori, ossia un gesto di apertura accompagnato o meno un *output* acustico (casi descritti in 1 e 2). Tra gli apprendenti si distingue, quindi, PI1 che solo in assenza di confine prosodico non realizza mai una vocale ma assimila la sequenza per il luogo di articolazione,

come fa il francofono; in presenza di confine realizza la sequenza con o senza gesto di apertura, e comunque senza *output* acustico (casi descritti in 2 e 3). Anche il parlante francofono non realizza mai una vocale in assenza di confine prosodico, assimilando la sequenza per il luogo di articolazione, mentre in presenza di confine realizza un gesto di apertura con schwa (caso descritto in 1) o assimila. I casi con gesto di apertura e *output* acustico motivano la durata maggiore delle sequenze, distinguendosi dai casi senza gesto di apertura e segmento vocalico, con durata più breve, e dai casi con gesto di apertura senza *output* acustico, con durata intermedia.

Dai risultati emerge, quindi, chiaramente l'influenza sia della velocità di eloquio che della struttura prosodica. L'inserimento della vocale si ha maggiormente a velocità normale di eloquio, mentre a velocità sostenuta si ha una maggiore variabilità. Per le produzioni a velocità sostenuta, soprattutto per gli italofoni PI2 e PI3, si nota un maggiore numero di inserimenti, soprattutto in presenza di un confine intonativo. Per quanto riguarda il tipo di sequenza (postalveolare-alveolare e alveolare-postalveolare), la vocale d'appoggio si riscontra in entrambe le sequenza sebbene si noti che, a velocità sostenuta di eloquio e sempre per gli apprendenti PI2 e PI3, percentuali di inserimento maggiori per la sequenza postalveolare-alveolare piuttosto che per la sequenza opposta.

In sostanza, per tutti i parlanti, i risultati articolatori mostrano che il segmento vocalico inserito è di tipo "*targetless*" poiché non è stato possibile individuare un *target* specifico sulla traiettoria del dorso della lingua. I risultati relativi al segmento vocalico inserito dal parlante francofono suggeriscono che derivi proprio dal rilascio acustico della consonante (Rialland, 1986), dovuto alla velocità normale di eloquio e alla presenza di confine intonativo a velocità sostenuta. Per quanto riguarda gli apprendenti, i tre tipi di realizzazione indicano una certa variabilità articolatoria legata, verosimilmente, ai tentativi nel realizzare nessi marcati nella lingua nativa per cui l'*output* acustico sembrerebbe dovuto ad un vero e proprio *gestural mistiming*.

In conclusione, le produzioni degli apprendenti italofoni risultano qui influenzate dalle caratteristiche fonetiche e fonologiche della lingua nativa per cui, com'è noto, il fine dell'inserimento vocalico è di riparare un nesso non nativo, marcato. Questo studio suggerisce, inoltre, che l'inserimento vocalico sia dovuto a un *gestural mistiming*, quindi legato a questioni articolatorie con particolare riferimento alla scarsa coordinazione tra le due consonanti: la produzione delle consonanti creerebbe uno spazio articolatorio sufficiente perché si realizzi un segmento vocalico d'appoggio, benché questo non corrisponda ad un vero bersaglio. In futuro, oltre ad un'analisi percettiva da parte di parlanti nativi delle produzioni relativamente all'inserzione vocalica e/o alla realizzazione di altri fenomeni coarticolatori, si prevede di incrementare il campione degli apprendenti, come anche quello dei parlanti nativi, al fine di poter maggiormente confermare i risultati qui presentati.

Riferimenti bibliografici

Boersma, P., Weenink, D. (2010). PRAAT: Doing phonetics by computer. Version 5.2. htttp://www.praat.org.

Browman, C.P. (1995). Dynamics and articulatory phonology. In *Minds as motion*, 175-193.

Browman, C.P., Goldstein, L. (1986). Towards an articulatory phonology. In *Phonology Yearbook*, 3, 219-252.

Browman, C.P., Goldstein, L. (1987). Tiers in Articulatory Phonology with some implications for casual speech. In *Haskins Laboratories Status Report on Speech research*, SR-92, 1-30.

Browman, C.P., Goldstein, L. (1989). Articulatory gestures as phonological units. In *Phonology*, 6, 201-251.

Browman, C.P., Goldstein, L. (1992). "Targetless" Schwa: An Articulatory Analysis. In Docherty, G.J., Ladd, D.R. (Eds.), *Papers in Laboratory Phonology II: Gesture, segment, prosody*. Cambridge: CUP, 26-56.

Browman, C.P., Goldstein, L. (2007). The gestural phonology model, Speech production: motor control. In *Brain research and fluency disorders*, 57-71.

Byrd, D., Choi, S. (2006). At the juncture of prosody, phonology and phonetics – the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In *Proceedings of the 10th Conference on Laboratory Phonology*.

Byrd, D., Kaun, A., Narayanan, S. & Saltzman, E. (2000). Phrasal signature in articulation. In Broe, M.B., Pierrehumbert, J.B. (Eds.), *Papers in Laboratory Phonology V.* Cambridge University Press, 70-87.

Byrd, D., Tan, C.C. (1996). Saying consonant clusters quickly. In *Journal of Phonetics*, 24, 209-244.

d'Apolito, S. (2012). La coarticolazione: Studio acustico, cinematico e percettivo di sequenze di sibilanti della lingua francese nelle produzioni di studenti italofoni. Tesi di Dottorato.

d'Apolito, S., Gili Fivela, B. (2013). Place assimilation and articulatory strategies: the case of sibilant sequences in French as L1 and L2. *Interspeech*, Lyon, France.

Davidson, L. (2005). Addressing phonological questions with ultrasound. In *Clinical Linguistics & Phonetics*, 19(6/7), 619-633.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. In *Journal of Phonetics*, 34, 104-137.

Davidson, L., Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: An Ultrasound study. In Garding, G., Tsujimura, M. (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics* (WCCFL). Somerville, MA: Cascadilla Press, 165-178.

Eckman, R.F. (2008). Typological markedness and second language phonology. In Hansen Edwards, J.G., Zampini, M.L. (Eds.), *Phonology and second language acquisition*. Philadelphia: John Benjamins, Ch. 4, 95-115.

Engwall, O. (2000). Dynamical aspects of coarticulation in Swedish fricatives: a combined EMA & EPG study. In *TMH Quarterly Status and Progress Report*, 4, KTH, Stockholm, 49-3.

Farnetani, E., Busá, M.G. (2004). Italian clusters in continuous speech. In *Proceedings of the 3rd ICSLP*, Yokohama, Japan, vol. 1, 359-362.

Gafos, A. (2002). A grammar of gestural coordination. In *Natural Language and Linguistic Theory*, 20, 269-337.

Hall, N. (2003). Gestures and segments: Vowel intrusion as overlap. PhD Dissertation, University of Massachusetts, Amherst.

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. In *Phonology*, 23, 387-429.

Hall, N. (2011). Vowel epenthesis. In van Oostendorp, M., Ewen, C.J., Hume, E. & Rice, K. (Eds.), *The Blackwell Companion to phonology.* Malden, MA & Oxford: Wiley-Blackwell, 1576-1596.

Kühnert, B., Nolan, F. (1999). The origin of coarticulation. In Hardcastle, W.J., Hewlett, N. (Eds.), *Coarticulation: Theory, Data and Techniques.* Cambridge: Cambridge University Press, 7-30.

Levin, J. (1987). Between epenthetic and excrescent vowels. In *Proceedings of the West Coast Conference on Formal Linguistics*, 6, 187-201.

Lyche, C. (2016). Approaching variation in PFC: the schwa level. In Sylvain, D., Durand, J., Laks, B. & Lyche, C. (Eds.), *Varieties of spoken French*. Oxford University Press, ch. 27, 352-362.

Muliačić, Z. (1973). *Fonologia della lingua italiana*. Bologna: Il Mulino.

Niebuhr, O., Lancia, L. & Meunier, C. (2008). On place assimilation in French sibilant sequences. In *Proceedings of the VII ISSP*, Strasbourg, France, 221-224.

Nespor, M., Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Perkell, J.S. (1986). Coarticulation strategies: Preliminary implications of a detailed analysis for lower lip protrusion movements. In *Speech Communication*, 5, 47-68.

Rialland, A. (1986). Schwa et syllables en français. In *Studies in compensatory lengthening*. Dordrecht: Foris Publications, 187-226.

Ridouane, R., Fougeron, C. (2011). Schwa elements in Tashlhiyt word-initial clusters. In *Journal of LabPhon*, 2, 275-300.

Sigona, F., Stella, A., Grimaldi, M. & Gili Fivela, B. (2015). Mayday: A Software for Multimodal Articulatory Data Analysis. In *Atti del X Convegno AISV*, 173-184.

Shanin, K., Blake, S.J. (2004). A phonetic study of schwa in St'at'imcets (Lillooet Salinsh). In Gerdts, D., Matthewson, L. (Eds.), *Studies in Salinsh linguistic in honor of M. Dale Kinkade*. Missoula: University of Montana, 311-327.

Tiede, M., Shattuck-Hufnagel, S., Johnson, B., Ghohs, S., Mattheis, M., Zandipur, M. & Perkell, J. (2007). Gestural phasing in /kt/ sequences contrasting within and cross word contexts. In *Proceedings of the XVI ICPhS*, Saarbrücken, August 6-10, 521-524.

DOI: 10.17469/O2102AISV000005

BIANCA SISINNI, BARBARA GILI FIVELA, MIRKO GRIMALDI

L'elaborazione preattentiva del tratto di durata in suoni non nativi: uno studio elettrofisiologico

Native Italian listeners do not use vowel duration to distinguish word meaning. Nonetheless, similarly to Spanish listeners (e.g., Chládková et al., 2013), they are supposed to be able to use duration to discriminate vowel pairs that belong to nonnative or second language (L2) systems, especially when similar to one native sound (Desensitization Hypothesis; Bohn, 1995). The present study investigates how Italian listeners elaborate duration in different L2 vowels, i.e., similar and new with regard to the native systems, by means of the Mismatch Negativity, an electrophysiological component that reflects a pre-attentive change detection of a deviant sound in a sequence of frequent sounds. The results show that Italian listeners can make use of the duration cue, irrespectively of the vowel sound. In addition, the use of double deviants, i.e., vowel formant frequency and duration differences, carried by a single stimulus (e.g., Sussman, Winkler, 2001) allowed us to observe a sequential processing of both duration and frequency.

Key words: Second language perception, vowel duration, Mismatch Negativity, double deviants.

Introduzione

La relazione di similarità/differenza fra il sistema nativo (L1) e un sistema non nativo (NN) o una seconda lingua (L2) è una delle cause principali delle difficoltà nel percepire suoni NN/L2 che sono caratterizzati da correlati non presenti o utilizzati in maniera differente nella L1 (Escudero, Benders & Lipski, 2009). Ad esempio, i parlanti nativi del giapponese hanno difficoltà nel discriminare i fonemi /r-l/ in quanto non sono in grado di percepire la differenza delle terza formante (F3) fra le due consonanti (Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann & Siebert, 2003).

Spesso i parlanti nativi di L1, nel discriminare contrasti NN/L2, si basano su dimensioni acustiche che per i parlanti nativi di L2 sono ridondanti o secondarie. Ad esempio, vi sono sistemi linguistici in cui il tratto di durata (consonantica e/o vocalica) ha una valenza fonologica e, pertanto, vengono definiti *quantity systems*. Ad essi si oppongono i *quality systems*, nei quali la differenza di durata non determina variazioni di significato (Nenonen, Shestakova, Huotilainen & Näätänen, 2005). L'uso della durata nella discriminazione di contrasti vocalici sembra differire fra i parlanti di lingue *quality* vs. i parlanti di lingue *quantity*. Ad esempio, i parlanti nativi dell'inglese distinguono le vocali /iː/-/ɪ/ principalmente o esclusivamente sulla base delle differenze formantiche fra le due vocali mentre è stato osservato che parlanti nativi dello spagnolo, del cinese mandarino (Flege, Bohn & Jang, 1997), del portoghese (Rauber, Escudero, Bion & Baptista, 2005), del catalano (Cebrian, 2006), del polacco (Bogacka, 2004), del russo (Kondaurova, Francis, 2008) e del giapponese (Morrison, 2002) basano la discriminazione del contrasto, assi-

milato alla vocale nativa /i/, sul tratto di durata, nonostante esso non abbia una valenza fonologica nei rispettivi sistemi nativi. Questi risultati sembrerebbero confermare l'ipotesi formulata da Bohn (1995), secondo la quale i parlanti sarebbero meno o non affatto sensibili (*desensitization hypothesis*) a differenze spettrali fra due vocali NN o L2 in regioni dello spazio acustico che contengono una sola vocale nativa. Conseguentemente, baserebbero la discriminazione della coppia NN o L2 sulla differenza di durata, essendo una dimensione psico-acusticamente saliente.

La rilevanza della durata nella discriminazione di contrasti nativi o non nativi è stata osservata anche attraverso lo studio della *Mismatch Negativity* (MMN), una componente elettrofisiologica che riflette la discriminazione preattentiva di stimoli uditivi, il cui picco massimo in ampiezza si registra attorno ai 100-250ms dalla presentazione dello stimolo. La MMN si elicita, fra gli altri, nel classico paradigma *oddball* in cui una serie di stimoli frequenti, detti standard, viene interrotta da stimoli meno frequenti, detti devianti, che differiscono dallo standard per una (o più) dimensione, come le frequenze formantiche o la durata.

Ylinen, Houtilainen & Näätänen (2005), in uno studio sull'elaborazione della durata consonantica in parlanti nativi del finlandese (*quantity system*), hanno ipotizzato l'esistenza di meccanismi neurali indipendenti e paralleli per la qualità e la quantità (durata) consonantica. Infatti, in risposta ad uno stimolo deviante che differisce dallo standard sia per qualità che per durata, gli autori hanno, riscontrato la presenza di una MMN unica che è risultata pari alla somma delle MMN generate dai devianti singoli per qualità o per durata (MMN additiva). In questo modo, hanno confermato la presenza di processi neurali indipendenti che si sommano fra loro per l'elaborazione del doppio deviante (si veda successivamente la definizione di doppio deviante).

Per ciò che concerne i suoni non nativi, Lipski, Escudero & Benders (2012) hanno confrontato le MMN in parlanti nativi dello spagnolo che discriminano il contrasto olandese /a/-/ɑ/, caratterizzato da differenze spettrali, e il contrasto /aː/-/a/, caratterizzato dalla sola differenza di durata (dove /a/ è un fonema nativo). Le MMN osservate per il contrasto spettrale hanno un'ampiezza minore rispetto al contrasto di durata e gli autori hanno quindi dedotto che l'elaborazione del tratto di durata nella discriminazione di vocali simili avviene anche a livello preattentivo.

La stessa elaborazione preattentiva sembra essere modulata dal rapporto di similarità/dissimilarità fra i suoni L2 e il sistema nativo. Sia Nenonen et al. (2005) che Chládková, Escudero & Lipski (2013) hanno osservato che le MMN relative all'elaborazione della durata sono maggiori per stimoli vocalici che differiscono dal sistema nativo, ovvero non sono categorizzabili rispetto ad esso e possono essere considerati *nuovi*, che per stimoli *simili*, riconducibili invece a categorie native. In entrambi gli studi viene ipotizzato che in parlanti di un *quality system* il sistema nativo inibisce l'elaborazione del tratto di durata per quei suoni che possono essere categorizzati rispetto al sistema nativo stesso, mentre non esercita alcuna azione inibitoria per i suoni non categorizzabili. In sostanza, poiché nel sistema nativo il tratto di durata vocalica non è presente, esso non avrebbe alcun effetto durante l'elaborazione di vocali categorizzabili come native. Al contrario, la durata sarebbe facilmente elaborata per quei suoni che non sono categorizzati in maniera

"coerente" dal sistema nativo. Questo accadrebbe sia per parlanti con esperienza in L2 (Nenonen et al., 2005) che per parlanti monolingue, senza alcuna conoscenza della L2 (Chládková et al., 2013).

In entrambi gli studi citati, lo stimolo deviante differisce dallo standard per una sola dimensione, ovvero il tratto di durata. Nenonen et al. (2005) hanno confrontato le MMN di parlanti russi in risposta al contrasto finlandese simile (categorizzabile con il sistema nativo) /kɑ/-/kɑː/ con quelle in risposta al contrasto finlandese *nuovo* /kæ/-/kæː/, mentre Chládková et al. (2013) hanno comparato le risposte MMN ottenute, fra gli altri, in parlanti spagnoli che percepivano i contrasti estoni *simile* /æ/-/æː/ vs. *nuovo* /ɤ/-/ɤː/.

Diversamente dagli studi di Nenonen et al. (2005) e Chládková et al. (2013), Ylinen, Uther, Latvala, Vepsäläinen, Iverson, Akahane-Yamada & Näätänen (2010) utilizzano stimoli standard e devianti che differiscono fra loro non solo per la durata ma anche per valori formantici, ovvero le vocali inglesi /iː/-/ɪ/ (nelle parole /b_t/). L'obiettivo dello studio è osservare le risposte MMN in parlanti nativi del finlandese per verificare se un training percettivo possa modificare le modalità di elaborazione del tratto di durata. Oltre alla presenza di stimoli che differiscono fra loro in due aspetti, gli autori hanno previsto l'utilizzo invertito di standard e deviante, per cui in una condizione lo standard è la vocale L2 /iː/ e nell'altra la vocale L2 /ɪ/. Anche in questo caso, il grado di similarità/dissimilarità rispetto alle categorie native sembra giocare un ruolo determinante. Essendo la vocale L2 /iː/ simile alla vocale nativa /i/, le MMN elicitate quando la vocale L2 /iː/ è utilizzata come standard sono di ampiezza maggiore rispetto alle MMN elicitate quando lo standard è la vocale L2 /ɪ/. Il contesto di elicitazione, quindi, sembra avere un ruolo cruciale in quanto utilizzare stimoli standard simili alle categorie native facilita l'attivazione delle rispettive rappresentazioni presenti nella memoria a lungo termine necessarie per una più regolare elicitazione della MMN (si veda anche Näätänen, 2001).

Pertanto, utilizzare come standard uno stimolo riconducibile ad una categoria nativa sembra essere la condizione ideale per un'elicitazione più corretta della risposta MMN. Inoltre, tale utilizzo sembra rispecchiare meglio ciò che accade nella realtà, in quanto stimoli percettivi in entrata vengono comparati con le categorie native dei parlanti (e non con stimoli nuovi o non presenti).

Tuttavia, come accennato in precedenza, ciò che sembra essere rilevante nel lavoro di Ylinen et al. (2010) è il ricorso a "doppi devianti", *double deviants*, ovvero devianti che differiscono dallo standard non solo per durata ma anche per valori formantici.

Più specificamente, i *double deviants* (Sussman, Winkler, Ritter, Alho & Näätänen, 1999; Sussman, Winkler, 2001; Jaramillo, Ilvonen, Kujala, Alku, Tervaniemi & Alho, 2001; Wang, Datta & Sussman, 2005; Oceák, Winkler & Sussman, 2008) possono definirsi come due deviazioni i cui onset occorrono in un intervallo temporale di circa 200 millisecondi (ms), ovvero la finestra temporale di integrazione (*Temporal Window of Integration*, TWI)[1]. Sussman, Winkler (2001) affermano che, indipendentemente dal fatto che le deviazioni siano veicolate da un unico stimolo o da due stimoli differenti, i

[1] La TWI può variare fra i 150ms (Wang et al., 2005) e i 200-250ms (Horváth, Czigler, Winkler & Teder-Sälejärvi, 2007).

doppi devianti elicitano una sola MMN in quanto le due deviazioni vengono elaborate parallelamente ed, infine, integrate, come fossero un unico evento deviante. Tuttavia, quando i doppi devianti sono presentati in sequenze di stimoli nelle quali compaiono anche "devianti singoli", ovvero stimoli che differiscono dallo standard per una delle due deviazioni dei doppi devianti, allora saranno elicitate due MMN successive, in quanto la seconda deviazione coinciderebbe con un'informazione nuova rispetto al contesto (ovvero la presenza di devianti singoli) e, per questo, rilevante. La presenza/assenza di devianti singoli, pertanto, determinerebbe il modo in cui i doppi devianti vengono elaborati, se come un evento singolo o come due eventi successivi. La presenza di due MMN elicitate da due deviazioni veicolate da un unico stimolo in presenza di devianti singoli è stata riscontrata anche nel lavoro di Oceak et al. (2008), nella condizione che egli definisce *Isochronous Combined Deviation*. Al contrario, Czigler, Winkler (1996) sostengono che, indipendentemente dalla presenza di devianti singoli, se le deviazioni sono veicolate da uno stimolo unico, una sola MMN sarà elicitata.

Recentemente, Althen, Huotilainen, Grimm & Escera (2016) hanno verificato se le MMN elicitate in paradigmi *multi-feature* (Näätänen, Pakarinen, Rinne & Takegata, 2004)[2] differissero da quelle elicitate in classici paradigmi *oddball*, utilizzando devianti singoli in frequenza e intensità e doppi devianti in "frequenza+intensità". I risultati hanno dimostrato che le ampiezze e le latenze delle MMN ottenute nel paradigma *multi-feature* non differiscono da quelle elicitate nel paradigma *oddball*, caratterizzate in entrambe le condizioni dalla presenza di due picchi. Inoltre, hanno osservato la presenza di due MMN successive, a differenza di quanto osservato in altri studi in cui il doppio deviante "frequenza+durata", ha elicitato una sola MMN (Wolff, Schröger, 2001; Paavilainen, Mikkonen, Kilpelainen, Lehtinen, Saarela & Tapola, 2003).

L'obiettivo del presente lavoro è quello di indagare se il tratto di durata nella discriminazione di vocali viene elaborato preattentivamente da parlanti italiani (varietà di italiano salentino, IS). L'italiano è un sistema in cui il tratto di durata ha valenza fonologica per le consonanti ma non per le vocali. L'italiano è, quindi, una lingua *quantity*, in quanto comprende in sé il tratto di durata, ma non per le vocali. Un primo obiettivo di questo lavoro è verificare se il tratto fonologico di lunghezza porta i parlanti italiani ad elaborare le vocali lunghe come i parlanti nativi di lingue *quantity* "pure", come il finlandese (Ylinen, Houtilainen & Näätänen, 2005), per i quali è stata osservata la presenza di meccanismi neurali indipendenti per durata e qualità formantica, oppure se esso non ha effetto sull'elaborazione delle vocali, per le quali conseguentemente ci si aspetterà un'elaborazione unica e sequenziale delle due caratteristiche dello stimolo. Tale obiettivo è in linea con il ricorso ai doppi devianti, come si può leggere successivamente.

In linea con Nenonen et al. (2005), Lipski et al. (2012) e Chládková et al. (2013), ipotizziamo che il tratto di durata venga elaborato in funzione della relazione inter-linguistica fra gli stimoli devianti e lo stimolo standard. Lo stimolo standard corrisponde alla vocale nativa /a/ (utilizzata come standard per una migliore condizione di elicita-

[2] Nel paradigma *multi-feature*, i 5 devianti differiscono dallo standard per una sola caratteristica (ad es., intensità, durata, ecc.) e vengono presentati alternatamente ad esso così che ogni deviante possa rafforzare la rappresentazione dello standard rispetto al deviante successivo.

zione della MMN, Ylinen et al., 2010) e gli stimoli devianti sono le vocali dell'inglese americano (AE) /ɑ/ e /ʌ/. La vocale nativa /o/ è utilizzata come controllo. Studi precedenti (Escudero, Sisinni & Grimaldi, 2014; Sisinni, Escudero & Grimaldi, 2013) hanno osservato che la vocale L2 deviante /ɑ/ viene percepita da ascoltatori italiani come molto simile alla vocale nativa /a/, mentre la vocale L2 /ʌ/ viene percepita come differente, nuova, rispetto al sistema nativo (non categorizzata in maniera univoca con nessun fonema nativo, Best, Tyler, 2007; Elvin, Escudero, 2014). Pertanto, ipotizziamo quanto segue. Relativamente allo sole differenze formantiche fra la vocali devianti e la vocale standard, la vocale L2 simile /ɑ/ eliciterà MMN poco ampie (difficoltà di discriminazione rispetto alla vocale L1 /a/), la vocale L2 nuova /ʌ/ eliciterà MMN elevate (suono nuovo facilmente discriminabile rispetto al sistema nativo), la vocale L1 /o/ eliciterà le MMN maggiori (vocale nativa discriminabile rispetto alla vocale standard). Relativamente all'elaborazione della durata, invece, ipotizziamo che la vocale L2 /ɑ/ eliciterà MMN poco ampie, in quanto il sistema nativo inibirà l'elaborazione del tratto di durata per la vocale L2 simile, mentre l'inibizione non avrà effetto sulla vocale L2 nuova /ʌ/, per la quale le ampiezze della MMN saranno quindi maggiori (Nenonen et al., 2005; Chládková et al., 2013). Infine, non ci aspettiamo l'elaborazione del tratto di durata per la vocale nativa deviante /o/ (controllo), in quanto le differenze formatiche fra le due vocali dovrebbero essere sufficienti per la discriminazione del contrasto e l'elaborazione della durata dovrebbe essere inibita dal sistema nativo.

Inoltre, in una versione modificata del paradigma *multi-feature* (Althen et al., 2016), si utilizzeranno doppi devianti, in cui la distanza temporale fra gli onset delle due deviazioni è compresa nella TWI (< 200ms). Il ricorso ai doppi devianti (L2 /ɑ/-/ʌ/ e L1 /o/ con durate intermedie e lunghe rispetto alla durata dello stimolo standard L1 /a/, si veda 2.2. per i dettagli) e ai devianti singoli (L2 /ɑ/ e /ʌ/ e L1 /o/ con durata pari a quello dello standard) ci permetterà di verificare se le due deviazioni di "frequenze formantiche + durata" siano elaborate come un evento unico (una MMN), frutto dell'integrazione di due processi cognitivi indipendenti e paralleli (Ylinen et al., 2005), o come due eventi successivi (due MMN).

2. *Metodologia*

2.1 Soggetti

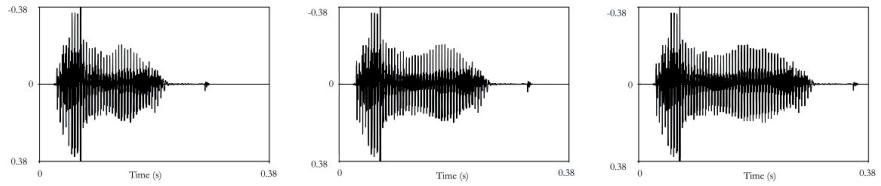
Quattordici soggetti hanno preso parte allo studio (8 uomini, età media 22,6; s.d. 1,6; 2 mancini) dopo aver preso visione e aver sottoscritto il modulo di consenso informato. Tutti i soggetti sono parlanti monolingue di italiano provenienti dal Salento, nati e cresciuti da genitori salentini. Ai soggetti è stato chiesto di indicare il loro livello di conoscenza di altre lingue, diverse dall'italiano, su una scala da 0 (nessuna conoscenza) a 7 (conoscenza nativa). In media, indipendentemente dalla lingua, hanno indicato il valore 4. Un soggetto è stato eliminato dall'analisi dei dati per problemi tecnici. Lo studio è in linea con la Dichiarazione di Helsinki (2008) ed è stato approvato dal comitato etico dell'Università del Salento.

2.2 Stimoli

Gli stimoli sono costituiti da pseudo-parole semi-sintetiche la cui struttura è /bVb/, dove V rappresenta le vocali dell'AE /ɑ/ e /ʌ/ e le vocali dell'italiano /a/ e /o/. La pseudo-parola originale è stata prodotta da una parlante dell'AE, ma le frequenze formantiche delle vocali sono state manipolate per ottenere le vocali target dell'AE e dell'IS, sulla base di valori formantici reali.

La durata della vocale standard è pari a 180ms e la durata totale dello stimolo è pari a 303ms. La durata delle vocali degli stimoli devianti, L2 /ɑ/, /ʌ/ e L1 /o/, è stata modificata allungando la parte stabile delle vocali, al di fuori delle transizioni, in modo da garantire che l'unica differenza percepibile sia quella della durata della vocale (Figura 1). Per ogni stimolo deviante, tre durate sono state ottenute: 180ms (standard), 218ms (intermedia), 255ms (lunga)[3]. I devianti la cui durata è pari a quella dello stimolo standard sono considerati come devianti "singoli" mentre i devianti che hanno durata intermedia e lunga come *double deviants*. La distanza temporale degli onset delle deviazioni nei doppi devianti è minore della TWI (180ms).

Figura 1 - *Da sinistra, la vocale L2 simile /ɑ/ con durata standard (180ms), intermedia (218ms) e lunga (255ms)*



La frequenza fondamentale è stata tenuta costante fra gli stimoli (202 Hz) cosi come l'intensità (70 dB).

I soggetti, ai quali non è stata specificata la lingua di appartenenza degli stimoli, non erano consapevoli che alcuni stimoli appartenessero all'inglese e altri all'italiano.

2.3 Procedura sperimentale

È stata utilizzata una versione modificata del paradigma *multi-feature optimum 3* (Nataanen et al., 2004). Lo stimolo contenente la vocale L1 /a/ di durata pari al 180ms è lo stimolo standard mentre gli altri 9 stimoli sono i devianti che differiscono dallo standard per i soli valori formantici (devianti singoli) o per formanti e durata (*double deviants*). Gli stimoli sono stati presentati attraverso casse audio posizionate ai lati della postazione del soggetto ad un volume accettabile.

[3] La durata standard è stata scelta sulla base dei valori medi della durata delle vocali italiane prodotte da una parlante nativa. La durata lunga è stata scelta sulla base della media di produzioni reali della vocale AE /ɑː/ di una parlante nativa dell'AE, giudicate come estremamente rappresentative da un gruppo di parlanti nativi. La durata intermedia è stata scelta di conseguenza.

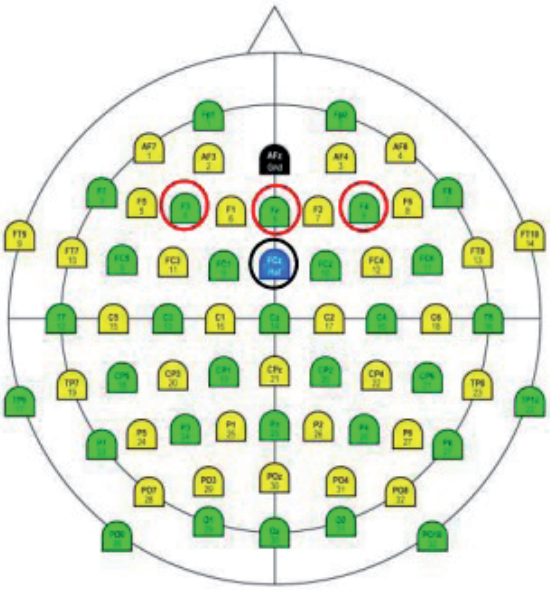
In una singola sequenza, gli stimoli standard (p = 50%) sono stati presentati in maniera alternata a ciascuno dei 9 devianti (p = 5%). La presentazione dei devianti è stata pseudo-randomizzata così che ogni deviante non fosse apparso mai in successione (Dev1– St – Dev2 – St – Dev3 – St – Dev4). Inoltre, anche due devianti contenenti la stessa vocale non sono mai apparsi in successione così come due devianti aventi la stessa durata (Dev_ɑ_180ms– St – Dev_ o_255ms – St – Dev_ʌ_218ms).

I 9 devianti sono stati presentati in maniera equiprobabile in una serie di 18 stimoli. Gli stimoli sono stati presentati in 6 sequenze da 6 minuti circa e ogni sequenza è cominciata con la presentazione di 5 standard, non inclusi nell'analisi finale. Ogni deviante è stato presentato per un totale di 162 volte (1456 devianti totali e 1456 standard). Il SOA (*stimulu onset asinchrony*) è pari a 800ms. Il tempo totale di registrazione è stato di 38 minuti circa.

2.3.1 Registrazione EEG e analisi dei dati

Durante le sessioni di registrazione dei dati elettroencefalografici (EEG), i soggetti sedevano su una comoda sedia in una stanza schermata acusticamente. È stato detto loro di ignorare gli stimoli uditivi e di prestare attenzione ad un film senza audio scelto da loro.

Figura 2 - *Configurazione degli elettrodi su una cuffia a 64 canali. In nero è evidenziato l'elettrodo utilizzato come referenza (FCz) e in rosso gli elettrodi frontali F3, Fz e F4 sui cui sono state svolte le analisi statistiche, poiché le ampiezze maggiori della MMN si registrano nella zona fronto-centrale dello scalpo*



L'attività cerebrale di ogni soggetto è stata registrata in maniera continua attraverso una cuffia a 64 elettrodi (Figura 2, BrainCap, Brain Products). La referenza era l'elettrodo FCz. I livelli delle impedenze sono stati tenuti sotto i 10 kOhm. In fase di registrazione,

effettuata ad una frequenza di campionamento di 500Hz, è stato applicato un filtro passa-banda 0.01-70 Hz. I movimenti oculari orizzontali e verticali sono stati monitorati, rispettivamente, con F7 e FP1. Dopo la registrazione, tutti gli elettrodi sono stati riferiti al naso e l'elettrodo FCz è stato utilizzato come attivo e non come referenza. I dati grezzi sono stati segmentati in epoche di 600ms, con un pre-stimolo di 100ms relativo all'onset della pseudo-parola. È stata quindi applicata una correzione oculare (Gratton, Coles, 1983) ed i trial contenti battiti delle palpebre sono stati esclusi dall'analisi, così come i trial con artefatti muscolari o altro rumore (rimozione degli artefatti +/- 150 μV). La correzione della baseline è stata applicata nell'intervallo pre-stimolo di 100ms. I primi cinque stimoli standard presentati all'inizio di ogni sequenza sono stati esclusi dall'operazione di *averaging*. Questa operazione consiste nel fare la media delle risposte cerebrali a ciascun tipo di stimolo, in questo caso agli stimoli standard e agli stimoli devianti, per ottenere così un'unica onda corrispondente al potenziale evocato dallo stimolo stesso. In questo modo, il potenziale correlato all'evento (ERP, *event related potential*), cioè la presentazione dello stimolo, emerge dalla registrazione elettroencefalografica continua. I dati sono stati filtrati a 1-20Hz e le MMN sono state ottenute per ciascun soggetto sottraendo dalle onde in risposta a ciascuno dei 9 stimoli devianti le onde in risposta allo stimolo standard L1 /a/. L'ampiezza delle 9 MMN è stata analizzata nell'elettrodo centrale Fz in un intervallo di 20ms intorno al picco massimo, in due finestre temporali differenti, chiamate d'ora in poi *range*, calcolate dall'inizio della/e deviazione/i, ovvero la prima deviazione relativa alla differenza di valori formantici (100-250ms; *range* 1, R1) e la seconda deviazione relativa all'inizio della deviazione di durata (280-430 ms; *range* 2, R2).

Gli elettrodi presi in considerazione per l'analisi statistica sono gli elettrodi frontali F3 (emisfero sinistro), Fz (linea mediana) e F4 (emisfero destro), poiché la maggiore ampiezza della MMN si registra nella zona fronto-centrale dello scalpo (Näätänen, Paavilanen, Rinne & Alho, 2007).

È stata eseguita un'Analisi della Varianza a Misure Ripetute con Vocale (/ɑ/, /ʌ/, /o/) x Range (R1, R2) x Durata (standard 180ms, intermedia 218ms, lunga 255ms) x Lateralità (F3, Fz, F4). La correzione Greenhouse-Geisser è stata applicata quando necessario. Il post hoc utilizzato è quello Bonferroni.

3. *Risultati*

L'analisi statistica ha evidenziato una significatività per il fattore Vocale [$F(2,22) = 5{,}356$ $p < 0{,}05$]. Il post hoc, tuttavia, mostra una tendenza alla significatività ($p = 0{,}07$) nel confronto fra le vocali L2 /ɑ/ *vs.* L1 /o/, la cui differenza in ampiezza viene comunque confermata dall'ispezione degli intervalli di confidenza (I.C.) al 95% delle ampiezze delle due vocali: l'ampiezza della vocale L2 /ɑ/ (-0,67 μV) è (appena) al di fuori dell'I.C. della vocale L1 /o/ (-1,34 μV... -0,68 μV), e l'ampiezza media della vocale L1 /o/ (-1,05 μV) è al di fuori dell'I.C. della vocale L2 /ɑ/ (-0,91 μV...-0,43 μV). Pertanto, si può ipotizzare che la vocale L2 /ɑ/ abbia elicitato una MMN minore rispetto alla vocale L1 /o/. L'ampiezza della componente elicitata dalla vocale L2 /ʌ/, invece, risulta essere equi-

valente sia a quella della vocale nativa /o/ che a quella della vocale L2 /ɑ/. Tali risultati sembrano suggerire che la vocale L2 /ɑ/ sia stata la più difficile da discriminare, la vocale L2 /ʌ/ abbia generato un livello intermedio di difficoltà e che la vocale L1 /o/ sia stata la più semplice da discriminare. Questo è in linea con la nostra ipotesi relativa alle sole differenze formantiche fra le vocali per cui la vocale L2 simile avrebbe elicitato MMN minori rispetto a quella nativa, così come quella nativa avrebbe elicitato MMN maggiori rispetto alla vocale L2 nuova.

Anche il fattore Durata è risultato essere significativo [$F(2,22) = 7{,}184$ $p < 0{,}05$] in quanto le ampiezze elicitate dalla vocali aventi durata lunga sono maggiori rispetto alle ampiezze delle vocali aventi durata pari allo stimolo standard e durata intermedia ($p < 0{,}05$). Questo risultato confermerebbe che il tratto di durata possa essere elaborato preattentivamente anche in parlanti che, almeno per quanto riguarda le vocali, fanno riferimento ad un sistema *quality*.

Non è stata riscontrata significatività per l'interazione Vocale x Durata ($p > 0{,}05$) è ciò sembra suggerire che la durata sia elaborata preattentivamente indipendentemente dalla vocale.

Figura 3 - *Grand average (N= 13) dell'onda di differenza (MMN) nell'elettrodo Fz delle vocali L2 simile /ɑ/, L2 nuova /ʌ/ e L1 /o/ delle durate standard (180ms), intermedia (218ms) e lunga (255ms) nel R1 (100-250ms) e R2 (280-430ms). Le mappe di voltaggio sono riferite al picco massimo di ampiezza delle MMN*

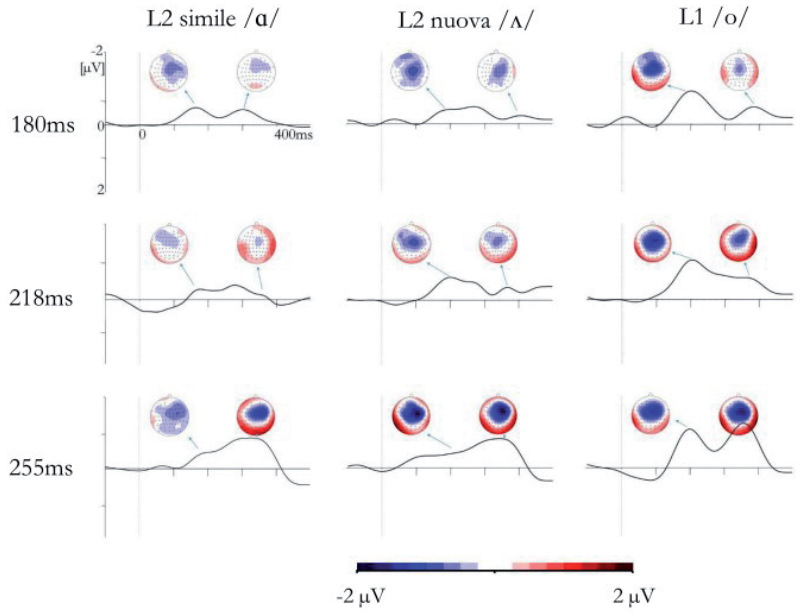


Tabella 1 - *Valori medi delle ampiezze in μV (deviazione standard fra parentesi) dei picchi delle MMN in Fz nei due range: R1 (100-250) e R2 (280-430) per le vocali di durata standard (180ms), intermedia (218ms) e lunga (255ms)*

	R1			*R2*		
	180	218	255	180	218	255
a	–0,84 (0,70)	–0,62 (0,45)	–0,81 (0,80)	–0,67 (0,58)	–0,67 (0,47)	–1,24 (0,91)
ʌ	–0,86 (0,64)	–0,92 (0,56)	–1,07 (0,62)	–0,56 (0,62)	–0,94 (0,61)	–1,51 (0,90)
o	–1,21 (0,76)	–1,36 (0,85)	–1,36 (1,18)	–0,80 (0,62)	–1,17 (0,99)	–1,61 (0,99)

Il fattore Lateralità [$F(2,22) = 28{,}552$ $p < 0.05$] deve la sua significatività alla ampiezze minori nell'elettrodo di sinistra (F3) rispetto ai siti centrale e destro ($p < 0{,}05$). La distribuzione delle componenti mostra, quindi, una minore attivazione nell'emisfero sinistro.

L'interazione Range x Durata [$F(2,22) = 7{,}923$ $p < 0{,}05$] evidenzia che le vocali con durata standard (devianti singoli) hanno ampiezza maggiore nel R1 rispetto al R2 ($p < 0{,}05$), mentre nel R2 le vocali di durata lunga (devianti doppi) hanno ampiezza maggiore delle vocali di durata standard e intermedia (devianti doppi) ($p < 0{,}05$). Pertanto, coerentemente con la durata degli stimoli, si evince la presenza di una componente anche nel R2 per i devianti doppi di durata lunga (Figura 2).

L'interazione Durata x Lateralità x Range [$F(4,48) = 3{,}119$ $p < 0{,}05$] mostra che, nel R2, nell'elettrodo centrale (Fz) le vocali con durata lunga elicitano MMN maggiori rispetto alle vocali di durata standard e intermedia ($p < 0{,}05$) mentre, nel sito di destra (elettrodo F4), le ampiezze delle vocali di durata intermedia eguagliano quelle di durata lunga ($p > 0{,}05$).

Inoltre, in entrambe le finestre temporali, le componenti hanno un'ampiezza minore a sinistra rispetto al centro e a destra ($p < 0{,}05$) indipendentemente dalla durata e, nel R2, la durata intermedia e lunga elicita ampiezza maggiori nel sito centrale-destro rispetto al sito sinistro ($p < 0{,}05$). Ciò sembra suggerire che l'aumento di durata nei devianti doppi, che sia intermedio o lungo, viene elaborato nel sito di destra.

4. *Discussioni e conclusioni*

Il presente lavoro ha come obiettivo lo studio dell'elaborazione preattentiva del tratto di durata vocalica in parlanti nativi dell'italiano (varietà di italiano salentino), il cui sistema fonologico prevede la presenza del tratto di durata sebbene esclusivamente per le consonanti. È stata scelta la vocale /a/ dell'italiano come standard e sono state comparate le ampiezze delle risposte MMN (componente indice dell'elaborazione e discriminazione preattentiva di stimoli linguistici) elicitate da vocali in-

glesi che vengono categorizzate con modalità differenti rispetto alle vocali italiane, ovvero la vocale L2 *simile* /ɑ/ e la vocale L2 *nuova* /ʌ/. A queste vocali inglesi, è stata aggiunta anche la vocale nativa /o/ come controllo. Le tre vocali devianti, inoltre, hanno durate differenti (standard, intermedia e lunga) per verificare se l'elaborazione della durata sia correlata con la modalità di categorizzazione dei suoni L2 rispetto a quelli L1, come dimostrato in studi precedenti (e.g., Nenonen, 2005; Lipski et al., 2012; Chládková et al., 2013). Inoltre, gli stimoli devianti che differiscono dallo stimolo standard sia per frequenze formantiche che per durata sono stati utilizzati per osservare come i devianti doppi, *double deviants*, vengano elaborati in presenza di devianti singoli in un paradigma *multi-feature* (Althen et al., 2016), ovvero se come stimolo unico o eventi distinti.

I risultati hanno dimostrato, in linea con le ipotesi iniziali relative alle sole differenze formantiche, che le vocali L2 /ɑ/ e /ʌ/ e la vocale L1 /o/ (controllo) elicitano MMN differenti in relazione alle modalità di categorizzazione rispetto al fonema nativo /a/ (Escudero, Sisinni & Grimaldi, 2014; Sisinni, Escudero & Grimaldi, 2013). Infatti, la vocale L2 simile /ɑ/ ha elicitato la componente con ampiezza minore, la vocale L2 nuova ha elicitato la componente con ampiezza intermedia mentre la vocale nativa /o/ ha elicitato la MMN con ampiezza maggiore. In effetti, la discriminazione di suoni simili, o *whitin-category*, sembra essere più difficile rispetto a suoni differenti o nuovi, o *across-category*, sia per bambini (Dehaene-Lambertz, Baillet, 1998) che per adulti (Winkler, Kujala, Tiitinen, Sivonen, Alku, Lehtokoski & Näätänen, 1999). I risultati del presente lavoro sono, quindi, in linea con gli studi precedenti in quanto la vocale L2 simile /ɑ/ può essere considerata come *whitin-category* rispetto alla vocale nativa /a/, pertanto ha elicitato MMN minori rispetto alle vocali *across-category* L2 /ʌ/ e L1 /o/. La MMN intermedia per la vocale L2 /ʌ/, a sua volta minore della MMN elicitata dalla vocale nativa /o/, è imputabile alla sua categorizzazione modale: sebbene non sia categorizzata in maniera univoca con nessuna vocale L1, la sua categorizzazione modale è con la vocale nativa /a/ (ad esempio, in Escudero et al., 2014, viene categorizzata come L1 /a/ nel 68% dei casi e come L1 /o/ nel 32%). Questa sovrapposizione percettiva fra la vocale L2 e quella L1 spiegherebbe la discriminazione intermedia (Van Leussen, Escudero, 2015; Tyler, Best, Faber & Levitt, 2014 per le previsioni sulla discriminazione di contrasti non nativi con parziale sovrapposizione percettiva).

Relativamente all'effetto della durata, le componenti maggiori sono state elicitate dalla durata lunga (255ms), più ampie rispetto a quelle elicitate dalla durata intermedia (218ms) e standard (180ms), indipendentemente dalla vocale. Ciò implicherebbe che la durata (lunga) viene elaborata a prescindere da come il suono venga categorizzato rispetto al sistema nativo. Questo risultato non è in linea con gli studi precedenti che hanno osservato una modulazione dell'elaborazione della durata relativa ai rapporti di similarità/differenza dei suoni L1/L2. Nel presente lavoro non si osserva l'azione inibitoria del sistema nativo quando percepisce la vocale L2 simile e l'azione opposta quando percepisce la vocale nuova (Nenonen et al., 2005; Chládková et al., 2013). I risultati qui ottenuti sembrano suggerire che

l'elaborazione della durata avvenga per tutte le vocali, anche per quella nativa /o/, e che l'elaborazione riguardi solo la variazione maggiore da noi proposta (durata lunga rispetto allo standard) e non quella intermedia (durata intermedia rispetto allo standard). Tuttavia, si tenga presente che in questo lavoro non è stato utilizzato un paradigma *oddball* con stimoli standard e devianti della L2 che differiscono fra loro solo in durata. Come stimolo standard è stato utilizzato uno stimolo nativo e come devianti, doppi devianti contenenti le vocali L2 e la vocale nativa. I paradigmi sperimentali, quindi, non sono strettamente comparabili. Conseguentemente, anche i risultati possono non essere facilmente paragonabili fra loro.

La difficoltà di comparazione emerge anche con il lavoro di Ylinen et al. (2010). Come nel presente lavoro, anche Ylinen e colleghi hanno utilizzato doppi devianti ma non in paradigmi *multi-feature* e non in presenza di devianti singoli. Ylinen et al. (2010) hanno trovato una MMN unica in risposta ai doppi devianti, diversamente dalle nostre MMN che presentano due picchi (come si discuterà anche in seguito). Studi successivi, pertanto, si rendono necessari per omologare quanto più possibile le procedure sperimentali per avere una visione più coerente dei risultati.

Il fatto che sia la durata lunga ad elicitare MMN maggiori e non la durata intermedia sembra poter suggerire l'esistenza di una soglia di salienza della durata. I nostri dati non ci permettono di identificare una soglia ben precisa, sebbene sembrino suggerire che una soglia minore o uguale ai 38 ms (differenza fra durata standard e intermedia) non abbia un effetto significativo, al contrario di una soglia maggiore o uguale a 75ms (differenza fra durata standard e lunga). Nell'emisfero destro, tuttavia, le ampiezze maggiori sono elicitate anche dalla durata intermedia, e non solo da quella lunga: ciò suggerirebbe che l'emisfero destro è più sensibile alle differenze di durata, anche con durate "intermedie". L'elaborazione a destra della durata intermedia e lunga è in linea con i risultati ottenuti da Nenonen et al. (2005) che ipotizzano che la MMN in risposta a stimoli caratterizzati dal tratto della durata, elicitata in parlanti di sistemi qualitativi, non può riflettere un'elaborazione fonologica ma, piuttosto, un'elaborazione acustica, come accade per suoni non linguistici (Giard, Lavikahen, Reinikainen, Perrin, Bertrand, Pernier & Näätänen, 1995; Paavilainen, Alho, Reinikainen, Sams & Naatanen, 1991). Nenonen et al. (2005) suggeriscono un'ulteriore ipotesi, ovvero che la maggiore attivazione nell'emisfero destro possa essere ricondotta ad un'attivazione frontale che maschera l'attivazione temporale sinistra (Giard, Perrin, Pernier & Bouchet, 1990; Opitz, Rinne, Mecklinger, von Cramon & Schröger, 2002; Rinne, Alho, Ilmoniemi, Virtanen & Näätänen, 2000). Tuttavia, considerando che i parlanti da noi analizzati non posseggono il tratto fonologico di durata vocalica, si può ipotizzare che si tratti di un'elaborazione di tipo meramente acustico (Näätänen, 2001). Alternativamente, non si può escludere che la durata vocalica sia elaborata da un punto di vista prosodico.

Il ricorso ai doppi devianti ci ha permesso di osservare come due deviazioni all'interno della TWI siano elaborate. Studi precedenti (e.g., Czigler, Winkler, 1996) hanno rilevato che due deviazioni presenti in uno stesso stimolo deviante sono elaborate come un unico evento anche quando presenti devianti singoli, la cui

presenza dovrebbe causare l'elaborazione della seconda deviazione che, in tale contesto, sarebbe rilevante. Dai nostri risultati parrebbe che le due deviazioni vengano elaborate come eventi sequenziali e non come un evento unico frutto di integrazione di processi paralleli: sia nel R1 che nel R2, per tutte e tre le durate, è stata osservata la presenza di una componente. Tuttavia, nel R2, per le vocali con durata standard, vi è una componente con ampiezze minori rispetto al R1, mentre per le vocali con durata intermedia e lunga, le componenti hanno pari ampiezza in entrambi i range. Osservando la morfologia delle componenti (Figura 2), inoltre, si evince chiaramente la presenza di due picchi. Ciò è particolarmente evidente per la vocale L1 /o/ di durata 255ms, ma anche per le vocali L2 aventi durata lunga, che nel R2 presentano un picco relativo all'elaborazione della seconda deviazione, la durata appunto (mentre nel R1 presentano un picco più basso).

I nostri dati, pertanto, suggeriscono che due deviazioni, sebbene veicolate dallo stesso stimolo, possono essere elaborate come eventi distinti e sequenziali. La presenza di devianti singoli (con durata pari a quello dello standard) fungerebbe da contesto informativo (Sussman, Winkler, 2001): la seconda deviazione presente nei devianti doppi veicolerebbe una nuova informazione che verrebbe, conseguentemente, elaborata.

Il presente studio differisce, quindi da, ad es., Czigler, Winkler (1996) ed è in linea con gli studi di Oceak et al. (2008) e Althen et al. (2016) che hanno osservato un'elaborazione sequenziale delle due deviazioni presenti nei devianti doppi. In entrambi questi studi i doppi devianti differiscono dallo standard per "frequenza+intensità"; nel presente studio, invece, differiscono per "frequenza+durata", deviazioni per le quali altri studi (Wolff, Schröger, 2001; Paavilainen et al., 2003) hanno osservato un solo picco con stimoli non linguistici.

Non si può escludere che la natura degli stimoli qui utilizzati abbia influito sul presente risultato. È necessario, infatti, interpretare i risultati ottenuti in una prospettiva prettamente linguistica: i nostri soggetti, parlanti nativi dell'italiano, *quantity system* per ciò che concerne le sole consonanti, e non per le vocali, sembrano aver elaborato i valori formantici prima, determinanti per il processo di categorizzazione e, solo successivamente, la durata. Ciò suggerirebbe che l'elaborazione della durata vocalica nei parlanti italiani non avviene come nei parlanti nativi di *quantity systems* "puri", per i quali si ipotizza l'esistenza di due meccanismi neurali indipendenti e paralleli, uno per la dimensione formantica e l'altro per la durata (Ylinen et al., 2005). Ulteriori ricerche sono necessarie per verificare se l'elaborazione delle durata consonantica sia in linea con quella dei parlanti *quantity*, avendo valenza fonologica per i parlanti italiani.

In conclusione, l'uso di suoni L2 simili e nuovi, così come l'uso di devianti doppi, sembra suggerire che quando i parlanti nativi dell'italiano (*quantity system* per le consonanti, ma non per le vocali) percepiscono suoni vocalici, elaborano i valori formantici e la durata come eventi separati e sequenziali: in primo luogo elaborano i valori formantici, fondamentali per il processo di categorizzazione e, solo successivamente, elaborano la loro durata, indipendentemente dalla vocale che percepiscono.

Riferimenti bibliografici

Althen, H., Huotilainen, M., Grimm, S. & Escera, C. (2016). Middle latency response correlates of single and double deviant stimuli in a multi-feature paradigm. In *Clinical Neurophysiology*, 127(1), 388-396.

Best, C.T., Tyler, M.D. (2007). Nonnative and second language speech perception: Commonalities and complementarities. In Munro, M.J., Bohn, O.-S. (Eds), *Second Language speech learning: the role of language experience in speech perception and production*. Amsterdam: John Benjamins, 13-34.

Bogacka, A. (2004). On the perception of English high vowels by Polish learners of English. In Daskalaki, E., Katsos, N., Mavrogiorgos, M. & Reeve, M. (Eds.), *CamLing 2004: Proceedings of the University of Cambridge second postgraduate conference in language research*. Cambridge: Cambridge University Press, 43-50.

Bohn, O.-S. (1995). Cross language speech production in adults: First language transfer doesn't tell it all. In Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 279-304.

Cebrian, J. (2006). Experience and the use of duration in the categorization of L2 vowels. In *Journal of Phonetics*, 34, 372-387.

Chládková, K., Escudero, P. & Lipski, S.C. (2013). Pre-attentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. In *Brain and language*, 126(3) 243-252.

Czigler, I., Winkler, I. (1996). Preattentive auditory change detection relies on unitary sensory memory representations. In *NeuroReport*, 7(15-17), 2413-2418.

Dehaene-Lambertz, G., Baillet, S. (1998). A phonological representation in the infant brain. In *NeuroReport*, 9.8, 1885-1888.

Escudero, P., Benders, T., Lipski, S.C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. In *Journal of Phonetics*, 37(4), 452-465.

Escudero, P., Sisinni, B. & Grimaldi, M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. In *The Journal of the Acoustical Society of America*, 135(3), 1577-1584.

Flege, J.E., Bohn, O.-S. & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. In *Journal of Phonetics*, 25, 437-470.

Giard, M.H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J. & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. In *Journal of Cognitive Neuroscience*, 7(2), 133-143.

Giard, M.H., Perrin, F., Pernier, J. & Bouchet, P. (1990). Brain Generators Implicated in the Processing of Auditory Stimulus Deviance: A Topographic Event-Related Potential Study. In *Psychophysiology*, 27(6), 627-640.

Horváth, J., Czigler, I., Winkler, I. & Teder-Sälejärvi, W.A. (2007). The temporal window of integration in elderly and young adults. In *Neurobiology of aging*, 28(6), 964-975.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y.I., Kettermann, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. In *Cognition*, 87(1), B47-B57.

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M. & Alho, K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds?. In *Cognitive Brain Research*, 459-466.

Kondaurova, M., Francis, A. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. In *Journal of the Acoustical Society of America*, 124, 3959-3971.

Lipski, S.C., Escudero, P. & Benders, T. (2012). Language experience modulates weighting of acoustic cues for vowel perception: An event-related potential study. In *Psychophysiology*, 49(5), 638-650.

Morrison, G.S. (2002). Perception of English /i/ and / I / by Japanese and Spanish listeners: Longitudinal results. In Morrison, G.S., Zsoldos, L. (Eds.), *Proceedings of the north west linguistics conference 2002*. Burnaby, BC, Canada: Simon Fraser University Linguistics Graduate Student Association, 29-48.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). In *Psychophysiology*, 38(1), 1-21.

Näätänen, R., Paavilanen, P., Rinne, T. & Alho, K. (2004). The mismatch negativity (MMN) in basic research of central auditory processing: A review. In *Clinical Neurophysiology*, 118, 2544-2590.

Näätänen, R., Pakarinen, S., Rinne, T. & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. In *Clinical Neurophysiology*, 115(1), 140-144.

Nenonen, S., Shestakova, A., Huotilainen, M. & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. In *Brain and language*, 92(1), 26-32.

Oceák, A., Winkler, I. & Sussman, E. (2008). Units of sound representation and temporal integration: A mismatch negativity study. In *Neuroscience letters*, 436(1), 85-89.

Opitz, B., Rinne, T., Mecklinger, A., Von Cramon, D.Y. & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. In *NeuroImage*, 15, 167-174.

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M. & Näätänen, R. (1991). Right hemisphere dominance of different mismatch negativities. In *Electroencephalography and clinical neurophysiology*, 78(6), 466-479.

Paavilainen, P., Mikkonen, M., Kilpelainen, M., Lehtinen, R., Saarela, M. & Tapola, L. (2003). Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: a human auditory event-related potential study. In *Neuroscience Letters*, 379, 79-82.

Rauber, A.S., Escudero, P., Bion, R. & Baptista, B.O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. In *Proceedings of Interspeech 2005*, 2913-2916.

Rinne, T., Alho, K., Ilmoniemi, R.J., Virtanen, J. & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. In *Neuroimage*, 12(1), 14-19.

Sisinni, B., Escudero, P. & Grimaldi, M. (2013). Salento Italian listeners' perception of American English vowels. In *Proceedings of Interspeech 2013*, 2091-94.

Sussman, E., Winkler, I., Ritter, W., Alho, K. & Näätänen, R. (1999). Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. In *Neuroscience letters*, 264(1), 161-164.

Tyler, M.D., Best, C.T., Faber, A. & Levitt, A.G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. In *Phonetica*, 71(1), 4-21.

Van Leussen, J.-W., Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. In *Frontiers in Psychology*, 6, 1-12.

Wang, W., Datta, H. & Sussman, E. (2005). The development of the length of the temporal window of integration for rapidly presented auditory information as indexed by MMN. In *Clinical neurophysiology*, 116(7), 1695-1706.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A. & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. In *Psychophysiology*, 36(5), 638-642.

Wolff, C., Schröger, E. (2001). Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. In *Neuroscience Letters*, 311, 37-40.

Ylinen, S., Houtilainen, M. & Näätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. In *Neuroreport*, 1857-1860.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R. & Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. In *Journal of Cognitive Neuroscience*, 22(6), 1319-1332.

Aspetti prosodici

DOI: 10.17469/O2102AISV000006

DEBORA VIGLIANO, ELISA PELLEGRINO, MASSIMO PETTORINO

L'apprendimento della prosodia dell'italiano in contesto LS: uno studio su apprendenti giapponesi

This study aims to test the effectiveness of self-imitation technique to develop a native-like prosodic competence. Seven Japanese learners of Italian (NNSs) and 2 Italian native speakers (NSs) were asked to read aloud and record two sentences in Italian conveying different pragmatic functions. The utterances of NNSs received the suprasegmental features of NSs' productions through the transplantation technique. NNSs imitated their own voice previously modified to match the reference NSs and recorded the new performance. 17 native Italian listeners rated pre- and post-training productions for pragmatic function and accentedness. The same productions were compared to the NSs' performance by utterance duration and length of vocalic and consonantal intervals. The results showed that self-imitation improves NNSs communicative effectiveness. After the training, utterance duration and vocalic intervals length better match the target duration.

Key words: L2 prosody training, self-imitation, prosodic transplantation technique.

Introduzione

L'acquisizione degli aspetti fonetico-fonologici e ritmico-prosodici rappresenta, per gli apprendenti adulti di una seconda lingua (L2), uno dei principali ostacoli al raggiungimento di una competenza linguistico-comunicativa comparabile a quella dei parlanti nativi (Birdsong, 2006). Il parlato in una L2 è, infatti, caratterizzato dalla presenza di deviazioni segmentali e soprasegmentali dalla pronuncia nativa, che lo rendono straniero all'orecchio del parlante madrelingua (Moyer, 2013). Quanto le due componenti concorrano alla formulazione del giudizio di accento straniero è tuttora materia di aperti dibattiti. Studi appositamente concepiti per testare il contributo specifico delle deviazioni segmentali e soprasegmentali nella percezione dell'accento non nativo offrono risultati contrastanti. In alcuni studi si comprova la preminenza percettiva degli errori segmentali (Boula de Mareüil, Marotta & Adda-Decker, 2004); in altre ricerche quella delle alterazioni di velocità, pause, ritmo e andamento intonativo (Anderson-Hsieh, Johnson & Koehler, 1992; Boula de Mareüil, Vieru-Dimulescu, 2006; Carmichael, 2000; Magen, 1998). Si rimanda a Ulbrich, Mennen (2015) per una trattazione esaustiva dell'argomento.

In ambito glottodidattico, un'attenzione specifica è stata dedicata alla corretta pronuncia dei segmenti (Pennington, Richards, 1986), mentre è stato a lungo trascurato l'insegnamento della prosodia (Derwing, Munro, 2015). Le ragioni di tale lacuna nella pratica didattica vanno ricercate in una pluralità di cause: la mancanza di strumenti e risorse didattiche adeguate, la modesta preparazione specifica degli

insegnanti di lingue e, per finire, la scarsa considerazione in cui vengono generalmente tenuti in conto gli aspetti paralinguistici della comunicazione.

Al giorno d'oggi, tuttavia, tanto i docenti quanto gli apprendenti di lingue straniere possono contare su una gamma di risorse più vasta rispetto al passato per migliorare l'accento e l'intelligibilità delle produzioni in L2 (Chun, Hardison & Pennington, 2008). Come per altri aspetti dell'apprendimento e insegnamento linguistico, la tecnologia e le risorse web hanno offerto proposte anche nel campo della didattica della fonetica della L2, mediante i cosiddetti sistemi di insegnamento della pronuncia assistiti dal computer o Computer Assisted Pronunciation Teaching (CAPT). Tali sistemi, che presentano innumerevoli vantaggi per chi intende lavorare sulle abilità orali, propongono infatti percorsi didattici: 1. personalizzati, 2. erogabili in ambienti di apprendimento rilassati e meno ansiogeni rispetto alla classe di lingua, 3. calibrati sui tempi e i modi di apprendimento dell'utente (Neri, Cucchiarini, Strik & Boves, 2002). Non meno rilevante è la possibilità offerta dai CAPT di fornire agli utenti feedback correttivi mirati (Hismanoglu, 2011).

Benché tali sistemi non vadano considerati come sostituti degli insegnanti, essi risultano particolarmente utili nell'apprendimento delle lingue in contesto di lingua straniera (LS), dove le possibilità di esposizione alla lingua target sono ridotte, e gli insegnanti non madrelingua, pur essendo ineccepibili sul piano morfo-sintattico, lessicale, testuale e pragmatico, possono conservare nella loro pronuncia tracce della lingua materna (Seferoglu, 2005; Neri, Mich, Gerosa & Giuliani, 2008; Levis, 2007).

I sistemi CAPT non esauriscono il loro campo d'azione all'insegnamento delle peculiarità segmentali della lingua oggetto di studio; il loro dominio applicativo si estende anche alla didattica degli aspetti soprasegmentali, tra cui l'accento di parola, l'accento di frase, il ritmo e l'intonazione (Donaldson, 2009). Relativamente alla specifica dimensione prosodico-intonativa, ricerche condotte nel settore della tecnologia della voce applicata all'apprendimento linguistico hanno dimostrato che non sono sufficienti semplici attività di ascolto e di imitazione delle produzioni native, affinché gli enunciati dei madrelingua possano trasformarsi in modelli prosodici interiorizzabili e riproducibili da parte dell'apprendente. Perché un tal tipo di input possa trasformarsi in *intake*, l'apprendente dovrebbe piuttosto essere esposto ad una voce quanto più simile alla propria dal punto di vista prosodico. I risultati dello studio di Probst, Ke & Eskenazi (2002) hanno infatti evidenziato che quanto migliore è l'abbinamento tra le voci degli apprendenti e quelle dei parlanti nativi in termini di velocità di articolazione e di F0, tanto più efficace risulta l'esercitazione. Tali risultati hanno indotto Felps, Bortfeld & Gutierrez-Osuna (2009) ad ipotizzare che il 'golden speaker' più efficace per apprendere la pronuncia di una L2 sarebbe proprio la voce dell'apprendente rimodellata su quella nativa. Le attività di esercitazione della pronuncia dovrebbero quindi riorientarsi verso l'auto-imitazione: far ascoltare agli apprendenti di L2 le proprie voci che riproducono enunciati con accento nativo, avvalendosi di metodi di conversione prosodica quali il trapianto prosodico-intonativo (Charpentier, Moulines, 1989).

L'efficacia pedagogica dell'auto-imitazione per l'apprendimento della prosodia è stata precedentemente testata su apprendenti giapponesi di inglese (Nagano, Ozawa, 1990), su apprendenti italiani di tedesco (Bissiri, Pfitzinger & Tillmann, 2006) e su apprendenti inglesi di cinese mandarino (Peabody, Seneff, 2006). Relativamente all'italiano, la validità del trapianto prosodico come tecnica di insegnamento della prosodia è stata accertata sperimentalmente su apprendenti cinesi in contesto L2 (Pettorino, De Meo & Vitale, 2012; De Meo, Vitale, Pettorino, Cutugno & Origlia, 2013; De Meo, Vitale & Pellegrino, 2016). A seguito dell'esercitazione basata sulla tecnica dell'auto-imitazione, gli studenti cinesi erano riusciti a produrre richieste, affermazioni e comandi con profili prosodico-intonativi conformi alle aspettative degli ascoltatori madrelingua italiani.

1. *Lo studio: obiettivi e partecipanti*

Il principale obiettivo della ricerca è stato quello di estendere le indagini sperimentali sull'efficacia pedagogica della tecnica dell'auto-imitazione per il miglioramento della competenza prosodica in italiano, coinvolgendo partecipanti con caratteristiche individuali diverse da quelle testate nei lavori precedenti.

Per tale scopo, lo studio non è stato condotto su apprendenti di italiano in contesto L2, bensì su studenti giapponesi di italiano in contesto LS, con rari contatti con parlanti madrelingua al di fuori del contesto universitario. I soggetti coinvolti (2 maschi e 5 femmine), di età compresa tra 21 e 28 anni, avevano studiato l'italiano per 5-6 anni, conseguendo un livello di competenza linguistica pari al B2 del Quadro Comune Europeo di Riferimento (Council of Europe, 2001). Tutti avevano studiato lingua e linguistica italiana per un anno in Italia. Al momento della ricerca, nessuno aveva deficit uditivi o linguistici.

Due parlanti nativi italiani (1 maschio e 1 femmina), di 27 e 25 anni, hanno preso parte alla ricerca come "donatori" dei loro parametri prosodici agli apprendenti giapponesi, considerati come "riceventi". Al momento della ricerca, i due madrelingua italiani risiedevano in Giappone.

Ulteriore obiettivo della ricerca è stato indagare in via preliminare i cambiamenti nella modulazione dei parametri ritmico-prosodici da parte degli apprendenti giapponesi a seguito dell'auto-imitazione.

2. *Articolazione del lavoro*

Il lavoro è stato articolato in cinque fasi progressive e interdipendenti: Preesercitazione, Manipolazione del segnale acustico, Auto-imitazione, Test percettivo e Analisi Acustica – ciascuna delle quali basata su specifiche attività, sinteticamente descritte in Tab. 1.

Tabella 1 - *Fasi e attività della ricerca*

Fase	*Attività svolte*
1. Pre-esercitazione	– Lettura e registrazione di due frasi in italiano con tre intenzioni comunicative da parte degli studenti giapponesi e degli italiani madrelingua.
2. Manipolazione del segnale acustico	– Trapianto delle caratteristiche prosodico-intonative degli enunciati dei 'donatori' nativi sui corrispondenti enunciati prodotti dai 'riceventi' giapponesi.
3. Auto-imitazione	– Esercitazione basata sull'ascolto e sull'auto-imitazione degli enunciati modificati prosodicamente sul modello nativo. – Registrazione del corpus di enunciati in fase di post-esercitazione.
4. Test percettivo	– Valutazione percettiva degli enunciati prodotti in fase di pre- e post-esercitazione per 1) intenzione comunicativa e 2) accento straniero, ad opera di ascoltatori italiani madrelingua. – Analisi dei risultati del test percettivo per verificare l'efficacia della tecnica dell'auto-imitazione.
5. Analisi acustiche	– Analisi comparative delle durate degli enunciati e dei relativi intervalli vocali e consonantici prodotti dai parlanti nativi e dagli studenti giapponesi nelle fasi di pre- e post-esercitazione.

Nella presentazione del lavoro verranno quindi ripercorse le tappe di sviluppo della ricerca.

2.1 La fase 1: pre-esercitazione

Per esigenze di comparabilità degli enunciati prodotti dai parlanti nativi con quelli dei non nativi ai fini della procedura di manipolazione, i soggetti sono stati coinvolti in un'attività di parlato letto. Come testo di input sono state selezionate due frasi ("Accendi la radio" e "Chiudi la finestra"), ciascuna da produrre con tre diverse funzioni comunicative (concessione, comando e richiesta). Le stesse frasi sono state prodotte dai due parlanti nativi.

Questo tipo di attività risulta particolarmente impegnativa per gli apprendenti giapponesi dato che la loro lingua materna, a differenza dell'italiano, ricorre prevalentemente ad elementi sintattici e morfologici per variare la modalità dell'enunciato (D'Imperio, 2002; Abe, 1998). Tale attività quindi rispondeva alla specifica esigenza di testare la capacità degli studenti giapponesi di veicolare intenzioni comunicative diverse, variando esclusivamente i parametri prosodico-intonativi, senza aver ascoltato in precedenza alcun modello né aver ricevuto indizi su come svolgere la prova.

Per garantire la corretta esecuzione del compito, le frasi e le funzioni comunicative da veicolare erano state precedentemente tradotte in giapponese. Di seguito si riporta la modalità con cui sono state presentate le frasi.

Frase 1
RICHIESTA Accendi la radio?
質問 ラジオつけてくれない?
Shitsumon rajio tsukete kurenai?[1]
COMANDO Accendi la radio!
命令 ラジオつけて!
Meirei rajio tsukete![2]
CONCESSIONE Accendi la radio.
譲歩 ラジオつけていいよ
Jōho rajio tsukete ii yo.[3]

Frase 2
RICHIESTA Chiudi la finestra?
質問 窓閉めてくれない?
Shitsumon mado shimete kurenai?
COMANDO Chiudi la finestra!
命令 窓閉めて!
Meirei mado shimenasai!
CONCESSIONE Chiudi la finestra.
譲歩 窓閉めていいよ
Jōho mado shimete ii yo

Dopo una breve fase di lettura silenziosa delle frasi, gli studenti hanno registrato il corpus di enunciati. Le registrazioni sono avvenute in una stanza silenziosa della Tokyo University of Foreign Studies, in sessioni singole, mediante il software Audacity 2.0.5, ad una frequenza di campionamento di 44.100 Hz. Lo stesso pro-

[1] Ci sono vari modi di fare richieste in giapponese a seconda del grado di gentilezza e/o di quanto si voglia essere indiretti. In generale, al verbo principale coniugato alla forma sospensiva in –te si aggiunge un secondo verbo (/kureru/, /morau/ o le loro versioni onorifiche /kudasaru/, /itadaku/). Il secondo verbo può anch'esso essere coniugato per esprimere diversi livelli di formalità. Nella frase 'Accendi la radio?' /tsukeru/ 'accendere' diventa tsuke-te+ kureru+nai (forma neg.) lett. 'Non accenderesti la radio per me? Le richieste possono essere formulate anche in maniera più diretta ed esplicita, aggiungendo alla forma in –te l'imperativo dell'onorifico /kudasaru/ovvero /kudasai/.

[2] Anche nel caso dei comandi la lingua giapponese ricorre a diverse coniugazioni, a seconda del livello di gentilezza che si vuole mostrare. Come regola generale al verbo principale si aggiungono suffissi o verbi secondari che forniscono il significato di comando. Procedendo per grado crescente di cortesia, si passa, ad esempio, dalla coniugazione imperativa di base meireikei /tsukero/, dall'accezione piuttosto negativa, alla forma V-nasai /tsuke-nasai/ che grazie all'aggiunta del verbo onorifico /nasaru/ rende il comando meno diretto e più gentile. Tuttavia le forme imperative sono usate raramente poiché considerate rudi durante normali conversazioni. Al loro posto è più comune usare la summenzionata forma in –te + kudasai che trasforma il comando in una più gentile richiesta. L'omissione del /kudasai/ rende la frase più colloquiale. Da qui la scelta nel presente studio della sola forma in –te per rendere il comando in giapponese.

[3] Per fare concessioni il giapponese ricorre alla già citata forma sospensiva in –te con l'aggiunta della aggettivo /ii/ 'buono'/'bene' e della particella di fine frase /yo/(V-te + ii+ yo), seguendo pertanto lo stesso pattern dell'inglese /OK, you can turn on the radio/. La particella /yo/, molto frequente nel giapponese standard conversazionale, è comunemente descritta come 'modale' in quanto esprime l'attitudine o l'opinione del parlante nei confronti dell'interlocutore.

tocollo di registrazione è stato usato con i parlanti nativi. Il corpus di parlato letto è risultato composto da 54 enunciati, di cui:

- 42 in italiano LS (7 parlanti giapponesi * 2 enunciati * 3 intenzioni comunicative);
- 14 richieste, 14 comandi, 14 concessioni;
- 12 in italiano L1 (2 parlanti italiani * 2 enunciati * 3 intenzioni comunicative);
- 4 richieste, 4 comandi, 4 concessioni.

2.2 La fase 2: manipolazione del segnale acustico

Per somministrare l'attività di auto-imitazione, e quindi consentire a ciascuno studente giapponese di ascoltare e imitare la propria voce con prosodia nativa italiana, è stato necessario trapiantare i parametri soprasegmentali delle produzioni native sui corrispondenti enunciati realizzati dai giapponesi. La procedura di manipolazione, basata sull'algoritmo PSOLA - *Pitch- Synchronous Overlap and Add* (Charpentier et al., 1989) implementato nel software Praat (Boersma, Weenink, 2016), è basata su una serie di operazioni fisse, riportate brevemente in Tab. 2.

Tabella 2 - *Procedura per il trapianto prosodico*

1. Segmentazione manuale degli enunciati in italiano L1 e LS in intervalli vocalici e consonantici
2. Trattamento manuale delle anomalie e allineamento dei segmenti tra gli enunciati in italiano L1 e LS
3. Trapianto delle durate
4. Sovrapposizione del contorno intonativo

Il trattamento manuale delle anomalie è stato effettuato parallelamente sui file audio degli enunciati in italiano L2 e i corrispondenti file prodotti dal donatore nativo. Dato l'elevato livello di competenza linguistica dei partecipanti, gli enunciati erano privi di pause piene, quindi si è proceduto direttamente alla rimozione delle pause silenti inter-frasali prodotte dagli apprendenti, seguendo la metodologia descritta in Pettorino, Vitale (2012: 12):

> [...] the silences not corresponding to the silent pauses of the donor have to be eliminated from the utterance produced by the receiver. At the same time, intervals of silence must be added inside the receiver's utterance in correspondence with the remaining donor's silent pauses.

Il trapianto delle durate e la sovrapposizione del contorno intonativo sono stati effettuati automaticamente mediante l'algoritmo PSOLA.

Per l'abbinamento donatore-ricevente è stato seguito il criterio della corrispondenza di genere. La voce del parlante nativo è stata appaiata alle voci dei due parlanti giapponesi di sesso maschile. La voce della parlante nativa, invece, è servita come modello per gli enunciati prodotti dalle 5 apprendenti giapponesi.

Al termine della fase di manipolazione degli enunciati è stato ottenuto un nuovo corpus di 42 enunciati, successivamente utilizzato per l'attività di esercitazione, basata sull'auto-imitazione.

2.3 La fase 3: auto-imitazione

Nella fase di auto-imitazione gli studenti giapponesi hanno ascoltato in cuffia i propri enunciati modificati prosodicamente sul modello nativo e si sono esercitati a ripeterli ad alta voce, con l'obiettivo di avvicinarsi il più possibile al modello prosodico-intonativo proposto. Raggiunto un grado di avvicinamento giudicato soddisfacente dagli stessi apprendenti, ciascuno studente ha registrato nuovamente gli enunciati esercitati, seguendo lo stesso protocollo utilizzato nella fase di pre-esercitazione.

È importante sottolineare che nel periodo intercorso tra la fase di pre-esercitazione e quella di auto-imitazione (circa 3 settimane) gli apprendenti non hanno frequentato corsi specifici per il miglioramento della prosodia e dell'intonazione in italiano. Il corpus di enunciati a seguito dell'esercitazione (da ora post-esercitazione) è costituito da 42 produzioni, di cui 14 richieste, 14 comandi, 14 concessioni.

2.4 La fase 4: il test percettivo

Il set di enunciati sottoposto a valutazione percettiva consisteva complessivamente di 84 enunciati in italiano LS, di cui 42 letti e registrati nella fase di pre-esercitazione e 42 prodotti dopo l'auto-imitazione.

Gli 84 enunciati sono stati proposti in sequenza randomizzata e suddivisi in tre gruppi di 28 frasi l'uno, intervallati da una pausa di 10 minuti per evitare il sovraccarico cognitivo.

Il test è stato somministrato in modalità *online* mediante il software *SurveyGizmo* ed è stato svolto individualmente, attraverso l'ascolto in cuffia degli enunciati. Ai partecipanti era stato detto che avrebbero ascoltato delle produzioni di apprendenti di italiano che stavano esercitandosi nella pronuncia. Nessuna specifica indicazione era stata fornita sulle finalità dello studio. Gli item del test sono stati valutati sia per il grado di accento straniero (scala da 1 a 5 punti: 1 = accento nativo; 5 = accento straniero molto forte), sia per funzione comunicativa veicolata (cinque alternative, di cui tre attese – richiesta (R), comando (C), concessione (Cn) – e due distrattori – affermazione (A), altro). I file potevano essere ascoltati soltanto una volta.

Alla valutazione percettiva hanno partecipato 17 ascoltatori nativi italiani, di provenienza campana, studenti dell'università di Napoli L'Orientale, di età compresa tra i 23 e i 30 anni; tutti avevano familiarità con diversi accenti stranieri ma nessuno conosceva il giapponese. Le lingue straniere conosciute erano prevalentemente l'inglese, il francese, lo spagnolo e il cinese.

2.4.1 Analisi dei risultati

Prima di procedere al commento dei risultati del test percettivo, è opportuno chiarire i criteri adottati per l'analisi dei dati. Verrà inizialmente esaminata la relazione tra intenzioni attese e intenzioni percepite per gli enunciati prodotti nelle fasi di pre-esercitazione (Tab. 3) e post-esercitazione (Tab. 4). Per valutare in maniera più puntuale la validità dell'auto-imitazione verranno successivamente confrontate:

1. la percentuale di risposte corrette ottenute prima e dopo l'esercitazione (Tab. 5);
2. la percentuale di corretto abbinamento tra intenzione comunicativa attesa e percepita per fase di esercitazione e atto linguistico (Tab. 6).

Infine è stata considerata la validità dell'auto-imitazione per l'attenuazione dell'accento straniero.

Il rapporto tra le intenzioni comunicative attese e quelle percepite nella fase di pre-esercitazione, rappresentato dalla matrice di confusione in Tab. 3, conferma i dati ottenuti nello studio condotto su apprendenti cinesi di livello A2 (De Meo et al., 2016). La richiesta è la funzione comunicativa veicolata più accuratamente dagli studenti giapponesi, anche senza esercitazione, poiché riconosciuta dalla maggior parte degli ascoltatori italiani (52,74%). La percentuale di corretta identificazione scende al di sotto del 40% per i comandi, che vengono per lo più confusi con le richieste (32,35%). Le concessioni vengono riconosciute come tali solo da una minima percentuale di ascoltatori (8,44%), mentre quasi il 50% le scambia per comandi (47,68%).

Tabella 3 - *Matrice di confusione tra intenzioni attese e percepite nella fase di pre-esercitazione*

Pre		*Intenzioni percepite*				
Intenzioni attese		*Richiesta*	*Comando*	*Concess.*	*Affermaz.*	*Altro*
	Richiesta	52,74%	16,88%	5,06%	11,81%	13,50%
	Comando	32,35%	39,92%	10,92%	13,87%	2,94%
	Concess.	20,25%	47,68%	8,44%	18,57%	5,06%

Le prestazioni degli studenti giapponesi migliorano considerevolmente a seguito dell'esercitazione. Come mostrato in Tab. 4, la confusione tra intenzioni attese e intenzioni percepite decresce considerevolmente per ordini e concessioni. Il comando è riconosciuto correttamente quasi dal 60% degli ascoltatori, mentre la concessione dal 47,06%. La richiesta si riconferma l'intenzione più chiaramente riconosciuta dagli ascoltatori (75,21%), subendo un incremento di circa il 20% nella percentuale di corretta identificazione rispetto alla fase di pre-esercitazione (Richiesta: Pre 52,74%; Post 75,21%).

Tabella 4 - *Matrice di confusione tra intenzioni attese e percepite nella fase di post-esercitazione*

Post		*Intenzioni percepite*				
Intenzioni attese		*Richiesta*	*Comando*	*Concess.*	*Affermaz.*	*Altro*
	Richiesta	75,21%	12,61%	4,20%	5,88%	2,10%
	Comando	14,29%	57,98%	11,34%	14,71%	1,68%
	Concess.	17,23%	11,34%	47,06%	17,23%	7,14%

I dati delle Tabb. 5 e 6 consentono di comprendere l'efficacia della tecnica di auto-imitazione ai fini del miglioramento della competenza comunicativa. La percentuale di corretta corrispondenza tra intenzioni attese e quelle percepite nella fase di post-esercitazione supera in maniera statisticamente significativa quella ottenuta nella fase pre-esercitazione, con uno scarto di quasi il 30% (ANOVA a misure ripetute [$F(1,32) = 65.18$, $p < .001$].

Altrettanto significative le differenze nelle percentuali di corretto riconoscimento delle tre funzioni comunicative nella fase precedente e successiva alla esercitazione (Tab. 6) (ANOVA a misure ripetute [$F(2; 32) = 32.13$, $p < 0.001$]. Questi risultati suggeriscono pertanto che l'attività di auto-imitazione migliora la capacità degli apprendenti giapponesi di modulare i parametri prosodico-intonativi in maniera conforme alle aspettative degli ascoltatori nativi.

Tabella 5 - *Percentuale di corretta identificazione tra intenzione comunicativa attesa e percepita per fase di esercitazione*

	Pre-esercitazione (A)	*Post-esercitazione (B)*	*Differenza (B – A)*
Media	33,61%	60,04%	+ 26,43

Tabella 6 - *Percentuale di corretta identificazione tra intenzione comunicativa attesa e percepita per atto linguistico e fase di esercitazione*

	Pre-esercitazione (A)	*Post-esercitazione (B)*	*Differenze (B – A)*
Richieste	52,52%	75,21%	22,69
Comandi	39,92%	57,98%	18,06
Concessioni	8,40%	47,06%	38,66

Tuttavia, la tecnica di auto-imitazione esercita un'influenza diversa a seconda della funzione comunicativa da veicolare. L'analisi statistica dei dati rivela infatti l'esistenza di un'interazione significativa tra fase di esercitazione * intenzione comunicativa [$F(2;32) = 3.51$, $p < 0.005$]. Infatti, come mostrato dalla quarta colon-

na della Tab. 6 (Differenze B-A), il miglioramento più cospicuo è ottenuto dalle concessioni. La percentuale di corretta identificazione infatti varia da 8,4% nella fase pre-esercitazione a 47,06% nella fase post-esercitazione. Tale risultato non sorprende, data la minore frequenza d'occorrenza delle concessioni nell'input e della scarsa attenzione verso la realizzazione prosodica degli atti linguistici. Grazie all'auto-imitazione si dà agli apprendenti la possibilità di migliorarsi su versanti delle competenza comunicativa, a cui non avrebbero altrimenti accesso.

Per quanto riguarda la validità dell'auto-imitazione ai fini dell'indebolimento dell'intensità dell'accento straniero, i dati indicano che il giudizio di forestierismo non subisce sensibili variazioni prima e dopo l'esercitazione (Tab. 7). Nonostante l'elevato livello di competenza linguistica degli apprendenti, il loro eloquio risulta ancora appesantito da marcate deviazioni segmentali, la cui salienza percettiva non viene evidentemente scalfita da una produzione prosodicamente affine a quella dei madrelingua.

Tabella 7 - *Media di accento nelle fasi di pre- e post-esercitazione (1= accento nativo; 5= accento straniero molto forte)*

	Media di accento
Pre-esercitazione	3,43
Post-esercitazione	3,53

2.5 La fase 5: analisi acustica

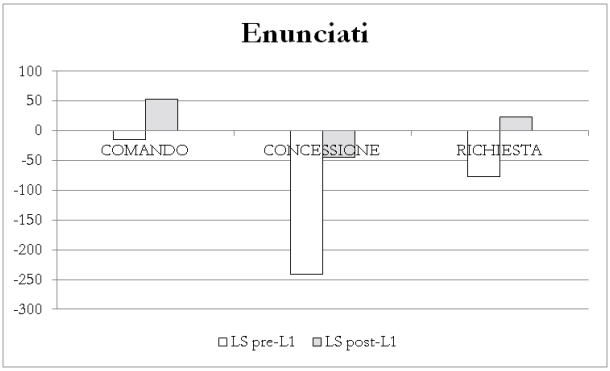
Al fine di indagare i cambiamenti nella modulazione dei parametri ritmico-prosodici a seguito dell'auto-imitazione, il corpus di enunciati prodotti nelle fasi di pre- e post-esercitazione è stato oggetto di analisi spettro-acustica (Boersma, Weenink, 2016). Date le differenze ritmiche tra l'italiano e il giapponese (Ramus, Nespor & Mehler, 1999), in questo primo studio, tra i diversi correlati fonetici che contribuiscono alle variazioni prosodiche, l'attenzione è stata concentrata in via preliminare sulla durata.

Sulla base delle segmentazioni effettuate per l'attuazione del trapianto prosodico, di ciascun enunciato realizzato dai parlanti nativi e dagli apprendenti giapponesi nelle fasi di pre- e post-esercitazione sono state ricavate automaticamente la durata complessiva e la lunghezza dei relativi intervalli vocalici e consonantici. Per le produzioni degli studenti giapponesi è stato calcolato il valore medio degli enunciati suddivisi per funzione comunicativa (richiesta, comando e concessione) e fase di esercitazione (pre e post). Nel caso dei madrelingua italiani, invece, è stato computato il valore medio ottenuto dai due parlanti nei comandi, nelle richieste e nelle concessioni.

Per ciascuna funzione comunicativa, inoltre, sono stati computati i valori differenziali di durata tra:

– gli enunciati prodotti dagli apprendenti giapponesi (LS) nelle fasi di pre- e post-esercitazione e i corrispondenti enunciati prodotti dai parlanti nativi (LS pre – L1; LS post – L1)(Fig. 1)
– gli intervalli vocalici e consonantici prodotti dagli apprendenti giapponesi (LS) nelle fasi di pre- e post-esercitazione e i corrispondenti intervalli prodotti dai parlanti nativi (L1)(Figg. 2-3)

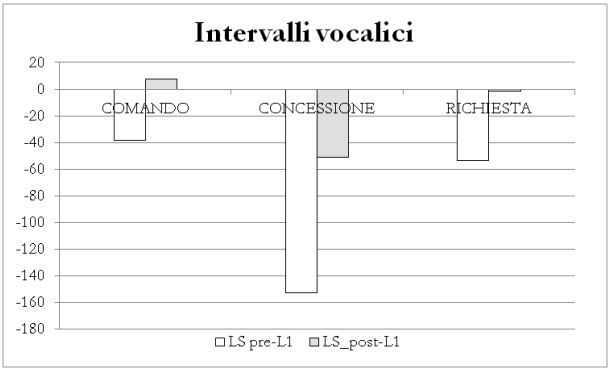
Figura 1 - *Confronto tra la differenza delle durate medie degli enunciati in LS rispetto alle durate in L1 per atto linguistico e per fase di esercitazione (0 = stessa durata del modello)*



Analizzando le differenze nella durata media degli enunciati prodotti nelle fasi di pre- e post-esercitazione con la durata media degli enunciati equivalenti prodotti dai parlanti nativi (Fig.1), è emerso che l'attività di auto-imitazione ha favorito l'avvicinamento delle produzioni degli apprendenti a quelle dei modelli. I valori negativi dei comandi, delle richieste e delle concessioni ottenuti nella fase di pre-esercitazione indicano che i corrispondenti enunciati sono stati prodotti ad una velocità superiore rispetto a quella dei parlanti madrelingua. L'auto-imitazione ha quindi determinato un rallentamento dell'eloquio, piuttosto che un'accelerazione. L'avvicinamento più consistente è stato ottenuto nelle concessioni, l'atto più complesso da veicolare senza specifiche istruzioni. Tale dato concorrerebbe a spiegare le ragioni dell'accresciuto riconoscimento di questo atto linguistico sul piano percettivo a seguito dell'esercitazione (Cfr. Tab.6).

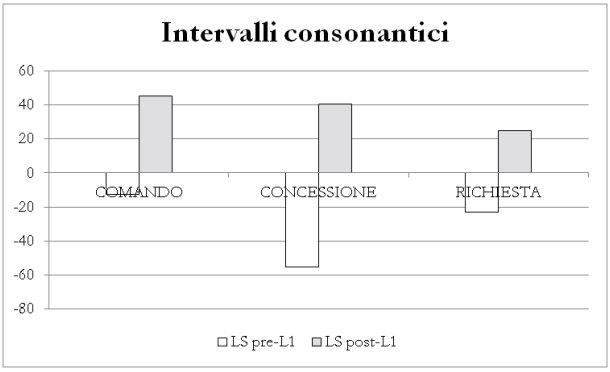
Per quanto riguarda il contributo delle porzioni vocaliche e consonantiche alla riduzione della distanza rispetto al modello, l'analisi dei dati mostra che il ruolo maggiore, in questo senso, è stato giocato dalle vocali. Come appare evidente in Fig.2, la produzione delle vocali migliora nettamente dopo l'esercitazione prosodica, approssimandosi così alla durata del valore di riferimento (0 = stessa durata del parametro target).

Figura 2 - *Confronto tra la differenza delle durate medie degli intervalli vocalici in LS rispetto alle durate in L1 per atto linguistico e per fase di esercitazione (0 = stessa durata del modello)*



Le consonanti tendono invece ad allungarsi nella fase di post-esercitazione, risultando più lunghe rispetto a quelle dei nativi (Fig.3).

Figura 3 - *Confronto tra la differenza delle durate medie degli intervalli consonantici in LS rispetto alle durate in L1 per atto linguistico e per fase di esercitazione (0 = stessa durata del modello)*



3. *Conclusioni*

Il presente studio ha dimostrato la validità dell'esercitazione basata sull'auto-imitazione ai fini del miglioramento dell'efficacia comunicativa in apprendenti giapponesi di italiano in contesto LS. La percentuale di corretto riconoscimento delle tre funzioni comunicative aumenta, infatti, in maniera significativa a seguito dell'esercitazione prosodica. Il miglioramento riguarda in particolar modo le concessioni. Le durate medie di tutti gli enunciati in L2 si avvicinano a quelle del modello nativo, grazie alla riduzione delle lunghezze vocaliche. Si può, dunque, ipotizzare che le variazioni di durata vocalica dovute all'auto-imitazione abbiano rivestito un ruolo significativo ai fini della corretta identificazione delle intenzioni comunicative. Per approfondire ulteriormente la relazione tra parametri soprasegmentali e giudi-

zi percettivi, nelle prossime fasi della ricerca le analisi acustiche saranno estese alle variazioni cross-linguistiche (italiano L1, italiano LS) di tono, intensità e contorni intonativi.

Per quanto riguarda l'efficacia dell'auto-imitazione per la riduzione dell'accento straniero, la sessione di esercitazione non ha giocato un ruolo altrettanto determinante. Così come riscontrato nello studio di De Meo et al. (2016), nessuna differenza sensibile è stata riscontrata tra le fasi di pre- e post- esercitazione. Si può quindi migliorare l'intelligibilità, pur mantenendo tracce di accento straniero. Tali considerazioni sono in linea con quanto riportato in letteratura sull'indipendenza (almeno parziale) tra le due dimensioni su cui si fonda la valutazione percettiva del parlato in L2: il giudizio di forestierismo (accentdness), inteso come "a particolar pattern of pronunciation that is perceived to distinguish members of different speech communities" e l'intelligibilità, intesa come "the degree of match between a speaker's intended message and the listener's comprehension" (Derwing, Munro, 2015: 5).

Alla luce di tali considerazioni è possibile riesaminare il ruolo giocato dal piano segmentale e soprasegmentale nella percezione dell'accento straniero. I dati di questo studio quindi inducono ad affermare che le deviazioni di pronuncia determinano tendenzialmente il giudizio di accento straniero (accentdness), mentre quelle soprasegmentali minano l'intelligibilità del parlante L2, generando fraintendimenti e incomprensioni deleteri per il buon esito degli scambi comunicativi esolingui.

Riferimenti bibliografici

Abe, I. (1998). Intonation in Japanese. In Hirst, D., Di Cristo, A. (Eds.), *Intonation Systems. A survey of twenty languages*. Cambridge: Cambridge University Press, 363-378.

Anderson-Hsieh, J., Johnson, R. & Koehler, K. (1992). The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentais, Prosody, and Syllable Structure. In *Language Learning*, 42(4), 529-555.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. In *Language Learning*, 56, 9-49.

Bissiri, M.P., Pfitzinger, H.R. & Tillmann, H.G. (2006). Lexical Stress Training of German Compounds for Italian Speakers by means of Resynthesis and Emphasis. In *Proceedings of the 11th Australian International Conference on Speech Science & Technology*. New Zealand: University of Auckland, 24-29.

Boersma, P., Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.19. http://www.praat.org/Ultimo accesso 13.06.16.

Boula de Mareüil, P., Marotta, G. & Adda-Decker, M. (2004). Contribution of prosody to the perception of Spanish/Italian accents. In *Proceedings of Speech Prosody*, 2, Nara, Japan, 681-684.

Boula de Mareüil, P., Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. In *Phonetica*, 63, 247-267.

Carmichael, L. (2000). Measurable degrees of foreign accent: a correlational study of production, perception, and acquisition. MA thesis, University of Washington.

Charpentier, F., Moulines, E. (1989). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. In *Proceedings of the First European Conference on Speech Communication and Technology – Eurospeech*. Paris: European Speech Communication Association, 2013-2019.

Chun, D.M., Hardison, D.M. & Pennington, M.C. (2008). Technologies for prosody in context. Past and future of L2 research and practice. In Hansen Edwards, J.G., Zampini, M.L. (Eds.), *Phonology and Second Language Acquisition*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.

De Meo, A., Vitale, M., Pettorino, M., Cutugno, F. & Origlia, A. (2013). Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian. In Levis, J., LeVelle, K. (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, Ames, IA: Iowa State University, Ago. 2012, 90-100.

De Meo, A., Vitale, M. & Pellegrino, E. (2016). Tecnologia della voce e miglioramento della pronuncia in una L2: imitazione e autoimitazione a confronto. Uno studio su cinesi apprendenti di italiano L2. In *Atti del XV Convegno Nazionale dell'Associazione Italiana di Linguistica Applicata (AItLA)*, «Linguaggio e apprendimento linguistico: metodi e strumenti tecnologici». Lecce: Università del Salento, 13-25.

Derwing, T.M., Munro, M.J. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 teaching and research*. Amsterdam: John Benjamins.

D'Imperio, M. (2002). Italian Intonation: an overview and some questions. In *Probus*, 14(1), 37-69.

Donaldson, J.P. (2009). *Literature Review: Computer Aided Pronunciation Training.* http://www.wou.edu/~donaldsj/TestWebsitePortfolio2/TestWebsitePortfolio2/portfolioartifacts/ResearchWriting/Jonan%20Donaldson%20ED%20633%20Final%20Literature%20Review.pdf/Ultimo accesso 28.06.16.

Felps, D., Bortfeld, H. & Gutierrez-Osuna, R. (2009). Foreign accent conversion in computer assisted pronunciation training. In *Speech Communication*, 51, 920-932.

Hismanoglu, M. (2011). Computer Assisted Pronunciation Teaching: From the Past to the Present with its Limitations and Pedagogical Implications. In *Frontiers of Language and Teaching*, *Proceedings of the 2011 IOLC*, 2, 193-202.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. In *Annual Review of Applied Linguistics*, 27, 184-202.

Magen, H. (1998). The perception of foreign-accented speech. In *Journal of Phonetics*, 26, 381-400.

Moyer, A. (2013). *Foreign Accent. The Phenomenon of Non-native Speech*. Cambridge: Cambridge University Press.

Nagano, K., Ozawa, K. (1990). English speech training using voice conversion. In *1st Internat. Conf. on Spoken Language Processing (ICSLP 90)*, Kobe, Japan, 295-308.

Neri, A., Cucchiarini, C., Strik, H. & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. In *Computer Assisted Language Learning*, 15(5), 441-467.

Neri, A., Mich, O., Gerosa, M. & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. In *Computer Assisted Language Learning*, 21(5), 393-408.

Peabody, M., Seneff, S. (2006). Towards automatic tone correction in nonnative mandarin. In *Chinese Spoken Language Processing: 5th International Symposium, ISCSLP 2006*, 602-613.

Pennington, M.C., Richards, J.C. (1986). Pronunciation revisited. In *TESOL Quarterly*, 20(2), 207-225.

Pettorino, M., De Meo, A. & Vitale, M. (2012). La competenza prosodico-intonativa nell'italiano L2. Analisi e sintesi del segnale fonico di cinesi, giapponesi e vietnamiti. In *La linguistica educativa. Atti del XLIV Convegno SLI*, 329-342.

Pettorino, M., Vitale, M. (2012). Transplanting native prosody in second language speech. In Busà, M.G., Stella, A. (Eds.), *Methodological Perspectives on Second Language Prosody Papers from ML2P 2012*. Padova: Cleup, 11-16.

Probst, K., Ke, Y. & Eskenazi, M. (2002). Enhancing foreign language tutors - In search of the golden speaker. In *Speech Communication*, 37, 161-173.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. In *Cognition*, 73(3), 265-292.

Seferoglu, G. (2005). Improving students' pronunciation through accent reduction software. In *British Journal of Educational Technology*, 36(2), 303-316.

Ulbrich, C., Mennen, I. (2015). When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. In *International Journal of Bilingualism*, 1-28.

DOI: 10.17469/O2102AISV000007

RICCARDO ORRICO, VIOLETTA CATALDO, RENATA SAVY, LINDA BARONE

Transfer, fossilization and prosodic drift in Foreign Language Learning

This article aims to shed light on the processes governing the development of prosodic competence in Italian learners of English as a Foreign Language (EFL). We have analysed the intonational contours of yes-no questions read by five groups of speakers (A, B, C, E, and P). Groups A, B, and C are learners of EFL at three different levels (beginner, intermediate, and advanced), who have never spent significant periods of time in English-speaking countries. Group E includes advanced learners who have English language experience in the UK. Group P includes Italian professors of EFL, who represent a sample of speakers with both acquisition experience and high proficiency in English. Our results highlight the total lack of improvement from Group A to C, allowing us to suggest the occurrence of prosodic transfer and drifts. On the contrary, speakers who have acquired the language in natural environments show improvements in their FL prosodic competence, but this is identified only in some of the prosodic cues analysed. The cues for which we did not observe improvements are thought to be more vulnerable to fossilization.

Key words: prosodic transfer, yes-no questions, English as a Foreign Language.

Introduction

Prosody is responsible for conveying information at three different levels. Firstly, it carries *linguistic* information, that is information about the pragmatics of the utterances, their information structure, the placement of stresses and accents, and so on; secondly, it conveys *paralinguistic* information, concerning the attitude and the emotional state of the speaker; and finally it communicates *extralinguistic* information about the speaker's identity, such as sex, age, and so on (Wagner, 2008).

Prosody plays a role in every spoken utterance, regardless of its length or the language in which it is produced. Since prosody is an inherent component of speech, the acquisition or learning of prosodic phenomena (such as intonation, stress, and rhythm) and how they are uniquely realized in a given language is essential in order to master a second language (L2) (Mennen, De Leeuw, 2014). The literature about second language acquisition (SLA) amply attests the difficulties that learners have to face when dealing with phonetics and phonology. These difficulties concern not only the reproduction and the perception of segments that differ from the learners' native language (L1), but also prosody. Indeed, several studies have documented that non-native prosody is highly

responsible for the perception of foreign accent as well as affecting negatively the intelligibility and comprehensibility of speech (Rasier, Hiligsmann, 2007; Derwing, Rossiter, 2003; Munro, Derwing, 1995). Moreover, some articles have also argued that prosodic deviations have a stronger effect on comprehensibility and accentedness than segmental errors (Anderson-Hsieh, Johnson & Koehler, 1992).

Nevertheless, researchers have so far focused mostly on the acquisition of segmental categories, while studies on the acquisition of prosody have been somewhat scarce (Mennen, 2004). Furthermore, the studies that have concentrated on prosody have failed to reach a good understanding of prosodic aspects of SLA, because most researchers have monitored the acquisition of intonational patterns in learners from different linguistic backgrounds, making it impossible to understand the processes governing the development of L2 prosodic skills (Mennen, Chen & Karlsson, 2010).

One common reflection emerging from these studies is, however, the occurrence of *prosodic transfer* in learners' interlanguages (IL), which means that learners' L1 has a strong influence on their productions in L2. Most studies demonstrating prosodic transfer in L2 have been carried out within the Autosegmental-Metrical (AM) approach. They attest that learners' L1 can influence interlanguages at different levels: phonological (that is the transfer of phonological categories of intonation) and phonetic (influence in how speakers realize a given category in terms of alignment, slope, shape of the contour, etc.). Additionally, transfer may involve the use of prosodic cues to express linguistic and paralinguistic meaning (Rasier et al., 2007; Mennen, 2015).

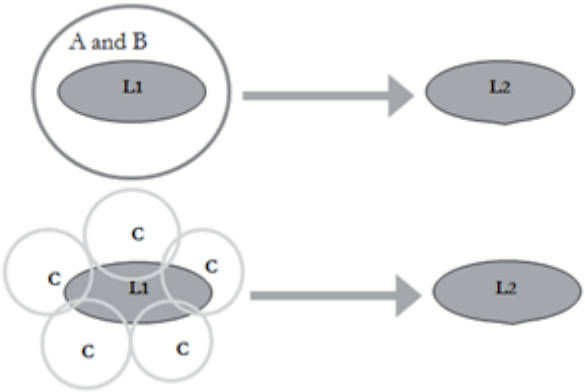
The concept of transfer has always been considered a major factor in the field of SLA, and not only with reference to phonology. Selinker (1972) extensively deals with *language transfer*, which implies that rules and systems at *all* linguistic levels are transferred from L1 to L2. The author includes this phenomenon among the processes that lead to linguistic errors in a learner's interlanguage and that keep the vast majority of learners from acquiring a native-like competence. Language transfer, along with other processes, is held responsible for the occurrence of fossilization. The concept of fossilization was first introduced in Selinker (1972), who argues that "fossilizable linguistic phenomena are linguistic items, rules and subsystems which speakers of a particular NL [native language] will tend to keep in their IL [...] no matter what the age of the learner or the amount of explanation and instruction he receives in the TL [target language]" (215). In other words, as pointed out by the same author some years later, "fossilization is the permanent cessation of IL learning before the learner has attained target language norm" (Selinker, Lamendella, 1978: 187). Since Selinker's theorization, various authors have dealt with fossilization, two of whom are particularly worth mentioning in the framework of the present study. The first is Long (2003), who invites researchers to exercise caution when talking about fossilization. The author states the difficulties related

to the testability of this phenomenon, since it is not easy to ascertain whether or not a learner is going to improve further. The second author is Han (2009; 2014). In her Selective Fossilization Hypothesis, Han states that fossilization is a phenomenon that does not affect all learners in the same way (inter-learner differential success or failure) and, most importantly, it does not affect all the linguistic structures of IL, i.e. only certain linguistic properties are vulnerable to fossilization (intra-learner differential success or failure). The vulnerability of a linguistic property to fossilization depends on the markedness of that property in the learner's L1 and on the robustness of the same property in the L2 input. The less marked the property is in L1 and the less robust it is in the input, the higher the likelihood that it will fossilize (Han, 2014).

Therefore, transfer and fossilization represent two phenomena of great importance in the process of language learning and acquisition. Linguistic interference between L1 and L2 affects every level of linguistic analysis, including prosody. While prosodic transfer has been extensively documented in the literature, to the best of our knowledge fossilization at the prosodic level has not been equally considered in L2 acquisition research. Even so, since fossilization is an inevitable stage of the learning process (Han, 2014: 52) and is closely linked to the phenomenon of transfer, we can argue that the subsequent stage of prosodic transfer is prosodic fossilization.

These assumptions constitute the starting point of the present contribution. The purpose of our study is to monitor the development of L2 prosodic competence in learners of English as a Foreign Language (EFL), whose linguistic competence has developed only in the classroom environment. As mentioned above, the idea underlying this project is that learners fail to reproduce the right intonational patterns in FL, transferring their L1 prosody into their interlanguages. The hypotheses we aim to test in this study stem from the results of previous research on FL prosody learning in Spanish by Italian students (Savy, Luque Moya, 2015). Results from Spanish show that none of the productions of the students heads towards the prosodic system of the TL. In particular, learners at earlier stages of language acquisition rely solely on their L1 prosodic system, while, productions spoken by more advanced learners shift from the L1 system, heading in different directions. The movement from L1 recorded in higher levels might denote an early sign of the development of a meta-prosodic awareness. However, the fact that they are not getting closer to the TL might be linked to the fact that they are not being given enough input or instructions necessary to reach a native-like prosody. We refer to this phenomenon as *prosodic drift*, in the sense that speakers move away from their L1, but in an aimless way and with a centrifugal movement (see Figure 1).

Figure 1 - *Expected pattern of FL prosody learning. A (beginner), B (intermediate), and C (advanced) refer to the levels of linguistic competence in FL. A and B gravitate around the L1 prosodic system; C randomly shifts towards different directions, but not towards L2*



1. *Methods*

We asked several Italian students from the University of Salerno to read two sets of questions, one set in their L1 and the other in English (the students' language of learning). The productions read by the students were then analysed with reference to both their mother tongue and productions of the same utterances read by English native speakers.

In this section of the paper we will discuss the dataset of productions analysed, the participants and the methodology used to carry out the analyses.

1.1 Dataset

We have analysed a small dataset of English and Italian questions taken from a larger corpus. The dataset consists of yes-no questions with the subject placed after the verbal phrase (henceforth VS yes-no questions), such as "Is there a millionaire?/C'è un bambino?". Each speaker repeated the same question several times, with different target words as the subject. The target words were chosen according to their stress patterns, the number of syllables and the structure of the syllables (only in the case of monosyllabic words). The choice of different target words was made in order to verify whether the intonational patterns changed along with the structure of the word.

In Italian, we chose 9 trisyllabic target words:

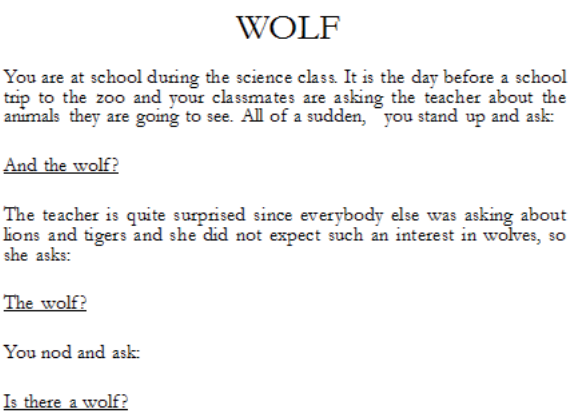
– 3 trisyllabic words with the stress on the antepenultimate syllable (*rondine, albero, dondolo*);
– 3 trisyllabic words with the stress on the penultimate syllable (*balena, bambino, budino*);
– 3 trisyllabic words with the stress on the last syllable (*pedalò, colibrì, lunedì*).

In English, we used trisyllabic and monosyllabic words, which were added since they are very frequent in English. For monosyllables, we chose words with different syllabic structures, one with multi-consonant coda (VCC) and one with multi-consonant onset (CCV):

- 1 monosyllable VCC (*wolf*);
- 1 monosyllable CCV (*crow*);
- 3 trisyllabic words with the stress on the antepenultimate syllable (*rectangle, triangle, pullover*);
- 3 trisyllabic words with the stress on the penultimate syllable (*December, September, fiancée*);
- 3 trisyllabic words with the stress on the last syllable (*millionaire, engineer, orangeade*).

To elicit the corpus, we made up situations in which the questions could be used in real life, as exemplified in the Figure 2 below. The speakers had to imagine themselves in those situations and read the questions.

Figure 2 - *Example of context. The speakers had to read the situation silently and then the underlined question out loud*



This method was used in order to increase the spontaneity of the productions and to keep the participants from misinterpreting the pragmatic structure of the questions.

Before the recording session, the students were asked to read carefully all the questions and the contexts in which they had been inserted, and were encouraged to ask clarifications about words or contexts. Then, they would start the recording with a dummy test, to make sure that the exercise had been understood as well as to test that the recording equipment was working properly. After that, the real recording would start.

The recording sessions took place at the Applied Linguistics Laboratory (LabL.A.) at the University of Salerno.

1.2 Participants

We chose five groups of Italian participants. All the participants were female.

The criteria used for the selection of the participants were the level of FL competence and their FL learning experience:

- Group A: beginner EFL students;
- Group B: intermediate EFL students;
- Group C: advanced EFL students;
- Group E: advanced EFL students with at least six months experience of the Erasmus Programme in English-speaking countries;
- Group P: Italian professors of EFL.

The participants in the first three groups are students at different FL levels and whose linguistic competence has developed by means of a *learning* process; the other two groups include participants who have had the opportunity to *acquire* the language in English-speaking countries, in addition to learning it at school and university.

In particular, the speakers in groups A, B, and C are all EFL students at the University of Salerno. All of them have lived in Salerno since childhood (as have their parents), so their L1 was Salerno Italian. The assessment of their FL level of competence in was based on their results in EFL exams. We selected 10 students from Group A, 7 from Group B, and 6 from Group C.

The students in Group E have the same requisites as the students in Group C, with the difference that they studied for at least one semester in an English-speaking country. We were able to select 7 Erasmus students. We chose these students as a Control Group, which allowed us to take into account the differences between FL classroom learning and acquisition in natural environments.

The 3 participants in Group P are professors of EFL at the University of Salerno. The participants of this Control Group ideally represent speakers with the highest EFL competence (learning experience abroad plus high levels of grammatical and communicative competences).

Group E and Group P were selected for a twofold purpose. On the one hand, they allow us to make considerations about the effects of the acquisition in natural contexts following the learning process. On the other hand, they add information about the possible development of a learner's prosodic competence from the earlier stages to the highest. Moreover, Group P was essential to verify the possibility of fossilization in higher stages.

The productions of these participants have been analysed with reference to the productions of native speakers of English (English L1). This was the group of reference used to verify the potential development of prosodic skills. English L1 speakers do not represent models of the entire language variety they speak, but just the model of

reference of the students chosen for the purpose of the research, since they were the teachers of those students. We had 3 speakers in this group.

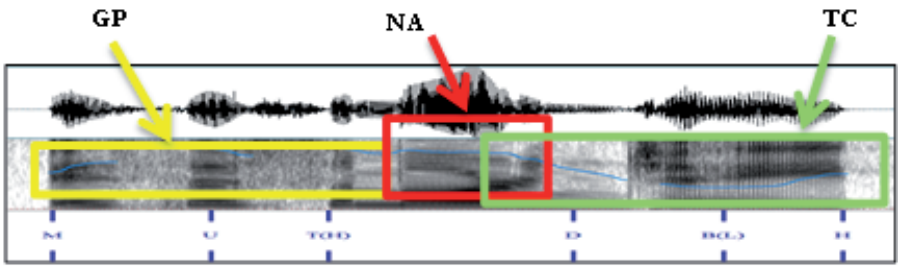
1.3 Methodology

The data were analysed using the INTSINT labelling system[1] (Hirst, Di Cristo, 1998) and employing a phonetic approach based on the description of the fundamental frequency (f_0) of the utterances.

Adopting a phonetic point of view when approaching the analysis of L2 prosody is a good option for the sake of data comparability, as well as for labelling and analysing interlanguages, whose variable nature implies the need to approach the study starting from the realization level. Since our goal was to have homogeneously labelled data in two different languages (English and Italian) and in the students' ILs, the INTSINT international labelling system was used. After the phonetic description of the curve of the productions, we proceeded to a more theoretical description of the utterances, by relating the phonetic realizations described to conceptual categories of the f_0 curve. Since these theoretical categories refer to a more abstract level of representation, they allowed us to make generalisations in the description of the observed data.
The categories considered are (Figure 3):

- *Global Profile* (GP), which is a description of the f_0 curve trend from its beginning to the nuclear accent. Several GP realizations are possible: rising, falling, flat, or combinations of such trends (for example rising-falling or falling-rising);
- *Nuclear Accent* (NA), which is the most prominent accent of the utterance, described according to whether it is a peak (H) or a valley (L), or whether it follows a rising (LH) or falling trend (HL). Again, the description of the accent is based on a phonetic approach, which provides information about its trend, without the description of its alignment to the syllable;
- *Terminal Contour* (TC), which is the terminal portion of the curve, from NA to its end. Three possible realizations of TC may occur, which are rising (H), falling (L), or flat (S).

Figure 3 - *Subdivision of the F_0 curve into the three parameters of GP, NA, and TC*



[1] The software used for pitch track analyses was Praat, using the script Prosomarker (Origlia, Alfano, 2012).

The study started by building the L1 models, both English and Italian, following the procedure described above.

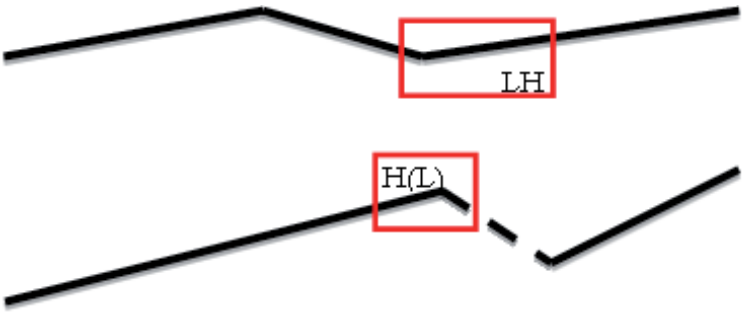
The VS questions are syntactically organised with the presence of a postponed subject, placed at the end of the utterance, after the verbal phrase, and with the expletive (*ci/there*) placed in the position of the subject. This particular question structure might have different prosodic realizations, both cross-linguistically and intra-linguistically, on the basis of its information and pragmatic interpretation. In particular, this utterance might allow at least two different interpretations based on the *presupposition* that speakers assume they share with the hearer[2].

In Italian, the question *C'è x?* ('Is there *x*?'), where *x* can represent any noun, can have a twofold interpretation: on the one hand, it can be uttered with the presupposition of the existence of the item *x*, which the speaker takes for granted is also assumed by the hearer. On the other hand, the utterance might be spoken without the assumption of the existence of *x* (zero-presupposition). On the contrary, the question *Is there x?* in English is appropriate only in the zero-presupposition case, when the presence of *x* is the purpose of the question, not the presupposition[3]. We analysed only those utterances that could be found in both languages (zero-presupposition).

Figure 4 shows the two stylised models of the utterances analysed. The L1 model of reference can be described as follows:

- P: Rising
- NA: H(L)
- TC: H

Figure 4 - *Stylization of L1s productions, Italian (top) and English (bottom)*



[2] Stalnaker (1978: 321) gives a definition of presupposition, stating, "a proposition is presupposed if the speaker is disposed to act as if he assumes or believes that the proposition is true, and as if he assumes or believes that his audience assumes or believes that it is true as well." From this definition, we can assume that speakers rely on the belief that there exists some common knowledge among those involved in the conversation, which does not need to be asserted and which can affect the information and the pragmatic status of the speech acts.

[3] As pointed out by Ward and Birner (2001), English existential *there*-sentences (*i.e. there*-sentences containing *be* as the main verb) are the result of the postposition of the logical subject, which has to be both hearer-new and discourse-new in order to be felicitous.

The label in brackets refers to the fact that in some cases the H tone of English NAs may be followed by a low tone. As a rule, the f_0 curve falls after the high accent and then rises again before the end; in cases where the target word is either monosyllabic or oxytone (that is when the NA occurs on the last syllable of the utterance), this fall begins within the accentual portion of the curve. This happens because of the lack of linguistic material between NA and TC, which causes the intonational phenomena to contract into a far shorter time interval, as shown in Figure 5a and 5b.

Italian L1 model is:

- P: slightly rising-falling
- NA: (L)H
- TC: H

Figure 5a - *Example of HL accent in an oxytone (Is there an engineer?)*

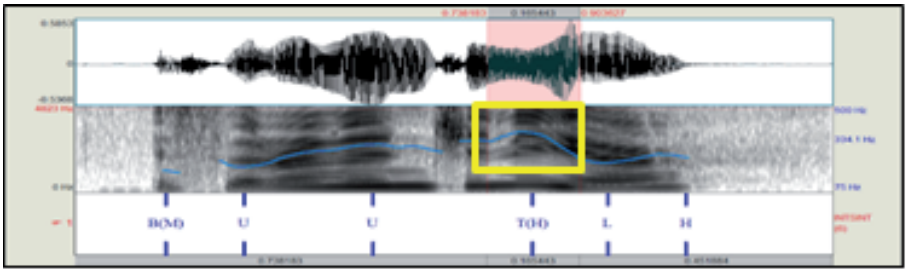
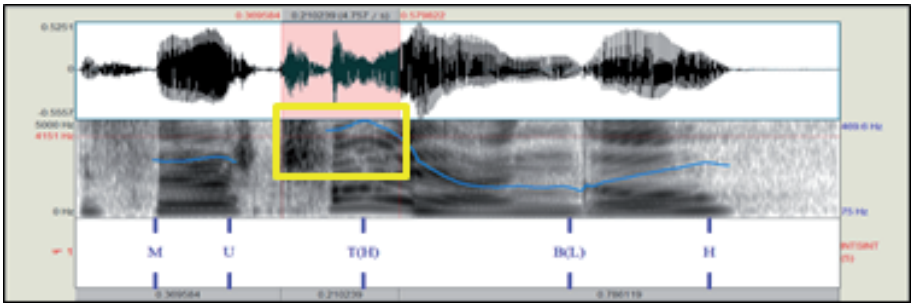


Figure 5b - *Example of H accent in a proparoxytone (Is there a triangle?)*



Besides the realization of the three parameters of P, NA, and TC, we noticed that the frequency range of the f_0 curves was systematically different from one language to another. In particular, utterances in English had an average range twice as wide as the utterances spoken in Italian. For each sentence, we calculated both the range of the whole phrase and the range of the TC; the mean values divided for languages are reported below:

English:

- Range: 226,9 Hz
- Range TC: 110,9 Hz

Italian:
– Range: 111,1 Hz
– Range TC: 58, 4 Hz

After the description of the models, we analysed the productions of the students' interlanguages, comparing them with both models, focusing on the aspects that appeared to differ systematically from one language to another, as reported in the following section.

2. *Results*

We have mainly focused on language-specific aspects, in order to verify whether or not prosodic transfer occurred in the students' IL. This contrastive comparison was always made with reference to the L1 learning models (both English and Italian) and to the control groups P and E.

The aspects taken into account are:
1. The information structure: presupposition vs zero-presupposition;
2. The trend of GP;
3. The range of frequencies of the utterances;
4. The realization of the accentual target.

In this paper, we will not present results regarding TC. In most of the cases, the terminal contour had the same trend (*i.e.* rising) in the two L1s, but with a much higher excursion in English. For now, this aspect will be accounted for in the section dealing with range. A similar, yet much more complicated situation is the realization of NA. The question of the nuclear accent has yet to be sufficiently analysed and the results shown in §2.4 are, therefore, not final.

2.1 Presupposition vs zero-presupposition

This aspect is not a purely prosodic one, but it greatly involves the prosodic realization of the utterances. As explained above, in Italian two prosodic realizations are allowed: questions with zero-presupposition structure, which have the NA on the subject and questions with the presupposition of the subject, which are realized with the NA on the verbal phrase. English, on the other hand, only allows questions with zero-presupposition and, therefore, only with the occurrence of NA on the subject (which is also the rightmost element). Whenever students realized the questions in English with NA on the expletive, it has been considered as prosodic transfer. The results for this aspect are reported in Table 1.

Table 2 - *Choice between the two possible informative interpretations*

	NA on the Noun phrase	*NA on the Verbal Phrase*
Group A	50.9%	49.1%
Group B	71.4%	28.6%
Group C	59.1%	40.9%
Group E	67.5%	32.5%
Group P	84.8%	15.2%

Results show that the prosodic cues related to this aspect are learned over time and even in FL classes. Indeed, these results point to a consistency between the level of FL competence and the correct use of these cues. Only the values registered in Group B do not follow this pattern. However, the high percentages of this group stem from the productions of only two speakers, without which we would have an average value of 51.5% of NAs placed on the subject. The positive results regarding this aspect might depend on the fact that the placement of the nuclear prominence is not a matter of prosody alone, but is closely linked to the informative-pragmatic interpretation of the utterance, of which prosody is only the phonetic interface.

Results presented below refer only to the utterances with NA on the subject.

2.2 Frequency range

The range of frequencies used to speak a language differs cross-linguistically and a narrower pitch range is a common mistake made by English language learners with different language backgrounds (Mennen, 2007).

Table 3 reports the results registered in the L1 models and in the interlanguages spoken by non-native-speakers of English.

Table 3 - *Values of the F_0 range registered in the different groups*

	Range	*Standard deviation*
English L1	226,9 Hz	60,3
Italian L1	111,1 Hz	36,2
Group A	155,2 Hz	56,7
Group B	136,1 Hz	47,5
Group C	134,4 Hz	54,4
Group E	234,4 Hz	82,0
Group P	251,7 Hz	80,5

We observed a striking contrast between native speakers of Italian and English in the use of frequency range, showing that the two languages are different in this perspective.

Groups A, B, and C show that the range of frequencies they use in their interlanguages is only slightly wider than the range they use in their L1 and far narrower than the results registered for their FL models. This result, which roughly equates all the students regardless of the level of competence, proves the occurrence of prosodic transfer from their L1 to L2. Therefore, we can argue that, although they might perceive higher peaks of frequencies (and lower valleys), they fail to reproduce them when speaking English, using the range of frequency they would use in Italian.

On the contrary, results from the control groups (P and E) give proofs of the fact that this prosodic feature can be acquired over time, and probably thanks to the immersion of the learner in a context where the language is spoken as L1. Groups P and E, indeed, implement ranges of frequencies that go even beyond the level of native-speakers (though the standard deviation shows that the phenomenon is not consistent among the speakers and their utterances).

This might allow us to conclude that the range of frequencies used in a language is a highly salient prosodic feature, noticed by learners when faced with a great deal of input language.

2.3 Realization of GP

Another contrastively significant aspect is the *Global Profile*, which, as explained above, gathers all the intonational phenomena of the curve before the nuclear accent, and is essential to describe the global trend of the curve in the whole utterance.

Table 4 - *Realization of GP*

	Rising	*Rising-falling*	*Flat*	*Falling*	*Others*
English L1	42.9%	28.6%	28.6%	–	–
Group A	28.6%	–	64.3%	7.1%	–
Group B	32.7%	9.1%	56.4%	1.8%	–
Group C	29.3%	14.6%	41.5%	14.6%	–
Group E	14.8%	13.0%	55.6%	7.4%	9.0%
Group P	32.1%	–	50.0%	17.9%	–

We encountered four main patterns in the realization of GP: rising, rising-falling, flat, and falling.

English L1 speakers realize their utterances with a rising GP in most of their productions (42.9%), while non-native speakers prefer a relatively flat pattern, regardless of their level of proficiency, or whether or not they have acquired English in natural contexts. These results might appear surprising, since we reported that control groups E and P used a wider range of f_0 than native speakers did. In any case, what we can deduce from this is that the realization of GP is a less salient feature of the curve, perhaps the salient one, and even when a proper range of frequencies is used, it is restricted to the final portion of the curve (NA and TC).

2.4 Nuclear accent

As mentioned above, the results registered from the realization of NA are not definitive and still require an in-depth examination.

The results found so far point to a complicated situation, which is nevertheless interesting from a researcher's point of view. Table 5 reports the frequency of the use of an H-targeted NA or an L-targeted NA in the different groups.

The H and L labels refer only to the accentual target; in other words, they refer to the main NA tone, regardless of the alignment to the syllable.

In contrast with the analyses of the L1s (both English and Italian), which showed more consistent realizations of NA, interlanguages had a much higher variation both among and within the groups. This additional and more general classification of NA, therefore, was needed in order to add greater homogeneity to the results and to facilitate their comparison.

Table 5 - *Realization of the target of the NA*

	H	L
English L1	83.3%	16.7%
Group A	39.2%	60.8%
Group B	47.3%	52.7%
Group C	48.8%	51.2%
Group E	74.1%	25.9%
Group P	67.9%	32.1%

As shown in the table above, in over 83% of the cases, the target of the accents in English L1 productions is H, while among non-native speakers both realizations can be found. Students A, B, and C realize a low-targeted NA, regardless of the level of English language proficiency, while Groups P and E, in most cases, have a high-targeted NA.

Accents in the Salerno Italian variety are phonetically realized as a rise from a low tone that aims to an H target; in English L1 NA aims to a high target. Thus, we did not expect L-targeted NA to occur frequently.

In most cases, NA in students' interlanguages, regardless of the level of FL competence, differs from both L1 and L2. In addition, no consistency in its realization has been found within the groups or within the different productions of one speaker (which is the reason why we opted for a more generalised description of NA). This phenomenon has been accounted for as *prosodic drift*: learners want to *leave* their L1, but they are not sure about what they should do because of the paucity of the input and the lack of direct instructions; thus, they drift away from the L2 norm in the attempt to reproduce something they believe is different from their L1.

As for the realization of NAs registered in the control groups, it is fairly close to the way native-speakers realize it. Nevertheless, it is not yet possible to state that they are realizing an "English" nuclear accent. A deeper analysis would be neces-

sary in order to state if this outcome represents a real movement towards the TL. However, by considering the study as pseudo-longitudinal, in which lower levels give a hint of how the higher ones used to speak and, symmetrically, higher levels represent the way lower ones will most probably speak, we can suppose that they realize the accent in a way that is closer to the native-speakers.

3. *Discussion*

Results show a twofold classification of the speakers. On the one hand we have students whose linguistic input come from their FL teachers alone (A, B, and C) and whose productions reveal either the occurrence of prosodic transfer in the realization of GP, use of a range of frequencies and prominence placement, or an opposite prosodic drift, in the case of NA. Either way, this confirms that students rely enormously on their L1. We expected this to happen, but what was not expected is the fact that it occurs regardless of the time of their exposure to the classroom input. The lack of improvement from one level to the next might be linked to the fact that current FL teaching methods do not impart the necessary prosodic knowledge, or even some sort of guidelines, to their students, and not even at higher levels.

On the other hand, there are Groups P and E, who either shift from their L1 and move towards L2, as in the case of the range of frequencies and prominence placement, or they transfer prosodic cues from L1 to IL, as in the case of the realization of GP or, perhaps, NA.

From this framework, we can assume that the different parameters analysed might not carry the same weight in terms of how they are perceived, whether they are noticed, and how they are employed when speaking in FL. For instance, GP is the one aspect that we can assume most people do not notice, since it is not implemented correctly by a large percentage of the informants. The range of frequencies, on the contrary, is mostly noticed and some speakers increase their own, in order to imitate native-speakers. However, this occurs only when they are faced with considerably more input than that restricted to the five-year English class at university. Moreover, the speakers do not seem fully aware of the use of this cue, since they wider the range of frequencies only in a portion of the curve. As for the accent, this is realized either incorrectly or ambiguously, which leads us to think that speakers perceive it as being different from their L1's, but they are not given enough instructions or input to move towards the target language norm.

The improvements made by students over the five-year classes do not regard prosody: the results are basically the same from level A to level C, meaning that a teaching programme for prosody needs to be developed. Secondly, some improvements are made when the learners spend time in the country where the language they are learning is spoken as L1. However, the lack of direct instructions makes them rely on what they perceive as important, leading to an erroneous use of prosodic cues in FL. In addition, these initial errors concerning some of the prosodic aspects are not improved over the years. This, might be considered a proof of prosodic fossilization.

Nevertheless, several studies on fossilization (Selinker, Mascia, 1999; Long, 2003) have argued that longitudinal studies are required in order to demonstrate the occurrence of fossilization. Since our study is only pseudo-longitudinal, we cannot state that speakers at the highest levels, namely Group P, are fossilized. However, it is safe to claim that those cues which are erroneously implemented by all the speakers regardless of the level of FL competence (GP and, in an ambiguous way, NA) are items which might be considered vulnerable to fossilization. On the contrary, aspects that improve over the years, such as the range of frequencies and the placement of the NA are items that are less likely to fossilize.

4. *Conclusion*

This study has aimed to monitor the process of prosody learning in EFL students. Results have shown that learners who have never had the opportunity to acquire the language in natural environments do not learn FL prosodic cues, even if they improve their general FL competence. Signs of improvement are registered only after they spend at least a semester in an English-speaking country.

This outcome clearly shows the link between the lack of improvement in prosodic competence and the FL teaching methods, which notoriously do not include instructions on phonetics or, least of all, prosody (Dalton, 1997; Munro, Derwing, 1999; Isaacs, 2009). Derwing, Rossiter (2003) demonstrate that general prosodic instructions positively affect the productions of non-native speakers, in terms of comprehensibility, accentedness, and fluency, which suggests that teaching prosody is useful and necessary.

Therefore, this article wants to stress the need for the development of such methods, which might represent the only way that FL learners have to improve their prosodic proficiency. This, however, is no simple task. In order to include prosody in the current teaching programmes, researchers need to gain a better understanding not only of the processes governing the acquisition/learning of prosodic structures but also of the prosodic cues which are primarily responsible for undermining the success of communication.

Bibliography

Dalton, D.F. (1997). Some techniques for teaching pronunciation. In *Internet TESL Journal*, 3. http://iteslj.org/Techniques/Dalton-Pronunciation.html/Accessed 31.05.16.

Derwing, T.M., Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. In *Applied Language Learning* 13/1, 1-17.

Han, Z.H. (2009). Interlanguage and fossilization: Towards an analytic model. In Cook, V., Wei, L. (Eds.), *Contemporary applied linguistics, Language teaching and learning*. London: Continuum, 1, 137-162.

Han, Z.H. (2014). From Julie to Wes to Alberto: Revisiting the construct of fossilization. In Han, Z.H., Tarone, E (Eds.), *Interlanguage: Forty Years Later*. Amsterdam: John Benjamins, 47-74.

Hirst, D.J., Di Cristo, A. (Eds.) (1998). *Intonation Systems. A survey of Twenty Languages*. Cambridge: Cambridge University Press.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. In *TESL Canada Journal*, 27, 1-12.

Long, M. (2003). Stabilization and fossilization in second language interlanguage development. In Doughty, C., Long, M., *The Handbook of Second Language Acquisition*. Oxford: Blackwell, 487-529.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. In *Journal of Phonetics*, 32, 543-563.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In Trouvain, J., Gut, U. (Eds.), *Non-native Prosody: Phonetic Descriptions and Teaching Practice*. Berlin: Mouton De Gruyter, 53-76.

Mennen, I. (2015). Beyond segments: towards a L2 intonation learning theory (LILt). In Delais-Roussarie, E., Avanzi, M. & Herment, S. (Eds.), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*. Berlin: Springer Verlag, 171-188.

Mennen, I., De Leeuw, E. (2014). Beyond segments. In *Studies in Second Language Acquisition*, 36(02), 183-194.

Munro, M.J., Derwing, T.M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. In *Language Learning*, 49, 285-310.

Origlia, A., Alfano, I. (2012). Prosomarker: a prosodic analysis tool based on optimal pitch stylization and automatic syllabification. In *Proc. of LREC*-2012, 997-1002.

Rasier, L., Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and Methodological Issues. In *Nouveaux cahiers de linguistique Française*, 28, 41-66.

Savy, R., Luque Moya, J.A. (2015). Aspectos prosódicos de las interrogativas en aprendientes italianos de ELE. In *NORMAS*, 7, 283-296.

Selinker, L. (1972). Interlanguage. In *IRAL*, 10(2), 209-231.

Selinker, L., Lamendella, J. (1978). Two perspectives on fossilization in interlanguage learning. In *Interlanguage Studies Bulletin*, 143-191.

Stalnaker, R.C. (1978). Assertion. In Cole, P. (Ed.), *Syntax and semantics, Pragmatics*. New York: Academic Press, 9, 315-332.

Wagner, A. (2008). A comprehensive model of intonation for application in speech synthesis. Unpublished PhD Thesis, Adam Mickiewicz University, Poznan, Poland.

Ward, G., Birner, B. (2003). Discourse and Information Structure. In *The handbook of discourse analysis*, 119, 119-137.

DOI: 10.17469/O2102AISV000008

DALIA GAMAL ABOU-EL-ENIN

L'intonazione sospensiva in arabo L1 e italiano L2. Analisi della prosodia e della portata interazionale in conversazioni semispontanee

This study presents an acoustic analysis of the non-final sentences. Analysis is conducted on conversations elicited by the 'Show-The-Difference-technique'. It consists of about 260 tone units in Italian L1 of Roman speakers, in Cairene Arabic and in Italian L2, produced by the same Egyptian informants. Analysis of a corpus in Italian L1 allows comparisons within a homogeneous material. ToBI transcription of the final contour has been conducted and then investigated phonetically by calculating both the slope on the last three syllables and the alignment of H target on the last accented syllable. The final rise is found to be the major clue of non-finality, while low target as a boundary tone appears more in L2. Differences between Arabic L1 and Italian L2 in slope and vowel duration show that even under the positive transfer hypothesis, speakers do not copy their mother tongue in the L2. However, Italian L1 and L2 show discrepancies in pitch accent types and in slope as well.

Key words: L2, Arabic, Italian, non-finality, communication.

Introduzione

L'enunciato sospensivo veicola un significato non compiuto, per cui viene denominato anche 'non finale'o 'non conclusivo'. La sospensione avverte l'ascoltatore del proseguimento di un discorso importante e nelle parole di Halliday (1992) rappresenta "una unità d'informazione che dipende da un'altra" (105).

Nella presente ricerca si intendono esaminare le sospensive che si riscontrano nel parlato, a prescindere dalla modalità: assertiva, interrogativa, iussiva, ecc.

VBit2_B098: e l'orecchio è abbastanza *grande*
Vai un poco a *sinistra*
VBit2_B150: dietro il piede destro del *ragazzino* c'è una piccola macchia?

L'enunciato sospensivo può essere rappresentato da segmenti non completi a livello della struttura:

VBit2_A077: non sto *parlando* | del piede sinistro;

o in frasi dal senso compiuto che però risultano parti di un costrutto più grande. In quest'ultimo caso solo la codifica prosodica sembra rendere la 'sospensione', chiedendo all'ascoltatore di aspettare un completamento o un ampliamento dell'idea. Tale sospensione si riscontra su:

1. voci non finali nelle enumerazioni

   VA1_A021: otto <eeh> c'abbiamo, allora uno due tre quattro *cinque* | poi *due* poi uno <mh>

2. componenti all'interno della proposizione:

   VA1_A055: [...] c'è uno *specchietto* | di fronte
   VA1_B014: c'è la statua di un *soldato* | con la *spada*, | sul *cavallo* |

3. una frase seguita da una coordinata:

   VA1_A137: la statua sta dentro un *cerchio* | <pb> e si circondano le piante

Negli ultimi due esempi la fine di ogni sintagma preposizionale costituisce un completamento della struttura grammaticale e potrebbe costituire, dunque, a livello conversazionale, un 'punto di rilevanza transazionale' (TRP: *Transition-Relevance Place*), che l'ascoltatore ha il diritto di decodificare come un punto possibile per il cedimento del turno (Sacks, Schegloff & Jefferson, 1974), ma solo con la codifica prosodica 'sospensiva' si può chiedere all'ascoltatore di aspettare un ulteriore completamento del discorso.

Con il presente contributo s'intendono esplorare le caratteristiche delle stringhe di discorso sospensive in arabo L1 e in italiano L2 dello stesso gruppo di parlanti egiziane. Viene inoltre analizzato un corpus di controllo in italiano L1 in modo da poter condurre un confronto di dati omogenei tra le diverse produzioni. Si mira ad approfondire tramite l'analisi acustica la resa fonetica delle sospensive in relazione alla loro funzione comunicativa e a raccogliere più dati, utili per l'esplorazione dell'interlingua degli egiziani e per la discussione dell'ipotesi del *transfer*.

1. *Studi sulle sospensive*

Malgrado le sospensive siano state sempre oggetto di osservazione per la loro 'appariscenza' uditiva, non c'è stato quasi mai uno studio ad esse esclusivamente dedicato. Vengono normalmente considerate in confronto alle conclusive, con cui presentano un contrasto uditivo, mentre alle modalità interrogativa, assertiva e direttiva si rivolge maggiore attenzione.

1.1 Sospensive in italiano L1

In italiano vari lavori descrivono il contorno melodico sospensivo come terminante in salita, in tenuta o in una lieve discesa. Lepschy (1978) propone il tono 3, piatto, come andamento che veicola l'incompletezza e/o l'esitazione e lo considera tipico delle enumerazioni e delle sospensive. Propone anche il tono 4, discendente-ascendente, per le sospensive enfatiche. Canepari (1985; 1986) individua la tonia sospensiva come una delle tre tonie fondamentali con la conclusiva e l'interrogativa; la descrive come neutra con tonalità media, tonica alta e postonica discendente da alta a media.

Chapallaz (1979) ritiene che il contorno discendente-ascendente (che denomina *Tune II*) sia tipico delle sospensive e delle enumerazioni e di alcuni tipi di domande; mentre il contorno ascendente-discendente (*Tune III*) è più comune nelle narrazioni e nelle enumerazioni.

Endo e Bertinetto (1997) conducono una analisi acustica sul profilo melodico delle dichiarative non-finali in nove località italiane del Nord, del Centro e del Sud, sottolineando che ad esse non sono stati dedicati lavori sperimentali. Gli autori rilevano due andamenti tipici, ma completamente diversi. Il primo, rilevato a Milano, Padova e Bologna. Viene denominato modulo 'a valle' e presenta una forte discesa sull'ultima tonica, seguita dopo la metà della tonica da un andamento finale ascendente. L'altro andamento, chiamato 'a monte', presenta una tonica finale alta seguita da una discesa notevole sulle postoniche, e si rileva a Macerata, Pisa, Roma e Cosenza.

Sorianello (1997) osserva che le sospensive nel parlato letto dei suoi soggetti cosentini presentano un movimento discendente sulla tonica seguito da una salita rappresentabile in termini autosegmentali con la stringa L* (o H+L*) H-H%; per contro, lo spontaneo manifesta una variazione di andamenti, con un accento nucleare L+H* e tono di confine a volte ascendente e a volte discendente: ciò solleva quindi la questione della differenza tra gli stili e dimostra ancora una volta la sensibilità della prosodia a tutte le varianti comunicative.

Studi sul fiorentino, il pisano e sul barese confermano la realizzazione di una salita finale a variazione dell'accento nucleare (Avesani, 1995; Gili Fivela, 2002; Grice, Savino, 1995; Savino, Refice, 1997).

Crocco (2003) dedica uno studio al profilo sospensivo in un *corpus* di parlato letto e semispontaneo (*Map Task*) di informatori pisani e napoletani, dove osserva due profili intonativi: uno progressivamente ascendente e un altro discendente con contorno finale ascendente.

1.2 Sospensive in arabo L1

Per l'arabo Kulk, Odé & Woidich (2003) svolgono un confronto tra l'arabo cairota e l'arabo damasceno e segnalano in ambedue le varietà, in produzioni spontanee, due profili terminali per gli enunciati sospensivi, quali la tenuta e la salita. Gli autori osservano infatti che il *pitch* non presenta differenze nei due dialetti, ma che nel damasceno, a differenza del cairota, le vocali finali di TU si triplicano rispetto alla loro durata media.

Altri studi sul parlato letto e semispontaneo e sul parlato radiofonico (Hellmuth, 2006; Rifaat, 2005a; 2005b) osservano, come andamento prevalente, una salita graduale a partire dalla tonica finale.

In un lavoro precedente (Gamal, 2005), al margine di uno studio sulla modalità pragmatica direttiva, l'ascesa finale è stata rilevata come il segnale più frequente di sospensione in arabo L1. Si è osservato anche che tale resa melodica rappresenta una scelta che arricchisce l'aspetto pragmatico-comunicativo degli atti linguistici diret-

tivi, in quanto garantisce l'esecuzione delle richieste di azione tramite il continuo richiamo d'attenzione e il mantenimento del turno di parola.

2. *Materiale e metodo*

Il *corpus* si compone di 104 unità tonali in italiano L2, 53 unità in italiano L1 e 100 unità in arabo L1, estratte da 8 conversazioni semispontanee elicitate con il metodo del 'gioco delle differenze' (Cerrato, 2007; Cutugno, 2007). Si è fatto uso delle vignette ideate nel quadro del progetto CLIPS, rilevabili all'URL: www.clips.unina.it. Le 53 TU sospensive in italiano L1 sono prese dalle conversazioni A01R e B04R registrate, nell'ambito del progetto CLIPS, con parlanti di Roma, 3 parlanti femmine e un maschio. Ho scelto il parlato di Roma perché la capitale è stata la destinazione delle informatrici egiziane che hanno soggiornato a Roma. Le conversazioni in italiano L2 e in arabo cairota sono state registrate al Cairo con apprendenti guidate di livello medio e avanzato (studentesse nel corso post-laurea e assistenti universitarie). In quattro conversazioni sono state registrate le stesse parlanti egiziane, utilizzando vignette diverse per l'arabo e l'italiano L2 (tabella 1).

Tabella 1 - *Si è cercato il più possibile di registrare le stesse informanti in arabo L1 e italiano L2*

Varietà	*Arabo L1*		*Italiano L2*		
Conversazione	VAar	VAar2	VA1	VB1	VBit2
Informanti	P1+P3	P4+P5	P1+P2	P1+P2	P4+P5

Gli informanti egiziani sono tutte di sesso femminile, del Cairo, per cui la varietà di lingua araba da esse parlata è la medesima. Sono apprendenti guidate che hanno imparato l'italiano in condizioni simili, poiché hanno iniziato l'apprendimento all'età di 17 anni, nella stessa facoltà. Le differenze tra loro, vista la differenza di età, sta soprattutto negli anni di studio della lingua e nel soggiorno in Italia, come si vede nella tabella 2.

Tabella 2 - *Età e sfondo linguistico delle informanti egiziane*

Informante	*Età*	*Lingue straniere*	*Soggiorno in Italia*	*Soggiorno in altri paesi*
P1	28	Inglese e francese	Roma: un anno	–
P2	28	Inglese e francese	Roma: un anno	–
P3	36	Inglese, francese e tedesco	Roma: 10 mesi	–
P4	23	Inglese e francese	–	–
P5	24	Inglese e francese	–	–

Dunque, le apprendenti possono essere classificate in due gruppi omogenei, in base alla variabile 'livello della lingua'.

Le unità tonali dalla melodia sospensiva sono state individuate uditivamente. Sono state escluse quelle che finiscono in pause piene, cioè di esitazione.

Osservando la struttura ritmica dell'arabo si è rilevato un gran numero di parole finali tronche. Perciò le TU in ciascun sottocorpus (italiano L1, arabo L1 e italiano L2) sono state classificate a seconda della posizione dell'accento sull'ultima parola lessicale. La tabella 3 ci presenta in percentuale la collocazione degli accenti lessicali nelle parole finali di TU.

Tabella 3 - *Occorrenza in percentuale dei tipi ritmici rilevati a fine TU*

Varietà linguistica	*Finali tronche*	*Piane*	*sdrucciole*
Arabo L1	44	55	1
Italiano L1	3	91	6
Italiano L2	9	86	5

Con l'ausilio del software *Praat 5.3.0.3.* sono stati misurati le durate delle ultime tre sillabe in ogni TU e i valori di f_0 ai confini dei movimenti. In generale viene considerato il movimento sul nucleo vocalico. Sono state evitate le porzioni sorde e un margine di 10-20 ms all'*offset* e all'*onset* delle vocali contigue. Il rilevamento dei valori frequenziali in corrispondenza delle consonanti sonore non è stato una prassi costante: veniva evitato nei casi di abbassamento notevole seguito da una ripresa dei valori sulla vocale seguente.

In arabo sono frequenti le sillabe pesanti, per cui sono frequenti i nessi consonantici in mezzo al contorno terminale. Nella tabella 4 esponiamo la tipologia delle ultime due sillabe di TU nel *corpus* di arabo L1.

Tabella 4 - *Frequenza dei vari tipi sillabici rilevati a fine TU nel corpus di arabo L1*

Tipi sillabici	*%*
cv cv	20%
cv cvc	38%
vc cv	3%
vc cvc	3%
cvc cv	17%
cvc cvc	11%
cvc cvcc	7%

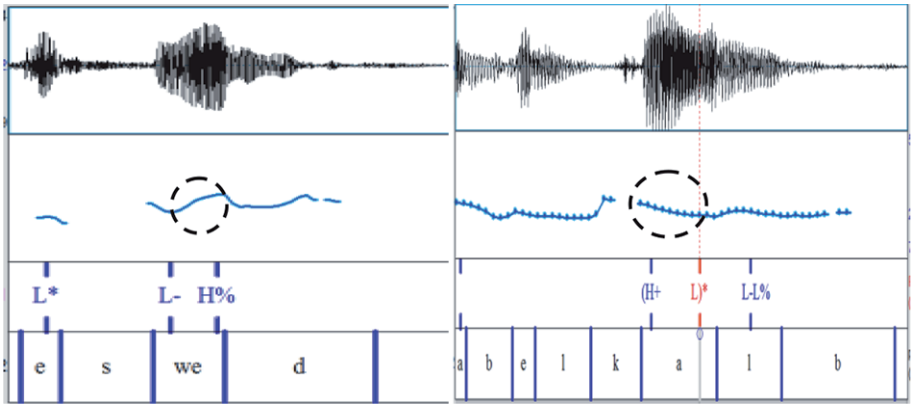
Le consonanti contigue, coda della penultima e testa dell'ultima, possono variare in funzione di sordità (R) e sonorità (N). La tabella 5 espone i dati percentuali.

Tabella 5 - *Le porzioni consonantiche vedono il susseguirsi di due foni sordi, R+R, due sonori, N+N, oppure un misto di sordo e sonoro, N+R e R+N*

sordo+sonoro	*%*
N+N	46%
N+R	18%
R+R	18%
R+N	18%

Sono stati osservati e segnalati i movimenti che presentano uno scarto dei valori frequenziali superiore ai 20Hz. Oltre all'andamento piatto, che non profila variazioni di *pitch* oltre i 7-10Hz, i movimenti sono classificabili, dunque, in due tipi principali: il movimento in salita (ascendente) e il movimento in discesa (discendente). Il movimento in salita presenta una crescita dei valori frequenziali sulla linea del tempo, la quale può interessare una sola sillaba o più di una sillaba. Il movimento in discesa, per contro, mostra un abbassamento dei valori di f_0 all'interno della sillaba o su più sillabe. Come inizio del movimento è stato considerato il massimo frequenziale nel caso del movimento discendente; il minimo, se si trattava di un movimento ascendente. Inoltre, non è stato considerato l'abbassamento relativamente lieve che si presenta a volte a fine fonazione e nemmeno sono stati considerati i valori alti coincidenti con la chiusura glottidale. Nella figura 1 diamo un esempio dei due tipi di movimento.

Figura 1 – *Esempio di un andamento ascendente (di salita) a sinistra sul nucleo finale in VAar_S01. A destra il contorno terminale è discendente sulla tonica prefinale in VAar_D05: i valori frequenziali scendono da 247Hz a 216Hz. I valori di f_0 sulle vocali che seguono le consonanti sorde* non *vengono rilevati all'immediato onset*



Per descrivere ulteriormente l'andamento di f_0 nella porzione finale di TU, è stata calcolata la pendenza su tutte le vocali prefinali e finali alla ricerca di una eventuale costanza nella tendenza del movimento. Sono stati trascritti gli accenti tonici di TU e la porzione seguente, adottando il metodo ToBI

(Pierrehumbert, 1987; Beckman, Hirschberg, 1994; Avesani, 1995; Marotta, Sorianello, 2001; Grice, D'Imperio, Savino & Avesani, 2005). Sono stati trascritti dunque l'ultimo accento intonativo (*Pich Accent*), che risulta nel *corpus* l'accento principale di TU, l'accento di sintagma (*Phrase Accent*) e il tono di confine (*Boundary Tone*). La determinazione del livello del bersaglio è stata considerata rispetto al contesto circostante e non rispetto all'intera TU né al *range* del parlante. Nell'assegnazione di accenti bitonali si è tenuto conto del fatto che la distanza tra i due bersagli non dovrebbe superare i 200 ms (Marotta, 2000). Quando la maggior parte della salita si consumava sulla sillaba postonica o iniziava sulla ultima testa postonica della TU, si è annotato l'accento L* e la salita finale come accento di sintagma e tono di confine L-H%. Nel caso di un movimento ascendente il primo bersaglio è stato annotato L; quando però si è rilevato un salto frequenziale positivo dopo un segmento sordo, l'accento di sintagma è stato annotato come H- invece di L-, a meno che la salita finale presentasse uno scarto notevole rispetto a quello realizzato in coincidenza della consonante sorda (esempi nelle figure 1 e 2). Questa scelta non è stata operata solo in base al dato acustico, ma anche considerando l'effetto percettivo e la frequenza del fenomeno.

3. *Dati e discussione*

3.1 Contorno terminale

Il contorno terminale **in arabo** è caratterizzato dalla salita che si verifica nel 96% dei casi. Come esposto sopra (tabella 3) il 55% del *corpus* in arabo finisce in parole piane e il 44% in parole tronche; a parte la posizione della tonica lessicale, il punto d'inizio della salita è la sillaba tonica. In generale la terzultima sillaba si presenta con contorno intonativo piatto o discendente.

Nelle Tu *finenti in parole piane* la salita inizia dalla tonica prefinale, con accento (L+H)* nel 63% delle TU, mentre il tono H* si riscontra nell'11%. L'accento intontivo basso L* si rileva nel 26% delle TU.

Nelle TU che *finiscono in parole tronche* l'accento intonativo è (L+H)* nell'89% dei casi e H* nel 7% (3 casi).

In italiano L2 le TU con parole tronche (9% del *corpus*) sono le monosillabiche, tutte trascritte (L+H)* H-H%.

Nelle TU con *parole piane* si profila una salita finale nell'87% dei casi. A differenza dell'arabo L1 il 13% delle TU segnala accento di sintagma e tono di confine bassi L-L%, preceduti per lo più dal PA (H+L)*.

L'accento intonativo è (L+H)* nel 63% delle TU e il tono H* nel 16%.

Le sospensive **in italiano L1** si chiudono con tono di confine alto H% nell'89% del *corpus*: L-H% in 18 TU e H-H% in 29 TU. L'accento di sintagma e il tono di confine H-H% presentano una tenuta dei valori frequenziali in 23 casi su 29.

La chiusura a tenuta bassa L-L% si riscontra 6 volte, di cui 4 nel parlante maschio.

Gli accenti intonativi e la loro occorrenza rappresentano il dato che distingue l'italiano L1 dal resto del *corpus*. La tabella 6 mette a confronto la frequenza degli accenti intonativi in L1 e L2.

Tabella 6 - *Occorrenza degli accenti intonativi finali nel corpus in arabo L1 e italiano L1 e L2. Per l'italiano sono presentati solamente i dati che riguardano le parole piane*

Accento	*L1-tronca*	*L1-piana*	*Italiano L1*	*Italiano L2*
H*	7%	11%	12%	16%
L*	2%	26%	55%	11%
(L+H)*	89%	63%	22%	63%
(H+L)*	2%	__	10%	10%

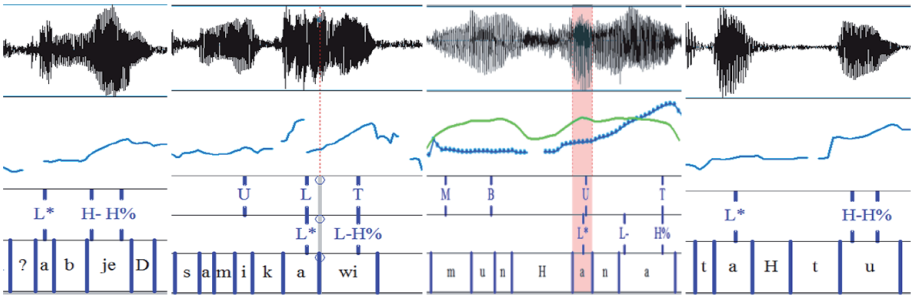
Si nota che l'occorrenza dei tipi accentuali in L2 presenta forti somiglianze con i dati in arabo nelle TU che portano l'accento in posizione prefinale. L'italiano L1, invece, si presenta molto diverso da due punti di vista: primo, gli accenti bitonali, che rivelano un movimento del *pitch*, sono meno frequenti in italiano L1 (il 32% rispetto al 73% in L2); secondo, l'italiano L1 mostra una preferenza per l'accento intonativo basso L*, sia rispetto all'arabo L1 che all'italiano L2, che ne presenta una minore occorrenza.

3.1.1 Discesa +salita

L'accento L* o (H+L)* seguito da salita si riscontra in 13 TU in italiano L2, mentre si rileva in 12 TU in arabo L1, nelle produzioni di una sola parlante.

In arabo L1 la vocale tonica che presenta un bersaglio basso L* è relativamente molto breve di durata rispetto alla postonica. Indipendentemente dal fatto che la tonica presenta una salita con poco scarto frequenziale o una tenuta, la grande pendenza si verifica sulla vocale postonica. Si ha un tracciato continuo con le consonanti sonore (testa della postonica ed, eventualmente, coda della tonica). Invece, in due casi su 10 si hanno delle consonanti sorde, per cui si presenta un salto dei valori frequenziali all'*onset* della vocale postonica, che si configura a sua volta come una salita graduale di f_0 (vedi figura 2).

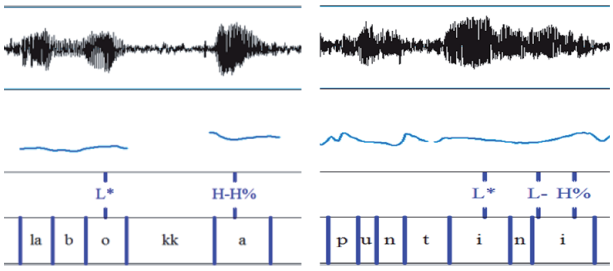
Figura 2 - *Quattro contorni terminali in arabo L1 con accento intonativo L*. Sono, da sinistra, le parole* أبيض *(bianco),* أوي *(molto),* منحنى *(curva) e* تحت *(sotto). L'ultimo esempio, a destra, presenta consonanti sorde tra i due nuclei*



In L2, sempre nei limiti degli esempi riscontrati, si osserva che nei casi di testa sorda della postonica, dopo una tonica aperta, la salita inizia sulla V tonica presentando poca differenza di *pitch* rispetto al tono L*, poi si rilevano valori alti all'*onset* della postonica che mostra una salita costante. L'inizio della salita sulla tonica ha varie posizioni, in 4 casi nella prima metà della vocale e in un caso solo si rileva verso l'*offset*. Tale profilo è stato trascritto L* H-H%.

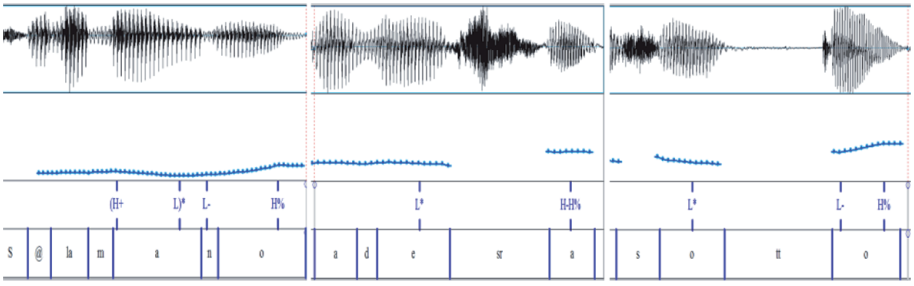
Nei casi di testa sonora della postonica /n/, /l/ l'inizio della salita si sposta più verso destra (5 casi), per cui si può avere la rappresentazione autosegmentale L* L-H% oppure (H+L)* L-H%. Non si può concludere però che il contesto fonetico sia l'unico fattore determinante di tali profili (vedi figura 3).

Figura 3 - *il contorno terminale con accento intontivo basso L* su /bokka/ in VA1_199 e / puntini/ in VB1_A061*



In italiano L1 si rileva un accento basso L* o bitonale (H+L)* seguito da bersaglio alto o movimento ascendente H-H%, L-H% in 32 TU su 53 (60%). Si vede nella figura 4 lo scarto frequenziale dopo una testa sorda. Nel caso di testa sonora di sillaba finale si osserva l'inizio della salita melodica sulla consonante.

Figura 4 - *Da sinistra della figura i turni B04R#p1_234: c'è la* mano*; A01R#p2_59: sulla* destra*; A01R#p2_33: due* sotto*. Il primo tracciato continuo mostra l'inizio della salita sulla testa nasale*



3.1.2 Pendenza

Il movimento di salita nel contorno terminale non è tipico solo delle sospensive, ma si presenta anche nella frase interrogativa, per cui bisogna cercare di esaminare ulteriormente la natura della salita da associare alla sospensione. La pendenza (*slope*)

è una caratteristica del movimento tonale (di salita o discesa) che si ottiene calcolando la differenza tra il valore massimo e minimo di f_0 diviso per la sua durata.

In *arabo* la pendenza varia nei due sottogruppi di TU (con parole finali piane o tronche):

a. sulle parole piane la posizione d'inizio della salita è relativamente arretrata. Si può schematizzare in questo modo il movimento sulle ultime tre sillabe della TU:

b. Nelle TU che finiscono in parole tronche invece tutta la salita è concentrata su una sola sillaba e quindi si segnala un movimento con maggiore pendenza

In *italiano L2* abbiamo ambedue i movimenti di salita e di discesa. Si osserva che i valori medi di pendenza sono più bassi rispetto all'arabo.

a. Parole finali piane

b. Parole finali toniche

Si riscontrano in 9 casi (9% ca. del *corpus* in L2) nelle TU monosillabiche (*poi, sì, tre*):

c. Pendenza della discesa finale

In *italiano L1* gli esempi di parole tronche e sdrucciole sono in tutto 5, per cui vengono qui presentate solo le TU che finiscono con parole piane. La media della pendenza nelle ultime due sillabe è più bassa negli informatori italiani:

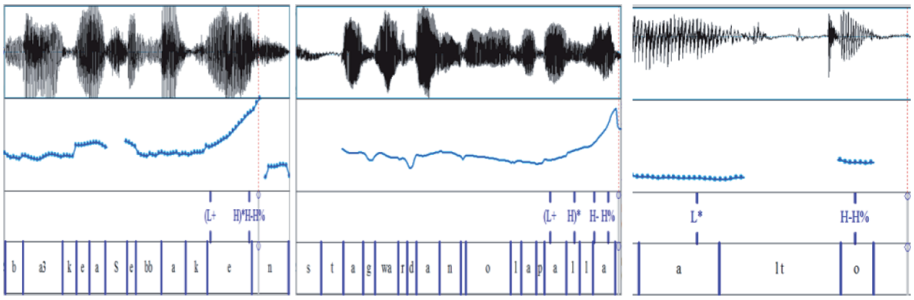
Intanto, il movimento discendente sulla tonica presenta una pendenza di 0,2.

Dunque, quanto alla pendenza la produzione delle informatrici egiziane in L2 si potrebbe collocare in una posizione mediana tra l'arabo L1 e l'italiano L1.

3.2 Allineamento del bersaglio H nell'accento intonativo

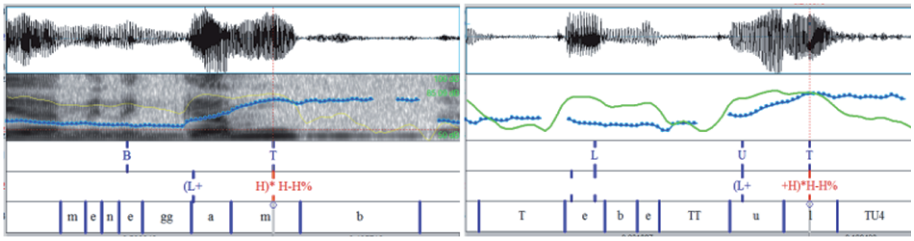
Nel *corpus* gli accenti (L+H)* presentano nella maggior parte dei casi una salita costante che non presenta picchi. Anche se si segnalano spesso avvallamenti locali sulle consonanti sonore, la salita continua all'*onset* della vocale finale, ad un valore di *pitch* vicino a quello rilevato all'*offset* del nucleo precedente. Il tono basso in italiano L1 si presenta spesso in una tenuta o una lieve discesa che non presenta un cambiamento di direzione successivo (vedi figura 5).

Figura 5 - *Tracciato della f_0 in arabo L1 (VAar_M02: poi due finestre) e in italiano L2 (VBit2_A013: sta guardando la palla) di due parlanti diverse. Si osserva sulla sillaba finale tonica una salita continua. In italiano L1 il tracciato viene interrotto in A01R#p1_43: uno poco più* alto*, in coincidenza della testa sorda*



Nelle parole finali tronche nel *corpus* in arabo, dove si accumulano l'accento intonativo, l'accento di sintagma e il tono di confine (*tonal crowding*), si raggiunge il bersaglio H sulla coda sonora, mentre il nucleo stesso non presenta un picco (figura 6).

Figura 6 - *Bersagli raggiunti su coda sonora in L1 (VAar_S18-tu3; S24)*



Ma con la coda sorda il picco rientra nei confini del nucleo (figura 7).

Nei casi di parole finali piane, che presentano un picco prima dell'*offset* del nucleo tonico, in arabo L1 e in italiano L2 è stata rilevata la posizione del picco rispetto all'*onset* e all'*offset* dei nuclei tonici prefinali. I valori medi sono rappresentati nei grafici della figura 8.

Figura 7 - *Il picco entro i confini del nucleo nelle sillabe con coda sorda (VAar_S18-TU1; S21)*

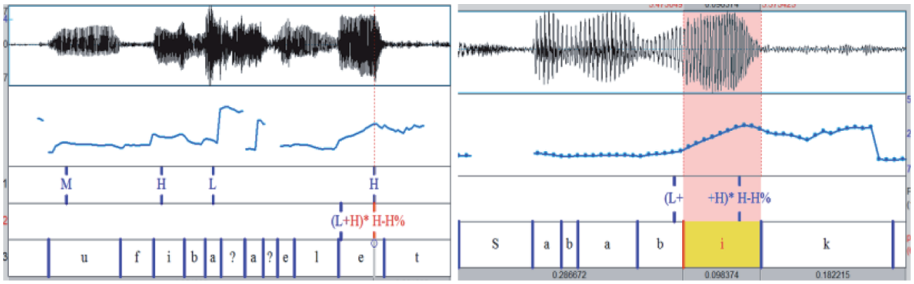
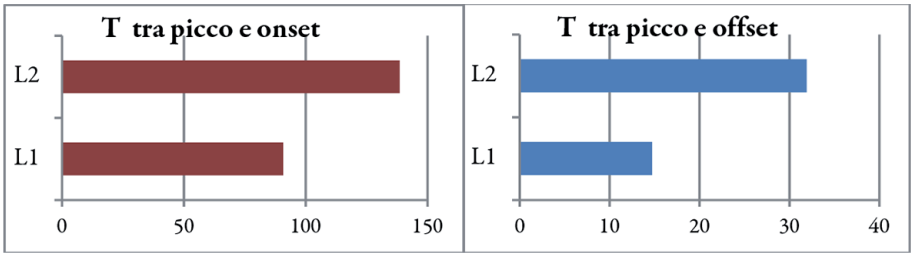


Figura 8 - *All'interno di ogni grafico si osserva la differenza tra L1 e L2 nella posizione del bersaglio H rispetto ai confini dei nuclei prefinali tonici*



La figura 8 presenta i valori assoluti medi, ma infatti si nota che la durata media dei nuclei vocalici prefinali tonici in L1 è minore rispetto alla L2 (rispettivamente: medie 101ms e 149ms; deviazione standard 36 e 52).

Per calcolare la posizione del bersaglio tonale relativamente alla durata del nucleo (*alignment ratio*) è stata adoperata l'equazione (H-S0)/(S1-S0) dove H sta per la posizione del bersaglio f_0 sulla scala del tempo, S0 inizio sillaba, S1 fine sillaba (Chen, Mennen, 2008). I valori medi in arabo e italiano L1 e in italiano L2 sono riportati nella tabella 7, in cui si osserva che la posizione non varia notevolmente rispetto alla durata nucleare.

Tabella 7 - *Allineamento proporzionale alla durata del nucleo prefinale tonico in L1 e L2*

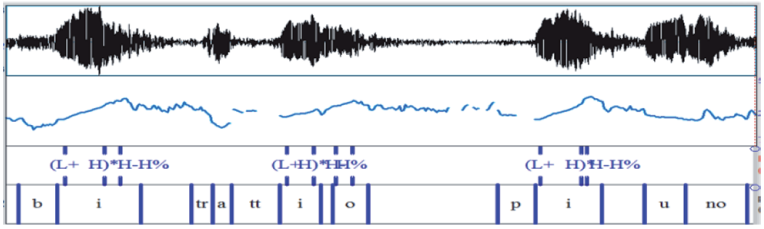
Allineamento	*media*	*dev.st*
Arabo L1	0,84	0,12
Italiano L1	0,82	0,1
Italiano L2	0,80	0,09

3.3 Serie di sospensive

Si è osservato un fenomeno costante a livello macroprosodico: quando si susseguono più unità tonali sospensive, il livello del *pitch* del bersaglio alto di confine (e a volte l'accento intonativo H*) aumenta man mano che si avvicina la fine dell'enunciato o, eventualmente, la TU conclusiva (figura 9). Se l'aumento non è costante, il valore

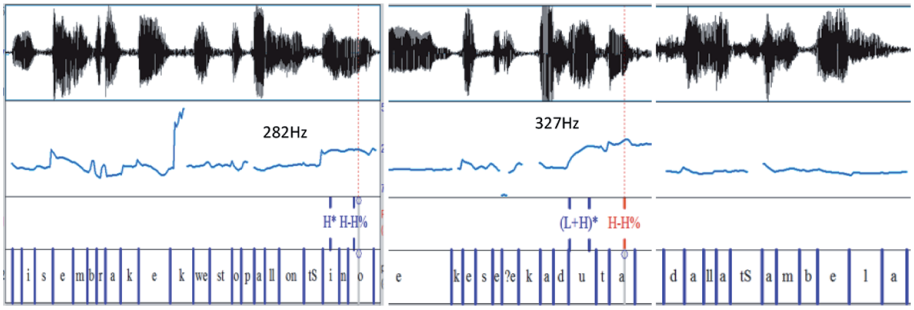
di *pitch* sull'ultima sospensiva risulta il più alto (figura 10). Di questo fenomeno abbiamo 15 esempi in italiano L2 e 10 in arabo L1.

Figura 9 - *Tracciato di 3 sospensive che si chiudono con una conclusiva (VB1_A075). I valori dei picchi nella figura sono 362Hz, 346Hz poi 371Hz sull'ultimo picco*



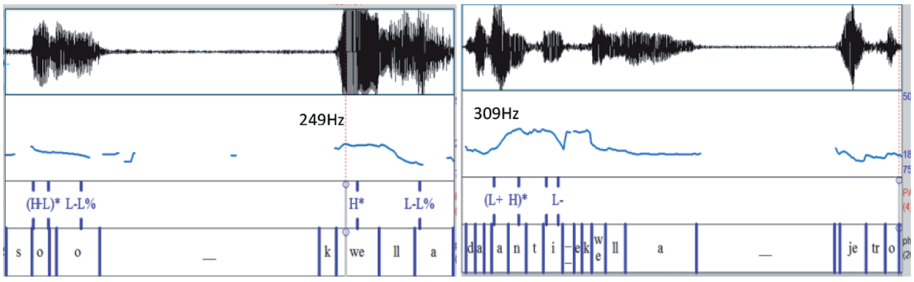
Le esitazioni in mezzo alle TU sospensive (escluse dalle analisi) non rompono il filo conduttore:

Figura 10 - *L'escalation si profila anche quando ci sono esitazioni in mezzo: VB1_B002: mi sembra che questo* palloncino *è la parte<ee> è una parte che è<ee> è* caduta*<eeh> dalla ciambella*



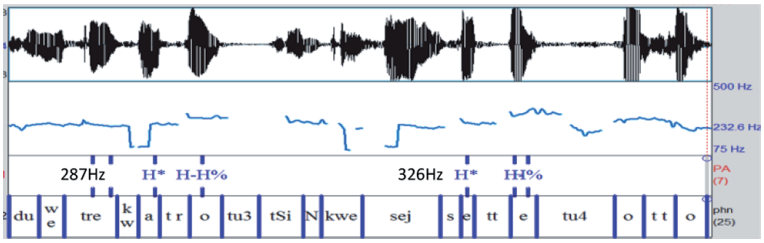
Il fenomeno si estende su più turni, come mostra la figura 11.

Figura 11 - *I massimi di f_0 sulle sospensive concatenate nei due turni VA1_B104:* sono quella*; VA1_B106:* davanti *e quella<aa> dietro. Malgrado la divisione dell'enunciato in due turni dialogici la melodia funge da 'connettivo' e rivela la connessione semantica dei turni della parlante*



Nelle enumerazioni l'andamento globale è piatto (vedi *infra*, § 3.4.1), ma quando la serie di parole enumerate è scandita in più TU, ciascuna con la finale sospensiva, si rileva una gradazione dei bersagli alti (figura 12)

Figura 12 - *I valori finali di f_0 nelle TU sospensive in VBit2_B048: uno due tre* quattro | *cinque sei* sette | *otto*



Lo stesso fenomeno si segnala in arabo (figura 13).

Figura 13 - *Serie sospensive nei turni VAar_S18: la* casa *che gli sta* davanti; *D01: una linea* dentro, *una* dentro

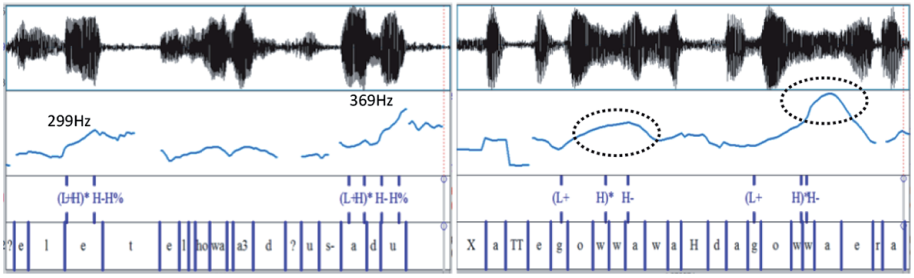
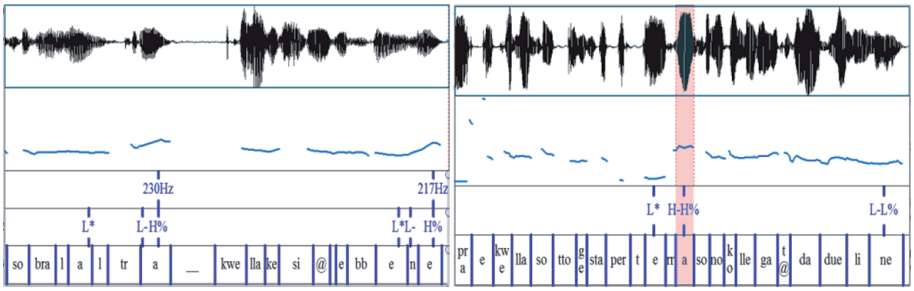


Figura 14 - *Due TU sospensive nel turno A01R_p2#189: no la gamba del tizio dico è messa a cavalcioni sopra* l'altra <*sp*> *quella che si vede* bene. *I valori di f_0 non presentano un aumento graduale. Nel turno B04R_p2#33: le palle, quella sopra* <*sp*> *e quella sotto che sta per* terra *sono collegate da*<*aa*> *due linee, l'innalzamento del pitch su 'terra' preannuncia una TU conclusiva*



In tali contesti non è stato segnalato un ruolo palese della durata o dell'intensità. Dunque, il *pitch* si profila come il parametro significativo nell'espressione della sospensione; e più il *pitch* diventa alto, più si cerca di mantenere la parola. Tale aumen-

to graduale nelle serie sospensive si dimostra un mezzo per far aspettare l'interlocutore annunciandogli che restano cose ancora più importanti da aggiungere e anche per annunciare la prossimità della conclusione.

In italiano L1 non abbiamo una realizzazione simile. Invece di una serie di due o più sospensioni che profilano un aumento graduale del *pitch*, si rileva il preannunciarsi della TU conclusiva. La figura 14 mostra due esempi del finale alto nella TU che precede la TU finale, a conclusione dell'idea.

3.4 Contesti pragmatici della sospensione

La non finalità dell'enunciato colpisce i vari tipi di frase (interrogativa, assertiva, iussiva, ecc.). Basta che il parlante per scelta propria o incertezza scandisca la frase in unità sospensive.

3.4.1 L'enumerazione

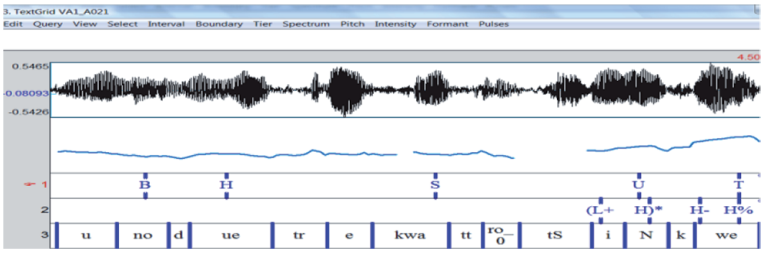
Negli studi sulle sospensive, soprattutto quelli uditivi, le enumerazioni sono l'esempio per eccellenza della sospensione. Nel nostro *corpus*, in arabo L1 e italiano L2, si osserva che l'andamento piatto indicato, tra gli altri, da Halliday (1992) e Lepschy (1978), è riscontrabile solo nelle enumerazioni, altrimenti la salita finale è l'indicatore prevalente della sospensione.

Gli elementi delle enumerazioni presentano un contorno tonale piatto, a meno che l'ultimo elemento sia lasciato in sospeso con salita finale, preannunciando una continuazione conclusiva. Nel nostro *corpus* di L2 ci sono 10 turni dialogici che comprendono una elencazione numerica.

L'andamento globale di tali TU è completamente piatto, nel senso che spesso non presenta la realizzazione della declinazione intontiva; invece, nel resto del *corpus* di L2 le TU presentano un movimento in discesa, mentre poi nella parte finale si verifica una salita. *Nelle parlanti egiziane* il contorno terminale delle enumerazioni varia tra salita, tenuta e discesa.

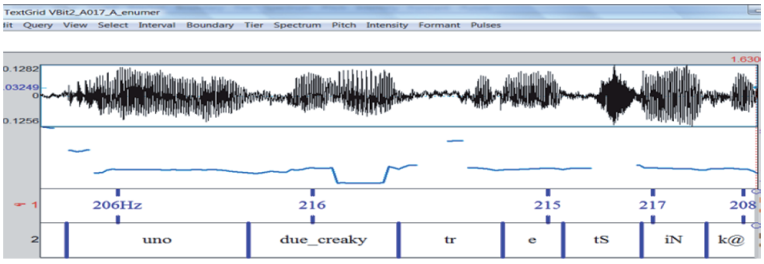
a. La salita finale si presenta, come nella maggior parte delle TU del *corpus*, con tono di confine alto che arriva al massimo frequenziale della TU, sempre nella fascia frequenziale alta del *range* delle parlanti ((L+H)* H-H%).

Figura 15 - *La TU di enumerazione nel turno VA1_A021: otto #<B022><eeh> c'abbiamo#, allora* uno due tre quattro cinque *poi due poi uno che presenta una salita finale a partire dalla tonica finale. Il minimo frequenziale è arretrato come si vede dalla stringa di trascrizione INTSINT (Hirst, Di Cristo & Espesser, 2000)*



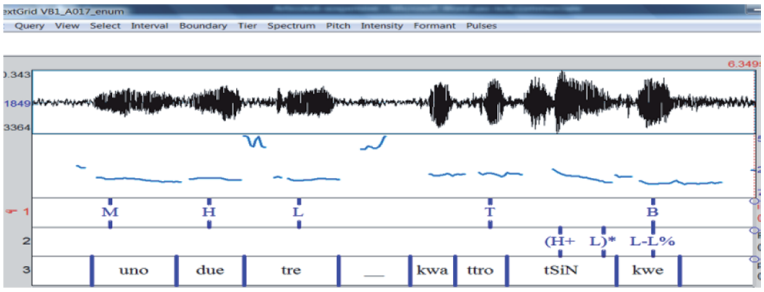
b. La tenuta si rileva alla conclusione dell'enumerazione. Tale tenuta è seguita, o meno, da una conferma: *uno due tre quattro cinque, sì*. Il valore finale di f_0 potrebbe essere il minimo, ma in questo caso si rileva un altro minimo in posizione arretrata.

Figura 16 - *Andamento globale piatto e contorno finale piatto dell'enumerazione in VBit2_A017: uno due tre cinque sì. I valori frequenziali sono esposti sotto il tracciato*



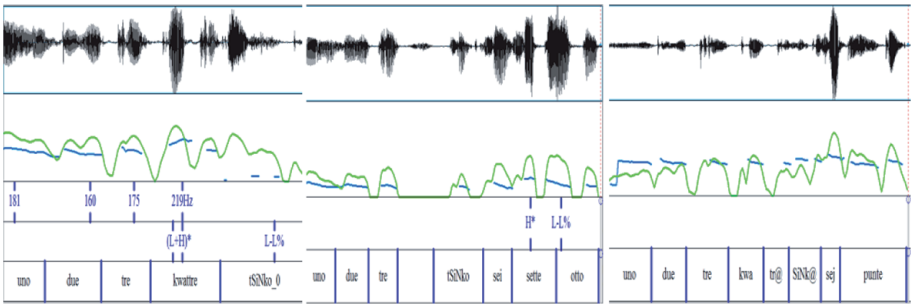
c. Alla fine dell'enumerazione, sia alla chiusura dell'intero turno sia prima del passaggio a un altro argomento, si ha il contorno finale discendente, rappresentabile con (H+L)* L-L% con il tono di confine al minimo frequenziale nella TU, il quale profila uno scarto notevole rispetto al *pitch* sulla tonica prefinale.

Figura 17 - *L'enumerazione conclude il turno in VB1_A017: ok nella<aa> nella *ciambiatella ci sono* uno due tre quattro cinque



Nelle due conversazioni in italiano L1 gli elenchi numerici sono stati riscontrati 19 volte (al di fuori delle 53 TU del *corpus*). L'andamento è generalmente piatto quando l'enumerazione è sussurrata. Nel resto dei casi si rileva un aumento di I e f_0 in coincidenza della penultima parola lessicale (simile al caso C, cfr. *supra*), e ciò in ogni caso, sia l'ultima parola un numero, un oggetto della numerazione o un aggettivo. La figura 18 ne mostra alcuni esempi.

Figura 18 - *L'enumerazione in tre turni in italiano L1 A01R-p2#31: ... uno due tre quattro e cinque; B04R-p1#332: ...<sp> uno due tre <sp> cinque sei sette otto; A01R-p2#125: ... uno due tre quattro cinque sei punte. Il tracciato verde e continuo dell'intensità, con il tracciato di f_0 mostra un aumento dei valori sulla penultima parola, in prossimità della conclusione*



3.4.2 Discesa finale vs salita

Tra le TU sospensive in L2 che finiscono con il tono di confine basso si rilevano 5 TU, prelevate da due dialoghi diversi, che introducono un elemento seguito da una pausa breve e poi contengono una domanda su tale elemento, come in:

VA1_A251: i lacci delle *scarpe*<pb>come sono ? sono solo tre lacci così?
VBit2_B150: dietro il piede destro del *ragazzino* c'è una piccola macchia?

In arabo L1 una terza parlante produce lo stesso profilo con la stessa struttura (nuovo elemento + domanda):

VAar_D05: i baffi del *cane*, quanti puntini sono?

L'altro tipo di sospensione con finale basso si profila su una idea che viene poi contrastata o corretta, come in:

VAar_A16: non sono tre *alberi*, ma tre strati;
VAar_A34b: il cane non *dorme*, cioè è accucciato.

Gli esempi sono però in tutto quattro e perciò non se ne possono trarre conclusioni sicure, ma si può avanzare qualche osservazione.

Nei limiti del *corpus* a nostra disposizione si può desumere che il tono alto mantenga il turno di parola preannunciando altre informazioni, per cui all'interlocutore è richiesto che rimanga 'ascoltatore'. Per contro, il tono basso, che lascia pure l'interlocutore in attesa, preannuncia una richiesta di partecipazione al *partner* che in questo caso sta per diventare 'parlante' attivo, non solo con il *feedback*, ma con aggiunte proprie.

3.5 Salita sospensiva vs interrogativa: osservazioni preliminari

In questo paragrafo presentiamo alcune osservazioni sulla realizzazione fonetica della salita finale. Confrontiamo i dati ottenuti dal presente *corpus* di sospensive

con alcuni dati rilevati in un *corpus* di frasi interrogative in L2 (Gamal, 2012)[1]. Ci limitiamo alla pendenza, all'allineamento e alle durate nucleari nel contorno terminale nelle TU che finiscono in parole piane. Del tipo interrogativo riportiamo nelle tabelle 8 e 9 i dati sulle *query-yn* e *check*.

Tabella 8 - *Pendenza del movimento di salita su sillaba tonica e postonica in TU sospensive e interrogative prodotte in italiano L2 da parlanti egiziani*

	Tonica	*Finale*
Sospensiva	0,43	0,64
Sì/no	0,43	0,51
Check	0,51	0,48

Si osserva che la maggiore differenza risiede nel movimento sulla sillaba finale che nella sospensione presenta una pendenza più alta.

Le durate dei nuclei prefinali tonici e finali presentano delle differenze nelle sospensive e nelle interrogative, essendo le medie notevolmente più alte nelle TU sospensive.

Tabella 9 - *Durate in ms dei nuclei prefinali e finali nelle sospensive e in due tipi interrogativi in italiano L2*

	Tonica	*Finale*
Sospensiva	149	167
Sì/no	132	118
Check	131	124

Infine, l'allineamento proporzionale del bersaglio H nell'accento finale di TU è più arretrato nelle interrogative (media 0,67, deviazione standard 0,1 *vs* 0,84 nelle sospensive).

I dati a disposizione dimostrano una flessibilità prosodica degli apprendenti e una certa capacità di adattamento alla situazione comunicativa in L2. Resta però da esaminare, tra altre cose, la realizzazione fonetica del contorno terminale delle interrogative in arabo L1 per verificare se tale differenziazione tra i due tipi è uguale nella lingua madre o meno.

[1] Tale confronto presenta il vantaggio di avere due informatrici, P1 e P2, nel presente contributo e anche nel lavoro sulle interrogative.

3.6 Il transfer

Rispetto alla lingua d'arrivo i dati ottenuti dal *corpus* in L2 si presentano in consonanza con gli studi acustici sull'italiano L1 che hanno rilevato una salita finale a variazione dell'accento intonativo (vedi *supra*, § 1.1.). Tale affinità ci fa pensare al '*transfer* positivo', per cui le affinità tra L1 e lingua d'arrivo fanno d'aiuto all'apprendente quando riprende in L2 strutture della lingua materna che risultano corrette nella lingua d'arrivo (Major, Kim, 1999). Tuttavia, il *corpus* di controllo in italiano L1 esaminato nel presente lavoro ci chiarisce le differenze riguardanti soprattutto la tipologia degli accenti intonativi (§ 3.1, tabella 6) e la pendenza. Anche se si tratta di un aumento finale della frequenza fondamentale, le informatrici egiziane non producono mai tenute a livello alto sulla sillaba finale. Inoltre, i loro accenti intonativi sono prevalentemente bitonali, il che rispecchia la maggiore tendenza al 'movimento', un dato sostenuto dalla maggiore pendenza nelle loro produzioni.

Detto ciò, la prosodia sospensiva delle informatrici in arabo L1 non risulta identica alla loro L2. Se potessimo, o dovessimo, supporre un grande influsso della lingua madre nel caso dell'affinità tra le due lingue, come potremmo spiegare le sottili differenze rilevate nella pendenza, nelle durate vocaliche e nell'allineamento?

Una delle cause di diversità, osservata nel *corpus*, ma che richiede un ulteriore approfondimento in vari tipi pragmatici, è la differenza nella struttura ritmica tra arabo e italiano. La resa in arabo delle sillabe aperte e chiuse, dal nucleo lungo e breve ha prodotto una distribuzione diversa degli accenti lessicali e quindi una frequenza diversa degli accenti intonativi (§ 3.1). È noto il ruolo del ritmo nella prosodia: Cruttenden (1986: 7), per esempio, ha spiegato come la prominenza lessicale, veicolo di una eventuale prominenza prosodica, sia "*the backbone of intonation*".

Si osserva che le TU in arabo con parole piane presentano una tipologia accentuale più vicina all'italiano L2. Da una parte, ciò potrebbe essere riconducibile al *transfer*, ma dall'altra parte rimane chiaro che la pendenza dei movimenti è diversa.

L'occorrenza dei toni di confine alti e bassi presenta un altro punto di divergenza tra L1 e L2 per la variazione della frequenza d'uso.

Nel nostro *corpus* tale andamento in L2 presenta maggiore frequenza d'uso rispetto all'arabo L1 e si dimostra legato ad alcune regolarità strutturali e comunicative che il *corpus* semispontaneo non offre simmetricamente nelle conversazioni nelle due lingue (§ 3.4.2).

Quindi, si deve supporre che intervengano anche fattori pragmatici e comunicativi. La struttura con dislocazione 'elemento nuovo (con sospensione bassa o discendente) + domanda' non risulta tra le scelte strutturali ricorrenti nel *corpus* di L1, dove le nostre informatrici hanno scelto strategie comunicative diverse, tra cui prevalgono le assertive di descrizione. Si osserva in generale che le parlanti hanno dimostrato più capacità di descrivere tutto senza inibizione: ciò ha aumentato il numero delle assertive, scandite in unità sospensive, con cui si mira a mantenere il turno il più possibile per completare la descrizione.

Conclusioni

Il presente contributo è stato dedicato agli enunciati non finali o sospensivi. La caratteristica della sospensione non è una modalità determinata, ma può interessare vari tipi strutturali e pragmatici. Utilizzando registrazioni delle stesse informatrici in arabo L1 e italiano L2 con lo stesso metodo e il medesimo strumento di elicitazione adoperato per il *corpus* in italiano L1, si intendeva raccogliere dati confrontabili in un ambito ancora aperto all'esplorazione. Ogni studio nel campo delle L2 apre ampi orizzonti di ricerca. Restano, per esempio, tra le molte altre cose, ancora da indagare le caratteristiche che eventualmente differenzino il contorno ascendente della frase interrogativa da quello della sospensione. Inoltre, una analisi più approfondita delle domande in arabo L1 per il confronto con i dati disponibili in italiano L2 e L1 è ancora necessaria.

Lo studio sistematico del ruolo del contesto segmentale, in particolare la struttura ritmica, nella determinazione delle caratteristiche fonetiche della melodia a variazione di competenza in L2 sembra essere una nuova grande e promettente sfida.

Ringraziamenti

"Chi non ringrazia la gente non ringrazia Dio": vorrei ringraziare sentitamente i miei professori al Cairo, i revisori, gli organizzatori del convegno, il pubblico nella sede del convegno. Un ringraziamento particolare a Ester Paone.

A Renata Savy un ringraziamento 'eterno'.

Riferimenti bibliografici

Avesani, C. (1995). ToBIt: un sistema di trascrizione per l'intonazione italiana. In *Atti delle 5e Giornate del Gruppo di Fonetica Sperimentale*, 85-98.

Beckman, M.E., Hirschberg, J. (1994). The ToBI Annotation Conventions. Ohio State University. http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html/ Accessed 23.12.03.

Canepari, L. (1985). *L'intonazione*. Napoli: Liguori.

Canepari, L. (1986). *Italiano standard e pronunce regionali*. Padova: CLEUP.

Cerrato, L. (2007). Sulle tecniche di elicitazione di parlato semispontaneo. Technical Report, progetto CLIPS. http://www.clips.unina.it/Accessed 11.05.09.

Chapallaz, M. (1979). *The Pronunciation of Italian: a Practical Introduction*. London: Bell & Hyman.

Chen, A., Mennen, I. (2008). Encoding interrogativity intonationally in a second language. In *Speech Prosody 2008*, 513-516.

Crocco, C. (2003). Analisi prosodica di un campione di profili sospensivi. In *Atti delle 14e Giornate del Gruppo di Fonetica Sperimentale (Viterbo, 4-6 dicembre 2003)*, 221-226.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Cutugno, F. (2007). Criteri per la definizione delle mappe, esempi di mappe e di vignette per il gioco delle differenze. Technical Report, progetto CLIPS. http://www.clips.unina.it/ Accessed 13.04.09.

Endo, R., Bertinetto, P.M. (1997). Aspetti dell'intonazione in alcune varietà dell'italiano. In *Atti delle 7ª Giornate del Gruppo di Fonetica Sperimentale*, 27-49.

Gamal, D. (2005). L'intonazione in italiano L2 di arabofoni. Studio sociolinguistico e analisi prosodica. Tesi di Dottorato non pubblicata, Università di Ain Shams, Il Cairo.

Gamal, D. (2012). L'intonazione interrogativa in italiano L2 di parlanti egiziane. In *Philology (Rivista della Facoltà Al-Alsun, Università di Ain Shams)*, 58, n. 2, 165-198.

Gili Fivela, B. (2002). L'intonazione della varietà pisana di italiano: analisi delle caratteristiche principali. In *Atti delle 12ª Giornate del Gruppo di Fonetica Sperimentale*, 103-110.

Grice, M., D'Imperio, M.P., Savino, M. & Avesani, C. (2005). Strategies for intonation labelling across varieties of Italian. In Jun, S.-A. (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 362-389.

Grice, M., Savino, M. (1995). Intonation and communicative function in a regional variety of Italian. In *Phonus*, 1, 19-32.

Halliday, M.A.K. (1992). *Lingua parlata e lingua scritta*. Firenze: La Nuova Italia. Tit. orig. *Spoken and Written Language*. Victoria: Deakin University (1985).

Hellmuth, S. (2006). Intonational pitch accent distribution in Egyptian Arabic. Tesi di Dottorato non pubblicata, School of Oriental and African Studies, University of London, London.

Hirst, D., Di Cristo, A. & Espesser, R. (2000). Levels of representation and levels of analysis for intonation. In Horne, M. (Ed.), *Prosody: Theory and Experiment*. Dordrecht: Kluwer, 51-87.

Kulk, F., Odé, C. & Woidich, M. (2003). The intonation of colloquial damascene Arabic: a pilot study. In *Institute of Phonetic Sciences Proceedings*, 25, 15-20.

Lepschy, G.C. (1978). Note su accento e intonazione con riferimento all'italiano. In Id., *Saggi di linguistica italiana*. Bologna: Il Mulino, 111-126.

Major, R.C., Kim, E. (1999). The Similarity Differential Rate Hypothesis. In Leather, J. (Ed.), *Phonological Issues in Language Learning*. Oxford: Blackwell, 151-183.

Marotta, G. (2000). Allineamento e trascrizione dei toni accentuali complessi: una proposta. In *Atti delle 10ª Giornate del Gruppo di Fonetica Sperimentale*, 139-149.

Marotta, G., Sorianello, P. (2001). La teoria autosegmentale nell'analisi dell'intonazione interrogativa in due varietà di italiano toscano (Lucca e Siena). In *Dati empirici e teorie linguistiche, Atti del XXXIII Congresso della SLI*. Roma: Bulzoni, 177-204.

Pierrehumbert, J.B. (1987). *The Phonology and Phonetics of English Intonation*. Bloomington, Indiana: Indiana University Linguistics Club publications.

Rifaat, K. (2005a). The phonology of Colloquial Egyptian Arabic intonation. Abstract submitted to PAPI 2005, Barcelona.

Rifaat, K. (2005b). The structure of Arabic intonation: a preliminary investigation. In Alhawary, M.T., Benmamoun, E. (Eds.), *Perspectives on Arabic Linguistics XVII-XVIII: Papers from the seventeenth and eighteenth annual symposia on Arabic linguistics.* Amsterdam-Philadelphia: John Benjamins, 49-67.

Sacks, H., Schegloff, E.A. & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking. In *Language*, 50, 4, Part 1, 696-735. http://www.jstor.org/Accessed 12.05.12.

Savino, M., Refice, M. (1997). L'intonazione dell'italiano di Bari nel parlato letto e in quello spontaneo. In *Atti delle 7e Giornate di Studio del Gruppo di Fonetica Sperimentale*, 79-88.

Sorianello, P. (1997). Dal parlato letto al parlato spontaneo: indici prosodici a confronto. In *Atti delle 7e Giornate del Gruppo di Fonetica Sperimentale*, 89-110.

DOI: 10.17469/O2102AISV000009

PATRIZIA SORIANELLO, ANNA DE MARCO

Sulla realizzazione prosodica delle emozioni in italiano nativo e non nativo[1]

This study explores the expression of vocal emotions by Russian, Spanish and Tunisian learners of L2 Italian. The research is firstly intended to contribute to the studies on the emotional speech in L2, through an analysis of the acoustic profiles of vocal emotions produced by the learners both in Italian and in their own L1. Furthermore, the study aims to determine whether and to what extent the prosodic realization of L2 emotional speech is influenced by transfer phenomena from the learners' cultural background and/or reveals more universal tendencies due to L2 acquisition process. Results show a fuzzy and unexpected picture. On the one hand, Tunisian learners reproduce in L2 Italian intonational contours very similar to their mother tongue. On the other hand, both Russian and Spanish learners show hybrid prosodic patterns which do not adhere to their mother tongue and diverge both from the Tunisian and the native speakers of Italian.

Key words: emotions, prosody, learners of Italian, prosodic transfer.

1. *Introduzione*

L'incontro di culture diverse si colloca in una dimensione ibrida, la cui interazione si fonda su equilibri delicati spesso alterati da lacune nella comunicazione, interpretazioni errate del comportamento altrui e valutazioni costruite sulla base delle proprie credenze. Come sostiene Hall (1976: 67) la cultura in cui una persona cresce condiziona la selezione di certe informazioni rispetto ad altre, senza che questa ne avverta coscienza. Pertanto, negli apprendenti di una seconda lingua (L2), le prime difficoltà nella comunicazione sorgono non solo a livello verbale, ma anche e soprattutto sui piani non-verbale e paraverbale. In questa dimensione si inserisce l'espressione emotiva di ogni cultura, di sovente associata a modalità di manifestazione vocale differenziate. Alcuni studi sull'acquisizione di una lingua seconda hanno posto l'accento sull'effettiva difficoltà sperimentata dagli apprendenti nel gestire la dimensione emozionale in una cultura altra (Dewaele, 2005). L'obiettivo del lavoro che presentiamo è quello di contribuire, in primo luogo, agli studi sul parlato emotivo in L2 attraverso un'analisi dei profili acustici del parlato emotivo prodotto dagli apprendenti in italiano e nelle loro rispettive lingue materne (L1). L'intento è quello di verificare sul piano acustico le differenze nell'espressione dell'emotività fra apprendenti con diversa L1. Lo studio mira a verificare se e in che misura il parlato emotivo dei soggetti interessati riveli

[1] La ricerca è frutto di una collaborazione continua tra le autrici; tuttavia PS è responsabile dei §§ 2.1, 4.3, 5, 5.1 e 6, mentre ADM dei §§ 1., 2., 3., 3.1, 3.2 e 7. I paragrafi 4, 4.1, 4.2 sono comuni.

aspetti riconducibili alla loro cultura d'origine. Prima di presentare il nostro studio affronteremo gli aspetti ad oggi più indagati sul parlato emotivo esaminando gli approcci e le teorie più recenti che hanno ispirato gli studi sia nella direzione del confronto interlinguistico sia nell'ambito più specifico dell'apprendimento di una lingua seconda.

2. *Il parlato emotivo: approcci e teorie*

La ricerca sulle emozioni vocali ha esplorato i due processi che sottendono al parlato emotivo:

Il processo di *encoding*, ovvero l'analisi dei parametri acustici coinvolti nell'espressione delle emozioni in cui la prospettiva assunta è quella del parlante (*speaker-centered*).

Il processo di *decoding*, ovvero lo studio dell'aspetto percettivo degli stati emotivi in cui il punto di vista adottato è quello dell'ascoltatore (*listener-centered*).

In linea generale, gli studi che hanno esplorato entrambe le prospettive sono esigui e, inoltre, la disparità dei metodi impiegati non ha garantito l'affermarsi di una metodologia condivisa. Uno degli aspetti maggiormente discussi riguarda la natura stessa delle emozioni indagate (simulate vs. spontanee) e il metodo di elicitazione adottato (simulazione, induzione sperimentale). Alcuni studi hanno impiegato emozioni vocali spontanee o semi-spontanee, coinvolgendo parlanti *naif* all'interno di un contesto naturale. Altri studi, invece, hanno optato per un parlato emotivo simulato, avvalendosi di attori o parlanti ingenui.

Le ricerche condotte da Scherer (1989) insieme ad altri studiosi hanno fornito un contributo fondamentale alla codifica acustica delle emozioni, sia in ambito intralinguistico sia – come vedremo – in ambito cross-linguistico. Proponiamo una sintesi dei correlati acustici di cinque emozioni di base (collera, paura, gioia, tristezza e disgusto) emersi da una serie di ricerche (cfr. Scherer, 1989; Pittam, Scherer, 1993; Banse, Scherer, 1996).

Secondo Banse, Scherer (1996), le alterazioni fisiologiche innescate dalle emozioni coinvolgono la respirazione, la fonazione, l'articolazione e determinano le variazioni del segnale vocale, avvalorando l'esistenza di una continuità filogenetica nei pattern acustici dell'espressione vocale di tipo emotivo (615).

Le conclusioni tratte da queste analisi mostrano che emozioni caratterizzate da un grado di attivazione simile condividono parametri acustici analoghi. In particolare, tale associazione risulta evidente in presenza di stati emotivi dotati di un'elevata attivazione come paura, collera e gioia (Banse, Scherer, 1996). Queste emozioni presentano, infatti, degli andamenti intonativi che si estendono nelle fasce frequenziali più alte (indicando un coinvolgimento emotivo altrettanto importante). Anche i valori relativi all'intensità riflettono l'incremento generale degli indici soprasegmentali. La durata delle espressioni, invece, si riduce visibilmente.

Per quanto riguarda gli stati emotivi caratterizzati da una bassa attivazione, come la tristezza e il disgusto, il loro contorno intonativo appare piuttosto statico e si situa soprattutto nelle fasce di frequenza più basse. L'escursione melodica è ridotta, mentre la durata viene enfatizzata, soprattutto nelle espressioni di disgusto.

I correlati acustici individuati da Scherer e collaboratori hanno contribuito alla formulazione di due teorie antitetiche che sostengono rispettivamente la *discretezza* e la *scalarità* degli andamenti intonativi paralinguistici (Poggi, Magno Caldognetto, 2004). Il primo approccio interpreta le risorse soprasegmentali (frequenza fondamentale, intensità e durata) come un sistema di opposizioni, mentre il secondo inserisce le loro variazioni in un sistema scalare, che varia sia all'interno del medesimo stato emotivo sia in relazione a stati emotivi differenti. Pertanto, le ricerche attuali mirano ad affermare l'una o l'altra teoria, attraverso l'analisi dei *pattern* acustici associati a determinate emozioni.

Più recentemente, l'approccio dominante è quello multidimensionale (Scherer, 2003). Come spiegano Poggi e Magno Caldognetto, "la condivisione di indici acustici tra le diverse emozioni è stata considerata indicativa della condivisione delle dimensioni psicologiche costitutive delle emozioni, in particolare di quelle di attivazione, valutazione e potere" (Poggi, Magno Caldognetto, 2004: 16). La dimensione di "attivazione" caratterizza emozioni come sorpresa, collera e paura rispetto a stati emotivi come la tristezza e il disgusto. Essa infatti è strettamente connessa all'importanza dello scopo, pertanto emozioni ad alta attivazione come paura, collera ecc. presentano dei correlati acustici ben definiti: i valori relativi alla frequenza fondamentale (*f0*) e all'intensità sono più elevati, così come la velocità di eloquio risulterà maggiore e la presenza delle pause è drasticamente ridotta. Viceversa, le emozioni caratterizzate da una bassa attivazione esibiscono specularmente dei valori opposti.

La dimensione di "valutazione" discerne stati emotivi di natura positiva (gioia, speranza ecc.) da stati emotivi prevalentemente negativi (collera, paura). Tali emozioni sono generate rispettivamente dal raggiungimento o dalla compromissione di uno scopo. La collera, ad esempio, presenta dal punto di vista acustico, un incremento dell'intensità, riduzioni (o diminuzioni) veloci di *f0* e delle pause più lunghe (Poggi, Magno Caldognetto, 2004). La dimensione relativa alla "potenza" riflette la condizione di dominanza o sottomissione legata al raggiungimento di uno scopo, ossia il grado di controllo che ciascun individuo esercita nel perseguire i propri fini. Emozioni legate alla dominanza, come la rabbia, si distinguono ad esempio per un'intensità maggiore rispetto alla paura (legata ad una dimensione di sottomissione).

2.1 L'intonazione delle emozioni

Ad oggi, gli studi sulle emozioni hanno privilegiato l'analisi delle caratteristiche vocali. Riassumendo l'ampia e variegata letteratura sull'argomento, si osserva che i parametri esplorati sono essenzialmente di due tipi: da un lato sono stati considerati i descrittori globali, quali l'intensità, il livello e il *range* di *f0*, la velocità articolatoria e le pause, indici di per sé sufficienti a differenziare le emozioni; dall'altro sono stati spesso ispezionati anche i parametri di perturbazione vocale, come *shimmer* e *jitter*, e la

qualità della voce (Gobl, Ní Chasaide, 2003). In questo vasto e interdisciplinare filone di indagine non mancano le ricerche incentrate sulla discriminazione percettiva delle emozioni, effettuata soventemente mediante sintesi del segnale sonoro.

Diversamente, le ricerche tese all'approfondimento dell'intonazione delle emozioni sono ad oggi di numero ridotto. Il ruolo svolto dall'intonazione nell'identificazione di enunciati emotivi costituisce infatti una tematica controversa, sul versante metodologico ed interpretativo. In passato, l'esistenza di moduli intonativi direttamente associabili ai tipi emotivi è stata persino messa in dubbio. La questione di fondo è se il contorno intonativo di un enunciato possa o meno veicolare un significato emozionale, indipendentemente dalla sua struttura linguistica. In questa direzione, gli aspetti degni di nota non mancano, ad es. Juslin, Laukka (2003) passano in rassegna 104 ricerche svolte sull'espressione vocale delle emozioni pubblicate tra il 1938 e il 2001, rilevando la frequente ricorrenza di un contorno intonativo finale ascendente per collera, gioia e paura e di un contorno discendente per tristezza e tenerezza. La situazione però è ben più complessa. Non solo l'intonazione è notoriamente portatrice di svariati significati, linguistici e paralinguistici, ma va altresì osservato che uno stesso stato emotivo può essere realizzato con diversa modalità frasale. A questo proposito, Scherer, Ladd & Silverman (1984) notano che il significato 'espressivo' trasmesso da un contorno discendente cambia a seconda che l'enunciato sia una domanda totale, in questo caso sarà interpretato come atteggiamento di sfida, o una domanda parziale. Secondo questi autori, le variazioni di *f0* (livello e *range*) covariano con gli aspetti vocali e riflettono il grado di attivazione psicologica incorporata nell'emozione, mentre la forma del contorno intonativo convoglierebbe innanzitutto informazioni linguistiche. L'azione del *pattern* intonativo dunque contribuirebbe a definire l'emozione solo se considerata in combinazione con altri parametri, come l'escursione tonale, l'intensità, la durata, compresa la tipologia modale dell'enunciato (cfr. Ladd, Silverman, Tolkmitt, Bergmann & Scherer, 1985; Scherer, Feldstein, Bond & Rosenthal, 1985). Questo risultato è confermato dalla ricerca di Bänziger, Scherer (2005) incentrata sull'analisi quantitativa di emozioni prodotte con basso/alto grado di attivazione (*arousal)*. La somministrazione di un test di discriminazione uditiva, insieme alla successiva stilizzazione dei contorni intonativi, dimostrano che i valori globali di *f0* (media, livello ed escursione) variano in modo considerevole in funzione del grado di coinvolgimento espresso dall'emozione; il contorno di *f0* si prefigura invece come un indice più debole nel discriminare le stesse emozioni. Diversamente, in Rodero (2011), durante il processo di riconoscimento percettivo delle emozioni svolto da 100 uditori, il tipo di contorno intonativo è risultato più incisivo rispetto al livello frequenziale (*pitch level*).

Questo settore di indagine, per quanto controverso e a tratti insidioso, vista la difficoltà di poter scindere il peso dei diversi fattori in gioco, continua a fornire elementi degni di osservazione. Recentemente, alcune ricerche (Liscombe, 2007; Cao, Beňuš, Gur, Verna & Nenkova, 2014) hanno ad esempio indagato il ruolo dell'intonazione nella resa delle emozioni, adottando il modello teorico Autosegmentale e Metrico (AM). I risultati conseguiti sono interessanti, poiché provano come anche un approc-

cio teorico notoriamente discreto come quello AM possa concorrere a differenziare i tipi emotivi, fornendo informazioni, se non esclusive, almeno complementari allo studio acustico dei descrittori globali. Dagli esiti raccolti si evince come non tutte le emozioni mostrino il medesimo andamento melodico finale, ad es. la gioia è spesso restituita dal tono !H-L%, i *pattern* di collera e disgusto sono invece tipicamente discendenti e associati ai toni demarcativi L-L%.

Per quanto concerne l'italiano, parimenti a quanto osservato in altre lingue, la realizzazione delle emozioni è il risultato congiunto di parametri vocali e di indici intonativi. La gioia è trasmessa da un *pattern* tendenzialmente alto e variegato, privo di variazioni brusche, e discendente nella parte finale. Diversamente, paura e collera mostrano veloci variazioni di *f0*, combinate a intensità e velocità elocutiva elevate; una diversa attivazione della durata contribuisce infine a differenziare tristezza e disgusto, entrambe associate ad un contorno di bassa intensità realizzato in uno spazio melodico ristretto (cfr. Anolli, Ciceri, 1992; Magno Caldognetto, 2002; Magno Caldognetto, Cavicchio & Cosi, 2008). Queste tendenze sono sostanzialmente confermate anche da un particolare filone di indagine che indaga l'espressione delle emozioni in italiano a partire da campioni di parlato cinematografico o teatrale (cfr. Gili Fivela, Grimaldi & Stefàno, 2004; Pettorino, 2008)[2].

3. *Il parlato emotivo in una prospettiva cross-linguistica*

Buona parte delle ricerche sul parlato emotivo ha tentato di individuare tendenze comuni nell'interpretazione delle emozioni vocali a livello interculturale. Tale approccio dovrebbe fornire – in linea teorica – una prova indiretta dell'esistenza di elementi universali (o specificamente culturali) nella comunicazione emotiva. Esiste, tuttavia, una serie di studi, quantitativamente meno consistente, focalizzata sull'analisi dei parametri acustici e della loro variazione a livello cross-linguistico.

In ambito italiano, ricordiamo uno studio condotto da Anolli, Wang, Mantovani & De Toni (2008) finalizzato al confronto del parlato emotivo prodotto da parlanti italiani e cinesi nelle loro rispettive lingue. Dai risultati sono emerse differenze sostanziali tra i due *corpora* e tra tutte le emozioni considerate per entrambi i gruppi di parlanti. Le emozioni considerate sono state otto più l'eloquio neutro: *gioia, collera, paura, tristezza, vergogna, orgoglio, disprezzo* e *senso di colpa*. In relazione al parametro temporale, il parlato dei cinesi si è distinto per una velocità d'eloquio inferiore rispetto a quello degli italiani e anche le pause sono risultate più lunghe.

Sul piano frequenziale, è emersa una variazione minore del *pitch* nelle emozioni vocali dei parlanti cinesi rispetto alle espressioni neutre; per contro, nel parlato degli italiani tale variazione è risultata maggiore. Anche in relazione all'intensità, i risultati

[2] Nel dettaglio, dal confronto della *performance* di tre attori che recitano lo stesso personaggio di una medesima opera teatrale si evince che le emozioni indagate, collera e tristezza, sono espresse in modo diverso, talora variando la gamma del coinvolgimento, e comunque sfruttando gli indici vocali in modo individuale. I tratti acustici di norma associati alle due emozioni sono però nel complesso rispettati (Gili Fivela et al., 2004).

hanno evidenziato un andamento similare: il parlato degli italiani ha esibito valori più alti e variazioni più ampie rispetto a quello dei cinesi.

Le differenze riscontrate tra i due *corpora* hanno evidenziato modalità di espressione vocale strettamente connesse al contesto culturale dei parlanti. In base all'interpretazione data dagli autori, la cultura cinese – al pari delle altre culture asiatiche – "stresses relational harmony and invites individuals to take their proper place, discouraging them from occupying too much space in relatioships" (Anolli et al., 2008: 583). Tale impostazione culturale si rifletterebbe nella tendenza generale ad evitare gli eccessi, e a ridurre l'espressività nella comunicazione emotiva, poiché una simile attitudine contrasterebbe con le norme sociali che impongono una linea di condotta sobria e moderata, al fine di preservare lo spazio di ciascun membro.

Tra i diversi *corpora* finalizzati al confronto cross-linguistico, ricordiamo anche il DEMO (*Dutch Emotion*)/ KEMO (*Korean Emotion*) Corpus elaborato da Goudbeek, Broersma (2010). Il corpus comprende otto emozioni vocali espresse da parlanti di origine coreana e olandese[3]. Gli autori hanno confrontato la durata dei fonemi nelle rispettive lingue, ipotizzando una diversa gestione temporale dei segmenti fonetici dovuta alla specificità linguistica. I risultati hanno confermato la presenza di differenze sostanziali tra le due lingue. Il parlato dei coreani si è distinto, in media, per una durata inferiore rispetto a quello degli olandesi; i segmenti fonetici hanno mostrato dunque un pattern temporale diverso, sia nel confronto tra le due lingue sia all'interno delle emozioni espresse in una sola lingua. I risultati hanno portato gli autori ad affermare che "language as well as emotion has a significant influence on the duration of a segment. [...] these two factors often interact with one another, thus creating language specific effect of emotion on phonetic segment duration" (Goudbeek, Broersma, 2010: 4). Infine, occorre menzionare il *corpus* di parlato emotivo mistilingue costituito da quattro lingue europee, italiano, inglese, francese e tedesco, e validato attraverso test di dicriminazione uditiva (Galatà, 2010).

3.1 Il parlato emotivo in L2

L'attenzione verso la gestione del parlato emotivo in L2 è relativamente recente, ne consegue che anche la letteratura di riferimento risulti abbastanza esigua. Tuttavia, le ricerche condotte in questo ambito hanno prodotto risultati interessanti che portano a riconsiderare il peso della comunicazione emotiva all'interno del processo di apprendimento di una lingua seconda. Analizziamo nel dettaglio le problematiche evidenziate da alcuni di questi studi.

Tra le ricerche che hanno coinvolto specificamente apprendenti di L2, ricordiamo quella di Altrov (2013), finalizzata alla valutazione delle influenze culturali sulle abilità di decodifica in una lingua seconda. I soggetti coinvolti appartenevano a tre gruppi distinti: parlanti nativi di origine estone, apprendenti di origine russa residenti

[3] Ciascuna emozione è stata veicolata da una frase standard, la medesima per entrambe le lingue, elaborata *ad hoc*, in base a tre criteri: 1) sono stati selezionati fonemi presenti sia in coreano sia in olandese; 2) la frase è priva di significato in entrambe le lingue; 3) non contiene parole riconoscibili.

in Estonia con un livello di competenza elevato (C1), parlanti nativi russi senza alcuna conoscenza della lingua estone. A ciascun gruppo è stato somministrato un test uditivo contenente un certo numero di stimoli di natura emotiva (insieme all'eloquio neutro) in lingua estone. Dai risultati è emerso che soltanto gli apprendenti russi residenti in Estonia hanno identificato le emozioni oltre la soglia della casualità. In base alle conclusioni tratte da Altrov, gli apprendenti russi, vivendo a contatto con la cultura estone, avrebbero "imparato" a decodificare con maggiore accuratezza le emozioni vocali espresse dai nativi rispetto ai parlanti russi residenti in Russia. L'autore conclude affermando: "Our results [...] suggest that understanding emotions is dependent on cultural factors and social interactions. That is, the social norms of a culture are learnt during practical interaction" (Altrov, 2013: 172).

In relazione alla gestione del parlato emotivo in L2, Kim, Dorner (2013) hanno analizzato, mediante la conduzione di interviste e la somministrazione di questionari, le difficoltà espressive sperimentate da un gruppo di apprendenti inglesi di origine coreana. Nel dettaglio, è stato richiesto agli apprendenti di indicare: a) quali termini di natura emotiva – legati alla loro L1 – fossero difficilmente traducibili in inglese; b) in quali contesti sperimentassero maggiori difficoltà di espressione nella L2. I risultati ottenuti hanno evidenziato il disagio provato da alcuni soggetti per una serie di fattori, tra i quali l'assenza di un lessico emotivo (in inglese) idoneo alle loro esigenze comunicative. Infatti, gli apprendenti hanno dichiarato in molti casi che alcuni concetti legati alla cultura coreana non trovavano alcuna corrispondenza nella lingua inglese. Inoltre, il ricorso alla L2 per esprimere stati d'animo o per parlare di questioni intime e personali risultava più problematico e limitava la loro spontaneità; in alcuni casi, nelle interazioni con i nativi, preferivano evitare di trattare argomenti relativi alla sfera emotiva per i quali non avrebbero saputo fornire spiegazioni esaustive. Anche in termini di espressività, non soltanto sul piano linguistico ma anche su quello vocale non verbale, alcuni soggetti hanno dichiarato di essere meno comunicativi e di risultare più distaccati e freddi nelle interazioni con i nativi.

Sul piano prettamente prosodico, ricordiamo uno studio condotto da Komar (2005) che ha proposto un'analisi contrastiva del parlato prodotto da parlanti sloveni in inglese e da parlanti nativi inglesi. I risultati hanno evidenziato la tendenza degli sloveni ad impiegare un'intonazione "piatta", meno dinamica rispetto ai nativi, dovuta principalmente alle diversità dei due sistemi intonativi, ma anche allo stato d'ansia e di disagio che si avverte parlando una lingua straniera.

In ambito italiano, sono ancora piuttosto esigui gli studi che hanno indagato questo particolare aspetto della comunicazione. Ricordiamo uno studio recente condotto da Maffia, Pellegrino & Pettorino (2014) che ha analizzato il parlato emotivo prodotto da apprendenti cinesi in italiano. I dati emersi dall'analisi acustica hanno evidenziato, nel confronto con il parlato dei nativi, una serie di differenze sul piano temporale e su quello frequenziale. Le produzioni degli apprendenti si sono differenziate per una velocità d'eloquio molto più lenta e per un'intonazione meno dinamica (83).

Il ricorso a un'intonazione poco modulata e variabile nell'espressione delle emozioni sembra dunque essere una tendenza comune agli apprendenti di una lingua se-

conda. Tale dato è stato riportato anche in studi successivi (De Marco, Paone, 2015; 2016) condotti su studentesse indonesiane e polacche (le cui produzioni mostravano un contorno intonativo similare in tutte le emozioni considerate). A livello percettivo molti studi hanno confermato che gli apprendenti di culture distanti mostrano maggiori difficoltà rispetto agli apprendenti di lingue più vicine. Tuttavia, non sempre la differenza culturale è sufficiente a garantire una comprensione maggiore del parlato emotivo (De Marco, Paone, 2014). Lo studio di Altrov (2013) suggerisce che la comprensione degli stati emotivi sia condizionata dalla conoscenza delle norme sociali di una cultura che si apprende nel corso dell'interazione. Ogni sistema culturale opera delle scelte sul piano comunicativo e indirizza i membri che ne fanno parte verso l'adozione di uno stile comunicativo in linea con il contesto e le norme sociali. In particolare, il contesto, sembra rivestire un ruolo determinante, soprattutto nella definizione proposta da Hall, a cui faremo riferimento nel prossimo paragrafo.

3.2 Il modello di Hall

Per l'interpretazione dei dati del nostro studio e in particolare in relazione all'influenza esercitata dalle caratteristiche della cultura d'origine sulle produzioni in L2, abbiamo adottato il modello di Hall che interpreta le varie culture del mondo attraverso una scala che misura la rilevanza del contesto all'interno della comunicazione.

Per contesto si intendono le informazioni globali che racchiudono un evento, pertanto, secondo Hall il contesto è strettamente legato al significato dell'evento stesso (Hall, Hall, 1990).

Nelle culture ad alto contesto comunicativo (High Context: HC), l'informazione è captata in modo implicito, dunque la realizzazione prosodica degli enunciati è più contenuta e meno modulata attraverso la persona e il contesto fisico. Nelle culture HC, buona parte del significato all'interno di una conversazione viene tendenzialmente trasmesso attraverso canali non verbali, quali i gesti, la mimica facciale, questo per sopperire alla mancanza di esplicitezza del contenuto verbale. La conversazione infatti si sviluppa maggiormente sul piano del non detto, del sottinteso. I membri di questo tipo di cultura sono in sintonia con il proprio ambiente e pertanto riescono ad esprimere e ad interpretare le emozioni anche in modo non verbale (Samovar, Porter & McDaniel, 2007).

Al contrario, le culture a basso contesto comunicativo (Low Context: LC) presentano generalmente una grande diversità all'interno delle comunità; ciò comporta da parte degli interattanti un'attenzione maggiore verso tutti quei dettagli in grado di fornire informazioni precise sul contesto comunicativo. Il messaggio verbale viene espresso in modo esplicito, e poco viene affidato ai segnali non verbali o al contesto fisico.

In linea generale, ogni cultura presenta degli aspetti legati ad entrambi i tipi di comunicazione (ad alto e basso contesto comunicativo).

Culture ad alto contesto ← Giappone – Paesi arabi – Grecia – Spagna – Italia – Inghilterra – Francia – Nord America – Paesi Scandinavi – Germania → Culture a basso contesto (adattato da Hall, Hall, 1990: 6).

A supporto di tale interpretazione, uno studio recente ha preso in esame la distanza culturale tra polacchi e inglesi (Biel, 2004), attraverso un'analisi dell'aspetto paraverbale. Gli inglesi, cultura a basso contesto comunicativo, si esprimono soprattutto attraverso il linguaggio verbale e paraverbale, limitando l'espressività del corpo e del viso. Al contrario, "as a HC culture, Polish culture relies more on nonverbal non-vocal coding. This difference may be observed when Poles speak English: the British complain that Poles seem to be little involved in their speech and their voice is monotonous and flat" (Biel, 2004, 123).

4. *La ricerca*

4.1 I partecipanti

La ricerca ha coinvolto un gruppo di tre italiani nativi (1M, 2 F) e tre gruppi di apprendenti, rispettivamente di lingua russa (2M, 1 F), provenienti da Kazan e Mosca, spagnola (2M, 1F) provenienti da La Coruña, Gran Canaria, Madrid, e tunisina (1M, 2F) provenienti da Tunisi. Si tratta nello specifico di studenti Internazionali frequentanti l'Università della Calabria aventi un livello di competenza compreso tra B1 (russi) e B2 (spagnoli e tunisini), e un'età compresa tra 20 e 25 anni.

Ai fini dell'indagine, i gruppi sperimentali considerati sono quindi quattro per le lingue native, rispettivamente: italiani nativi (It-L1), russi (R-L1), spagnoli (S-L1), tunisini (T-L1), tre per l'it-L2, ovvero: apprendenti russi (R-it-L2), apprendenti spagnoli (S-it-L2) e apprendenti tunisini (T-it-L2).

4.2 Obiettivi

Per quanto sia difficile isolare, all'interno della comunicazione, ciò che è emotivo da ciò che non lo è, la ricerca si propone di esaminare l'espressione vocale di tre emozioni primarie, per la precisione collera, gioia e tristezza, prodotte da italiani nativi e da apprendenti di italiano L2. L'ipotesi di partenza è che le differenze nell'espressione delle emozioni da parte dei soggetti coinvolti siano riconducibili alla loro L1 e cultura di origine. In tal senso il ruolo del *transfer* agirebbe come filtro sulla struttura prosodica dei parlanti.

Per quanto riguarda gli aspetti caratteristici delle culture dei non nativi e, in riferimento al modello di Hall sopra citato (§ 3.2), le culture esaminate dovrebbero rispecchiare le tendenze che le caratterizzano come culture ad alto contesto o culture a basso contesto. Per questo motivo nel parlato emotivo degli apprendenti russi, similmente agli apprendenti polacchi, il contenuto paraverbale del messaggio dovrebbe incidere poco sulla trasmissione dell'informazione, rivestendo un ruolo secondario. Lo stesso vale per il gruppo dei tunisini che si affidano poco alla prosodia e molto alla mimica

facciale e a gesti. Quando discutono, i tunisini fanno largo uso della comunicazione non verbale e tendono a esprimere i loro sentimenti privilegiando la mimica facciale (Nishimura, Nevgi & Tella, 2009).

La produzione emotiva del parlante nativo dovrebbe presentare invece dei tratti più espliciti ed enfatizzati, posizionandosi nella parte centrale della scala di Hall. Nella cultura italiana, infatti, il significato della comunicazione è dato, quasi in egual misura, dal contesto e dal codice. Pertanto il messaggio viene veicolato sia dall'aspetto paraverbale, sia da quello non verbale (gesti, espressioni del viso, posizione del corpo ecc.). A ciò si deve dunque la generale vivacità che caratterizza il parlato emotivo degli italiani, spesso fraintesa dalle altre culture. Allo stesso modo gli apprendenti spagnoli che si posizionerebbero centralmente nella scala di Hall, dovrebbero esibire uno stile comunicativo più vicino a quello dei parlanti italiani e fare ampio uso sia del canale non verbale che di quello del paraverbale.

Nel nostro studio, la scelta delle emozioni non è stata casuale, essendo differenziate su più assi. Secondo il modello multidimensionale (Scherer, 2003), almeno in termini generali, un alto grado di attivazione contrappone la collera e la gioia alla tristezza. Gioia e collera a loro volta sono però caratterizzate da un diverso grado di valutazione, essendo positiva la prima, negativa la seconda. La collera, inoltre, emozione a elevato controllo, si distingue dalla tristezza, che si caratterizza invece per un basso controllo[4].

Le tre emozioni indagate sono state affiancate dalla produzione dell'eloquio neutro, assunto come riferimento. Per eloquio neutro intendiamo un parlato privo di sfumature emozionali di rilievo, che non sia caratterizzato da particolari valori elevati o ridotti relativi all'intensità o ad andamenti elocutivi particolarmente rapidi. Siamo tuttavia consapevoli dell'elusività intrinseca delle emozioni e della non trascurabile limitatezza della loro categorizzazione e riconosciamo che alcune emozioni siano meno ambigue rispetto ad altre, sia nella loro caratterizzazione acustico-prosodica, sia nella realizzazione minico-facciale (Matsumoto, 2009).

Dal punto di vista metodologico è stato adottato un approccio integrato che ha considerato sia i diversi stadi interlinguistici presenti sia le relative L1, al fine di comprendere quegli aspetti che non sono necessariamente dovuti a fenomeni di *transfer*, ma riconducibili a *pattern* di tendenza più universale.

4.3 Il materiale

Per l'elicitazione delle emozioni, ci siamo serviti del metodo della *simulazione*, un espediente impiegato già in altri studi (cfr. Anolli, Ciceri, 1992; Banse, Scherer, 1996; Anolli et al., 2008). Inizialmente abbiamo selezionato una frase bersaglio che si prestasse ad una realizzazione emotiva, per la precisione *hanno chiuso la scuola*. Con la collaborazione degli apprendenti, è stata scelta l'equivalente frase in spagnolo (*han cerrado*

[4] Beninteso, le possibilità combinatorie delle dimensioni psicologiche sono naturalmente maggiori rispetto a quanto da noi semplificato, e ricoprono tutte le sfumature espressive richieste da precipui contesti comunicativi. Di conseguenza, anche una stessa emozione può essere prodotta con gradi di coinvolgimento o di controllo diversi; motivo per cui ad es. la collera può avere un alto o anche un basso grado di attivazione.

la escuela), in russo (*zakryli školu*) e in tunisino (*sakru al madrasa*). Successivamente, sono stati ideati degli scenari dialogici tesi a riprodurre i diversi contesti emozionali al cui interno era presente la frase bersaglio. La frase è stata scelta poiché facilmente riproducibile nei contesti emotivi esplorati. La sua adattabilità contestuale ha ridotto le implicazioni sul piano semantico, permettendo una più agevole analisi delle differenze acustiche e prosodiche presenti negli enunciati. Per facilitare il compito della simulazione e della identificazione del contesto emotivo, gli apprendenti hanno ricevuto l'indicazione di provare a veicolare lo stato emotivo anche attraverso l'ausilio dell'espressione facciale (su questo si veda Laird, Strout, 2007). I parlanti non nativi hanno realizzato gli enunciati emotivi anche nella loro L1, al fine di consentire il confronto tra le produzioni in L1 e L2.

Le registrazioni digitali sono state effettuate presso l'Università della Calabria mediante un registratore professionale (*Zoom Recorder H4N*, formato .wav, 44000 Hz, 32 bit) e un microfono direzionale (*Dynamic supercardioid Audio Technica N/D767A*).

4.3 L'analisi spettro-acustica

Il materiale raccolto è stato oggetto di analisi acustica. Sono stati analizzati sia descrittori generali come la Velocità di Eloquio (da ora VE), calcolata partendo dalla durata totale dell'enunciato bersaglio, e l'intensità media (dB_{mean}), sia descrittori intonativi, tra cui il rilievo del valore iniziale (*onset*), finale (*offset*) e medio ($f0_{mean}$) di *f0*, l'Escursione Melodica dell'enunciato (da ora EM) calcolata in semitoni, ST. Per ogni parametro sono stati computati la media aritmetica (X) e la deviazione standard (DS). Al fine di neutralizzare le differenze tra locutori, i dati relativi ai descrittori globali sono stati normalizzati secondo la formula riportata in (1), uniformandoci così alla metodologia già impiegata in altre ricerche[5]:

(1) Valore normalizzato= (x-N)/N

in cui x è il valore assoluto di ogni parametro indagato, mentre N è il valore assoluto rilevato di volta in volta per il tipo neutro, la frase assunta quale riferimento. La formula produce un valore relativo, positivo o negativo, a seconda della differenza che si stabilisce in rapporto alla realizzazione neutra.

L'analisi acustica è stata condotta mediante il *software Praat*.

5. *I risultati: le lingue native*

Consideriamo innanzitutto ciò che avviene nelle lingue native, un passaggio, quest'ultimo, fondamentale per poter individuare gli aspetti divergenti, potenziali obiettivi di fenomeni di *transfer*[6]. A tal scopo abbiamo confrontato le diverse L1 per verificare se la

[5] Cfr. Anolli et al. (2008); Wang, Yong-Cheol & Ma (2016).

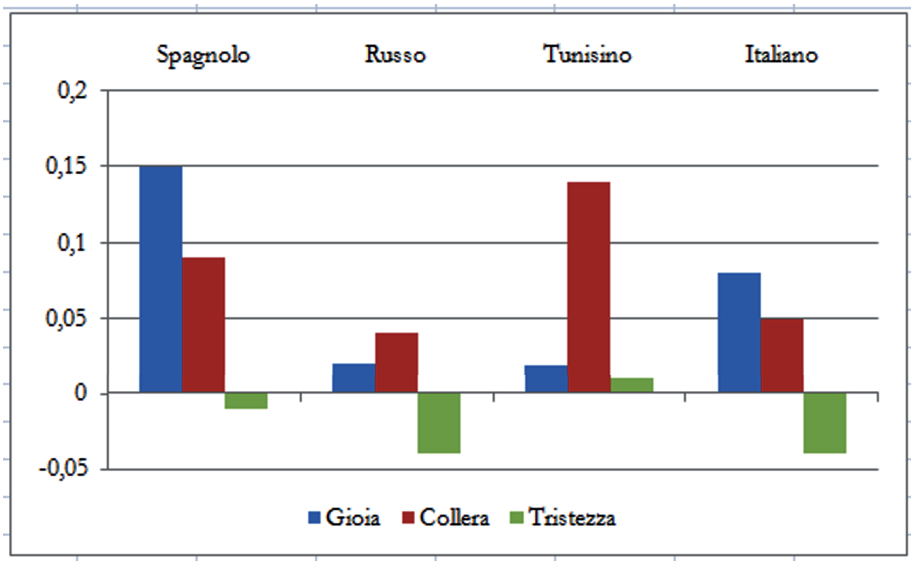
[6] La realizzazione prosodica delle emozioni è stata più volte argomento di studio in spagnolo (tra gli altri, Martinez-Castilla, Peppé, 2008; Rodero, 2011) e in russo, lingua in cui è disponibile il *database*

realizzazione degli enunciati emotivi mostrasse un condizionamento linguo-specifico. I risultati dell'analisi confermano alcune tendenze già presenti nella letteratura sul parlato emotivo, aggiungendo anche ulteriori informazioni sulla costituenza prosodica.

Le emozioni selezionate sono differenziate sia sul piano temporale che su quello intonativo. Rispetto alla frase neutra, gli enunciati emotivi manifestano una serie di peculiarità, alcune più stabili, altre suscettibili di maggiore libertà di variazione.

Il primo parametro che discuteremo concerne l'intensità. In tutte le lingue native esaminate si osserva una polarizzazione di quest'indice che raggiunge valori elevati in gioia e collera, bassi nella tristezza[7]. Dai valori normalizzati (Fig. 1), osserviamo che gioia e collera nello spagnolo, e limitatamente alla collera anche nel tunisino, si distanziano maggiormente dal valore neutro, il rapporto è positivo e ciò vuol dire che la loro intensità media è maggiore rispetto al punto neutro di riferimento. La stessa tendenza emerge, sebbene in modo più attenuato, in italiano e in russo. Significativamente, la tristezza si attesta su un valore negativo, ad eccezione del tunisino, essendo realizzata con un grado di energia inferiore.

Figura 1 - *Valori normalizzati dell'intensità media nelle lingue native*



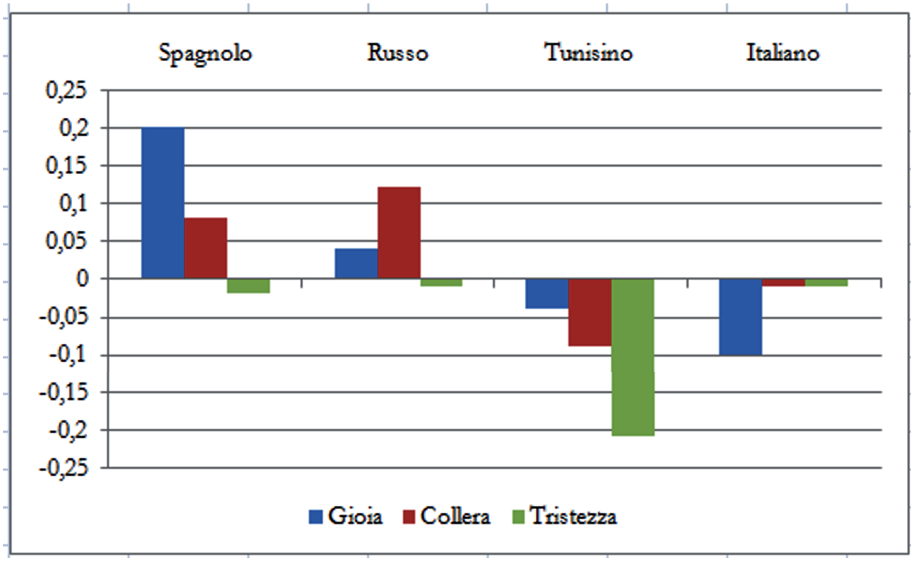
Per quanto riguarda la Velocità di Eloquio (Fig. 2), il quadro risultante è eterogeneo, in quanto i gruppi si comportano in modo diverso. In R-L1 e in S-L1, gioia e collera sono infatti articolate in modo più veloce rispetto al neutro. Nell'italiano

RUSLANA (*Russian Language Affective*, Makarova, Petrushin, 2012). Esigui sono invece gli studi disponibili su questa tematica per l'arabo tunisino (cfr. Maalej, 2007).

[7] L'intensità è notoriamente il parametro acustico più instabile, poiché influenzato da una serie di condizionamenti esterni, tra cui anche la distanza dalla fonte di registrazione, ed è pertanto da valutare con una certa cautela. Nella nostra ricerca, una buona omogeneità dei risultati sembra confermare il suo carattere discriminante.

nativo, solo la gioia si distingue dagli altri tipi emotivi, essendo realizzata in modo più rallentato rispetto al neutro (5,4 sill/sec); collera e tristezza mostrano invece una velocità simile a quella del neutro, facendo registrare un rapporto pari a 0. In tunisino, infine, la velocità delle tre emozioni è sempre minore rispetto a quella del neutro, soprattutto la tristezza.

Figura 2 - *Velocità Elocutiva normalizzata nelle quattro lingue native*



5.1 Le lingue native. Il profilo intonativo

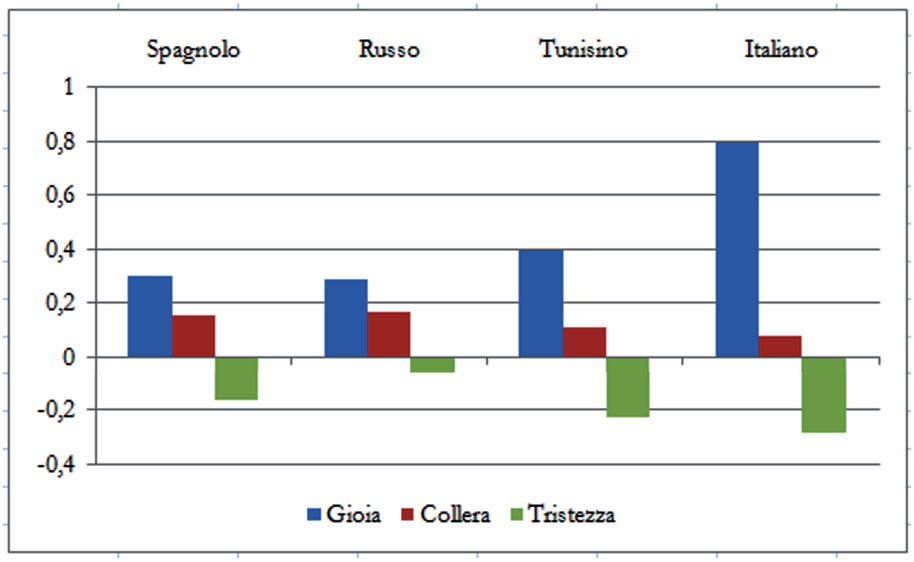
Più informativo è invece il dato inerente all'escursione tonale (Fig. 3), anche perché tutte le lingue seguono lo stesso *trend*.

Dai rapporti normalizzati si deduce che collera e soprattutto gioia sfruttano un'EM maggiore rispetto al tipo neutro. Questo comportamento è particolarmente evidente in it-L1 in cui la gioia mostra l'EM più elevata, in termini assoluti 13 ST. In modo diametralmente opposto, ma in linea con le nostre attese, nelle lingue considerate, la tristezza manifesta un indice negativo, segno della presenza di un'escursione più ristretta, rispetto al neutro, ma anche alle altre emozioni. In termini assoluti, rileviamo che l'EM della tristezza è compresa tra 5 e 6,6 ST, uno spazio melodico praticamente dimezzato se confrontato a quello della gioia.

Al fine di pervenire a un quadro più dettagliato, abbiamo computato anche la differenza in semitoni tra la *f0* media di ogni emozione e rispettivamente il valore iniziale (*onset*) e finale (*offset*) della curva intonativa dell'enunciato. Nell'it-L1 ancora una volta è la gioia a mostrare maggiori divergenze: il valore iniziale della curva intonativa si colloca infatti 3 ST (DS: 1,2) più in alto rispetto alla $f0_{mean}$ dello stesso tipo emotivo, segue la collera con 1,6 ST (DS: 1,3), mentre la tristezza ha un *onset* di poco inferiore alla media frequenziale (0,5; DS: 0,5). In tutti gli enunciati esaminati, in modo atteso, l'*offset* è più basso della *f0* media, tuttavia questa differenza è più

cospicua per collera (6,8 ST; DS: 3) e gioia (4,6 ST; DS: 2,4). Questo vuol dire che in questi tipi emotivi l'intero contorno è posizionato su un livello frequenziale più alto tanto da innalzare in modo significativo il valore di $f0_{mean}$[8]. Questo dato fornisce anche una risposta alla ridotta escursione melodica riscontrata per la collera; questa emozione ha un *range* ridotto poiché interamente collocata su alte frequenze.

Figura 3 - *Escursione Melodica normalizzata nelle quattro lingue native*



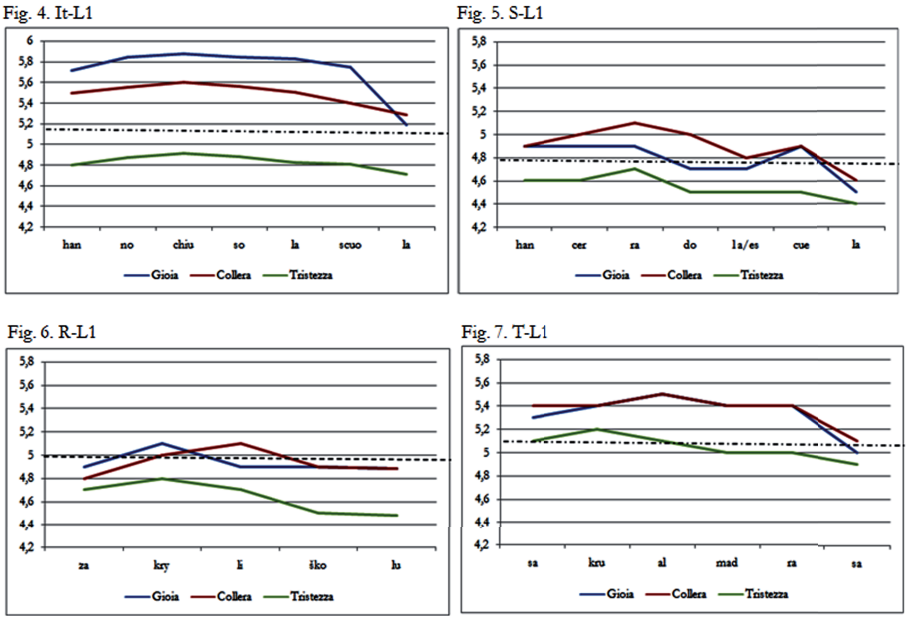
Anche nelle altre lingue native (R-L1, S-L1, T-L1), il contorno intonativo di collera e gioia si posiziona nella regione frequenziale più elevata. Queste emozioni, sebbene non sempre prodotte con una maggiore escursione tonale, mostrano una distintività melodica; il valore iniziale rilevato per la gioia e la collera è infatti più alto rispetto a quello della tristezza (già De Marco, Sorianello & Paone, in stampa). Ciò è particolarmente evidente nei russi e negli spagnoli, e in modo netto per la gioia. Il contorno melodico di gioia e collera inoltre non perviene a un valore minimo finale, l'*offset* è infatti maggiore rispetto alla $f0_{mean}$, ciò è palese nei russi e, con riferimento alla sola collera, negli spagnoli e nei tunisini.

Le dinamiche frequenziali finora discusse forniscono utili informazioni sulla realizzazione prosodica delle emozioni. Tuttavia, riteniamo che l'osservazione globale dei contorni possa fornire ulteriori aspetti di rilievo. A tale scopo, abbiamo ispezionato le curve melodiche nella loro interezza. Poiché i campioni delle lingue esaminate sono formati sia da femmine che da maschi, tutti i dati frequenziali sono stati convertiti in logaritmi naturali (*nlog*), al fine di neutralizzare le differenze imputabili al genere. Nelle Figure 4-7 sono illustrati i contorni intonativi delle frasi bersaglio

[8] Il posizionamento del contorno finale della gioia su un valore frequenziale intermedio è stato già rilevato per l'italiano da Anolli, Ciceri (1992).

prodotte nelle quattro lingue; in questo caso abbiamo ritenuto utile sovrapporre ai grafici il valore della $f0_{mean}$ ottenuto per la frase neutra (linea tratteggiata). La stilizzazione della curva è avvenuta riportando in ordinata l'asse frequenziale espresso in *nlog* e assegnando ad ogni sillaba dell'enunciato il valore medio di *f0* estratto in corrispondenza della sillaba medesima, in ascissa è invece riportato il tempo riprodotto attraverso la sequenza delle sillabe che compongono l'enunciato. La sovrapposizione dei contorni consente di evidenziare le differenze tra emozioni, sia all'interno dello stesso gruppo, sia tra gruppi.

Figure 4-7 - *Contorni intonativi stilizzati in nlog relativi alle lingue native, la linea tratteggiata indica* $f0_{mean}$ *del tipo neutro*



Complessivamente, da questo confronto emerge che nella produzione delle emozioni ci sono delle somiglianze che travalicano le differenze interlinguistiche; una brusca discesa di *f0* accomuna il contorno finale della gioia in italiano, spagnolo e tunisino, mentre il contorno della collera, pur essendo discendente finale, mostra un valore finale prossimo o superiore a quello del neutro. La tristezza esibisce un profilo intonativo sempre collocato entro un *range* basso. Va altresì aggiunto, poiché non desumibile dai grafici, che la durata dell'ultima vocale tonica, sede di accento intonativo nucleare, è in tutte le lingue significativamente più lunga rispetto alla stessa rilevata per la frase neutra. Sulla porzione finale dell'enunciato si manifestano importanti variazioni intonative che vanno oltre la mera forma del contorno. Ad

esempio, tutte le emozioni sono tendenzialmente discendenti finali, ma la realizzazione di questo movimento è parzialmente diversa, per allineamento e *scaling*[9].

In italiano buona parte del contorno intonativo della gioia è alto e sostenuto, privo di variazioni rilevanti, se non nella parte finale. È significativo notare che l'$f0_{mean}$ del neutro si prefigura come un valido *discrimen* tra le emozioni, separando nettamente la tristezza da gioia e collera. Nelle altre lingue, il quadro è meno nitido; in russo infatti l'$f0_{mean}$ interseca la realizzazione di gioia e collera, pur non sovrapponendosi, mentre in spagnolo collera e gioia si pongono al di sopra di essa, in modo più evidente nella parte iniziale. In controtendenza, i tunisini in cui la tristezza scivola al di sotto del valore medio di *f0* solo nella parte finale.

Nel complesso, dunque, ci sono degli indici più robusti, poiché stabilmente presenti nelle quattro lingue native, fra tutti l'EM o l'intensità, e altri che sembrano attivarsi in modo più libero, è il caso della Velocità di Eloquio[10]. Al termine di questo primo confronto, sembra emergere un interessante filo conduttore. In tutte le lingue native considerate, pur con intervalli di variazione differenti, l'espressione di gioia e collera è caratterizzata dalla presenza di una modalità che possiamo definire di *uptrend*, che innalza il registro di *f0*, l'$f0_{mean}$, il valore d'attacco del contorno e l'intensità. In modo speculare, l'espressione della tristezza mostra una modalità *downtrend* in cui l'abbassamento del registro di *f0* e la restrizione dello spazio melodico globale avvengono in modo congiunto, mentre l'intensità è mediamente bassa.

6. *Risultati: l'italiano L2*

Da quando emerso finora, ci aspettiamo che gli apprendenti quando si esprimono in it-L2 mantengano nel loro eloquio emotivo quegli aspetti sistematicamente presenti nelle lingue native.

Il quadro descrittivo risultante non è però sempre trasparente né omogeneo. In ogni gruppo di apprendenti emergono dei tratti atipici, per molti versi inattesi. In merito alla Velocità Elocutiva (Fig. 8, riquadro superiore), l'italiano di russi e tunisini è nel complesso più lento rispetto alla loro lingua nativa, mentre negli spagnoli non si notano divergenze significative, ad eccezione della gioia più veloce nella lingua nativa. La presenza di una minore VE è d'altra parte un dato che accomuna i primi stadi di apprendimento di una lingua straniera (tra gli altri, per l'italiano Maffia et al., 2014; De Marco et al., in stampa). Tuttavia, si osserva che negli apprendenti russi e tunisini la VE delle emozioni è decisamente inferiore rispetto a quella del tipo neutro; su questo punto torneremo in seguito.

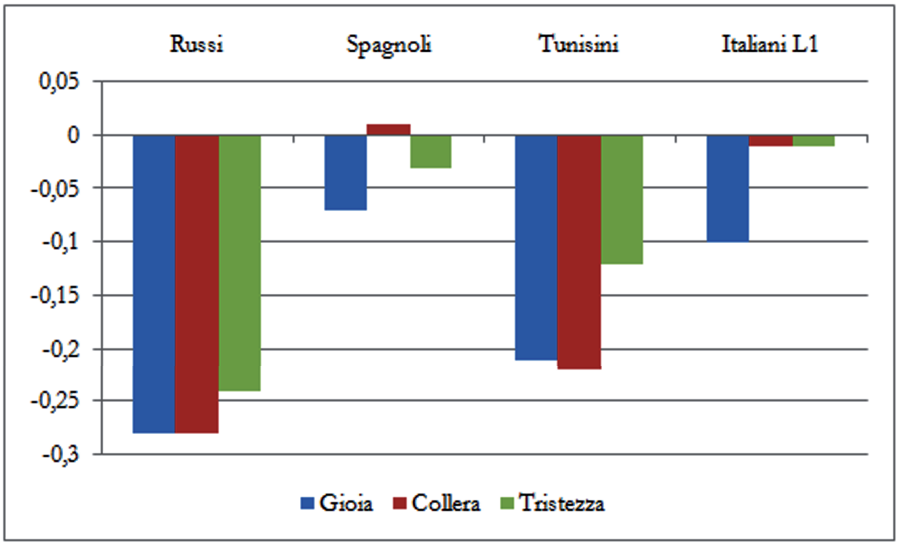
Relativamente all'Escursione Melodica (Fig. 8, riquadro inferiore), si nota invece come tale indice sia minore a un confronto tra le interlingue e le lingue native

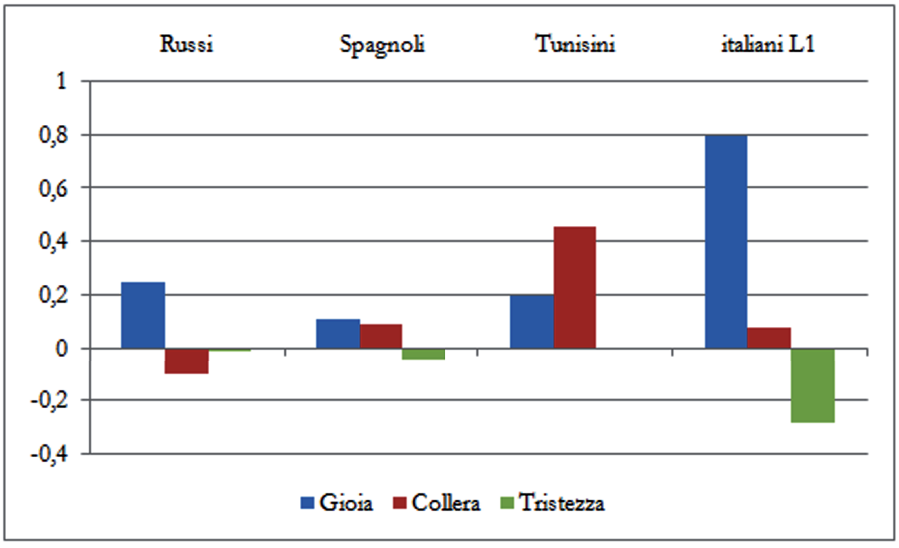
[9] Questi ultimi aspetti meritano un approfondimento specifico e pertanto non saranno discussi in questa sede.

[10] A questo proposito, per il russo nativo Makarova, Petrushin (2012) non trovano differenze significative tra i tipi emotivi, collera e gioia sono di norma prodotte più lentamente del 20%, ma osservano che ci sono modi diversi per esprimere la stessa emozione, soprattutto la paura.

degli apprendenti. Rispetto alla frase neutra, inoltre, si rilevano tendenze divergenti; ad esempio nei russi la collera ha un'EM inferiore al neutro, un aspetto non conforme né a quanto rilevato nella loro lingua nativa, né a quello della lingua *target*, cioè l'It-L1. Anche il comportamento della tristezza è atipico: nell'italiano di russi, spagnoli e tunisini, il parametro mostra una escursione frequenziale approssimabile a quella del neutro, privo dell'abbassamento di registro di *f0*.

Figura 8 - *Velocità Elocutiva (sopra) ed Escursione Melodica (sotto) normalizzate relative all'It-L2*





In realtà, la diversa escursione melodica è l'aspetto più macroscopico di una alterata realizzazione intonativa delle emozioni. Lo scenario si arricchisce infatti di nuovi particolari, se analizziamo i contorni intonativi nella loro interezza; si veda in merito quanto riprodotto nelle Figg-9-11.

Figure 9-11 - *Contorni intonativi stilizzati in nlog relativi all'It-L2 prodotti rispettivamente da apprendenti spagnoli, russi e tunisini. La linea tratteggiata indica f0$_{mean}$ della frase neutra*

Fig. 9 - S-It-L2

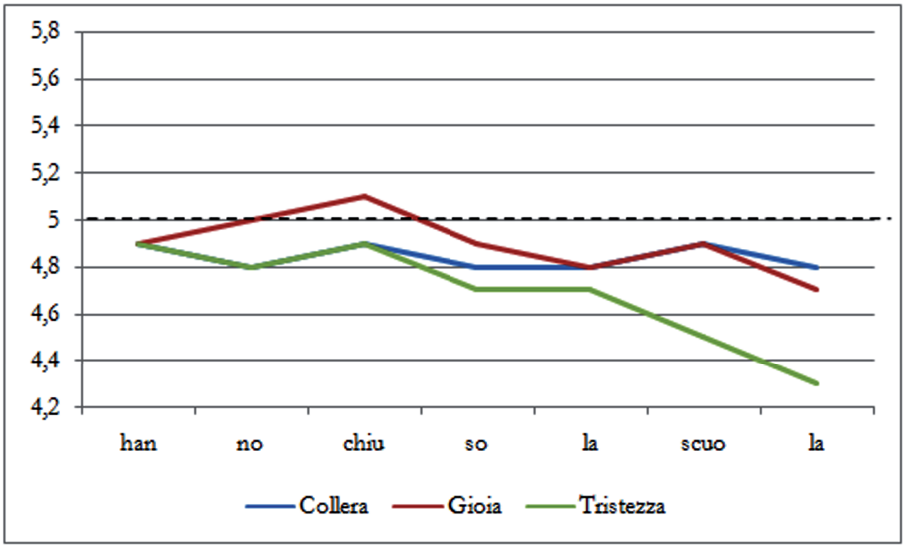


Fig. 10 - R-It-L2

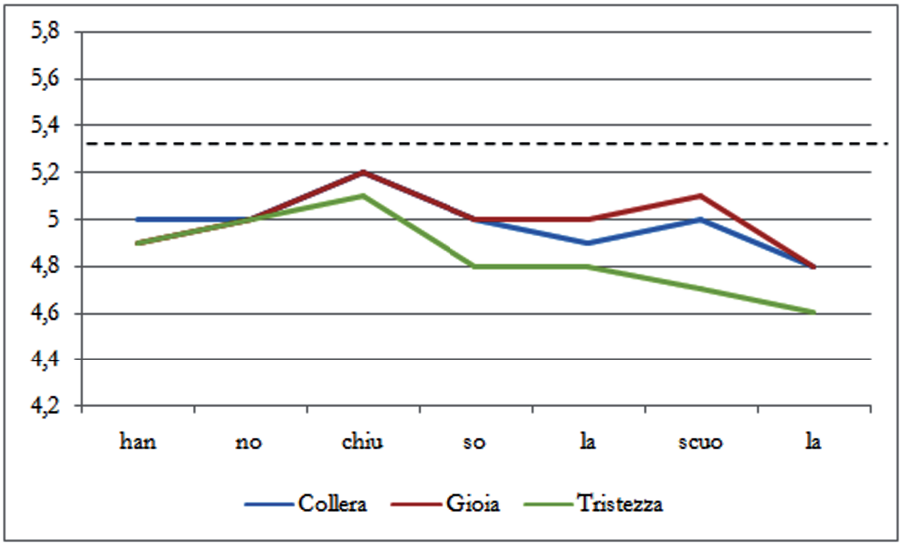
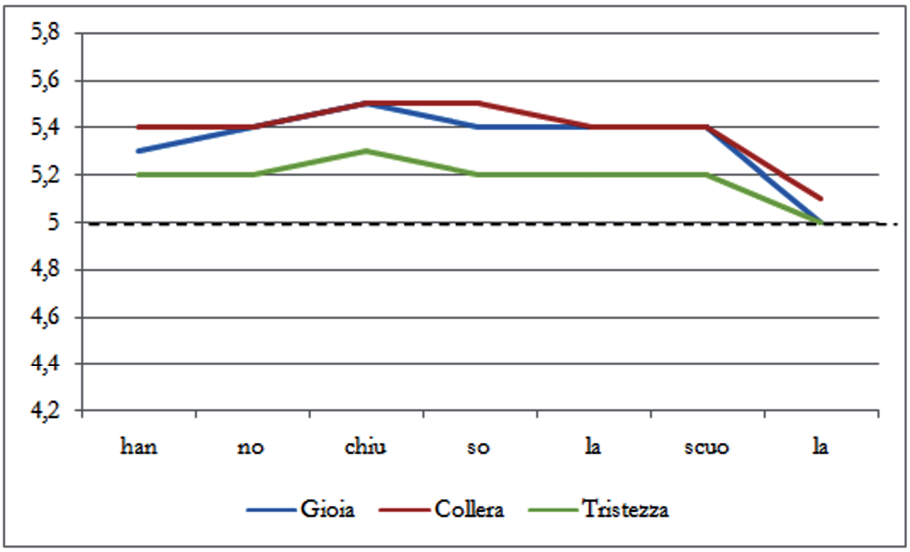


Fig. 11 - T-It-L2



Innanzitutto, osserviamo che la produzione dell'enunciato neutro, evidenziato nei grafici come linea tratteggiata, solleva alcune incertezze, essendo collocato negli apprendenti russi e spagnoli nel registro frequenziale più alto, al di sopra degli enunciati emotivi. In modo speculare, nell'italiano dei tunisini, $f0_{mean}$ del neutro si pone al di sotto degli altri tipi emotivi. È probabile che in buona parte degli apprendenti la produzione di una dichiarativa priva di sfumature emotive, in una lingua non nativa, sia stato un compito, se non difficile, quanto meno poco chiaro. Complessivamente, collera e gioia sono realizzate in modo adeguato, almeno nei tunisini; in questi ultimi le due emozioni conservano la presenza di un *onset* elevato, già rilevato nella lingua nativa, laddove nei russi (Fig. 10) si nota, per queste due emozioni, un *offset* avente una frequenza più alta rispetto alla tristezza. Un caso particolare riguarda gli apprendenti russi e spagnoli (Figg. 9 e 10) nei quali i contorni sono per buona parte sovrapposti e piuttosto monotoni, perdendo parte della loro distintività uditiva. Il contorno della tristezza acquisisce una certa autonomia, ma solo nella parte finale.

Non ci stupisce notare come gli apprendenti tunisini nel produrre gli enunciati emotivi in italiano riproducano contorni molto simili a quelli della loro lingua nativa, si confronti in merito quanto rappresentato nelle Figure 7 e 11. Russi e spagnoli manifestano invece delle forme ibride, a riprova della loro incertezza nel controllare schemi prosodici di una lingua diversa da quella nativa.

7. *Conclusioni*

Il presente studio ha inteso indagare l'espressione delle emozioni vocali in italiano da parte di apprendenti di diversa origine e si è proposto altresì di verificare in che misura la distanza culturale (e linguistica) dalla lingua *target* incidesse su questo

aspetto della comunicazione emotiva. In particolare abbiamo voluto esaminare gli effetti del *transfer* sui correlati acustici delle tre emozioni indagate.

Il quadro delineato dai risultati si presenta decisamente eterogeneo. La produzione in it-L2 è per certi versi confusa e questo denota senz'altro la difficoltà cui va incontro l'apprendente quando realizza in una lingua diversa un fenomeno paralinguistico complesso come quello delle emozioni. In questa situazione sperimentale (e artificiale), l'apprendente manifesta una sorta di incertezza che lo porterà a una sorta di mescidanza dei parametri e conseguentemente alla produzione di emozioni molto spesso non congruenti sul piano pragmatico a realizzare pienamente ciò che un parlante nativo italiano si attenderebbe in un certo contesto. Un limite della nostra ricerca, strettamente connesso a questa prima riflessione conclusiva, è quello di aver tralasciato la decodifica percettiva delle produzioni degli apprendenti da parte dei parlanti nativi. Questa analisi ci avrebbe permesso di poter confermare con una verifica più accurata quanto avviene sul piano della codifica delle emozioni. La dimensione ristretta del nostro campione, per ora limitata a soli tre apprendenti per gruppo, non ha inoltre consentito l'elaborazione statistica dei risultati, una procedura che ci avrebbe pemesso di verificare la significatività delle tendenze emerse, eliminando il rischio di eventuali condizionamenti aleatori.

Ciò che sembra emergere a livello di interferenza dalla L1 è piuttosto una strategia di elaborazione utilizzata dagli apprendenti per capire il funzionamento della L2 (Giacalone Ramat, 2003). Le strutture simili o avvertite tali dagli apprendenti non sono necessariamente causa di *transfer* da L1. In questa direzione sembra porsi, ad esempio, il comportamento degli spagnoli i quali non ripropongono gli schemi prosodici né della loro lingua nativa, né quelli della L2. Gli apprendenti russi invece modulano poco le diverse emozioni, rimanendo nel complesso lontani dal *target* dell'it-L1.

Per quanto la distanza culturale e linguistica possa giocare un ruolo importante nella codifica delle emozioni vocali, non è possibile affermare, in base ai risultati ottenuti, che ciò avvenga sempre. Ciascun gruppo di apprendenti si è confrontato in misura diversa con dubbi e incertezze, a prescindere dalla L1. Da un lato, la distanza dalla lingua *target* non sempre è stata determinante in tal senso, dall'altro, la maggiore vicinanza non sembra aver costituito un fattore in grado di agevolare la gestione del parlato emotivo in L2. Rispetto alla chiave interpretativa offerta dal modello di Hall che abbiamo presentato sopra, non sembra esserci una piena congruenza con le ipotesi relative alla caratterizzazione dei comportamenti attribuiti al grado di assimilazione alle culture di basso o alto contesto comunicativo. Sebbene alcuni parametri, specialmente in relazione agli apprendenti tunisini, sembrino confermare una certa congruenza con le modalità espressive tipiche delle culture ad alto contesto parzialmente riflesse anche nella produzione di L2, non ci sembra di poter affermare con la stessa chiarezza una corrispondenza di questo tipo per gli apprendenti russi e spagnoli. Una verifica percettiva delle abilità di decodifica degli apprendenti potrebbe chiarire se e in che misura le difficoltà sul piano della produzione si riflettano anche su quello della percezione e se la distanza culturale giochi un ruolo

in tal senso. Infine, una validazione percettiva delle produzioni da parte dei nativi, insieme a un numero più ampio di parlanti, può restituirci dei risultati più chiari, statisticamente interpretabili, sui quali poter avanzare delle osservazioni puntuali circa l'efficacia comunicativa del parlato emotivo in L2.

Riferimenti bibliografici

Altrov, R. (2013). Aspects of cultural communication in recognizing emotions. In *Trames. Journal of the Humanities and Social Sciences*, 17(2), 159. http://doi.org/10.3176/tr.2013.2.0/Accessed 29.11.2016.

Anolli, L., Ciceri, R. (1992). *La voce delle emozioni. Verso una semiosi della comunicazione vocale non-verbale delle emozioni*. Milano: FrancoAngeli.

Anolli, L., Wang, L., Mantovani, F. & De Toni, A. (2008). The voice of emotion in Chinese and Italian young adults. In *Journal of Cross-Cultural Psychology*, 39, 565-598.

Banse, R., Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. In *Journal of Personality and Social Psychology*, 70(3), 614-636. http://doi.org/10.1037/0022-3514.70.3.614/Accessed 29.11.2016.

Bänziger, T., Scherer, K.R. (2005). The role of intonation in emotional expressions. In *Speech Communication*, 46, 252-267.

Biel, L. (2004). *Emotional and social distance in English and Polish*. University of Gdańsk.

Cao, H., Beňuš, Š., Gur, R.C., Verna, R. & Nenkova, A. (2014). Prosodic cues for emotion: analysis with discrete characterization of intonation. In Campbell, N., Gibbon, D. & Hirst, D. (Eds.), *Proceedings of 7th International Conference on Speech Prosody*, Dublin, 130-134.

De Marco, A., Paone, E. (2014). L'espressione e la percezione delle emozioni vocali in apprendenti di Italiano L2: uno studio cross-linguistico. In *Educazione Linguistica, Language Education*, 9, 483-500.

De Marco, A., Paone, E. (2015). The acquisition of emotional competence in L2 learners of Italian. In Gesuato, S., Bianchi, F. & Cheng, W. (Eds.), *Teaching, Learning and Investigating Pragmatics: Principles Methods and Practices*. Cambridge: Cambridge Scholar Publishing, 441-468.

De Marco, A., Paone, E. (2016). Uno studio sui correlati acustici delle emozioni vocali in apprendenti di italiano L2. In Elia, A., Iacobini, C. & Voghera, M. (Eds.), *Livelli di analisi e fenomeni di interfaccia. Atti del XLVII Congresso Internazionale SLI 2013*. Roma: Bulzoni, 75-93.

De Marco, A., Sorianello, P. & Paone, E. (in stampa). L'acquisizione delle emozioni nell'italiano non nativo. Un percorso didattico longitudinale. In Gudmundson, A., Alvarez, L. & Bardel, C. (Eds.), *Multilingualism and language acquisition. Spoken Romance languages*. Frankfurt am Main: Peter Lang Publishing Group.

Dewaele, J.M. (2005). Investigating the psychologycal ad emotional dimensions in instructed language learning: obstacles ad possibilities. *The Modern Language Journal*, 89(3), 367-380.

Galatà, V. (2010). Produzione e percezione di emozioni vocali: uno studio cross-linguistico-culturale europeo. Tesi di Dottorato non pubblicata, Università della Calabria, Cosenza.

Giacalone Ramat, A. (Ed.) (2003). *Percorsi e strategie di acquisizione*. Roma: Carocci.

Gili Fivela, B., Grimaldi, M. & Stefàno, M. (2004). Emozioni a teatro: voce e gesto nello spazio. In Magno Caldognetto, E., Cavicchio, F. & Cosi, P. (Eds.), Atti del Convegno *Comunicazione Parlata e Manifestazione delle Emozioni*. Liguori Multimedia: Napoli, 383-423.

Gobl, C., Ní- Chasaide, A. (2003). The role of voice quality in communication emotion, mood and attitude. In *Speech Communication*, 40, 189-212.

Goudbeek, M., Broersma, M. (2010). The Demo/Kemo corpus: A principled approach to the study of cross-cultural differences in the vocal expression and perception of emotion. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M. & Tapias, D. (Eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, 2211-2215.

Hall, E.T. (1976). *Beyond culture* (1st ed.). Garden City N.Y.: Anchor Press.

Hall, E.T., Hall, M.R. (1990). *Understanding cultural differences*. Yarmouth: Intercultural Press.

Juslin, P.N., Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. In *Psychological Bulletin*, 129, 770-814.

Kim, S., Dorner, L.M. (2013). "I won't talk about this here in America:" Sociocultural context of Korean English learners' emotion speech in English. In *L2 Journal*, 5(2), 43-67.

Komar, S. (2005). The impact of tones and pitch range on the expression of attitudes in Slovene speakers of English. In Naidment, J.A. (Ed.), *Proceedings of PTLC2013*. London: University College London, 1-4.

Ladd, D.R., Silverman, K., Tolkmitt, F., Bergmann, G. & Scherer, K.R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. In *Journal of the Acoustical Society of America*, 78, 435-444.

Laird, D.J., Strout, S. (2007). Emotional behaviours as emotional stimuli. In Coan, J.A., Allen, J.B. (Eds.), *Handbook of emotion elicitation and assessment*. Oxford: Oxford University Press, 54-64.

Liscombe, J.J. (2007). Prosody and speaker state: paralinguistic, pragmatics and proficiency. PhD Thesis, Columbia University. http://www.cs.columbia.edu/speech/ThesisFiles/Jackson_liscombe.pdf/Accessed 29.11.16.

Maalej, Z. (2007). The embodiment of fear expressions in Tunisian Arabic: Theoretical and practical implications. In Sharifian, F., Palmer, G.B. (Eds.), *Applied cultural linguistics: Implications for second language learning and intercultural communication*. Amsterdam-Philadelphia: John Benjamins Publishing Company, 87-104.

Maffia, M., Pellegrino, E. & Pettorino, M. (2014). Labeling expressive speech in L2 Italian: the role of prosody in auto- and external annotation. In Campbell, N., Gibbon, N. & Hirst, D. (Eds.), *Proceedings of the 7th International Conference on Speech Prosody*. Dublin: SproSIG, 81-84.

Magno Caldognetto, E. (2002). I correlati fonetici delle emozioni. In Bazzanella, C., Kobau, P. (Eds.), *Passioni, emozioni, affetti*. Milano: McGraw-Hill, 197-213.

Magno Caldognetto, E., Cavicchio, F. & Cosi, P. (Eds.) (2008). *Comunicazione parlata e manifestazione delle emozioni.* Napoli: Liguori Editore.

Makarova, V., Petrushin, V. (2012). Phonetic tracing emotions in Russian vowels. In Makarova, V. (Ed.), *Russian language studies in North America: a new perspective from theoretical and applied linguistics*. London-New York: Anthem Press, 3-41.

Martinez-Castilla, P., Peppé, S. (2008). Intonation features of the expression of emotions in Spanish: preliminary study for a prosody assessment procedure. In *Clinical Linguistics and Phonetics*, 22, 363-370.

Matsumoto, D. (2009). The origin of universal human emotions. Retrieved October 2, 2014. http://davidmatsumoto.com/content/NG%20Spain%20Article_2.pdf/Accessed 29.11.16.

Nishimura, S., Nevgi, A. & Tella, S. (2009). Communication style and cultural features in High/Low context communication cultures: A Case study of Finland, Japan and India. In Kallioniemi, A. (Ed.), *Renovating and developing subject didactics. Proceedings of a subject-didactic symposium in Helsinki*, University of Helsinki, 2 February 2008, Part 2, 783-796.

Pettorino, M. (2008). Inglese, italiano e giapponese: analisi dei correlati acustici delle emozioni nel parlato cinematografico. In Magno Caldognetto et al. (Eds.), *Comunicazione parlata e manifestazione delle emozioni.* Napoli: Liguori Editore, 45-57.

Pittam, J., Scherer, K.R. (1993). Vocal expression and communication of emotion. In Lewis, M., Haviland, J. (Eds.), *Handbook of Emotions*. New York: Guilford Press, 185-197.

Poggi, I., Magno Caldognetto, E. (2004). Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici. In Albano Leoni, F., Cutugno, F., Pettorino, M. & Savy, R. (Eds.), *L'Italiano parlato.* Napoli: D'Auria Editore, CD-Rom.

Rodero, E. (2011). Intonation and emotion of pitch levels and contour type on creating emotions. In *Journal of Voice*, 25(1), e25-e34.

Samovar, L.A., Porter, R.E. & McDaniel, E.R. (2007). *Communication between cultures*. Belmont, CA: Wadsworth-Thomson Learning.

Scherer, K.R. (1989). Vocal correlates of emotion. In Manstcad, A., Wagner, H. (Eds.), *Handbook of psychophysiology: Emotion and social behavior*. London: Wiley, 165-197.

Scherer, K.N. (2003). Vocal communication of emotions: a review of research paradigm. In *Speech Communication*, 40, 227-256.

Scherer, K.R., Feldstein, S., Bond, R.N. & Rosenthal, R. (1985). Vocal cues to deception: A comparative channel approach. In *Journal of Psycholinguistic Research*, 14, 409-425.

Scherer, K.R., Ladd, D.R. & Silverman, K. (1984). Vocal cues to speaker affect: testing two models. In *Journal of Acoustical Society of America*, 76, 1346-1356.

Wang, T., Yong-Cheol, L. & Ma, Q. (2016). An experimental study of emotional speech in Mandarin and English. In *Proceedings of the 8th International Conference on Speech Prosody*, Boston, MA, USA: Boston University, 430-434.

DOI: 10.17469/O2102AISV000010

PAOLA ZANCHI, MARIAPAOLA D'IMPERIO, LAURA ZAMPINI, MIRCO FASOLO

L'intonazione delle narrazioni di bambini ed adulti italiani: un'analisi all'interno dell'approccio autosegmentale metrico

This study aims to compare the intonation used in narratives by Italian 3-year-old children and adults. Indeed, just few studies investigated the prosodic characteristic of complex language, such as discourse or narration, in childhood (e.g., Redford, Dilley, Gamache & Wieland, 2012; De Ruiter, 2014). The intonation was transcribed using the ToBI system, cast within the Autosegmental-Metrical theory. The results show that 3-year-old children can use the same nuclear pitch accent than adults, but they do not produce rising boundary tones in the same measure than adults. These results are discussed considering the possible relationships between intonation and cognitive development.

Key words: Intonation, Preschoolers, Language Development, AM approach, ToBI transcription.

1. *Background teorico*

La prosodia riveste un ruolo importante nell'acquisizione del linguaggio. Mentre le caratteristiche prosodiche dell'input linguistico fornito ai bambini e le capacità precoci di discriminarle sono state diffusamente studiate (per una rassegna si veda ad esempio Morgan, Demuth, 2014), la prosodia delle produzioni linguistiche infantili non è stata altrettanto frequentemente oggetto di studio. I primi lavori sull'argomento hanno analizzato longitudinalmente la prosodia delle produzioni preverbali infantili basandosi sull'analisi del contorno intonativo, ossia considerando la configurazione globale e la direzione del *pitch* piuttosto che il dettaglio dei diversi target tonali. La maggior parte di questi studi si è inoltre concentrata sulle prime produzioni dei bambini, come i pianti, le vocalizzazioni, la lallazione e le prime parole (per una rassegna si veda ad esempio Snow, Balog, 2002), mentre solo pochi studi si sono occupati dei primi enunciati formati da più più parole. Fra questi, uno dei primi è il lavoro di D'Odorico e Carubbi (2003): le due autrici hanno analizzato longitudinalmente la realizzazione prosodica degli enunciati delle prime combinazioni di parole (parole contenuto) prodotte da 24 bambini italiani di età compresa fra i 19 e i 24 mesi. I risultati hanno messo in evidenza come la complessità del legame intercorrente fra le due parole costituenti le prime combinazioni influenzi la realizzazione prosodica degli stessi; nello specifico, le combinazioni di parole vere e proprie, nelle quali le parole prodotte costituiscono un singolo atto linguistico, vengono prodotte con pattern intonativi più evoluti. Behrens e Gut (2005), studiando gli enunciati di due parole prodotti da un bambino tedesco osservato periodicamente nel periodo compreso fra i 24 e i 25 mesi, hanno almeno in

parte confermato questi risultati: tipi diversi di enunciati di due parole andavano incontro a sviluppi prosodici diversi, diventando prosodicamente "fluenti" in diversi momenti dello sviluppo; inoltre, la variabilità delle caratteristiche prosodiche (come la presenza e la lunghezza delle pause e il posizionamento dell'accento) nei primi enunciati di due parole risultava essere particolarmente elevata. D'Odorico e collaboratori (D'Odorico, Fasolo & Zanchi, 2010), analizzando acusticamente i primi enunciati con più argomenti prodotti da bambini italiani nel terzo anno di vita, hanno rilevato come la complessità sintattica influenzi la capacità del bambino di unire gli elementi del suo enunciato all'interno di un unico contorno intonativo. Il legame tra prosodia ed altre competenze linguistiche, come quella di realizzare frasi sintatticamente complesse oppure quella di raccontare delle storie, ha trovato conferma anche in recenti lavori condotti su bambini dallo sviluppo tipico (Zanchi, Zampini, Fasolo & D'Odorico, 2016) e atipico (Zampini, Fasolo, Spinelli, Zanchi, Suttora & Salerni, 2016).

Gli studi sull'intonazione degli enunciati di più parole prodotti dai bambini nelle prime fasi del loro sviluppo sintattico precedentemente descritti sono stati condotti attraverso un'analisi fonetica di tipo acustico (i.e., valori della frequenza fondamentale, durate). Tale strategia di analisi, tuttavia, li rende difficilmente confrontabili con i dati relativi alla prosodia degli adulti, generalmente descritta seguendo le convenzioni della teoria autosegmentale metrica, AM (Pierrehumbert, 1980; Ladd, 1996), ed utilizzando il sistema di trascrizione ToBI (*Tones and Breaks Indecies*) da esso derivato, per il quale un'analisi fonologica dell'intonazione dell'enunciato deve precedere quella acustica (e non viceversa). L'approccio AM è stato utilizzato nello studio di diverse lingue (fra le quali l'italiano; si vedano ad esempio, D'Imperio, 2002; Gili Fivela, Avesani, Barone, Bocci, Crocco, D'Imperio, Giornado, Marotta, Savino & Sorianello, 2015; Grice, D'Imperio, Savino & Avesani, 2005) ed è nel tempo diventato il più frequentemente impiegato nelle ricerche sull'intonazione. Alcune recenti ricerche hanno utilizzato l'approccio AM nella descrizione dell'intonazione dei bambini. Chen e Fikkert (2007), ad esempio, hanno analizzato le produzioni di tre bambini danesi nel periodo di tempo compreso fra i 16 e i 25 mesi. I dati di questi autori hanno mostrato come i bambini padroneggino l'inventario degli accenti melodici (*pitch accent*) nucleari utilizzati dagli adulti ad un'ampiezza di vocabolario di circa 160 parole e il set di accenti prenucleari successivamente, ad un'ampiezza di vocabolario di circa 230 parole, quindi dopo l'emergere delle prime combinazioni di parole. Inoltre, le prime combinazioni di parole verrebbero prodotte accentando entrambi gli elementi lessicali che le compongono, indipendentemente dalla loro relazione semantica. Frota e Vigário (2008), sempre utilizzando la trascrizione ToBI, hanno condotto uno studio longitudinale sulle produzioni preverbali, prime parole e prime frasi prodotte da due bambini portoghesi, evidenziando come, diversamente da quanto riscontrato da Chen e Fikkert (2007) per il danese, l'intonazione ed il *phrasing* vengano padroneggiati dai bambini prima dell'emergere delle combinazioni di parole. Prieto e collaboratori (Prieto, Estrella, Thorson & Vanrell, 2012) hanno dato conferma all'idea che l'intonazione sia pa-

droneggiata prima della produzione di combinazioni di parole, nello specifico già nella produzione di enunciati di una sola parola. Raramente e solo recentemente l'approccio AM è stato utilizzato nello studio di materiale linguistico complesso, come le narrazioni, prodotto da bambini: Redford e collaboratori (Redford, Dilley, Gamache & Wieland, 2012) hanno trovato come i toni di confine e le pause vengano utilizzati per marcare la continuazione *vs.* il completamento della frase nelle narrazioni di bambini tra i 5 e 7 anni in maniera simile ma non del tutto sovrapponibile a quanto fatto dagli adulti; anche De Ruiter (2014), analizzando le narrazioni di bambini di 5 e 7 anni ed adulti tedeschi, ha mostrato come i bambini di questa età siano in grado di marcare lo status informativo (*given vs. new*) degli elementi della narrazione in modo comparabile a quanto fatto dagli adulti, ma non siano ancora pienamente in grado di strutturare la narrazione attraverso l'intonazione.

Il presente studio si pone come principale obiettivo quello di analizzare, all'interno del *framework* AM, le narrazioni di adulti e bambini italiani di età prescolare. Alla luce della letteratura sopra descritta, esistono infatti pochi studi che abbiano avuto come oggetto di indagine l'intonazione di produzioni complesse in età evolutiva, i quali non sempre hanno portato a risultati univoci. Inoltre, l'utilizzo della trascrizione prosodica ToBI, a nostra conoscenza mai ad oggi utilizzata nel contesto italiano con partecipanti in età di sviluppo, consentirà un confronto con questi stessi recenti studi condotti su altre lingue.

2. *Metodo*

2.1 Partecipanti

I partecipanti allo studio sono 20 bambini (10 maschi e 10 femmine) di età compresa fra i 37 e i 48 mesi (Media = 44 mesi) e 8 adulti (3 maschi e 5 femmine) di età compresa fra i 18 e i 48 anni (Media = 26 anni), tutti monolingui italiani residenti nella regione Lombardia (prevalentemente a Milano e nel suo hinterland). I bambini partecipanti allo studio sono stati casualmente selezionati da un *corpus* dati più ampio, sulla base dell'età. Per tutti i bambini sono state chieste alcune informazioni ai genitori, quali la nascita pretermine, l'esposizione a lingue diverse dall'italiano, la presenza di deficit visivi e/o uditivo e la presenza di ritardi o difficoltà nello sviluppo. Questo ha consentito di selezionare, per il presente lavoro, 20 bambini nati a termine, monolingui italiani e dallo sviluppo tipico.

2.2 Procedura

Le narrazioni sono state elicitate mediante il racconto di una storia rappresentata attraverso 18 immagini, creata *ad hoc* per la valutazione delle abilità narrative in età prescolare (*Narrative Competence Task*, Zampini, Zanchi, 2015). I partecipanti sono stati invitati a sfogliare le illustrazioni della storia, in modo da potersene fare un'idea generale, e quindi a raccontarla, sfogliando nuovamente le immagini. Durante la somministrazione della prova l'esaminatore non interviene nel racconto prodotto, se

non attraverso rinforzi positivi, in modo da mantenere alta la motivazione al compito nei bambini. Le immagini facenti parte del *Narrative Competence Task* sono state realizzate da grafici esperti mediante l'impiego di figure semplici e lineari, facilmente riconoscibili e interpretabili da soggetti in età evolutiva. La storia proposta rispetta la grammatica tipica di un racconto (situazione iniziale, evento-problema, tentativi di risoluzione del problema e finale) e fa riferimento ad una situazione potenzialmente vicina all'esperienza dei bambini (i due bambini protagonisti della storia, un maschio e una femmina, si incontrano al parco e decidono di giocare insieme a palla; ad un certo punto, la bambina lancia la palla sopra i rami di un albero; i due bambini mettono quindi in atto diversi tentativi per recuperare la palla; infine, con l'aiuto di un vigile, riescono a riprendere la palla e a ricominciare a giocare).

Le narrazioni sono state audioregistrate mediante registratore digitale con microfono incorporato in un ambiente tranquillo (per i bambini, una stanza appositamente dedicata all'interno delle scuole dell'infanzia frequentate; per gli adulti, un laboratorio universitario).

2.3 Trascrizione dell'intonazione

L'intonazione è stata trascritta mediante le norme del sistema ToBI per l'italiano, utilizzando come riferimento il recente lavoro di Gili Fivela e collaboratori (Gili Fivela et al., 2015). Secondo l'approccio AM, da cui il sistema di annotazione ToBI trae origine, l'intonazione può essere descritta come una sequenza di toni alti (*High*, H) e bassi (*Low*, L). I *pitch accent* sono i toni associati alle sillabe prominenti all'interno dell'enunciato e vengono indicati attraverso il diacritico *. Tali toni possono essere semplici (H* oppure L*), oppure complessi (come nel caso dell'ultimo *pitch accent*, o *pitch accent* nucleare, tipico dell'enunciato dichiarativo *broad focus* in italiano, H+L*). I *pitch accent* (in particolare quelli nucleari) giocano un ruolo fondamentale nell'attribuzione dello status informativo degli elementi all'interno dell'enunciato (come ad esempio nel contrasto *new vs. given*) (Pierrehumbert, Hirshberg, 1990). I toni di frontiera (*boundary tones)* rappresentano invece i toni allineati al confine destro del sintagma intonativo (associati con il sintagma stesso) e vengono indicati con il diacritico %. Il sistema di annotazione ToBI, oltre alla descrizione dell'intonazione come susseguirsi di toni, permette di descrivere il livello di coesione percepito fra le parole dell'enunciato stesso. Tale unione viene annotata attraverso i *break*, gerarchicamente ordinati su 5 livelli (da 0 a 4), di cui i *break* 3 e 4 rappresentano i maggiori livelli di disgiunzione possibile tra le parole di uno stesso enunciato. Il *break* 3 rappresenta infatti il punto di disgiuntura tra una *intermediate phrase* e la successiva, il *break* 4 la separazione tra *intonation phrase* diverse.

Per gli scopi del presente studio, che vuole rappresentare un'indagine pilota dell'intonazione utilizzata nelle narrazioni di bambini italiani di età prescolare, si è deciso di prendere in considerazione le seguenti variabili (Figura 1):

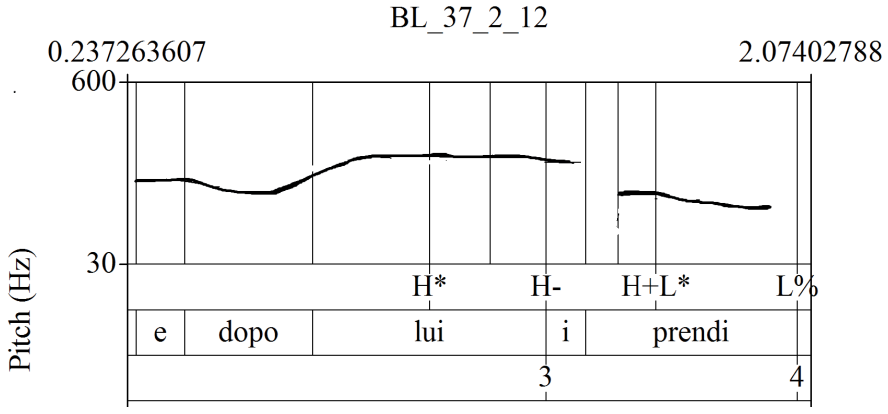
– *Pitch accent nucleare*, ossia l'ultimo individuabile all'interno del sintagma intonativo;

– *Boundary tone*, considerando solo l'opposizione H% *vs.* L% (i *boundary tone* complessi come LH% o HL% sono stati quindi semplificati rispettivamente come H% e L%);
– numero di *break* 3 e 4 presenti all'interno dell'enunciato (non è stato quindi conteggiato il *break* 4 normalmente previsto al confine destro dell'enunciato stesso).

Al fine di limitare l'influenza della sintassi sulla realizzazione intonativa, si è deciso di includere nelle analisi gli enunciati privi di verbo (ad es. "La bambina con la bicicletta rosa"), gli enunciati semplici (ad es. "La bambina pedala") e gli enunciati complessi con subordinata implicita (ad es., "I bambini provano a prendere la palla"). Tutti gli enunciati analizzati rientrano nella tipologia assertiva (sono stati escluse dalle analisi eventuali domande poste all'esaminatore nel corso della narrazione). Gli enunciati caratterizzati per la presenza di eccessivo rumore di sottofondo e sovrapposizione con la voce dell'esaminatore sono stati esclusi dalle analisi (9 enunciati su 440 totali, meno del 2% del totale).

Le annotazioni prosodiche sono state effettuate da uno degli autori mediante l'utilizzo del *software* PRAAT (Boersma, Weenink, 2005).

Figura 1 - *esempio di trascrizione per un partecipante, Leonardo, di 37 mesi*



Per il calcolo dell'accordo tra osservatori, due autori dello studio hanno analizzato indipendentemente il 20% degli enunciati. L'accordo, calcolato attraverso l'indice Kappa di Cohen, è risultato essere .98 per l'identificazione del *pitch accent* nucleare, .77 per il *boundary tone* e .96 per l'indicazione della presenza/assenza di break 3 o 4 all'interno dell'enunciato. In caso di disaccordo, l'enunciato è stato analizzato da un terzo autore per raggiungere un accordo.

3. *Risultati*

3.1 Analisi dei dati

Sono stati analizzati 431 enunciati totali, 334 per il gruppo di bambini (M = 16,7; *range* = 9-33) e 97 per il gruppo di adulti (M = 12.1; *range* 3-16). La media di enunciati analizzati per gli adulti è inferiore a quella del gruppo di bambini perché, durante le loro narrazioni, gli adulti utilizzano un numero maggiore di enunciati complessi con subordinate esplicite, esclusi dalle analisi.

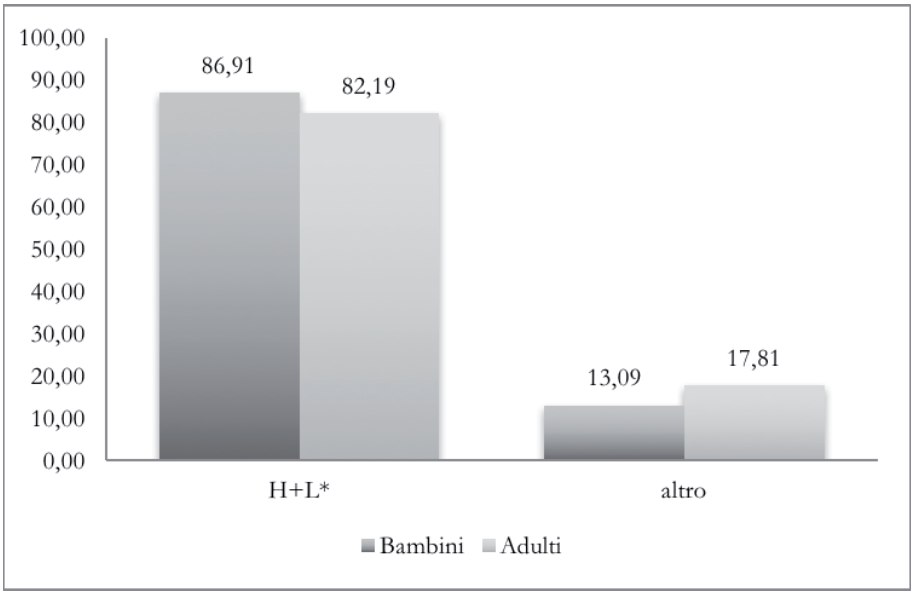
Per l'analisi dei dati, si è deciso di calcolare per ogni soggetto la percentuale di: enunciati con *pitch accent* nucleare H+L* vs. enunciati con altre tipologie di *pitch accent* nucleare; di enunciati con *boundary tone* H% vs. enunciati con *boundary tone* L%; di enunciati con *break* vs. senza *break*.

Le analisi sono state effettuate mediante un confronto tra campioni indipendenti non parametrico (Mann-Whitney).

3.2 *Pitch Accent* nucleare

Dal confronto effettuato non emergono differenze significative nel tipo di *pitch accent* nucleare utilizzato da bambini e adulti ($p > .05$, Grafico 1). Per entrambi i gruppi infatti la maggior parte degli enunciati è realizzata utilizzando il *pitch accent* nucleare H+L*, caratteristico dell'enunciato dichiarativo *broad focus* in italiano (Gili Fivela et al., 2015).

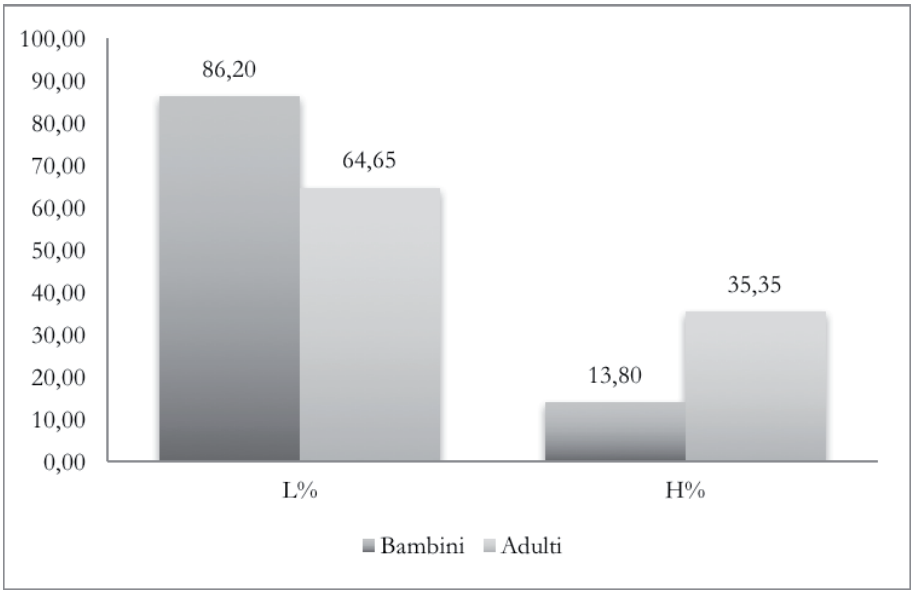
Grafico 1 - *percentuali di pitch accent nucleari H+L* vs. altri nel gruppo di bambini e di adulti*



3.3 *Boundary Tone*

Per quanto riguarda il tipo di *boundary tone* utilizzato (Grafico 2) emergono delle differenze significative tra bambini e adulti: gli adulti, infatti, utilizzano un maggior numero di *boundary tone* ascendenti (H%) rispetto ai bambini, fornendo quindi all'ascoltatore la corretta indicazione di continuità tra gli enunciati ($U = 124.5$; $p < .05$); complementarmente, i bambini utilizzano un numero di *boundary tone* discendenti maggiore rispetto agli adulti ($U = 35.5$; $p < .05$).

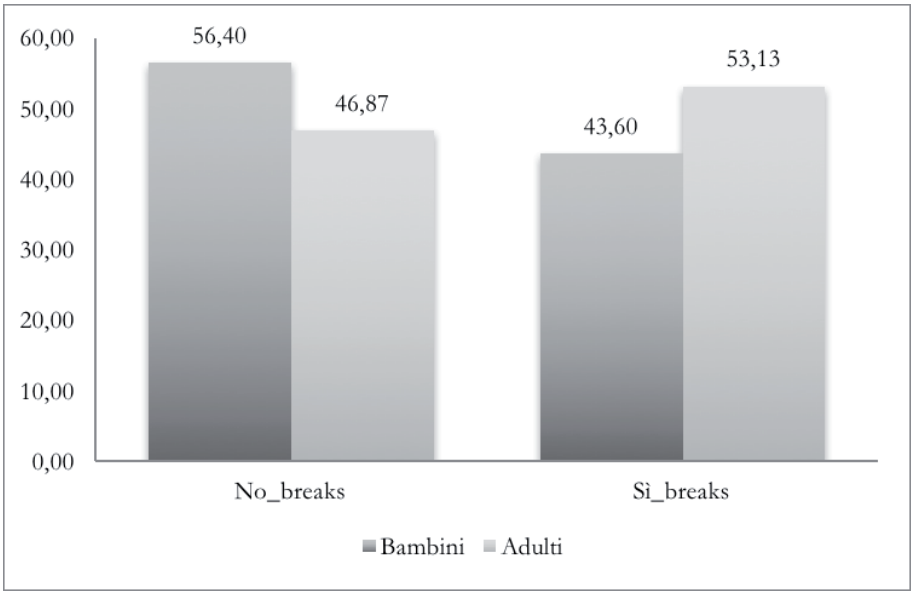
Grafico 2 - *percentuali di boundary tone H% vs. L% nel gruppo di bambini ed adulti*



3.4 Numero di *Break*

Nell'analisi dei *break* 3 e/o 4 presenti all'interno degli enunciati, è stata considerata la percentuale per ogni partecipante di enunciati caratterizzati dalla presenza di almeno un *break* al loro interno (denominati sì_breaks) *vs*. la percentuale di enunciati privi di *break* al loro interno (denominati no_breaks). Dall'analisi così condotta, non sono emerse differenze significative tra bambini e adulti ($p > .05$; Grafico 3).

Grafico 3 - *percentuali di enunciati sì_breaks vs. no_breaks nel gruppo di bambini e adulti*



4. *Discussione*

Il principale obiettivo dello studio è stato l'analisi ed il confronto dell'intonazione utilizzata nelle narrazioni da bambini di età prescolare (nello specifico, del primo anno di scuola dell'infanzia) e adulti monolingui italiani all'interno del *framework* AM. A nostra conoscenza non esistono studi italiani che abbiano impiegato tale metodologia nell'analisi della prosodia in età evolutiva, rendendo quindi difficile il confronto con i dati presenti in letteratura sulla prosodia degli adulti e quello con i recenti dati disponibili per l'età evolutiva nel panorama internazionale.

Dai risultati, emerge come, già a 3 anni di età, i bambini siano in grado di utilizzare correttamente, in modo simile agli adulti, il *pitch accent* nucleare tipico dell'enunciato *broad focus* in italiano, ossia H+L* (Gili Fivela et al., 2015). Questo dato è in linea con quanto riportato dalla letteratura disponibile per lingue diverse dall'italiano, come il danese (Chen, Fikkert, 2007), il portoghese e lo spagnolo (Frota, Vigário, 2008; Prieto et al., 2012), secondo la quale il pattern del *pitch accent* tipico della propria lingua sarebbe padroneggiato dal bambino molto presto nel corso dello sviluppo, per alcuni anche prima dell'emergere delle combinazioni di parole (Prieto et al., 2012).

Le narrazioni di adulti e bambini, invece, si differenziano rispetto all'utilizzo di un tono di confine ascendente, altrimenti chiamato *continuation rise* (Hirschberg, Pierrehumbert, 1986), necessario in alcune tipologie di produzione linguistica, come appunto le narrazioni, per fornire informazioni all'interlocutore circa la necessità di interpretare il significato dell'enunciato insieme al successivo. Enunciati assertivi caratterizzati da un contorno finale ascendente inducono infatti l'ascol-

tatore a legare tale enunciato al successivo, arrivando ad una interpretazione del significato che li prenda in considerazione entrambi. I bambini di 3 anni del nostro gruppo infatti utilizzano tale tono di confine solo in una piccola percentuale di casi, circa il 14% sul totale, a differenza degli adulti che mostrano di utilizzarlo nel 35% circa degli enunciati. Recentemente anche De Ruiter (2014) ha trovato delle differenze significative nell'utilizzo di toni di confine ascendenti in bambini di 5 e 7 anni ed adulti tedeschi, evidenziando un trend evolutivo che prevede un aumento nell'utilizzo di tale confine intonativo parallelo all'aumentare dell'età. Tale dato era stato precedentemente mostrato per la lingua inglese da Redford e collaboratori (Redford et al., 2012) con bambini della stessa età, i quali correttamente utilizzano toni di confine e pause per segnalare continuazione *vs.* completamento nelle narrazioni, ma in maniera non completamente sovrapponibile a quanto osservato negli adulti. Una possibile spiegazione del mancato utilizzo della *continuation rise* nei bambini di 3 anni potrebbe essere ricercata, a nostro parere, nel generale sviluppo cognitivo; infatti, l'abilità di utilizzare gli indizi prosodici che consentono all'ascoltatore una corretta interpretazione del significato di quanto narrato potrebbe essere legata alla consapevolezza nel bambino del punto di vista dell'ascoltatore stesso. Solo la consapevolezza che l'altro non possiede tutte le informazioni del parlante (ad esempio rispetto a ciò che il parlante ha intenzione di comunicare negli enunciati successivi) potrebbe permettere al parlante stesso di legare il significato di enunciati successivi, ma semanticamente relati, attraverso la prosodia. È noto infatti, anche se controverso, il rapporto tra teoria della mente e linguaggio (ad es. Milligan, Astington & Dack, 2007) ed il suo non ancora pieno sviluppo a 3 anni di età (ad es., Perner, 1999). Al fine di dare sostegno a questa ipotesi, sarebbe importante poter confrontare le performance del presente gruppo di bambini di 3 anni con quello di bambini di 5 anni, età in cui nello sviluppo tipico le prove di teoria della mente di primo livello vengono ormai generalmente superate, inoltre, sarebbe importante raccogliere contemporaneamente informazioni sulle competenze narrative e di teoria della mente in un campione di bambini di età prescolare e scolare, in modo da poter evidenziare in modo diretto le possibili relazioni ed influenze tra sviluppo cognitivo e linguistico in aspetti ad oggi ancora troppo raramente investigati in età evolutiva, come appunto l'intonazione.

Infine, i nostri dati non evidenziano delle differenze nella presenza di *break* 3 o 4 fra bambini e adulti. Tuttavia, questa prima analisi, condotta utilizzando come dato la percentuale di enunciati che presentano *break*, rappresenta solo un primo tentativo in tale direzione. Sarebbe interessante effettuare un'analisi più accurata, che permetta ad esempio di considerare l'influenza della lunghezza dell'enunciato sulla presenza di pause all'interno dell'enunciato stesso e di identificare i toni di confine associati alla fine dell'i*ntermediate phrase*, e, quindi, al *break* 3. Infatti, le analisi ad oggi effettuate non ci consentono di indagare il possibile legame tra utilizzo della *contonuation rise* e lo sviluppo della gerarchia prosodica (gli *intermediate phrase break* ed i *pitch accent* ad essi associati potrebbero essere più difficili da padroneggiare rispetto agli *intonative phrase break* ed ai toni di confine di sintagma).

Inoltre, sarebbe interessante effettuare un confronto fra adulti e bambini rispetto alla durata delle pause all'interno dell'enunciato, alla loro tipologia ed al loro posizionamento, gettando possibile luce sui legami tra linguaggio e pensiero. È possibile infatti ipotizzare che le strategie di "programmazione" della narrazione non siano le stesse tra adulti e bambini, ipotesi supportata dall'impressione qualitativa che i partecipanti di 3 anni programmino la loro narrazione "*on-line*", giustapponendo le parti del racconto passo dopo passo, anche all'interno dello stesso enunciato, al contrario degli adulti, per i quali il *planning* avverrebbe invece prevalentemente a priori. Una differenza nel *pre-planning* potrebbe inoltre in parte spiegare anche l'utilizzo di una percentuale diversa di toni di confine ascendenti da parte di adulti e bambini. Ulteriori analisi sono necessarie per gettare maggiore luce sui primi risultati riportati nel presente studio esplorativo.

Riferimenti bibliografici

Behrens, H., Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. In *Journal of Child Language*, 32(1), 1-34.

Boersma, P., Weenink, D. (2005). Praat: Doing phonetics by computer (version 5340). Retrieved from http://www.praat.org/Accessed 25.11.15.

Chen, A., Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In *Proceedings of the 16th International Congress of Phonetic Sciences* (vol. 3). Parrot GmbH Dudweiler, 1553-1556.

De Ruiter, L.E. (2014). How German children use intonation to signal information status in narrative discourse. In *Journal of child language*, 1(05), 1015-1061.

D'Imperio, M. (2002). Italian intonation: An overview and some questions. In *Probus*, 14(1), 37-69.

D'Odorico, L., Carubbi, S. (2003). Prosodic characteristics of early multi-word utterances in Italian children. In *First Language*, 23(1), 97-116.

D'Odorico, L., Fasolo, M. & Zanchi, P. (2010). *Prosodic characteristics of multi-argument utterances in Italian children*, Child Language Seminar, London, UK.

Frota, S., Vigário, M. (2008). The intonation of one-word and first two-word utterances in European Portuguese. In *XI International Conference for the Study of Child Language*, 1-4.

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giornado, R., Marotta, G., Savino, M. & Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian. In Frota, S., Prieto, P. (Eds.), *Intonation in Romance*. Oxford: OUP, 140-197.

Grice, M., D'Imperio, M., Savino, M. & Avesani, C. (2005). Strategies for intonation labelling across varieties of Italian. In Jun, S.-A. (Ed.), *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press, 55-83.

Hirschberg, J., Pierrehumbert, J. (1986). The intonational structuring of discourse. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 136-144.

Ladd, D.R. (1996). *Intonational phonology*. Cambridge Studies in Linguistics, 79. Cambridge: Cambridge University Press.

Milligan, K., Astington, J.W. & Dack, L.A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. In *Child development*, 78(2), 622-646.

Morgan, J.L., Demuth, K. (2014). *Signal to syntax*: *Bootstrapping from speech to grammar in early acquisition*. Philadelphia, PA: Psychology Press.

Perner, J. (1999). Theory of mind. In Bennett, M. (Ed.), *Developmental psychology: Achievements and prospects*. Philadelphia, PA: Psychology Press, 205-230.

Pierrehumbert, J. (1980). The phonetic and phonology of English intonation. Unpublished PhD Thesis, MIT.

Pierrehumbert, J., Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In Cohen, P.R., Morgan, J. & Pollack, M.E. (Eds.), *Intentions in communication*. Cambridge: MIT Press, 271-311.

Prieto, P., Estrella, A., Thorson, J. & Vanrell, M.D.M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. In *Journal of Child Language*, 39(2), 221-257.

Redford, M., Dilley, L., Gamache, J. & Wieland, E. (2012). Prosodic Marking of Continuation versus Completion in Children's Narratives. In *Interspeech*, 2494-2497.

Snow, D., Balog, H.L. (2002). Do children produce the melody before the words? A review of developmental intonation research. In *Lingua*, 112(12), 1025-1058.

Zampini, L., Fasolo, M., Spinelli, M., Zanchi, P., Suttora, C. & Salerni, N. (2016). Prosodic skills in children with Down syndrome and in typically developing children. In *International Journal of Language & Communication Disorders*, 51(1), 74-83.

Zampini, L., Zanchi, P. (2015). Lo sviluppo delle abilità narrative in età prescolare. Presentazione orale al XXVIII Congresso Nazionale Associazione Italiana di Psicologia – Sezione di Psicologia dello Sviluppo, Parma.

Zanchi, P., Zampini, L., Fasolo, M. & D'Odorico, L. (2016). Syntax and prosody in narratives: A study of preschool children. In *First Language*, 36(2), 124-139.

DOI: 10.17469/O2102AISV000011

CLAUDIA CROCCO, LINDA BADAN

L'hai messo DOVE il focus? Un'analisi prosodica delle domande eco wh-

In this paper we explore the prosody of regular and echo wh- questions in Este Italian. We analyzed 40 regular and 40 echo wh- questions introduced by *dove*, collected by means of a reading task and produced by 4 native speakers. Regular questions are phrased in two ϕ separated by a L-. The first phrase has an H+L* L- tune with the pitch accent (PA) associated to the verb. The second phrase can be realized with different tunes. Regular wh- questions in Este Italian share several features with their counterparts in other Italian varieties. Echo questions are also phrased in two units. The tune of the first phrase is L+¡H* H-H% with the PA associated to the wh-. The tune of the second phrase is L*+H H-H%. Echo questions are characterized by an expanded pitch range.

Key words: domanda wh- neutra, domanda wh- eco, focus, italiano di Este, interfaccia sintassi-prosodia, prominence, phrasing.

Introduzione

Questo lavoro si inscrive in un progetto più ampio di analisi all'interfaccia sintassi-prosodia in diversi tipi di domanda totale e parziale in una varietà settentrionale di italiano (Crocco, 2013; Badan, Crocco, in stampa).

Dal punto di vista teorico, lo studio si basa sulle premesse dell'approccio cartografico allo studio della sintassi (si vedano Rizzi, 1997; Cinque, 1999 e i numerosissimi studi successivi che su questi si fondano). Questo approccio ha come scopo quello di identificare un'articolata serie di posizioni sintattiche degli elementi funzionali della frase, specializzate nel segnalare alle interfacce proprietà rilevanti per il sistema interpretativo da un lato, per il sistema fonologico dall'altro (Rizzi, 1997; Belletti, 2004; Benincà, Poletto, 2004).

In particolare, tra le posizioni sintattiche individuate negli studi con approccio cartografico, la definizione della proiezione sintattica di focus (FocusP) ha avuto un ruolo cruciale nello sviluppo di modelli cartografici della periferia sinistra della frase e, più in generale, per lo studio dell'articolazione della struttura informativa della frase. In generale, il focus è concepito come quel costituente nominale che veicola l'informazione nuova e che è caratterizzato da particolari proprietà prosodiche. L'interesse suscitato dallo studio del focus, infatti, è legato anche al fatto che molte ricerche hanno dimostrato che l'attivazione della posizione sintattica di focus è definita come visibile al componente fonologico. Sono state infatti verificate determinate caratteristiche prosodiche stabilmente associate al sistema fonologico

all'attivazione della proiezione del focus in sintassi (Truckenbrodt, 1995; Brunetti, 2004; Bocci, 2013).

In questo articolo presentiamo alcuni dati prosodici relativi all'interrogativa parziale neutra e all'interrogativa parziale eco. Di seguito si riporta un esempio di domanda neutra e uno di domanda eco del tipo di quelle utilizzate per questo lavoro:

(1) Dove vendono le mandorle?

(2) Le vendono DOVE, le mandorle?

Dal punto di vista semantico, l'elemento wh- nelle domande neutre in italiano è generalmente considerato un focus informativo (Rizzi, 1997; 2001; 2004; Benincà, Poletto, 2004). La domanda wh-, cioè, richiede nella risposta l'identificazione di un elemento che costituisce l'informazione nuova all'interno di un insieme presupposto[1]:

(3) D: Dove vendono le mandorle?
R: Al mercato

Per contro, il focus nelle domande eco sembra evocare una qualche forma di contrasto o di violazione delle aspettative del parlante. In particolare, la domanda eco sembra esprimere una richiesta di ripetizione o chiarificazione di una frase pronunciata precedentemente che non viene capita (come nell'esempio in (4)) o interpretata come sorprendente dal parlante (come nell'esempio in (5)).

(4) A: Per il dolce caprese ti occorrono mandorle e cioccolato. Il cioccolato te lo posso dare io, le mandorle le vendono sotto al ponte di <*noise*>
B: Le vendono DOVE le mandorle?

(5) A: Per il dolce caprese ti occorrono mandorle e cioccolato. Il cioccolato te lo posso dare io, le mandorle le vendono sotto al ponte di Porcaballacca.
B: Le vendono DOVE le mandorle?

Sulla base delle premesse teoriche della cartografia, possiamo ipotizzare che le particolarità informative e sintattiche di questi due tipi di domanda wh- vadano di pari passo con le loro differenze prosodiche. L'analisi dettagliata delle proprietà sintattiche e interpretative del focus nelle domande eco verrà affrontata in un altro lavoro (Badan, Crocco, in stampa). Lo scopo di questo articolo è quello di presentare un confronto tra le principali caratteristiche prosodiche della domanda parziale neutra e di quelle della domanda eco in un quadro autosegmentale-metrico (Ladd, 2008 [1996]; per l'italiano: Avesani, 1995; Grice, D'Imperio, Savino & Avesani, 2005; Bocci, 2013; Gili Fivela, Avesani, Barone, Bocci, Crocco, D'Imperio, Giordano, Marotta, Savino & Sorianello, 2015). In particolare, l'analisi presentata in questo lavoro riguarda le caratteristiche intonative in un campione di domande introdotte da *dove*, lette da parlanti veneti dell'area di Este (Padova).

[1] Per le proprietà sintattiche delle domande neutre wh- si veda più avanti §1.

L'articolo è strutturato nel modo seguente: nel § 1 si introducono alcuni aspetti rilevanti per questo lavoro della sintassi e della prosodia delle domande wh- in italiano. In § 2 sono presentati la metodologia e il corpus sperimentale utilizzati per raccogliere e analizzare i dati. In § 3 sono esposti e discussi i risultati dell'analisi delle domande wh- neutre (§ 3.1) e di quelle eco (§ 3.2). In § 4, infine, sono presentate le conclusioni dello studio.

1. *Sintassi e prosodia della domanda wh- in italiano*

Secondo Rizzi (1997) l'elemento interrogativo nelle domande neutre in italiano subisce un movimento sintattico alla periferia sinistra della frase; in particolare è definito come mosso dalla sua posizione argomentale interna alla frase, alla posizione di focus (FocusP) in periferia sinistra. Si veda in (6) la rappresentazione del movimento sintattico dell'elemento wh- in un'interrogativa neutra alla posizione di focus (FocusP) in periferia sinistra:

(6) Dove vendono le mandorle?
$[_{LP}[_{FocusP}$ Dove]] $[_{IP}$ _ vendono le mandorle?]

In uno studio successivo, Rizzi (2001) ha individuato per l'italiano due classi di elementi interrogativi con proprietà sintattiche differenti. In particolare, mentre alcuni elementi, come *perché*, non devono obbligatoriamente essere adiacenti al verbo, altri, come *chi* e *dove* devono necessariamente collocarsi accanto al predicato:

(7) Perché Maria non è ancora arrivata?

(8) * Dove Maria andrà?

Questa condizione di adiacenza può essere violata solo a patto che l'enunciato presenti una prosodia particolare. Questo è appunto quello che avviene nelle domande eco:

(9) Hai visto CHI stamattina?

(10) Arriva DOVE il treno?

Diversamente dalle domande neutre, nelle domande eco l'elemento interrogativo *dove* si trova in posizione argomentale direttamente dopo il verbo all'interno della frase. Nella domanda eco in (11), l'elemento interrogativo *dove* viene definito *in situ* poiché non sottostà ad alcun movimento sintattico, ma viene realizzato (almeno apparentemente, si veda Badan, Crocco in stampa) nella sua posizione argomentale originale. In altre parole, l'elemento interrogativo nelle domande eco non si sposta in posizione di focus in periferia sinistra.

(11) Le vendono DOVE le mandorle?
$[_{LP}]$[Le vendono DOVE le mandorle?]

Dal punto di vista prosodico le caratteristiche della domanda wh- eco sono poco note. Una prima analisi della domanda eco incredula non *in situ* in diverse varietà di italiano si trova in Gili Fivela et al. (2015). Questo studio intende quindi contribu-

ire allo studio della prosodia dell'italiano fornendo dati per la descrizione di questo tipo di interrogativa.

Per quanto riguarda la domanda wh- neutra, gli studi autosegmentali sull'italiano (ad es., Marotta, Sorianello, 1999; Marotta, 2002; Bocci, 2013; Gili Fivela et al., 2015) hanno messo in luce alcune proprietà notevoli riguardo all'assegnazione della prominenza principale dell'enunciato. Come osservato da Marotta (2002, sull'italiano di Siena) le due classi di elementi wh- individuati da Rizzi presentano anche specifiche caratteristiche prosodiche. In particolare, secondo Bocci (2013; sempre sul toscano) i wh- che non richiedono adiacenza al verbo (come *perché*) attraggono la prominenza principale dell'enunciato, mentre nel caso dei wh- che devono essere adiacenti al verbo (come *chi* e *dove*), l'accento principale non è attratto dall'elemento interrogativo ma tende piuttosto a collocarsi sul predicato seguente oppure a occorrere alla fine dell'enunciato, in una posizione prosodicamente non marcata. In questo quadro, è importante verificare se la varietà atestina qui esaminata presenti tratti prosodici in linea con quelli delle altre varietà già esaminate in altri studi, in particolare con quella toscana, oppure se abbia caratteristiche intonative specifiche e costituisca quindi un controesempio.

2. *Corpus e metodo*

Per questo studio è stato utilizzato un campione di parlato letto raccolto in Veneto, in provincia di Padova. Gli informatori sono quattro parlanti nativi della varietà atestina, tre di sesso femminile (parlanti A, L, S) e uno di sesso maschile (parlante C), di età compresa tra i trenta e in quarant'anni e in possesso di una laurea o di un dottorato.

Per elicitare il materiale è stato utilizzato un compito di lettura elaborato sulla base dell'inchiesta usata per l'italiano per l'*Interactive Atlas of Romance Intonation* (IARI; Frota, Prieto, 2015; Gili Fivela et al., 2015). Il compito prevede che il parlante legga una frase *target*, che viene presentata in un contesto opportunamente pensato per elicitare l'intonazione desiderata. Di seguito forniamo due esempi di contesto utilizzato per elicitare l'interrogativa wh- neutra (12) e eco (13). Le frasi rilevanti sono in corsivo:

(12) A: Dobbiamo comprare bambole, dolci e dei vestiti per le figlie di mia cugina.
B: Va bene. *Dove vendono le bambole?*

(13) A: Per il dolce caprese ti occorrono mandorle e cioccolato. Il cioccolato te lo posso dare io. Le mandorle le vendono sotto al ponte di Porcaballacca.
B: *Le vendono DOVE le mandorle?*

L'inchiesta utilizzata per questo lavoro consiste di 78 contesti, dei quali solo 8 miravano ad elicitare interrogative wh- neutre e wh- introdotte da *dove*. Il resto dell'inchiesta era composto da *fillers* e da contesti utilizzati per elicitare enunciati con strutture prosodiche e sintattiche varie. Circa un terzo dei contesti dell'inchiesta completa ricalcano quelli del questionario italiano utilizzato per lo IARI. Poiché gli

studi sulla varietà veneta dell'area padovana sono nel complesso ancora pochi (ad es., Endo, Bertinetto, 1997), la raccolta di questi materiali ha consentito di avere informazioni supplementari sulla varietà. Gli enunciati elicitati esemplificano infatti le principali intonazioni della varietà in esame, costituendo quindi una base per possibili raffronti tra le domande wh- neutre e eco, ed enunciati con caratteristiche diverse (dichiarative, domande totali, altri tipi di domanda wh-, frasi con strutture sintattiche e informative marcate ecc.) prodotti in un contesto analogo dagli stessi parlanti.

Per gli esperimenti è stato utilizzato un registratore portatile Marantz PMD 620 collegato a un microfono Røde HS1-P. Le registrazioni sono state effettuate in un ambiente silenzioso, presentando l'inchiesta agli informatori con un Powerpoint. Dopo aver ricevuto alcune istruzioni sullo svolgimento del compito, durante la registrazione i parlanti sono stati lasciati soli. Ciascun parlante ha letto due volte l'inchiesta. Nel complesso, quindi, il campione utilizzato per questo lavoro consiste di 80 enunciati (8 enunciati interrogativi * 2 repliche *5 parlanti), di cui 40 sono domande wh- neutre e 40 domande eco. Le domande analizzate hanno la seguente struttura:

wh-	+ verbo trans.	+ oggetto diretto	
dove	+ *vendono*	+ *articolo det.* +	*nome trisill. sdrucciolo* *CVC.CV(C).CV* *es. "mandorle", "vongole"*

Tutti i materiali sono stati analizzati con Praat (Boersma, Weenink, 2016). Gli enunciati *target* sono stati annotati a mano a livello di parola, sillaba e fonema. Inoltre, sono stati annotati i punti di snodo (*target points*) degli accenti melodici e i movimenti tonali in corrispondenza della fine dell'enunciato. Attraverso uno *script* in R è stata generata una tavola di dati utilizzata per l'elaborazione statistica. Per l'analisi intonativa sono stati elaborati dati sull'allineamento dei *target* tonali, l'escursione tonale e la *slope* di accenti e toni di confine.

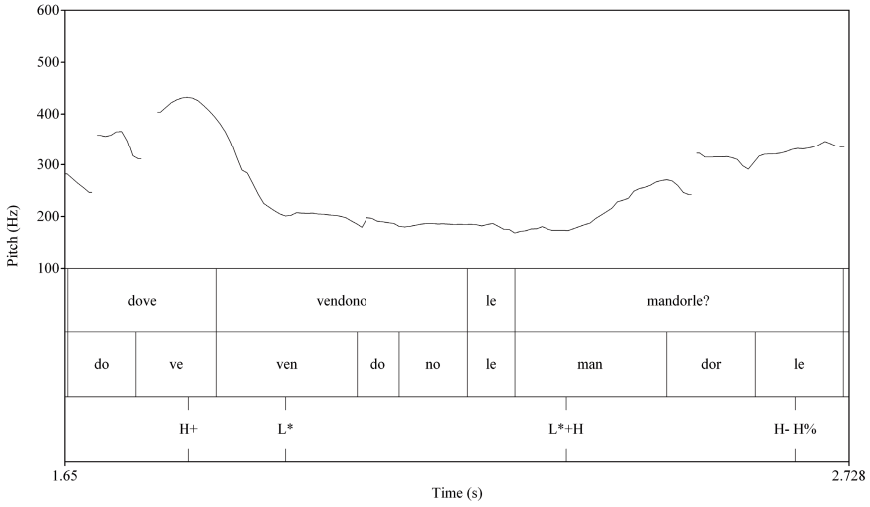
L'analisi prosodica presentata di seguito verte principalmente sulle caratteristiche degli accenti nell'area della parola wh- *dove* nelle due condizioni esaminate, cioè nella domanda neutra e nella domanda eco. Gli accenti che occorrono in fine di enunciato, in particolare quelli delle domande neutre, sono quindi descritti in minore dettaglio. Nel paragrafo seguente sono presentati i risultati relativi alle domande wh- neutre (§ 3.1) e alle domande wh- eco (§ 3.2).

3. *Risultati*

I risultati presentati in questa sezione mirano a fornire una descrizione intonativa degli accenti melodici e dei toni di confine caratteristici delle domande wh- neutre e eco nell'italiano parlato a Este. In particolare, per le domande neutre, sono presentati le misure relative all'allineamento e all'escursione della prominenza che si

colloca nell'area dell'elemento interrogativo *dove*, mentre nel caso delle domande eco sono presentati dati riguardanti sia l'accento nell'area del wh-, sia l'ultima prominenza e i confini melodici dell'enunciato. Sulla base dei dati ricavati dall'analisi sperimentale verrano proposte un'analisi fonologica e un'annotazione di tipo ToBI dei due tipi di domanda esaminati.

Figura 1 - *Domanda wh- neutra prodotta dalla parlante A*



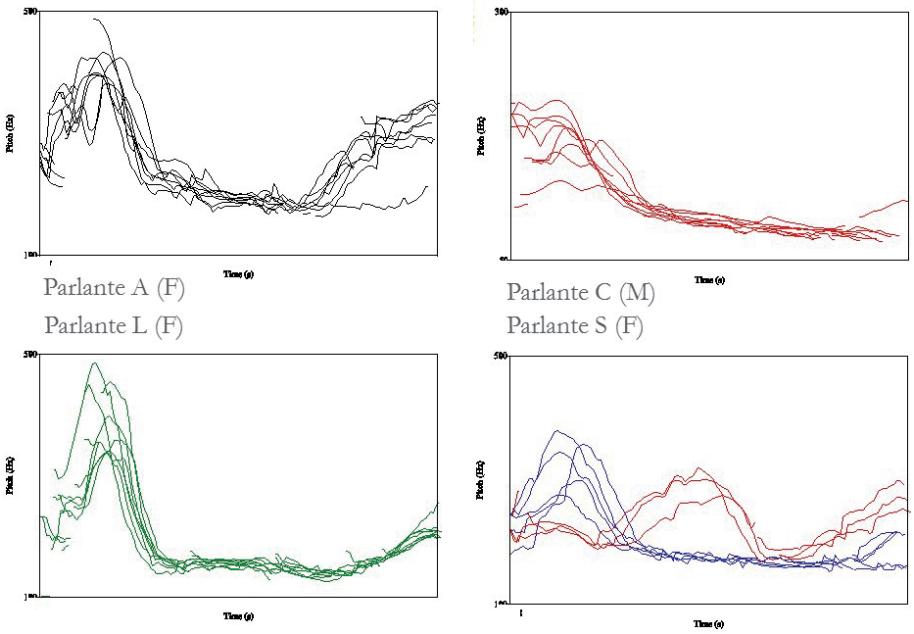
3.1 Domande wh- neutre

Le domande wh- neutre nella varietà atestina presentano un andamento melodico globale simile a quello riscontrato per enunciati analoghi in altre varietà (cfr. Marotta, 2002; Bocci, 2013), caratterizzato da un picco melodico nell'area della parola wh- seguito da un *plateau* basso e quindi da un movimento ascendente in fine di enunciato, che però non compare sistematicamente e, soprattutto, può avere inizio sull'ultima sillaba tonica o più tardi. Un esempio di domanda wh- neutra è presentato nella figura 1 (per altri esempi si veda anche la figura 4).

La figura 2 presenta tutte le domande wh- neutre prodotte dagli informatori. Come si vede, le domande prodotte dai diversi parlanti condividono un *pattern* globale con alcune varianti ricorrenti, come quella appena menzionata riguardante la presenza e le caratteristiche della risalita finale. Solo le domande prodotte della parlante S presentano due varianti nettamente differenti tra di loro: la prima, corrispondente alla variante sommariamente descritta in precedenza e prodotta anche dagli altri parlanti, presenta un ampio movimento melodico nell'area della parola wh-, e più specificamente tra la parola wh- e il successivo predicato. Nella seconda variante, invece, il movimento melodico è interamente collocato sul predicato, mentre l'elemento interrogativo non è interessato da movimenti di *pitch* rilevanti. Si noti che questa seconda variante, oltre a essere prodotta da una sola parlante, è anche

quantitativamente molto minoritaria, in quanto ricorre solo in 3 casi sul totale dei 40 analizzati.

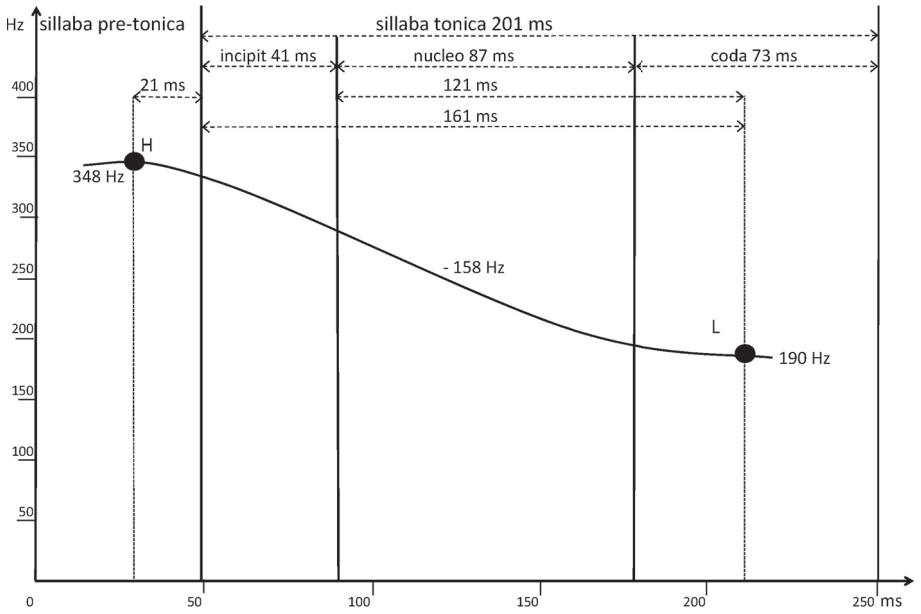
Figura 2 - *Sovrapposizione dei plot delle domande neutre wh- prodotte dai parlanti A, C, L e S. Le produzioni della parlante S presentano due diverse configurazioni: nella prima (plot in blu) è presente un picco tra la parola wh- e il successivo predicato; nella seconda (plot in rosso) il picco di colloca interamente sul predicato*



Notiamo qui cursoriamente che questa variante sembra più marcata regionalmente, per cui i due tipi potrebbero costituire altrettanti punti su un continuum sociofonetico. Tale considerazione tuttavia richiederebbe approfondimenti che non possono essere condotti in questa sede e sono per questo rimandati a un successivo studio.

Come si è detto poco più sopra, nel complesso il profilo melodico caratterizzato da un movimento tonale collocato tra la parola wh- e il predicato seguente è quello più frequente nel campione. Il movimento melodico caratteristico di questo profilo ha la forma di un picco la cui fase discendente si conclude appunto sul verbo ed è analizzabile come una sequenza di un tono alto (H) e di un tono basso (L) (cfr. figura 1; si vedano, più avanti, anche gli esempi in figura 4). La figura 3 presenta in maniera sintetica le durate medie di incipit, nucleo e coda della sillaba tonica del predicato ("*ven*dono") e i dati relativi all'escursione melodica e all'allineamento dei toni H e L rispetto all'attacco sillabico e a quello vocalico.

Figura 3 - *Allineamento tonale e escursione del pitch accent nell'area tra il wh- e la sillaba tonica "ven" del predicato nella domanda wh- neutra. I valori di frequenza sono riferiti alle parlanti donne*



Come si può vedere, mentre il *target* H tende a occorrere poco prima dell'attacco sillabico (in media 21 ms), e quindi sulla vocale finale della sillaba postonica dell'avverbio interrogativo *dove*, il *target* L si colloca in media a 121 ms dall'attacco vocalico, all'incirca a metà della coda sonorante. L'escursione tonale media per questo movimento discendente è di 158 Hz per le parlanti donne e di 78 Hz per il parlante uomo.

La sequenza tonale H L può essere analizzata come un *pitch accent* bitonale H+L*. Dei due toni H e L, infatti, L è più prominente (cfr. Prieto, D'Imperio & Gili Fivela, 2005), in quanto il movimento risulta uditivamente discendente. Inoltre, il tono L è foneticamente allineato all'interno della sillaba tonica del predicato, al contrario del tono H, che si colloca sulla sillaba precedente (cfr. figure 1 e 3). Dal punto di vista percettivo, l'accento H+L* associato al predicato risulta essere nella maggior parte dei casi la prominenza principale dell'enunciato.[2] Dal punto di vista ritmico non risultano confini interni uditivamente apprezzabili in questo tipo di enunciato. L'andamento piatto successivo al *pitch accent* H+L* è però compatibile con la presenza di un tono di confine L- marcato solo sul piano tonale e non su quello metrico (cfr. Bocci, 2013: 170).

Per completare la descrizione del profilo della domanda wh- neutra nell'italiano di Este presentiamo di seguito in maniera sintetica le tre configurazioni presenti nel secondo *intermediate phrase* dell'enunciato.

[2] Per un'analisi fonologica relativa agli enunciati con prominenza principale in posizione non finale di enunciato si vedano D'Imperio, 2001; Frota, 2000; Bocci, 2013).

La figura 4 illustra con esempi i *tune* osservati nella varietà. Questi *pattern* differiscono tra di loro per la forma dell'accento e per il tipo di contorno terminale. Come si è già notato in precedenza, nella domanda neutra wh- la risalita finale del *pitch* non è sempre presente e può cominciare sulla sillaba tonica oppure occupare solo l'ultima sillaba dell'enunciato. Inoltre, l'accento può essere costituito da un tono basso, oppure avere un andamento ascendente quando la risalita finale inizia sulla tonica. Sinteticamente i *pattern* osservati possono essere descritti come segue:

(14)

(14.a)	L*	L-	H%
(14.b)	L*	L-	L%
(14.c)	L*+H	H-	H%

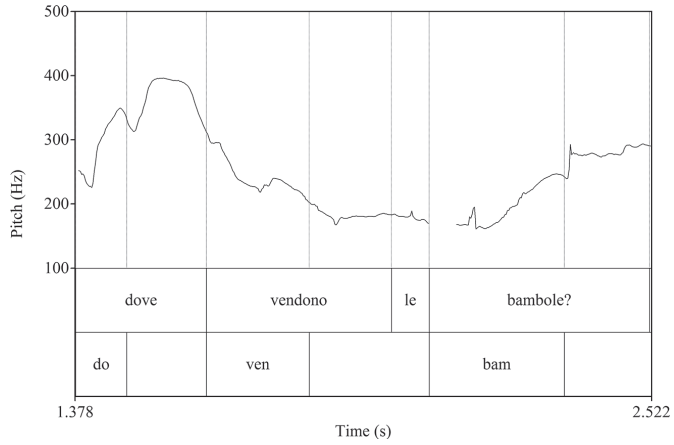
Profili simili a quelli qui descritti sono stati osservati in numerose varietà dell'italiano. In particolare, i profili (14.a) e (14.c) sono confrontabili con quelli caratterizzati da un accento nucleare H+L* riscontrati in diverse varietà meridionali, centrali e settantrionali. Si noti che anche nella varietà di Este l'accento L* dell'interrogativa wh- può occasionalmente alternarsi a un accento discendente H+L*. La tabella 1 è basata su una analoga presentata in Gili Fivela et al. (2015: 179); qui si inserisce nel quadro generale delle varietà indagate anche la varietà atestina presa in esame in questo lavoro.

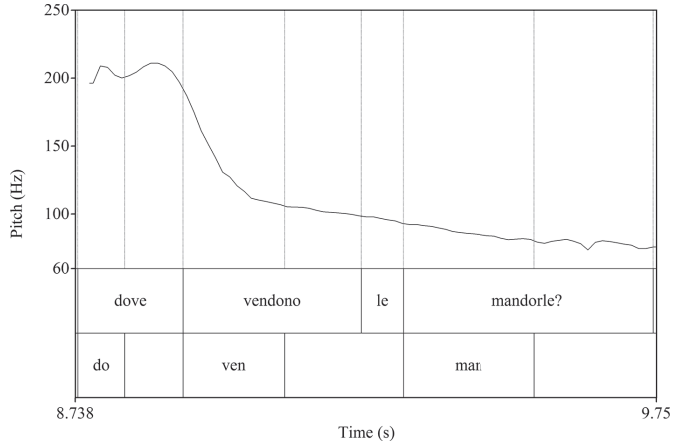
Tabella 1 - *Tune finale della domanda wh- neutra nella varietà atestina (in grassetto) e nelle varietà regionali esaminate in Gili Fivela et al. (2015)*

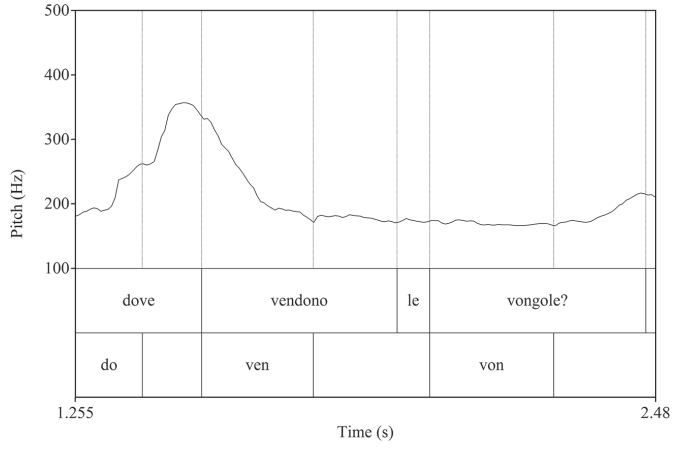
	LH%	**H%**	**L%**
(H+)L*	Bari, Cosenza, Este, Firenze, Lucca, Milano, Roma, Salerno, Siena, Torino		Bari, Cosenza, Este, Lecce, Lucca, Milano, Napoli, Pescara, Pisa, Roma, Salerno, Siena, Torino
L+H*		Cosenza, Este, Roma	Cosenza
L*+H		Pescara, Salerno	Pescara

La somiglianza tra i profili della varietà atestina e quelli osservati in altre varietà italiane può essere considerata un indice della presenza di caratteristiche intonative comuni – o per lo meno ricorrenti – in tutta l'area italiana, concorrendo a delineare un nucleo comune di tratti intonativi italiani per il tipo intonativo esaminato. La presenza di queste regolarità, ovviamente, non implica l'assenza di specificità fonologiche e soprattutto di variazione nella effettiva realizzazione fonetica dei profili.

Figura 4 - *Profili melodici della domanda wh- nella varietà di Este. Dall'alto verso il basso: enunciato prodotto dalla parlante di sesso femminile S (cfr. es. 14.a riportato nel testo); enunciato prodotto dal parlante di sesso maschile C (cfr.es. 14.b); enunciato prodotto dalla parlante di sesso femminile A (cfr.es. 14.c)*







Concludiamo questa sezione sulle domande wh- neutre presentando un quadro sintetico che riassume quanto detto in questa sezione riguardo ai fenomeni accentuali e di *phrasing* osservati:

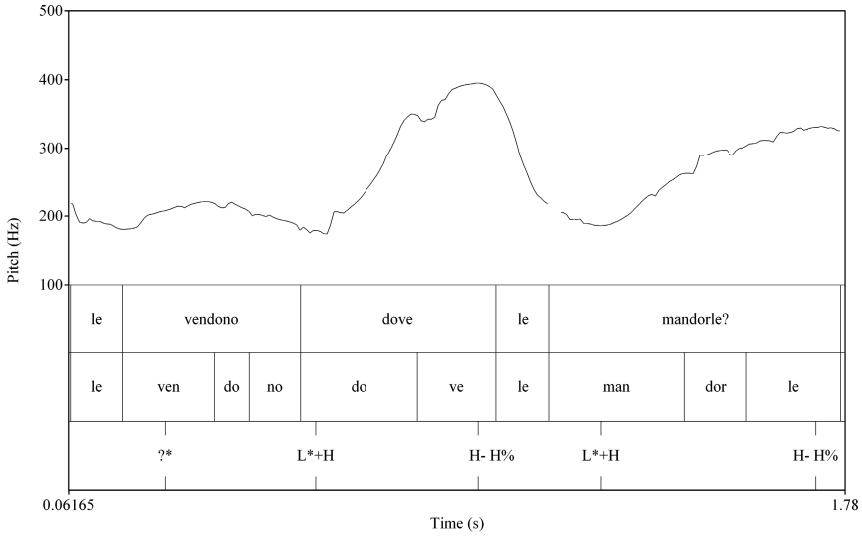
(15)			L*+H	H-H%
			L*	L-H%
	H+L*	L-	L*	L-L%
	[(Dove vendono)ɸ		(le mandorle?)ɸ]ι	

3.2 Domande wh- eco

Osservando il profilo melodico delle domande wh- eco si individuano due movimenti melodici molto evidenti, cioè un picco in corrispondenza dell'elemento interrogativo e un ulteriore movimento ascendente sull'ultima parola dell'enunciato. All'ascolto, questi enunciati risultano suddivisi in due unità minori da un confine ritmico oltre che melodico. La presenza di tale confine è attesa anche dal punto di vista fonologico, in quanto gli enunciati esaminati per questo lavoro contengono, sul piano sintattico, una dislocazione a destra con ripresa clitica dell'oggetto diretto.

Un esempio di domanda wh- eco è presentato nella figura 5, mentre la figura 6 mostra i *plot* di tutte le domande wh- neutre prodotte da ciascuno degli informatori. A differenza di quanto osservato per le domande neutre, nel caso delle domande eco non sono presenti varianti, ma tutte le domande sono state realizzate in modo molto simile dai diversi locutori.

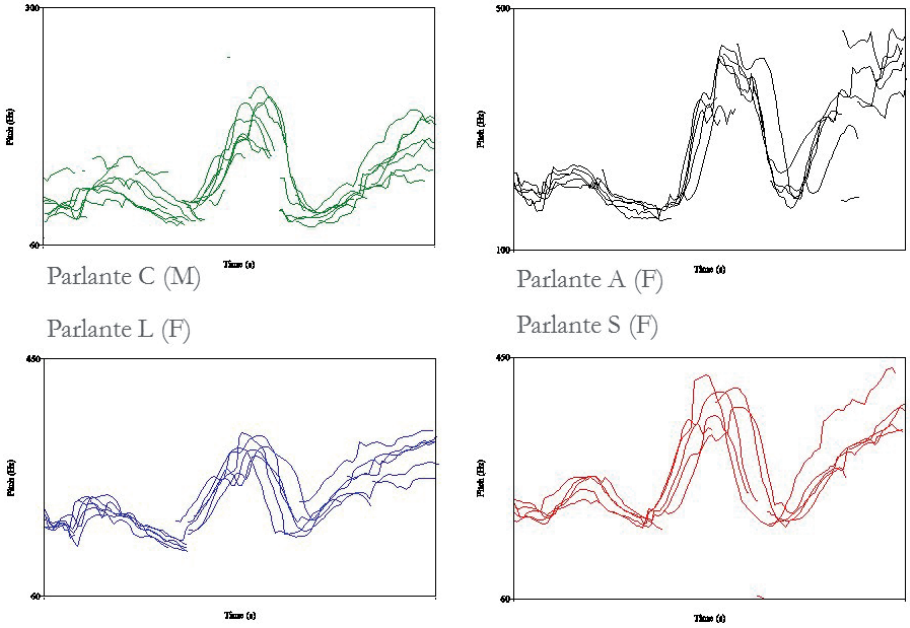
Figura 5 - *Domanda wh- eco prodotta dalla parlante S*



3.2.1 *Pitch accent* nell'area del wh-

Come mostra la figura 7, in corrispondenza dell'avverbio interrogativo wh- si trova una sequenza tonale ascendente. Il tono L si colloca sull'incipit, in media 28 ms prima dell'attacco vocalico ed è seguito da un movimento continuativamente ascendente che raggiunge il culmine in media 43 ms prima dell'*offset*. Il movimento ascendente si presenta molto ampio e ripido. L'escursione del *pitch* tra i due *target point* è in media di 202 Hz nelle parlanti di sesso femminile e 113 Hz nel parlante di sesso maschile. Per approssimare la *slope* della curva abbiamo misurato il rapporto Hz/s, che in media è di 780Hz/s nelle parlanti donne e 457Hz/s nel parlante uomo C. Questa misura sarà utilizzata più avanti per confrontare il movimento tonale nell'area del wh- e quello presente sull'ultimo elemento lessicale dell'enunciato.

Figura 6 - *Plot sovrapposti delle domande wh- eco prodotte dai parlanti C, A, L e S*

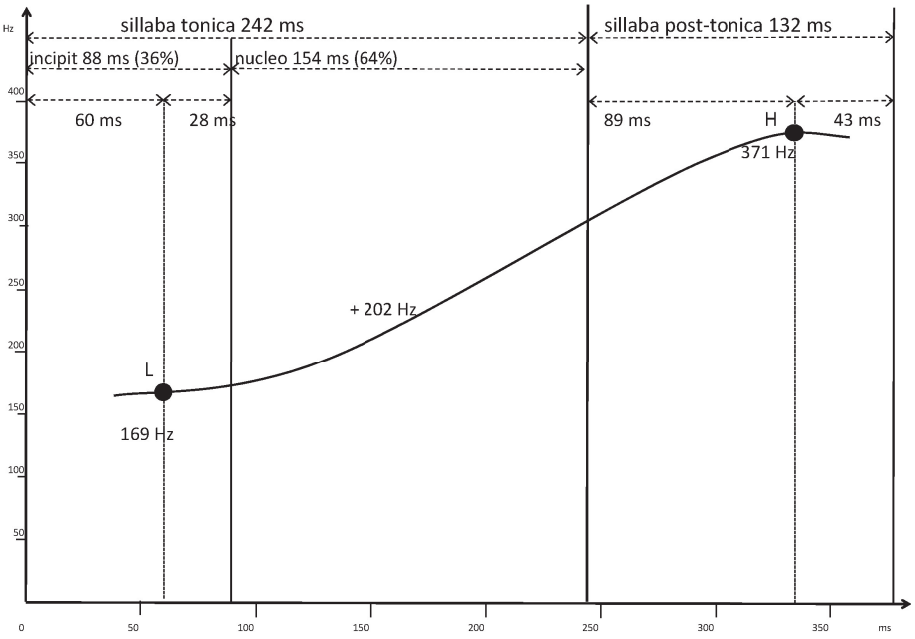


L'ampiezza del movimento tonale ascendente è una caratteristica già riscontrata in precedenza in altre varietà italiane nelle domande eco che esprimono incredulità e sorpresa dovute a una violazione delle aspettative del parlante (Gili Fivela et al., 2015: 181). Benché le domande eco esaminate in questo lavoro presentino una struttura sintattica particolare (con il wh- *in situ*), diversa da quelle delle domande wh- eco considerate da Gili Fivela et al. (2015), anche nel caso qui in esame si ritrova questa particolarità. Di fatto, l'espansione del *range* sembra essere una delle caratteristiche peculiari della marcatura prosodica che rende l'ordine delle parole in queste domande accettabile (cfr. § 1). Tuttavia sono necessari studi più ampi e sistematici per stabilire quale sia il ruolo dell'espansione del *range* in questo tipo di enunciati.

Nel complesso, il *tune* del primo *phrase* delle domande eco può essere analizzato come L+¡H* H-H%. Dato che il movimento accentuale e quello successivo relativo al confine formano una sequenza continua ascendente, non è possibile collocare con esattezza il punto in cui è raggiunto il *target* tonale del tono accentuale ¡H*. Tuttavia, l'ampiezza del movimento tonale e il suo andamento marcatamente ascendente sembrano ragioni sufficienti per supportare l'analisi di ¡H* come tono asteriscato. Questa analisi è inoltre in linea con quella proposta in Gili Fivela et al. (2015: 148, 181, 184) per le domande eco wh-.

Il *pitch accent* L+¡H* è uditivamente molto saliente e rappresenta la prominenza principale dell'enunciato. Inoltre, a differenza di quanto ossevato nelle domande wh- neutre, tale prominenza è associata all'elemento wh- e non al predicato.

Figura 7 - *Allineamento tonale e escursione del pitch accent associato al wh- nelle domande eco. I valori di frequenza sono riferiti alle parlanti donne*



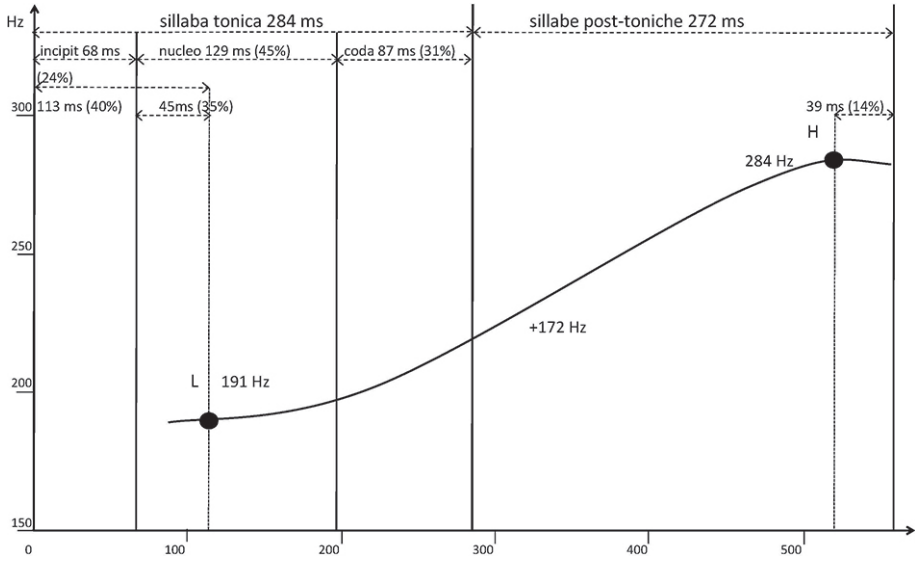
3.2.2 Pattern tonale del phrase finale

Nel *phrase* prosodico finale della domanda wh- eco si osserva un altro movimento melodico ascendente. Anche in questo caso, si ha un *target* L all'interno della sillaba tonica, seguito da un *target* H, che viene raggiunto poco prima del confine destro dell'enunciato. Più in dettaglio, come mostra la figura 8, L si trova nell'interno del nucleo vocalico in media a 113 ms dall'attacco sillabico e a 45 ms dall'attacco vocalico, mentre H è raggiunto in media 39 ms prima della fine dell'enunciato. L'allineamento di L risulta quindi diverso nel primo e nel secondo accento: sul wh- il tono basso si trova nell'incipit sillabico, mentre nel *phrase* finale esso occorre

nel nucleo. Non si può escludere tuttavia che la presenza di una coda nella sillaba accentata abbia un effetto sull'allineamento di L. Il movimento ascendente è anche qui ampio, con un'escursione tonale media di 172 Hz per le parlanti donne e di 74 Hz nel parlante uomo.

La somiglianza tra i movimenti tonali presenti nel primo e nel secondo sintagma prosodico dell'enunciato potrebbe indurre a invocare un meccanismo fonologico di copia, per cui il movimento presente nel secondo *phrase* costituirebbe una replica di quello realizzato dal parlante in corrispondenza del focus. Ci sono però elementi che suggeriscono di scartare, almeno in prima istanza, questa ipotesi, e di trattare i due movimenti come sequenze tonali indipendenti. Per quanto riguarda l'allineamento tonale, rispetto all'accento nell'area del focus il *target* L dell'accento nucleare occorre più tardi. Esso infatti non si colloca nell'incipit sillabico ma nel nucleo vocalico, a circa un terzo della vocale e poco prima della metà della sillaba, mentre nell'accento assegnato al focus L occorre alla fine del primo quarto della sillaba. Anche la pendenza del movimento presenta delle differenze. In media la *slope* del *pitch accent* del secondo sintagma intermedio è 573 Hz/sec per le parlanti di sesso femminile e 166 Hz/sec per il parlante di sesso maschile. Questi valori sono significativamente diversi da quelli riscontrati nell'accento del primo *intermediate phrase* (Wilcoxon rank sum test one sided: p <0.005, sia per le donne che per l'uomo). Queste differenze suggeriscono prudenza nell'analizzare il movimento finale come una copia di quello presente sul focus.

Figura 8 - *Allineamento tonale e escursione melodica nel phrase finale della domanda eco. I valori di frequenza sono riferiti alle parlanti donne*



I *pattern* del primo e del secondo *phrase* sembrano quindi distinti dall'*upstep rise* e dall'allineamento del tono L. Di seguito è presentata in modo schematico l'analisi delle domande wh- eco proposta in questo lavoro:

L+¡H* H-H%	L*+H H-H%
{[Le vendono dove]$_{\phi}$	[*le mandorle?*]$_{\phi}$}$_{\upsilon}$

4. *Conclusioni*

I risultati di questo lavoro mostrano che, alle diverse strutture sintattiche e proprietà interpretative della domanda wh- neutra e di quella eco corrispondono strutture prosodiche chiaramente differenziate.

La domanda wh- neutra presenta una struttura caratterizzata da due accenti melodici, uno nell'area del wh- e l'altro in corrispondenza dell'ultimo elemento lessicale dell'enunciato. Ciascun accento è il nucleo di un sintagma intermedio. I due sintagmi intermedi sono separati solo da un confine di tipo melodico (L-), mentre a livello ritmico non ci sono segni apprezzabili di disgiuntura. Il primo *pitch accent* è associato al predicato ed è, con poche eccezioni che sembrano marcate in senso diatopico, un H+L*. In generale, quindi, le caratteristiche di questi enunciati non differiscono in modo evidente da quelle osservate in altre varietà e in particolare in toscano (Marotta, 2002; Bocci, 2013; Gili Fivela et al., 2015). Anche nella varietà atestina, infatti, l'accento nucleare del *phrase* pur collocandosi in parte sull'avverbio interrogativo *dove* è piuttosto associato al predicato seguente. Gli enunciati wh-neutri presentano nel secondo *phrase* un accento nucleare di forma variabile, che può essere basso (L*) o ascendente (L*+H), e possono terminare con una risalita finale (L- H%, H- H%). *Pattern* nucleari simili sono stati osservati in numerose varietà italiane.

Le domande wh- eco presentano una disgiuntura interna, fonologicamente attesa data la presenza di una dislocazione a destra nell'enunciato. Inoltre, il primo *pitch accent* L+¡H* è chiaramente associato al wh-, mentre il predicato, che nella domanda eco *in situ* occorre prima e non dopo l'elemento interrogativo, non è interessato da movimenti melodici di rilievo. Infine, il *range* melodico appare espanso, sia sul wh- che sul materiale seguente, in modo simile a quanto osservato da Gili Fivela et al. (2015) su altri tipi di domanda eco.

In conclusione, questo studio ha fornito dati che indicano, per quanto riguarda la domanda wh- neutra, sostanziali somiglianze tra la varietà di italiano parlata a Este e le altre analizzate in letteratura. Inoltre, con questo lavoro si è contribuito a delineare le caratteristiche di un tipo intonativo meno noto, cioè la domanda wh-eco. I dati qui presentati costituiscono una prima base per lo studio delle proprietà sintattiche, prosodiche e interpretative di questi enunciati.

Riferimenti bibliografici

Avesani, C. (1995). ToBIt. Un sistema di trascrizione per l'intonazione italiana, in Metodologie di analisi e di descrizione delle caratteristiche prosodiche e intonative dell'italiano. In Lazzari, G. (Ed.), *Atti delle V giornate di studio del Gruppo di fonetica sperimentale (AIA)*, Povo, 17-18 novembre 1994. Trento: Servizio Editoria ITC, 85-98.

Badan, L., Crocco, C. (in stampa). Focus in Italian echo wh-questions: an analysis at syntax-prosody interface. In *Probus*.

Belletti, A. (2004). Aspects of the low IP area, The Structure of IP and CP. In Rizzi, L. (Ed.), *The cartography of Syntactic Structures*. New York: Oxford University Press, 16-51.

Benincà, P., Poletto, C. (2004). Topic, Focus and V2: defining the CP sublayers, The Structure of IP and CP. In Rizzi, L. (Ed.), *The cartography of Syntactic Structures*. New York: Oxford University Press, 52-75.

Bocci, G. (2013). *The syntax-prosody interface from a cartographic perspective: Evidence from Italian*. Amsterdam-Philadelphia: John Benjamins.

Boersma, P., Weenink, D. (2016). Praat: doing phonetics by computer. Computer program. Version 6.0.19. Retrieved from http://www.praat.org/Accessed 13.06.16.

Brunetti, L. (2004). *A Unification of Focus*. Padova: Unipress.

Cinque, G. (1999). *Adverbs and Functional Heads. A Cross-linguistic Perspective*. New York: Oxford University Press.

Crocco, C. (2013). Is Italian Clitic Right Dislocation grammaticalised? A prosodic analysis of yes/no questions and statements. In *Lingua*, 133, 30-52.

D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. In *Speech Communication*, 33, 339-356.

Endo, R., Bertinetto, P.M. (1997). Aspetti dell'intonazione in alcune varietà di italiano. In Cutugno, F. (Ed.), *Fonetica e fonologia degli stili dell'italiano parlato*. Atti delle VII giornate di studio del Gruppo di fonetica sperimentale (AIA), 14-15 novembre 1996, Napoli. Roma: Esagrafica, 27-49.

Frota, S. (2000). *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. London: Routledge.

Frota, S., Prieto, P. (Eds.) (2015). *Intonational Variation in Romance*. Oxford: Oxford University Press.

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M. & Sorianello, P. (2015). Varieties of Italian and their intonational phonology. In Frota, S., Prieto, P. (Eds.), *Intonational Variation in Romance*. Oxford: Oxford University Press, 140-197.

Grice, M., D'Imperio, M., Savino, M. & Avesani, C. (2005). Strategies for intonation labelling across varieties of Italian. In Jun, S.-A. (Ed.), *Prosodic typology. The phonology of intonation and phrasing*. Oxford: Oxford University Press, 362-389.

Ladd, R.D. (2008) [1996]. *Intonational phonology*. *Second edition*. Cambridge: Cambridge University Press.

Marotta, G. (2002). L'intonation des énoncés interrogatifs ouverts dans l'italien toscan. In Bel, B., Marlien, I. (Eds.), *Speech Prosody 2002*, Aix-en-Provence, Université de Provence, 475-478.

Marotta, G., Sorianello, P. (1999). Question Intonation in Sienese Italian. In Ohala, J.J. et al. (Eds.), *Proceedings of the XIV international congress of phonetic sciences*, San Francisco, 1-7 August 1999, Berkeley, University of California, 3 voll., vol. 2, 1161-1164.

Prieto, C., D'Imperio, M. & Gili Fivela, B. (2005). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. In *Language and Speech*, 48(4), 359-396.

Rizzi, L. (1997). The fine structure of the left periphery. In Haegeman, L. (Ed.), *Elements of Grammar*. Dordrecht et alibi: Kluwer, 281-337.

Rizzi, L. (2001). On the position "Int(errogative)" in the left periphery of the clause. In Cinque, G., Salvi, G. (Eds.), *Current Studies in Italian Syntax. Essays offered to Lorenzo Renzi*. Amsterdam: Elsevier North-Holland, 287-296.

Rizzi, L. (2004). *The Structure of CP and IP. The Cartography of Syntactic Structures 2*. New York: Oxford University Press.

Truckenbrodt, H. (1995). Phonological phrases: Their relation to syntax, focus and prominence. PhD Dissertation, MIT, Cambridge, MA.

DOI: 10.17469/O2102AISV000012

FRANCESCO OLIVUCCI, FILIPPO PASQUALETTO, MARIO VAYRA, CLAUDIO ZMARICH

Lo sviluppo dell'accento lessicale nel bambino in età prescolare: una prospettiva fonetico-acustica

Few studies have addressed the ability of young children to produce stressed and unstressed syllables. Most of them focused on English speaking children and have adopted an acoustic analysis limited to the parameters of duration, intensity and F0. Even less investigated is the development of lexical stress in Italian children, and of the few studies found in the literature, none has looked at parameters such as formant values or spectral emphasis. Our aim is to investigate how two years old Italian children produce and develop lexical stress, observing their productions longitudinally over a six months period. We will also compare children and adults' productions.

Key words: lexical stress, infants, Italian, development, acoustic analysis.

1. *Stato dell'arte*

Sono numerosi gli studiosi che hanno indagato le proprietà fonetico-acustiche che distinguono le sillabe toniche dalle sillabe atone nel parlato adulto in lingua italiana. I primi lavori che hanno esaminato le caratteristiche acustiche che rendono una sillaba prominente sulle altre nel parlato adulto si sono concentrati principalmente sui parametri acustici di durata vocalica, intensità e frequenza fondamentale. Per l'italiano, tra i primi studiosi ad analizzare a fondo il fenomeno attraverso i tre parametri citati sopra dobbiamo ricordare Bertinetto (1981; 1985), e, più di recente, D'Imperio, Rosenthall (1999).

Un'altra linea di studi ha invece esaminato gli effetti globali della prominenza oltre che sulla durata e/o sull'intensità anche sulla struttura spettrale della vocale, ossia sui suoi valori formantici (cfr. Farnetani, Kori, 1982, con rif.; Vayra, Fowler, 1987, con rif.; Albano Leoni, Cutugno & Savy, 1995; Savy, Cutugno, 1997, con rif.; Vayra, Avesani & Fowler, 1999, con rif.). Complessivamente i risultati mostrano un maggior grado di apertura delle vocali toniche rispetto alle atone per quanto concerne le vocali basse (che presentano una F1 più alta) e un maggiore avanzamento delle toniche rispetto alle atone nelle vocali anteriori (che presentano una F2 più alta).

Un ulteriore parametro acustico associato al grado di prominenza prosodica è stato recentemente identificato nell'*enfasi spettrale* (Tamburini, 2009; Bocci, Avesani, 2011). Per quanto non vi sia totale accordo sulle procedure sperimentali con cui misurare l'enfasi spettrale di una vocale, pare tuttavia che vocali prosodicamente più prominenti presentino maggiore intensità nelle frequenze alte dello spettro. I principali

parametri utilizzati dagli studiosi per indagare il grado di enfasi spettrale sono s*pectral balance* e s*pectral tilt*.

Con *spectral balance* (Sluijter, van Heuven, 1996) si intende la differenza, in dB, tra i valori medi di intensità all'interno di quattro bande di frequenza dello spettro, generalmente identificate con le bande 0-0,5 kHz, 0,5-1 kHz, 1-2kHz e 2-4kHz. La scelta dei limiti superiore e inferiore di ciascuna delle bande di frequenza è funzionale alla necessità di far rientrare in ciascuna di esse una formante. All'interno della prima banda si troverà dunque F0, nella seconda F1 e così di seguito.

Con *spectral tilt* invece si intende generalmente la differenza tra l'intensità dell'armonica più vicina al picco di F1 e quella dell'armonica più vicina al picco di F2 (cfr. Fulop, Kari & Ladefoged, 1998). Altri studiosi (cfr. Hanson, 1997) hanno calcolato lo *spectral tilt* come la differenza, in dB, tra l'intensità della prima armonica e l'intensità dell'armonica più vicina al picco di F3. In entrambi i casi gli Autori utilizzano una procedura di normalizzazione nel calcolo dello *spectral tilt*, volta ad eliminare possibili effetti della qualità acustica della vocale sull'enfasi spettrale. La normalizzazione agisce dunque sulla variazione nella distribuzione delle formanti in vocali di diverso colore spettrale.

Uno studio acustico sui correlati acustici dell'accento, che comprende l'analisi dell'enfasi spettrale insieme a durata, intensità e frequenza fondamentale, è quello di Ortega-Llebaria, Prieto (2011) sullo spagnolo e il catalano. Per quanto concerne lo *spectral tilt*, i valori ricavati attraverso una procedura di normalizzazione del tipo descritto sopra, vengono qui comparati con valori non normalizzati. Dal momento che in quest'ultimo caso emergono differenze significative tra vocali toniche e atone, le Autrici suggeriscono che differenze nello *spectral tilt* determinate da differenze nelle frequenze formantiche delle vocali possano essere neutralizzate dal processo stesso di normalizzazione adottato.

Infine, per quel che riguarda la produzione dell'accento di parola, altri studiosi hanno identificato sia indici palatografici associati alla prominenza accentuale (Farnetani, Vayra, 1996; Farnetani, 1997) sia parametri cinematici (Vayra, Fowler, 1992; Magno Caldognetto, Vagges & Zmarich, 1995; Avesani, Vayra & Zmarich, 2007).

Complessivamente questi dati confermano l'ipotesi che, nel parlato adulto, all'accento lessicale corrisponda un incremento globale dello sforzo articolatorio (Fowler, 1995), incremento che si riflette sulle caratteristiche sia acustiche che articolatorie delle toniche, differenziandole dalle atone.

Sul piano articolatorio, l'attività elettromiografica dei muscoli implicati nella fonazione risulta più intensa per le sillabe toniche. Anche l'attività dei muscoli intercostali è maggiore, causando un aumento del flusso espiratorio. I gesti con cui le vocali toniche sono prodotte hanno durata maggiore, maggiore ampiezza e maggiore velocità massima. Lo stadio stabile delle vocali, infine, è tenuto più a lungo quando la vocale è inserita in una sillaba tonica (cfr. Avesani, Vayra & Zmarich, 2009).

Sul piano acustico questo scenario si riflette in una maggiore durata vocalica e sillabica delle sillabe toniche. La frequenza fondamentale della vocale che compone il nucleo di una sillaba tonica è generalmente più alta rispetto a quella di una sillaba atona

e presenta intensità maggiore. Le vocali toniche hanno inoltre valori formantici meno centralizzati e valori maggiori di intensità alle alte frequenze dello spettro rispetto alle vocali atone.

Sono state inoltre individuate differenze tra sillabe toniche e atone nel grado di coarticolazione consonante-vocale. Nelle sillabe toniche, al *rafforzamento articolatorio* di cui sono soggette corrisponderebbe una maggiore *resistenza coarticolatoria* rilevabile nel minor grado di influenze reciproche tra consonante e vocale (cfr. Fowler, 1981; per un'analisi del fenomeno nella lingua italiana cfr. Zmarich, Avesani & Marchiori, 2007; Zmarich, Avesani, 2015).

Se i lavori volti allo studio delle proprietà acustiche che distinguono sillabe toniche da atone nel parlato adulto sono numerosi, lo stesso non si può dire dei lavori dedicati allo sviluppo dell'accento lessicale nel parlato dei bambini. I lavori disponibili riguardano principalmente la lingua inglese.

Pollock, Brammer & Hageman (1993) indagano le proprietà acustiche di sillabe toniche e atone in bambini anglofoni di 2, 3 e 4 anni di età attraverso i parametri di durata, intensità e frequenza fondamentale. Gli autori rilevano come a partire dai due anni di età i bambini utilizzino il parametro di durata per realizzare la prominenza lessicale, ma riescano a padroneggiare in produzione i parametri di intensità e frequenza fondamentale soltanto a partire dai tre anni. Questi risultati lascerebbero pensare ad uno sviluppo dell'accento lessicale compreso in un periodo che va dal secondo al terzo anno di vita del bambino.

Gli stessi parametri analizzati da Kehoe, Stoel-Gammon & Buder (1995) inducono invece a conclusioni diverse suggerendo come fin dai 18 mesi di età i bambini riescano a padroneggiare bene ciascuno dei tre parametri analizzati.

Si potrebbe pensare che tale differenza nei risultati dei due lavori presentati sopra sia dovuta alla procedura utilizzata: nel lavoro di Pollock e colleghi, infatti, i target analizzati sono stati ottenuti attraverso la ripetizione di non-parole, mentre nel lavoro di Kehoe e colleghi sono stati estratti dal parlato spontaneo dei soggetti. Tuttavia in uno studio successivo di Schwartz, Petinou, Goffman, Lazowski & Cartusciello (1996), in cui si è utilizzata una procedura simile a quella di Pollock, gli Autori rilevano come fin dai due anni di età i bambini di lingua inglese padroneggino correttamente ciascuno dei tre parametri analizzati.

Se non sono tanti i lavori che riguardano lo sviluppo dell'accento lessicale nel bambino di lingua inglese, ancora meno sono i lavori che si occupano di bambini italiani. Tra i pochi contributi disponibili possiamo citare il lavoro di Arciuli, Colombo (2016). Le Autrici analizzano le produzioni di bambini di età compresa tra i 3 e i 7 anni per delineare delle traiettorie di sviluppo nella capacità di produrre sillabe toniche e sillabe atone in parole con un inizio trocaico o giambico. L'obiettivo delle Autrici è di individuare eventuali differenze tra bambini e adulti nella produzione della prominenza lessicale in due diversi *pattern* ritmici: SW (forte-debole), tipico delle parole con incipit trocaico e WS (debole-forte), tipico delle parole con inizio giambico, presente nel più comune tra i trisillabi italiani (come ad esempio [pa'tata]).

Sebbene le Autrici non trovino differenze nella produzione dell'accento lessicale da parte di bambini e adulti, dati interessanti emergono dall'analisi della capacità dei bambini di produrre la prominenza accentuale in sillabe chiuse da consonante geminata: qui il comportamento dei bambini si distingue da quello degli adulti.

Per ultimo citiamo uno studio di Pasqualetto (2015), che condivide molte caratteristiche con il presente lavoro. L'autore analizza le produzioni lessicali, ottenute attraverso un compito di denominazione di figure (TFPI, cfr. Zmarich, Fava, Del Monego & Bonifacio, 2012), di 31 soggetti suddivisi in 5 fasce d'età dai 24 ai 37 mesi[1], selezionate per aver conservato la stessa struttura sillabica dei corrispondenti target adulti (es. [ʃiˈbaːe] per [tʃaˈbaːe]), ed escludendo le parole bisillabiche. Per ogni parola vengono misurati i valori relativi a durata vocalica, durata sillabica, picco di F0, picco di intensità, sia per la sillaba tonica che per l'adiacente sillaba atona. Il parametro di durata si rivela fin dai 24 mesi il più utilizzato, e statisticamente il più significativo, nella distinzione tra sillabe toniche ed atone. I dati relativi al picco di frequenza e di intensità delle vocali in sillaba tonica ed atona risultano invece meno sistematici e non significativi.

Quello che emerge dalla letteratura sullo sviluppo fonetico-fonologico dell'accento è dunque che, mentre per quanto concerne bambini anglofoni è disponibile un certo numero (seppure non vastissimo) di studi, sono pochi i lavori su bambini aventi l'italiano come L1.

Inoltre tutti i lavori che trattano dello sviluppo dell'accento si concentrano sui parametri di durata, intensità e frequenza fondamentale che, sebbene importanti, sono soltanto alcune delle proprietà acustiche considerate oggi rilevanti dalla letteratura sulla prominenza prosodica nell'adulto. Più specificamente, nessun lavoro sul bambino fa riferimento a valori formantici o enfasi spettrale delle vocali.

Un'ultima considerazione, non meno rilevante delle precedenti, riguarda l'età dei bambini esaminati. L'età studiata da Arciuli, Colombo (2016) è relativamente avanzata (3-7 anni): in linea di principio, quindi, eventuali discontinuità rispetto ai *pattern* acustici del parlato adulto potrebbero essere scomparse nell'arco di tempo considerato, pur essendo presenti in età più precoci.

L'obiettivo del presente lavoro è di studiare la capacità dei bambini, nel periodo intorno ai due anni di età, di produrre sillabe toniche e atone. Esamineremo, attraverso un confronto con il parlato degli adulti, il ruolo svolto da parametri acustici come durata, intensità, valori formantici ed enfasi spettrale, al fine di individuare eventuali differenze con il *target* adulto.

[1] Il campione normativo, costituito da 31 soggetti di età compresa tra i 25 e i 37 mesi con sviluppo linguistico tipico, è stato selezionato da materiale già registrato (attraverso la somministrazione del TFPI) per studi precedenti. In particolare dai lavori di: Zmarich, Fava, Del Monego & Bonifacio (2012), 30 soggetti; Bossetti (2012) e Seccafien (2013), 20 soggetti; Pigato (2014), 2 soggetti.

2. *Metodo*

2.1 Soggetti

Per il presente lavoro ci si è avvalsi di materiali raccolti in precedenza e già utilizzati in studi precedenti (Zmarich, Bonifacio, 2004; 2005). Gli Autori dei suddetti lavori, in fase di raccolta dati, si erano assicurati che i soggetti registrati avessero avuto uno sviluppo psicomotorio regolare documentato da un pediatra. Inoltre, a 18 mesi ciascun soggetto era stato sottoposto ad esami delle funzionalità uditive, volti ad escludere deficit uditivi e ad una visita logopedica e alla somministrazione del questionario genitoriale *Primo Vocabolario del Bambino* (PVB, Caselli, Casadio, 1995). Le registrazioni erano state fatte su parlato spontaneo elicitato attraverso la denominazione di giocattoli o di immagini presenti in libri per bambini e avevano una durata che andava dai 30 ai 45 minuti.

Tra i soggetti di tale corpus ne sono stati selezionati cinque tra i più loquaci e le cui registrazioni avessero una buona qualità di audio, in modo da poter contare su un buon numero di target validi anche per le tappe di sviluppo più precoci. Per ciascuno dei cinque soggetti, tre dei quali triestini (VL, BS e FS), e due padovani (ZD e ZA) sono state digitalizzate le registrazioni relative ai 15, 18, 21, 24 e 27 mesi: tali registrazioni erano infatti ancora in formato analogico su nastri magnetici (per i soggetti di TS) e cassette audio (per i soggetti di PD). In fase di selezione dei target idonei secondo le modalità descritte di seguito, ci si è accorti della scarsità di target validi per alcuni soggetti alle tappe dei 15 e 18 mesi. Secondo i criteri che ci eravamo imposti per valutare valida una produzione, infatti, a 15 mesi non vi erano target validi per nessuno dei soggetti selezionati. Anche a 18 mesi, perfino i soggetti più loquaci presentavano un numero esiguo di produzioni considerabili valide. Anziché scegliere di modificare i nostri criteri di selezione dei target si è quindi deciso di limitare il nostro studio sui cinque soggetti in questione alle sole tappe di sviluppo di 21, 24 e 27 mesi.

Per poter contare su un confronto con il parlato adulto si è proceduto a selezionare e a registrare 4 soggetti adulti che fungessero da gruppo di controllo. I soggetti adulti, tutte donne (quindi con una F0 più simile, rispetto agli uomini, alla F0 di bambini) di età media 31 anni (la più giovane aveva 22 anni, la più "anziana" 52), sono stati scelti in modo da essere di origine triestina (CC e GG) e di origine padovana (MO e RG). In questa maniera si è mantenuto anche nel gruppo di controllo l'eterogeneità relativa alle varietà regionali parlate dai soggetti del nostro studio. Ai soggetti adulti è stato sottoposto un questionario per accertare l'assenza di disturbi linguistici. Tutti i soggetti del gruppo di controllo hanno dichiarato di avere avuto uno sviluppo linguistico nella norma. Uno solo dei soggetti (CC) ha dichiarato di non aver mai imparato il suo dialetto locale, mentre gli altri hanno dichiarato di utilizzare saltuariamente, di preferenza nelle situazioni familiari, il loro dialetto di origine.

2.2 Materiali e procedura

Le registrazioni relative ai bambini sono state attentamente esaminate e all'interno dell'alto numero di produzioni che ciascuna presentava sono state selezionate come

target validi per il nostro studio soltanto le produzioni che rispettavano alcune proprietà. Innanzitutto la produzione doveva essere spontanea: questo ha portato ad escludere dallo studio qualsiasi produzione infantile avvenisse su ripetizione immediata del parlato adulto. Inoltre la produzione doveva coincidere, per struttura accentuale e per numero di sillabe, con il target adulto. Allo stesso modo, sono state scartate tutte le parole in cui la struttura di una delle sillabe che la componevano fosse diversa dalla corrispondente pronuncia adulta. Nel caso la parola in questione fosse un bisillabo è stata accettata soltanto se non era prodotta in isolamento (ovvero soltanto se era seguita da un'altra parola). Questa scelta è stata dovuta al fatto che si voleva poter sempre contare su un confronto tra una sillaba tonica e una atona nella stessa parola. Poiché nel caso dei bisillabi, tutti piani, il confronto sarebbe stato fatto obbligatoriamente tra la penultima sillaba, tonica, e l'ultima, atona, si voleva evitare che quest'ultima fosse influenzata dalla presenza di un confine di enunciato che avrebbe causato allungamento della vocale in questione e un profilo discendente di F0 (cfr. Snow, 1997). Sono state invece accettate parole isolate di lunghezza superiore a due sillabe. In conclusione il numero di target validi per il presente studio individuati alle varie tappe di ogni soggetto va da un minimo di 3 target per i 21 mesi di FS ad un massimo di 76 target per i 27 mesi di BS. Le parole individuate sono in prevalenza trisillabi (piani o sdruccioli), molti bisillabi e pochi quadrisillabi (piani e sdruccioli). Trattandosi di parlato spontaneo le parole ritenute valide per l'analisi variano da soggetto a soggetto. Benché sia possibile individuare alcune parole ricorrenti, ogni registrazione relativa ad una tappa di sviluppo di ciascun soggetto presenta un numero diverso di target validi e le parole target di ogni soggetto sono diverse.

Per poter allestire un corpus relativo al gruppo di controllo che fosse comparabile a quello dei bambini sono state individuate 47 parole, scelte tra bisillabi e trisillabi piani e sdruccioli, nell'insieme delle parole più frequenti pronunciate dai bambini. In questa maniera si sarebbe potuto contare su un confronto diretto tra il gruppo di controllo e ogni soggetto per ciascuna tappa di sviluppo: vi era infatti sempre almeno una parola per ogni tappa di ogni soggetto presente anche nel corpus delle 47 del gruppo di controllo. I soggetti adulti sono stati registrati in camera anecoica nella sede dell'ISTC-CNR di Padova con registratore Edirol R-09. Le parole da pronunciare erano proiettate dallo schermo di un computer e l'avanzamento della presentazione era controllato direttamente dal soggetto. Per ogni soggetto è stato randomizzato l'ordine di presentazione delle parole. Per evitare gli effetti del confine di enunciato sulla sillaba precedente, i bisillabi sono stati inseriti in frasi del tipo "Dico X chiaramente". Ai soggetti veniva chiesto di leggere a voce alta ciò che compariva sullo schermo in maniera chiara, scandendo bene e a voce alta. Dopo una breve frase di addestramento cominciava il test vero e proprio. Ogni soggetto è stato sottoposto a tre sessioni di registrazione, separate da una breve pausa: in questa maniera, anche se una delle tre registrazioni fosse stata di scarsa qualità, si sarebbe potuto comunque contare su almeno due diverse registrazioni di ciascun target per ogni soggetto.

Il materiale così raccolto, sia per il gruppo di studio che per il gruppo di controllo, è stato segmentato e annotato a mano attraverso Praat.

Le misure acustiche a cui fa riferimento il presente lavoro sono state estratte automaticamente dalle registrazioni attraverso *script* di Praat.

2.3 Misure acustiche

In una fase precedente del presente lavoro, che ha portato alla discussione della tesi di laurea magistrale del primo autore, si erano indagati i tre parametri di durata, intensità ed F0 delle vocali toniche e atone.

I risultati ottenuti, tuttavia, hanno condotto ad alcune considerazioni. Innanzitutto i risultati relativi al parametro di F0 ci hanno suggerito che, anche per il bambino, i valori di F0 associati alla sillaba accentata fossero condizionati dall'intonazione globale dell'enunciato. Abbiamo ritenuto pertanto che lo studio del parametro F0 associato all'accento lessicale non potesse prescindere da un'analisi intonativa dell'enunciato. Non essendovi i prerequisiti per individuare all'interno del nostro corpus di dati un numero di enunciati sufficientemente alto per un'analisi statistica con profili intonativi assimilabili, alla fine si è scelto di escludere il parametro di F0 dalla nostra analisi.

Inoltre, in accordo con la bibliografia esistente sul parlato adulto, si è scelto di aggiungere nella fase del lavoro successiva alla discussione di tesi di laurea le analisi dei valori formantici e dell'enfasi spettrale.

2.3.1 Durata

Attraverso uno *script* di Praat creato dal primo autore sono stati estratti, da ciascuna registrazione segmentata e annotata a mano, due diversi valori relativi alla durata: la durata di ciascuna vocale e quella di ciascuna sillaba. In questa fase si presenteranno soltanto i risultati relativi alla durata vocalica, studiata anche da Arciuli e Colombo (2016), in quanto quelli relativi alla durata sillabica (Olivucci, 2015) davano risultati praticamente sovrapponibili ai primi in termini di differenza fra valori relativi ai contesti di sillaba tonica e sillaba atona.

2.3.2 Intensità

Oltre ai valori di durata, per ogni vocale è stata calcolata l'intensità attraverso due diverse procedure. Una prima misura era rappresentata dal picco di intensità all'interno dei confini vocalici, mentre una seconda misura era rappresentata dalla media dell'intensità in una frazione della vocale pari al 40% della sua durata totale centrata sul picco di intensità. Come nel caso dei valori relativi alla durata, anche per quanto riguarda l'intensità le due misure hanno dato risultati sovrapponibili. In questa sede, dunque, presenteremo soltanto i risultati relativi al picco di intensità nella vocale.

2.3.3 Valori formantici

Per ciascuna vocale, tonica o atona, del nostro corpus, sono stati estratti i valori di F1 ed F2. Tali valori sono stati calcolati come media, in Hz, della frequenza relativa a ciascuna formante all'interno di un intervallo centrato sul punto medio della vocale e di durata pari al 40% del totale.

2.3.4 Enfasi spettrale

Lo *script*, nella sua forma definitiva, è stato progettato per estrarre due misure di enfasi spettrale relative alla pendenza dello spettro: lo *spectral balance* e lo *spectral tilt*. Per entrambe si è fatto riferimento a Bocci, Avesani (2011).

Lo *spectral balance* è una misura della differenza fra l'intensità dello spettro alle bande a bassa frequenza e quelle ad alta frequenza. Lo spettro viene tradizionalmente diviso in bande contenenti F0 e le prime tre formanti. I confini di ogni banda usati nella bibliografia, tuttavia, fanno riferimento ai valori medi delle formanti nelle vocali del parlato adulto. Per il parlato infantile era necessario apportare alcune modifiche. Facendo quindi riferimento ad Assmann, Katz (2000), si sono individuati, per ciascuna delle sette vocali della lingua italiana, nuovi confini di frequenza validi per le prime tre bande dello spettro (contenenti quindi F0, F1 e F2). All'interno di ciascuna banda è stata poi calcolata l'intensità media dello spettro (ottenendo in questa maniera B1, B2 e B3).

Senza entrare nel dettaglio delle modalità con cui è stato calcolato lo *spectral tilt* (modalità per le quali si rimanda a Fulop et al., 1998) basterà dire che esso rappresenta una misura di enfasi spettrale normalizzata rispetto alla qualità vocalica. Poiché la posizione delle formanti sull'asse delle frequenze influenza l'enfasi spettrale (più alti sono i valori delle formanti sull'asse delle frequenze, più alti saranno i valori di intensità, dato che si associano alle alte frequenze dello spettro) per il calcolo dello *spectral tilt* si procede in tre fasi distinte. Innanzitutto si calcola la differenza reale (A1-A2), in dB, tra l'intensità relativa all'armonica più vicina al picco di F1 e quella più vicina al picco di F2. In un secondo momento si stima un valore ipotetico di A1-A2 calcolato sulla base di un modello che tiene conto soltanto dell'apporto delle formanti. La differenza tra la misurazione reale e la stima ipotetica rappresenta lo *spectral tilt*, che risulta in questa maniera normalizzato sui valori delle formanti. Tuttavia la necessità di adattare le formule utilizzate in Fulop et al. (1998) al segnale acustico del parlato infantile ha richiesto di apportarvi alcune modifiche. Innanzitutto Fulop e colleghi utilizzano nelle loro formule larghezze di banda relative alle formanti stimate sul segnale acustico del parlato adulto. In particolare stimano una larghezza di banda di 30 Hz per F1, 80 Hz per F2 e 150 Hz per F3. Tali valori, tuttavia, non risultano validi nel caso le formanti si trovino a frequenze maggiori rispetto a quelle considerate standard per gli adulti, come nel caso del parlato infantile. Si è preferito dunque, in questa sede, non utilizzare valori fissi di larghezza di banda, ma calcolarli per ciascuna formante di ciascuna vocale attraverso la formula riportata in Fant (1971):

(1)

$$B_n = \frac{Fn}{2\pi}$$

Dove B_n è la larghezza di banda e F_n la frequenza della formante in questione.

Un'ulteriore modifica apportata alle formule di Fulop et al. (1998) riguarda la frequenza di risonanza del tratto vocale. Infatti gli Autori considerano una lunghezza di 17,5 cm per il tratto vocale, tipica del tratto vocale adulto, con una frequenza di risonanza pari a 506 Hz. Dovendosi applicare al parlato di bambini di circa 2 anni di età la frequenza di risonanza doveva essere riconsiderata a partire da una lunghezza del tratto vocale pari a 10

cm (cfr. Boë, Granat, Badin, Autesserre, Pochic, Zga, Henrich & Ménard, 2006) e quindi essere modificata a 875 Hz.

3. *Risultati*

3.1 Durata

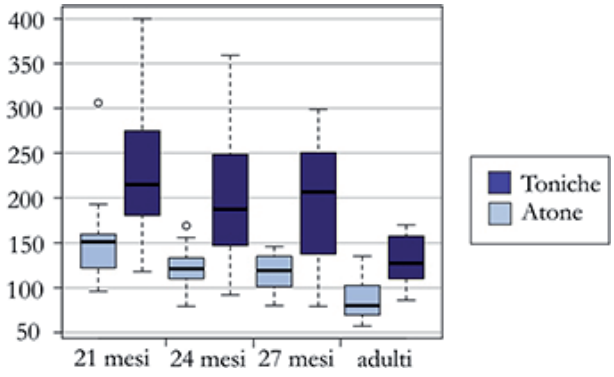
In una fase preliminare sono state indagate eventuali differenze tra le durate vocaliche in sillaba tonica aperta e chiusa. Per ciascuna tappa di sviluppo di ogni soggetto e per ciascuna registrazione di ogni adulto del gruppo di controllo, è stato calcolato attraverso un *two sample t-test* il grado di significatività della differenza tra le durate vocaliche in sillaba tonica aperta e chiusa. Per quanto riguarda i bambini le sillabe toniche aperte presentavano sempre durata vocalica maggiore rispetto alle sillabe chiuse: per i 21 mesi tale differenza era mediamente pari a 21 ms, per i 24 mesi era pari a 34 ms e per i 27 mesi 18 ms. Tuttavia la differenza di durata vocalica nei due diversi contesti non era significativa per nessuno dei soggetti analizzati a nessuna delle tappe di sviluppo. Per quanto riguarda gli adulti la situazione era simile: dei quattro soggetti analizzati uno soltanto presentava una differenza statisticamente rilevante tra sillabe toniche aperte e chiuse (GG, durata vocalica sillabe toniche aperte: 160 ms, durata vocalica sillabe toniche chiuse: 123 ms; t(85,402)=-5,267, *p-value*<0,001**), mentre gli altri tre soggetti non presentavano differenze statisticamente rilevanti. Almeno per quanto riguarda i soggetti adulti ci si sarebbe potuti aspettare differenze più rilevanti nei due diversi contesti (cfr. Vayra, Avesani & Fowler, 1999). Ulteriori studi sarebbero necessari per indagare le cause dei nostri risultati. Una possibile spiegazione potrebbe essere da ricercare nella varietà regionale dei nostri soggetti, tutti provenienti da aree che presentano forti fenomeni di degeminazione (cfr. Telmon, 1993). Dato che, all'interno del nostro corpus, nella maggioranza schiacciante dei casi la chiusura della sillaba era da attribuire alla presenza di una consonante geminata, è ipotizzabile che fenomeni come l'indebolimento della consonante geminata abbiano influito sul grado di accorciamento vocalico e, a livello statistico, abbiano indebolito la significatività della differenza di durata vocalica tra i contesti aperto/chiuso. Ad ogni modo, per il presente studio, al fine di evitare un impoverimento dei dati disponibili, si è scelto di incorporare i dati relativi a sillabe aperte e quelli relativi a sillabe chiuse.

Dai nostri dati emerge che fin dai 21 mesi le sillabe toniche presentano durata vocalica maggiore rispetto alle sillaba atone. Un primo calcolo del grado di significatività di tale differenza è stata un'analisi della varianza con misure ripetute (ANOVA), calcolata utilizzando il software SYSTAT, in cui ogni soggetto è rappresentato dalla media delle durate vocaliche, il fattore *between* è l'accento (tonica vs atona) e il fattore *within* la tappa di sviluppo (21 vs 24 vs 27 mesi). L'analisi della varianza ha confermato un alto grado di significatività nella differenza tra sillabe toniche e atone (F-ratio=36,263; *p-value*<0,001**. Significativo è anche il fattore *within* (tappe di sviluppo, F-ratio= 9,135, p-value=0,001**): significatività probabilmente dovuta al fatto che la velocità di locuzione è significativamente più alta a 27 mesi che a 24 e a 21 mesi. Non era significativa, invece, l'interazione dei due fattori.

Tuttavia, data l'eterogeneità dei dati, si è resa necessaria un'ulteriore conferma di questo risultato che tenesse conto dell'alto numero di fattori coinvolti dall'elaborazione di un *linear mixed model* (LMM, cfr. Baayen, Davidson & Bates, 2008; Jaeger, 2008). Utilizzando il software R si è stimato un modello della durata vocalica con il fattore "accento" e "tappa di sviluppo" come *fixed effects* e "soggetto", "tipo sillabico" (sillaba aperta vs chiusa) e "posizione della sillaba nella parola" come *random effects*. Il modello stimato conferma che vocali in sillaba tonica sono, nei bambini, mediamente più lunghe di 83 ms rispetto alle atone. La significatività di questo risultato è confermata dal confronto tra il modello reale e un ulteriore modello nullo (che quindi non tiene conto del fattore "accento", cfr. De Boeck, Bakker, Zwitzer, Nivard, Hofman, Tuerlinckx, Partchev, 2011), attraverso il quale emerge che il modello reale descrive meglio i nostri dati con un *p-value*<0,001**.

Anche per gli adulti la differenza in termini di durata vocalica tra sillabe toniche e atone risulta significativa. In particolare un'analisi della varianza in cui ogni soggetto è rappresentato dalla media dei valori di durata vocalica ha confermato che i valori relativi alle sillabe toniche sono maggiori rispetto a quelli relativi alle sillabe atone con una significatività di p<0,001** (F-ratio=9,135). Un LMM elaborato utilizzando come *fixed effect* il fattore "accento" e come *random effects* i fattori "soggetto", "tipo sillabico" (aperto vs chiuso) e "posizione della sillaba nella parola" ha stimato che le sillabe toniche hanno vocali mediamente più lunghe di 52 ms rispetto alle atone. Il confronto con un modello nullo (che non utilizza quindi l'accento come fattore) ha confermato che il modello reale descrive meglio i dati con p<0,001**.

Figura 1 - *Valori medi e deviazione standard della durata vocalica (in ms) in sillabe toniche e atone di adulti e bambini suddivisi per tappa di sviluppo*



3.2 Intensità

I valori relativi al picco di intensità all'interno del nucleo sillabico, divisi per tappe di sviluppo, si sono rivelati sistematicamente maggiori in sillaba tonica piuttosto che in sillaba atona

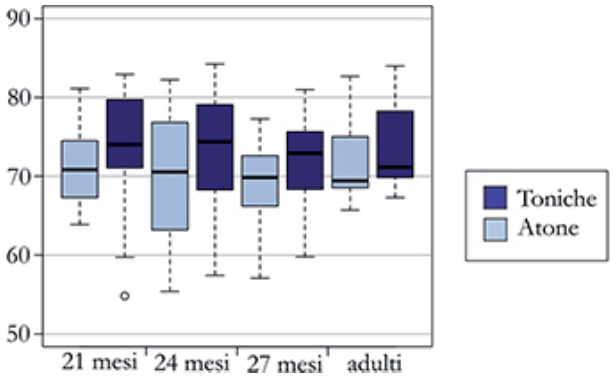
Un'analisi della varianza con misure ripetute, operata sui dati relativi ai bambini utilizzando come fattore *between* l'accento e come fattore *within* le tappe di sviluppo, ha rivelato la significatività di tale differenza (F-ratio=5,182; p-value=0,03*). Il fattore "tappe di svi-

luppo" non si è invece rivelato significativo, così come non significativa era l'interazione tra i due fattori.

Per tenere conto di tutti i fattori coinvolti si è elaborato, anche per i dati relativi all'intensità, un LMM, utilizzando come *fixed effect* l'accento e come *random effects* il soggetto in questione, la tappa di sviluppo, il tipo sillabico (aperto vs chiuso) e la posizione della sillaba all'interno della parola. Dal modello reale è emerso che nei bambini le sillabe toniche presentano mediamente un'intensità maggiore rispetto alle sillabe atone di 2,8 dB. Il confronto con il modello nullo ha confermato che il modello reale descrive meglio i risultati con p=0,002**.

Anche per quanto riguarda gli adulti l'analisi della varianza a misure ripetute ha confermato che le sillabe toniche hanno intensità significativamente maggiore rispetto alle atone (F-ratio=18,499; p-value<0,001**). L'elaborazione di un LMM con *fixed effect* l'accento e *random effects* il soggetto, il tipo sillabico e la posizione della sillaba all'interno della parola ha confermato che le sillabe toniche negli adulti presentano intensità mediamente superiore di 1,2 dB rispetto alle sillabe atone, mentre il confronto con il modello nullo ha evidenziato che il modello reale descrive meglio i nostri dati con un p=0,01*.

Figura 2 - *Valori medi e deviazione standard dell'intensità vocalica (in dB) in sillaba tonica e atona in adulti e bambini divisi per tappa di sviluppo*



3.3 Valori formantici

Fino a questo punto abbiamo analizzato i dati relativi alla durata e all'intensità senza tener conto del tipo di vocale in questione. Essendo il nostro un corpus di parlato spontaneo ed essendo quindi impossibile elicitare le stesse produzioni ad ogni tappa, si è scelto di unire i dati relativi a diverse vocali al fine di contare su un ampio numero di dati su cui impostare le nostre analisi statistiche. Questo era ovviamente impossibile per quanto riguarda i valori formantici, valori che risentono fortemente del tipo di vocale prodotta. La necessità di dividere i dati e analizzarli separatamente per ciascuna vocale ha inevitabilmente impoverito i dati a disposizione obbligandoci a scartare le analisi relative alle vocali meno frequenti del corpus. Le vocali più frequenti e quindi quelle su cui si è concentrata la nostra analisi sono le vocali /a/ e /i/.

I dati relativi a F1 e a F2 ad ogni tappa di sviluppo e relativi agli adulti sono i seguenti.

Tabella 1 - *Medie e deviazioni standard (tra parentesi) dei valori di prima e seconda formante per le vocali /a/ e /i/ divisi per tappa di sviluppo e prominenza*

		/a/		/i/	
		F1	F2	F1	F2
21					
	A	1017,96 (217,88)	2000,94 (142,60)	584,22 (163,75)	3306,83 (308,80)
	T	1175,74 (131,03)	2015,21 (132,08)	688,50 (153,40)	3096,44 (354,02)
24					
	A	834,18 (125,53)	1949,30 (168,57)	593,80 (117,16)	3019,71 (388,06)
	T	1077,02 (93,55)	1948,62 (141,36)	565,58 (114,74)	3248,43 (201,37)
27					
	A	919,83 (96,73)	2023,78 (159,10)	643,75 (399,64)	2856,22 (508,47)
	T	1203,65 (243,76)	1921,52 (211,78)	485,91 (70,02)	3318,74 (236,93)
adulti					
	A	760,67 (96,44)	1671,05 (89,62)	395,21 (57,42)	2423,87 (219,15)
	T	897,00 (120,03)	1590,14 (64,10)	362,14 (37,41)	2493,73 (254,35)

Dalla tabella 1 si può notare, sia per i bambini che per gli adulti, come la prima formante di /a/ presenti valori più alti in sillaba tonica che in sillaba atona. Tale tendenza si è rivelata significativa attraverso lo sviluppo di un LMM avente come *fixed effect* l'accento e come *random effects* il soggetto, il tipo sillabico (sillaba aperta vs chiusa), la tappa di sviluppo e la posizione all'interno della parola. Nei bambini l'LMM ha confermato per la vocale /a/ valori di F1 mediamente maggiori di 307 Hz quando erano in sillaba tonica, con una significatività, dovuta al confronto con il modello nullo, di $p<0{,}001^{**}$. Nei dati relativi agli adulti la vocale /a/ presenta una F1 mediamente maggiore di 178 Hz quando si trova in sillaba tonica, con una significatività dovuta al confronto con il modello nullo di $p<0{,}001^{**}$. L'analisi di F2 per la vocale /a/ ha invece restituito valori diversi per bambini e adulti: i bambini non mostrano differenze significative in termini di F2, mentre per gli adulti le sillabe toniche presentano valori significativamente inferiori rispetto alle atone ($p<0{,}001^{**}$).

Anche per quanto riguarda la vocale /i/ sono state rilevate differenze in termini di valori formantici tra adulti e bambini. I bambini non presentano a nessuna tappa differenze significative in termini di F1, mentre esibiscono valori di F2 maggiori nelle toniche ($p=0{,}03^{*}$). Per gli adulti, al contrario, i valori di F1 risultano minori in sillaba tonica ($p=0{,}045^{*}$), mentre non vi sono differenze significative in termini di F2.

3.4 Enfasi spettrale

Anche per le misure di enfasi spettrale, *spectral balance* e *spectral tilt*, è stato necessario separare i dati relativi a ciascuna vocale.

Per *spectral balance* si intende la differenza, in dB, fra i livelli di intensità distribuiti in tre bande contigue dello spettro (B1, B2 e B3). Calcolare la differenza algebrica tra B1 e

B2 e poi tra B2 e B3 avrebbe portato ad avere due diversi valori per rendere conto di un'unica misura. In questa sede si è preferito procedere calcolando, attraverso il metodo della regressione lineare, la pendenza della retta che meglio intercettasse i valori di intensità a B1, B2 e B3: tale pendenza avrebbe rappresentato una stima della pendenza dello spettro e avrebbe reso conto, con un unico valore, della *spectral balance*. Non per tutte le vocali erano disponibili dati in quantità sufficiente per un'indagine approfondita. Ad esempio, come si può notare sotto, nei dati non compaiono mai i valori relativi alle vocali toniche medio-basse /ɛ/ e /ɔ/: per queste vocali, infatti, si disponeva di uno scarso numero di esemplari per un confronto in sillaba atona (anche se nelle varietà regionali in questione è possibile reperire i due tipi vocalici sopra indicati anche in contesto di sillaba atona finale, cfr. Telmon, 1993). Inoltre, a causa di un fenomeno di innalzamento della vocale in sillaba tonica, tipico delle varietà regionali in questione (cfr. Telmon, 1993) anche in sillaba tonica le vocali sopra indicate risultano estremamente rare. Per ogni soggetto sono stati esclusi i dati relativi ad una vocale laddove non vi fossero almeno due differenti valori, l'uno relativo alla vocale tonica e l'altro relativo alla vocale atona.

L'analisi della *spectral balance* ha evidenziato interessanti differenze nei valori di enfasi spettrale tra le vocali in sillaba tonica e quelle in sillaba atona. Alla figura (3) sono riportati i valori dei coefficienti angolari delle rette ottenute da B1, B2 e B3 attraverso regressione lineare. È importante tenere presente che a valori maggiori dei coefficienti angolari corrispondono rette con pendenza inferiore e quindi con maggior enfasi alle alte frequenze dello spettro. Dai grafici in questione possiamo vedere come le vocali in sillaba tonica presentino tendenzialmente, sia per i bambini che per gli adulti, coefficienti angolari più alti rispetto alle vocali in sillaba atona, ovvero maggior intensità alle alte frequenze dello spettro. L'analisi della significatività di tale differenza nei dati riguardanti i bambini, calcolata attraverso l'LMM, con *fixed effects* l'accento e la tappa di sviluppo e con *random effects* i soggetti, il tipo di vocale e la posizione della parola all'interno dell'enunciato, ha fornito un *p-value*, ottenuto dal confronto con un modello nullo, inferiore a 0,001**. Per quanto riguarda i dati relativi al gruppo di controllo adulto l'analisi con LMM ha confermato valori di *spectral balance* più alti in sillabe toniche di 1,5 punti. Il confronto con il modello nullo ha restituito un *p-value*=0,006**.

L'analisi dello *spectral tilt* (ST) non ha richiesto una regressione lineare come nel caso dello *spectral balance*: Lo ST, infatti, rappresenta in un unico valore la differenza fra due diverse armoniche dello spettro, quelle più prossime rispettivamente a F1 e a F2. L'elaborazione di un LMM che utilizza come *fixed effect* l'accento e come *random effetcs* il soggetto, la tappa di sviluppo, il tipo di vocale e la posizione della parola all'interno dell'enunciato non ha evidenziato, nel confronto con il modello nullo, alcuna differenza statisticamente significativa, né per i bambini né per gli adulti.

Quindi sebbene i valori di *spectral balance* delle vocali differiscano tra sillaba tonica e sillaba atona, lo stesso non si è potuto dire dello *spectral tilt*, dove, oltre a non esservi differenze significative tra sillabe toniche e atone, non si è potuta neppure individuare una netta tendenza delle prime ad avere valori maggiori o minori rispetto alle seconde.

Figura 3 - *Valori medi e deviazione standard di spectral balance divisi per tipo di vocale. I grafici si riferiscono ai 21 mesi (in alto a sinistra), 24 mesi (in alto a destra), 27 mesi (in basso a sinistra) e agli adulti (in basso a destra)*

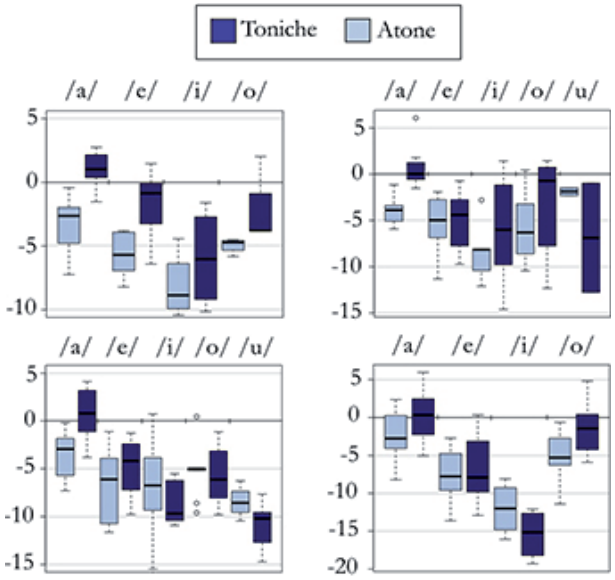
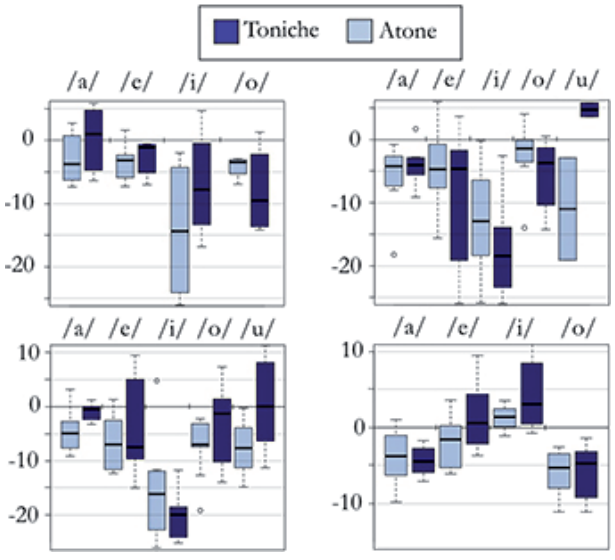


Figura 4 - *Valori medi e deviazione standard di spectral tilt divisi per tipo di vocale. I grafici si riferiscono ai 21 mesi (in alto a sinistra), 24 mesi (in alto a destra), 27 mesi (in basso a sinistra) e agli adulti (in basso a destra)*



4. *Discussione*

I risultati osservati relativi ai bambini dai 21 ai 27 mesi ci portano a concludere che, fin dalla tappa più precoce, i bambini possiedano una buona capacità di produrre sillabe toniche e sillabe atone distinte fra loro. Il confronto con il gruppo di controllo adulto ha infatti evidenziato come, astraendo dalle differenze di durata assoluta legate alla diversa velocità di elocuzione dei bambini, non vi siano differenze sostanziali in termini di parametri acustici tra le produzioni adulte e quelle infantili, e attraverso le varie tappe di sviluppo esaminate.

Il parametro di durata vocalica è senza dubbio quello che ha fornito i risultati più netti: le vocali in sillaba tonica presentano durata vocalica nettamente maggiore rispetto a quelle in sillaba atona, sia nei bambini a partire dai 21 mesi che negli adulti. I nostri dati trovano conferma anche nel citato lavoro di Pasqualetto (2015). Questo studio indaga le differenze di durata, intensità e frequenza fondamentale in sillabe toniche e atone in un gruppo di soggetti più ampio rispetto al nostro, ed aumenta le fasce d'età considerate includendo come tappe di sviluppo quelle di 24, 27, 30, 33 e 36 mesi. Lo studio di Pasqualetto non è tuttavia di tipo longitudinale, in quanto i soggetti variavano ad ogni tappa in funzione dei dati disponibili. Questo lavoro conferma i nostri dati relativi alla differenza tra sillabe toniche e atone, trovando differenze altamente significative in termini di durata fin dalla tappa più precoce, a volte con un rapporto di 2:1, mentre i dati relativi al picco di frequenza e di intensità delle vocali in sillaba tonica ed atona risultano meno sistematici e non significativi. Nello studio in questione l'analisi della varianza (ANOVA) evidenzia che i soggetti mostrano valori statisticamente significativi di durata sia per il fattore *within* (tonica vs atona, a prescindere dai gruppi di età) che per il fattore *between* (il fattore "gruppi di età", significativo a prescindere dal fattore "accento").

Risultati meno netti, ma comunque altamente significativi, sono stati ottenuti per il parametro di intensità massima all'interno della vocale. Anche in questo caso i dati relativi ai bambini non differiscono da quelli relativi agli adulti: entrambi i gruppi risultano capaci di differenziare la produzione di sillabe toniche e sillabe atone per quanto concerne il parametro intensità.

Ben più complessa risulta la discussione dei risultati relativi ai valori formantici. Per quanto riguarda la vocale bassa /a/, i risultati ottenuti relativi a F1 erano certamente attesi: trova conferma infatti la tendenza delle vocali basse ad essere associate ad una posizione della mandibola più bassa quando accentate (cfr. Avesani, Vayra & Zmarich, 2009, per uno studio cinematico su sillaba e accento in parlanti toscani; cfr. anche lo studio cinematico di Magno Caldognetto, Vagges & Zmarich, 1995 sui movimenti articolatori nella produzione di vocali toniche e atone in parlanti del nord-Italia). Questa tendenza vale sia per i bambini che per gli adulti, indice, questo, che il controllo del grado di apertura è già maturo a 21 mesi. L'analisi della seconda formante ha invece evidenziato come, negli adulti, ad un maggior grado di prominenza corrisponda una vocale bassa più arretrata (in armonia con quanto osservato in Zmarich, Avesani, 2015). Tale differenza tra vocali toniche e atone non è stata invece rilevata nei dati relativi ai bambini.

I valori relativi alla vocale /i/ indicano invece come gli adulti producano le vocali anteriori alte toniche come più alte, presentando in tale contesto valori minori di F1 (e nessuna differenza nei valori di F2), mentre i bambini come più avanzate, presentando valori maggiori di F2 (ma nessuna differenza nei valori di F1).

Come è noto, l'effetto dell'accento sul grado di apertura di una vocale alta è complesso. Esso infatti chiama in causa una compensazione articolatoria fra il movimento di abbassamento della mandibola associato alla fase di apertura della sillaba, e il compito di raggiungere il target fonologico (in questo caso il tratto [+alto] della /i/). Dai nostri dati relativi agli adulti emerge che le vocali alte tenderebbero ad alzarsi in sillaba tonica, mentre i bambini incontrerebbero maggiori difficoltà a controllare la coordinazione gestuale necessaria a raggiungere il target. Nella prospettiva della *Articulatory Phonology* il problema può essere formulato in termini di competizione fra gesti cooccorrenti: abbiamo una competizione fra il gesto di apertura orale, associato alla sillaba e ampliato in sillaba tonica, e il gesto di innalzamento del dorso della lingua associato alla vocale alta /i/. Questi gesti entrano in competizione per quanto riguarda la mandibola, che è un articolatore (secondario) condiviso da entrambi.

Dunque, il bambino incontrerebbe maggiore difficoltà a controllare la coordinazione gestuale necessaria per produrre una vocale accentata alta, sia rispetto all'adulto, sia rispetto alla produzione di una vocale accentata bassa come /a/. Con /i/ l'abbassamento (fonetico) della mandibola associato alla fase di apertura della sillaba tonica prevarrebbe sul gesto di innalzamento dorsale associato alla vocale alta accentata.

Sorprende invece la tendenza dei bambini a produrre come più avanzata la vocale anteriore alta /i/ in posizione tonica, quando negli adulti, forse nuovamente per un regionalismo, /i/ tonica non presenta alcun fenomeno di *enhancement* dell'anteriorità. Una spiegazione potrebbe risiedere nella diversa dimensione e configurazione del tratto vocale, come ci è stato opportunamente suggerito da un revisore. Si tratta senz'altro di un tema che merita un ulteriore approfondimento.

Per quanto riguarda i risultati relativi all'enfasi spettrale, i nostri dati presentano differenze significative tra sillabe toniche e atone in termini di *spectral balance*, ma non di *spectral tilt*. Questo dato potrebbe essere dovuto al fatto che, essendo lo *spectral tilt* una misura acustica normalizzata in funzione della posizione delle formanti sull'asse delle frequenze nello spettro di una vocale, tale normalizzazione potrebbe avere neutralizzato le differenze significative trovate in termini di *spectral balance*. Secondo questa ipotesi, tali differenze sarebbero dovute alla diversa struttura formantica delle vocali in sillaba tonica vs atona.

Da questa propettiva risulta estremamente interessante il confronto con l'analisi dei valori di *spectral tilt* in vocali toniche e atone dello spagnolo e del catalano, proposta recentemente in Ortega-Llebaria e Prieto (2011). Questo studio mostra che, sebbene non compaiano differenze significative tra gradi diversi di prominenza quando lo *spectral tilt* viene "corretto" (normalizzato), in funzione delle differenze fra frequenze formantiche, tuttavia differenze statisticamente significative, associate

a singoli tipi vocalici, emergono qualora nel calcolo della pendenza dello spettro si eviti il processo di normalizzazione.

In sostanza, i nostri risultati si allineano perfettamente con quelli discussi in Ortega-Llebaria e Prieto (2011), dove le differenze di *spectral tilt* sono interpretate come dipendenti principalmente da differenze nella distribuzione spettrale delle formanti, associate alla riduzione della vocale in sillaba atona.

Un'analisi più approfondita sulla relazione fra i valori formantici e i valori di enfasi spettrale in ciascuna vocale potrebbe dunque aiutarci a far luce sulla ratio sottostante agli esiti diversi delle due misurazioni.

In generale possiamo concludere che i bambini, fin dai 21 mesi, riescono a produrre l'accento di parola in maniera simile all'adulto. Alcune differenze tra adulto e bambino sono tuttavia riscontrabili nei valori formantici. Al fine di spiegare le differenze fra adulto e bambino per quanto concerne la dimensione dell'"altezza" abbiamo avanzato l'ipotesi della "complessità articolatoria" intrinseca alla produzione di una vocale accentata alta: qui è presente una competizione fra il gesto di abbassamento della mandibola, associato alla fase di apertura del ciclo sillabico e il gesto di innalzamento del dorso della lingua, associato alla vocale alta e "rinforzato" in sillaba tonica. Tale competizione potrebbe implicare un grado di coordinazione gestuale ancora non raggiunto dal bambino nell'età qui studiata. Per quanto concerne invece le differenze fra bambino e adulto entro la dimensione dell'anteroposteriorità, la questione è senz'altro apertissima: un'ipotesi da verificare con attenzione è che la produzione di vocali più anteriori nel bambino rispetto all'adulto sia connessa a differenze nella dimensione del tratto vocale.

Ringraziamenti

Il presente lavoro non sarebbe stato possibile senza la pazienza e la disponibilità di alcune persone.

Desideriamo innanzitutto ringraziare Cinzia Avesani, che con lunghe discussioni ha contribuito, per pura e sincera passione per la scoperta, a far crescere questo lavoro: è grazie a lei, che con regolarità ha alimentato il frutto delle nostre fatiche consigliandoci sui parametri da utilizzare, sulle analisi statistiche da effettuare e sulle interpretazioni da attribuire ai nostri risultati, che questo lavoro è maturato, passando dalla prima acerba fase di tesi di laurea alla forma presente.

Un ulteriore sentito ringraziamento va a Vincenzo Galatà, che ha seguito e stimolato i nostri primi goffi tentativi di realizzazione degli *script* utilizzati nel presente lavoro.

Infine, un sentito ringraziamento ai revisori, che ci hanno aiutato a migliorare questo lavoro sia nella forma che nei contenuti.

Riferimenti bibliografici

Albano Leoni, F., Cutugno, F. & Savy, R. (1995). The vowel system of Italian connected speech. In *Proceedings of XIIIth ICPhS*, Stockholm, 13-19 agosto 1995, 4, 396-399.

Arciuli, J., Colombo, L. (2016). An acoustic investigation of the developmental trajectory of lexical stress contrastivity in Italian. In *Speech Communication*, 80, 22-33.

Assmann, P.F., Katz, W.F. (2000). Time-varying spectral change in the vowels of children and adults. In *The Journal of the Acoustical Society of America*, 108(4), 1856-1866.

Avesani, C., Vayra, M. & Zmarich, C. (2007). On the articulatory bases of prominence in Italian. In *Proceedings of XVIth ICPhS*, Saarbrücken, Germany, 6-10 agosto 2007, 2, 981-984.

Avesani, C., Vayra, M. & Zmarich, C. (2009). Coordinazione vocale consonante e prominenza accentuale in italiano. La sfida della Articulatory Phonology. In *Linguistica e modelli tecnologici di ricerca. Atti del XL Congresso Internazionale SLI*. Roma: Bulzoni, 353-386.

Baayen, R.H., Davidson, D.J. & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. In *Journal of memory and language*, 59(4), 390-412.

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

Bertinetto, P.M. (1985). A proposito di alcuni recenti contributi alla prosodia dell'italiano. In *Annali della Scuola normale superiore di Pisa. Classe di lettere e filosofia*, 15(2), 581-643.

Bocci, G., Avesani, C. (2011). Phrasal Prominences do not need Pitch Movements: Postfocal Phrasal Heads in Italian. In *Interspeech*, 1357-1360.

Boë, L.J., Granat, J., Badin, P., Autesserre, D., Pochic, D., Zga, N., Henrich, N. & Ménard, L. (2006). Skull and vocal tract growth from newborn to adult. In *7th International Seminar on Speech Production*, Belo Horizonte, Brazil, dicembre 2006, 75-82.

Bossetti, S. (2012). Test fonetico della prima infanzia: un nuovo test fonetico per bambini dai 18 ai 36 mesi. Tesi di Laurea, Università di Padova, non pubblicata.

Caselli, M.C., Casadio, P. (1995). *Il Primo Vocabolario del Bambino*. Milano: Franco Angeli.

De Boeck, P., Bakker, M., Zwitzer, R., Nivard, M., Hofman, A., Tuerlinckx, F. & Partchev, I. (2011). The estimation of Item Response Models with the lmer Function from the lme4 Package in R. In *Journal of Statistical Software*, 39(12), 1-28.

D'Imperio, M., Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. In *Phonology*, 16, 1-28.

Fant, G. (1971). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*, vol. 2. Berlin: Walter de Gruyter.

Farnetani, E. (1997). The phonetic word: the articulation of stress and boundaries in Italian. In *Quaderni del Centro di Studio per le Ricerche di Fonetica*, 16, 158-172.

Farnetani, E., Kori, S. (1982). Lexical stress in spoken sentences: a study on duration and vowel formant pattern. In *Quaderni del Centro di Studio per le Ricerche di Fonetica del CNR*, 1, 106-133.

Farnetani, E., Vayra, M. (1996). The role of prosody in the shaping of articulation in Italian CV syllables, From Control Strategies to Acoustics. In *Proceedings of the 1st ESCA Tutorial and Reasearch Workshop on Speech Production Moeling*, 9-12.

Fowler, C. (1995). Acoustic and kinematic correlates of contrastive stress accent in spoken English. In Bell-Berti, F., Raphael, L. (Eds.), *Producing Speech. A Festschrift for Katherine Safford Harris*. Woodbury: AIP Press, 355-373.

Fulop, S.A., Kari, E. & Ladefoged, P. (1998). An acoustic study of the tongue root contrast in Degema vowels. In *Phonetica*, 55(1-2), 80-98.

Hanson, H.M. (1997). Glottal characteristics of female speakers: Acoustic correlates. In *The Journal of the Acoustical Society of America*, 101(1), 466-481.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. In *Journal of memory and language*, 59(4), 434-446.

Kehoe, M., Stoel-Gammon, C. & Buder, E.H. (1995). Acoustic correlates of stress in young children's speech. In *Journal of Speech, Language and Hearing Research*, 38(2), 338-350.

Magno Caldognetto, E., Vagges, K. & Zmarich, C. (1995). Visible articulatory characteristics of the italian stressed and unstressed vowels. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 1, 366-369.

Olivucci, F. (2015). Lo sviluppo dell'accento lessicale nel bambino: una prospettiva fonetico-acustica. Tesi di Laurea, Università di Bologna, non pubblicata.

Ortega-Llebaria, M., Prieto, P. (2011). Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. In *Language and Speech*, 54(1), 73-97.

Pasqualetto, F. (2015). L'acquisizione dell'accento di parola in bambini italofoni da 24 a 36 mesi d'età: uno studio acustico. Tesi di Laurea, Università di Padova, non pubblicata.

Pigato, G. (2014). Applicazione di un nuovo Test Fonetico a soggetti con Disturbo Specifico di Linguaggio e a Parlatori Tardivi. Tesi di Laurea, Università di Padova, non pubblicata.

Pollock, K.E., Brammer, D.M. & Hageman, C.F. (1993). An acoustic analysis of young children's productions of word stress. In *Journal of Phonetics*, 21, 183-203.

Savy, R., Cutugno, F. (1997). Ipoarticolazione, riduzione vocalica, centralizzazione: come interagiscono nella variazione diafasica. In Cutugno, F. (Ed.), *Fonetica e fonologia degli stili dell'italiano parlato. Atti VII Giornate di Studio del GFS*, Napoli, 14-15 novembre 1996. Roma: Esagrafica, 177-194.

Seccafien, G. (2013). Inventari fonetici e processi fonologici nel primo vocabolario di bambini con sviluppo tipico valutati con un nuovo test fonetico. Tesi di Laurea, Università di Padova, non pubblicata.

Schwartz, R.G., Petinou, K., Goffman, L., Lazawski, G. & Cartusciello, C. (1996). Young children's production of syllable stress: An acoustic analysis. In *The Journal of the Acoustical Society of America*, 99(5), 3192-3200.

Sluijter, A.M., Van Heuven, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. In *The Journal of the Acoustical society of America*, 100(4), 2471-2485.

Snow, D. (1997). Children's acquisition of speech timing in English: a comparative study of voice onset time and final syllable vowel lengthening. In *Journal of Child Language*, 24, 35-56.

Tamburini, F. (2009). Prominenza frasale e tipologia prosodica: un approccio acustico. In Ferrari, G. (Ed.), *Linguistica e modelli tecnologici di ricerca. Atti del XL Congresso internazionale di studi della Società di linguistica italiana (SLI)*, Vercelli, 21-23 settembre 2006. Roma: Bulzoni, 437-455.

Telmon, T. (1993). Varietà regionali. In Sobrero, A. (Ed.), *Introduzione all'italiano contemporaneo. La variazione e gli usi, II*. Roma-Bari: Laterza, 93-149.

Vayra, M., Avesani, C. & Fowler, C. (1999). On the phonetic bases of vowel-consonant cooordination in Italian: a study of stress and compensatory shortening. In *Proceedings of 14th ICPhS*, San Francisco, USA, 1-7 agosto 1999, 495-498.

Vayra, M., Fowler, C. (1987). The word-level interplay of stress, coarticulation, vowel height and vowel position in Italian. In *Proceedings of the XIth International Congress of Phonetic Sciences*, Tallinn, 1-7 agosto 1987, 4, 24-27.

Vayra, M., Fowler, C. (1992). Declination of supralaryngeal gestures in spoken Italian. In *Phonetica*, 49(1), 48-60.

Zmarich, C., Avesani, C. (2015). L'influenza della durata consonantica sulla coarticolazione della sillaba CV con gradi diversi di prominenza prosodica. In Romano, A., Rivoira, M. & Meandri, I. (Eds.), *Aspetti prosodici e testuali del raccontare: dalla letteratura orale al parlato dei media*. Alessandria: Edizioni dell'Orso, 305-318.

Zmarich, C., Avesani, C. & Marchiori, M. (2007). Coarticolazione e accentazione. In Giordani, V., Bruseghini, V. & Cosi, P. (Eds.), *Scienze Vocali e del linguaggio – Metodologie di valutazione e risorse linguistiche, Atti del III Convegno Nazionale AISV*, Povo, Trento, 29 novembre - 1 dicembre 2006. Torriana: EDK editore, 5-15.

Zmarich, C., Bonifacio, S. (2004). Gli inventari fonetici dai 18 ai 27 mesi d'età: uno studio longitudinale. In Albano Leoni, F., Cutugno, F., Pettorino, M. & Savy, R. (Eds.), *Atti del Convegno Nazionale "Il Parlato Italiano"*, Napoli, 13-15 febbraio 2003. Napoli: M. D'Auria, CD-ROM, 1-20.

Zmarich, C., Bonifacio, S. (2005). Phonetic inventories in Italian children aged 18-27 months: a longitudinal study. In *Proceedings of INTERSPEECH*, Lisboa, 4-8 settembre 2005, 757-760.

Zmarich, C., Fava, I., Del Monego, G. & Bonifacio, S. (2012). Verso un "Test fonetico per la prima infanzia". In Falcone, M., Paoloni, A. (Eds.), *La voce nelle applicazioni, atti dell'VIII convegno AISV*. Roma: Bulzoni, 51-66.

PARTE II

STRUMENTI E TECNOLOGIE PER L'APPRENDIMENTO E LA DIDATTICA DELLE LINGUE

DOI: 10.17469/O2102AISV000013

PIERO COSI, RON COLE

Mindstar books – An imaginative new generation of intelligent tutoring systems in science and in reading

MindStar Books represents an imaginative new generation of intelligent tutoring systems in science and in reading. Great strides are seeking in the quest to immerse students more effectively in multimedia learning activities in which they are challenged, motivated and empowered to acquire the knowledge and skills to learn reading and science.

Key words: Tutoring system, Reading.

Introduction

MindStar Books (MSB) are designed to scaffold effective science learning with the following aims:

- They will enable students, especially including English, Italian, and Spanish language learners, to acquire the prerequisite vocabulary and concepts to listen to and understand science texts that are read aloud to them, eventually by a virtual tutor, while they view illustrations that help them visualize the science being explained.
- They will assess students' understanding of the science through spoken presentation of deep reasoning questions, challenging answer choices representing common misconceptions, and immediate formative feedback on their answer choices.
- They will engage students in activities that lead to accurate, fluent and expressive reading of grade-level texts; skills that correlate highly with reading comprehension and future reading success (Baker, Smolkowski, Katz, Fien, Seeley, Kame'enui & Beck, 2008; Fuchs, Fuchs, Hosp & Jenkins, 2001; LaBerge, Samuels, 1974; Perfetti, 1985; Reynolds, 2000; Samuels, 1997; Stanovich, 2000).

These are important and exciting aims based on prior research and development, and they are within grasp. As for "Scientific Foundations", MindStars Books are based on theory and evidence indicating that a student's ability to read and understand a text – their reading comprehension ability – consists of two component skills: listening comprehension and word reading automaticity. Listening comprehension is an individual's ability to listen to a text and answer spoken questions about it. Reading fluency is the ability to recognize words accurately and effortlessly. Research shows that students' reading comprehension abilities can be accurate-

ly predicted by independent measures of their listening comprehension skills and their ability to recognize words accurately and rapidly (Gough, Hoover & Peterson, 1996; Gough, Tunmer, 1986; Hoover, Gough, 1990).

We should however underlined that the knowledge that the skill of reading comprehension is based on only these two components in the reading model called "Simple View Reading" (SVR) (Hoover, Gough, 1990), that is the listening comprehension and the ability to recognize the words has been frequently criticized in the literature. As an example, Harrison (Hall, Goswami, Harrison, Ellis & Soler, 2010) criticizes this model as a too simplistic view that ignores some important aspects such as fluency, vocabulary, cognitive flexibility, and morphology. For example, it argues that the recognition of the written word and its sound are two related but different processes (225) and that being able of decoding is not the same as being able to read (226). Even Pressley (Pressley, 2000) says that this model is not sufficiently explanatory: "Although skilled and eventually fluent word recognition certainly facilitates comprehension, it is not enough."

- SVR has often been erroneously associated to the definition of reading (Uppstad, Solheim, 2011); "the SVR was as a model to predict reading comprehension by means of two factors: decoding and linguistic comprehension. Over time, the SVR has acquired the status of a definition of reading, and it counts as a starting point for both research and teaching programs for reading."
- Even the authors of the SVR model underlined that the two cited components are necessary but not sufficient.
- Hoover and Gough (Hoover, Gough, 1990) showed that the children's needs on the two dimensions of SVR are subject to change as the children become more fluent in the reading of the words and the reading process becomes automatic.
- In a study of Wolf and Bowers (Wolf, Bowers, 1999), it was shown that also the speed with which we process linguistic elements has its own importance.
- Joshi and Aaron (Joshi, Aaron, 2000) have confirmed that this is a component that allows you to make predictions on students of elementary school level.

In other words, SVR is not the only model and it is surely lacking of completeness, however, its finding that students' reading comprehension abilities can be predicted by independent measures of their listening comprehension skills and their ability to recognize words accurately and rapidly is still a valid conclusion and MSBs will exploit this finding.

1. *Design and Organization of MindStars Books*

The MindStars Books Toolkit was developed to provide an easy to use authoring environment for developing the listening comprehension activities in MS Books, and publishing the book in a library.

Each MSB consisted of three independent activities:

- Narrated multimedia science explanations,

- Multiple choice questions for assessing students' knowledge and providing them with immediate feedback on their answer choices, and
- Reading practice, which used automatic recognition of children's speech while reading aloud to provide them with feedback on their reading.

Narrated multimedia presentations of science produce optimal learning, as measured by both short-term retention of information and by transfer of learning to new tasks. During narrated science explanations, each "page" of a book is presented to a student as a sequence of pictures. The student looks at each picture (or a collage of pictures) while the voice of the intelligent agent presents information related to the picture.

Multiple Choice Questions (MCQs) were presented at logical stopping points within the book to assess students' understand of science presented in the preceding several pages, and to provide students feedback on their answer choices. Each multiple-choice question concluded with an expansion of the correct answer choice, with the goal of helping students master concepts and build on them during the remainder of the book.

The goal of reading practice in MSBs was to improve students' confidence in their ability to read text passages about science, by enabling them to practice reading them fluently, with both support and feedback on their oral reading fluency. Creating a reading passage in the MSB Editor consists of the author a) typing the text passage into the Editor, b) recorded each sentence in the text, c) recording each word in the sentence individually, and d) optionally importing a picture associated with each sentence. When the author "builds" or publishes the book, the typed and recorded text is automatically transformed into the complete set of oral reading fluency activities students can do within the MindStars Book.The MSB Editor was used to create each of these activities.

The MSB Editor is an authoring tool that was used to create, test and refine the books developed over the course of the project by various teachers. The Editor is an intuitive, flexible and powerful tool that could be used by individuals with no programming experience for creating, testing and publishing MSBs. Creation of a book within the MSB Editor produces a published book with all of the interactive science learning and reading activities, which included automatic feedback to students on their oral reading fluency. Working from a completed script, a complete MSB can be developed and published in a single day.

In particular, the MSB Editor is a tool that enables an author to:

a. type in each sentence MSB will say, eventually by a virtual tutor;
b. record the sentences in English and record the Italian and Spanish translation of each sentence;
c. select a picture that will be presented with each narrated sentence (portions of pictures are highlighted using Photoshop);
d. include optional sound files into the narration;

e. design one or more multiple choice questions, with optional illustrations, that are presented after the page has been narrated;
f. record the questions and answer choices in English and both Italian and Spanish. Once the listening comprehension activities have been developed, the oral reading fluency training activities, which follow listening comprehension, are generated automatically, using the text that is narrated, eventually by a virtual tutor, during listening comprehension training.

2. *Listening Comprehension*

In MSBs, each page of a science text is narrated, eventually by a virtual tutor, while the student views illustrations that help them visualize the science. The narration is self-paced in alignment with research that indicates that self-paced presentations improve learning (Baker, 2003; Cole, Van Vuuren, Pellom, Hacioglu, Ma, Movellan, Schwartz, Wade-Stein, Ward & Yan, 2003; Cole, Halpern, Ramig, Van Vuuren, Ngampatipatpong & Yan, 2007). Students can stop and resume the narration after each sentence is spoken, and have the sentence repeated in English or say an Italian or Spanish translation of the sentence.

Figure 1a - *Screen shot in MSB Editor of the first page of a book called "The Bird, The Brain and The Train." The Picture at the top of the page is associated with the first sentence. The pictures associated with the second and third sentence are shown before these sentences*

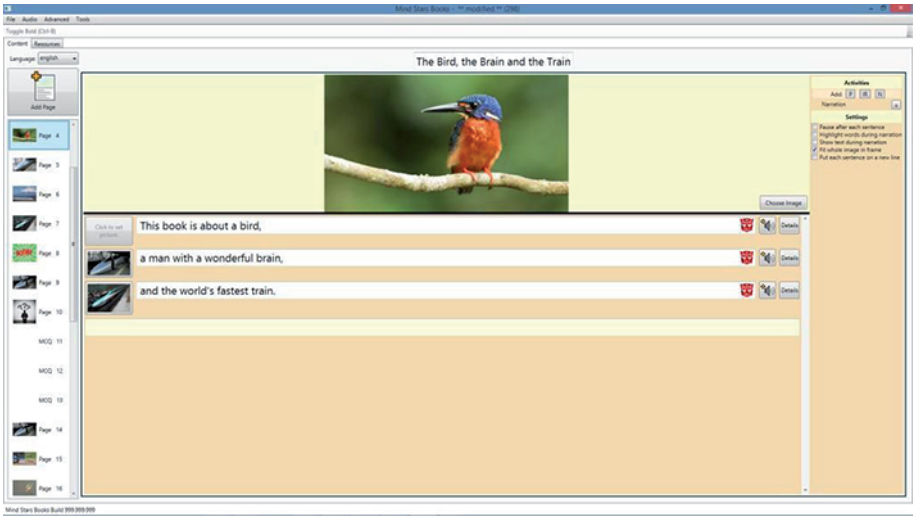


Figure 1a shows a single page of an MSB called "The Bird, The Brain and The Train." in the Editor. Figure 1b shows the first two screens presented to the student, corresponding to the first two lines of the page in the Editor.

Figure 1b - *Screen shots of first two pages for students. Students see pictures only while the agent narrates the sentences shown in the MSB Editor*

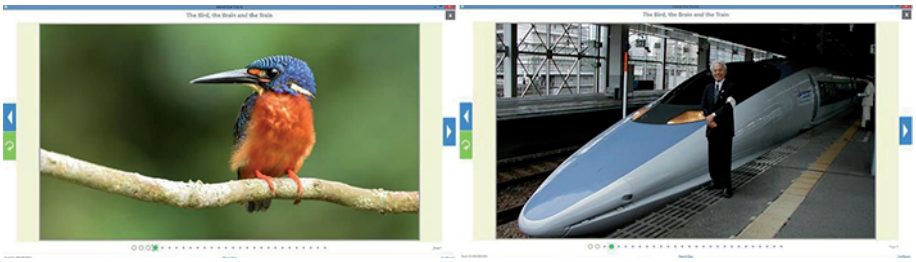
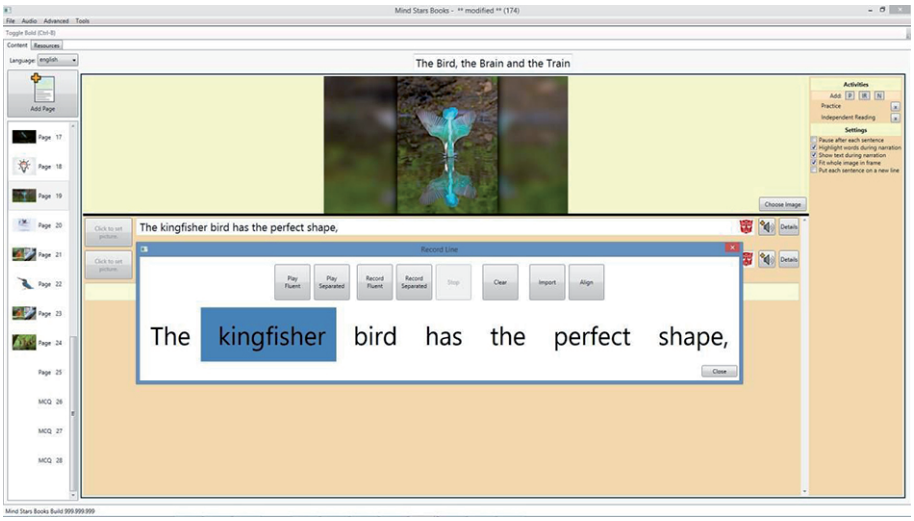


Figure 2 - *Recording function within MSB editor. The author has recorded the sentence that will be spoken to the student while looking at a picture on a page, and is watching the words highlight during playback of the recording to assure accurate synchronization of her speech to each highlighted word*



Differently from the Editor (Figure 2), we emphasize that words are not displayed on screen during narrated multimedia presentations, as the goal is to have students listen to the agent's voice while looking at the pictures, enabling them to construct rich multimodal mental representations of the science. Research has shown that putting words on the screen during a narrated multimedia presentation impedes learning (relative to not having words on the screen), as students will switch attention between the printed words and the pictures, resulting in poorer recall of the presented information.

After listening to one or more pages of text, students are presented with multiple choice questions (MCQs) to assess their understanding of the vocabulary and concepts. These are deep reasoning questions with challenging answer choices that represent common misconceptions. Students can listen to

the question and answer choices either in English or in Italian or Spanish as often as they like.

After selecting an answer, the student receives immediate feedback about the answer they selected. A positive feedback is obviously provided to a correct answer. If the student selects an incorrect answer choice, learning is supported by providing a hint; e.g., that spider has 8 legs, so it can't be an insect. After two tries, the correct answer is presented to the student, along with an explanation as to why the answer is correct. During listening comprehension activities, words are not presented on the page, as the goal is to have students listen carefully while viewing illustrations; research indicates that printed words can distract the student's attention from the illustrations and reduce learning (Cole, Wise & Van Vuuren, 2007).

Figure 3 displays the Editor interface (Figure 3a) for developing multiple-choice questions (Figure 3b). It shows slots for typing or importing questions, pictures, answer choices, and feedback on answer choices. The Editor uses the same recording tool for recording each question and spoken answer choices.

Figure 3a - *Editor for Multiple Choice Questions in the MSB Editor. The top of the page shows the question; the first line is the answer choice and feedback. The pairs of sentences, more broadly, show answer choices to the MCQ's followed by feedback after selection of the answer*

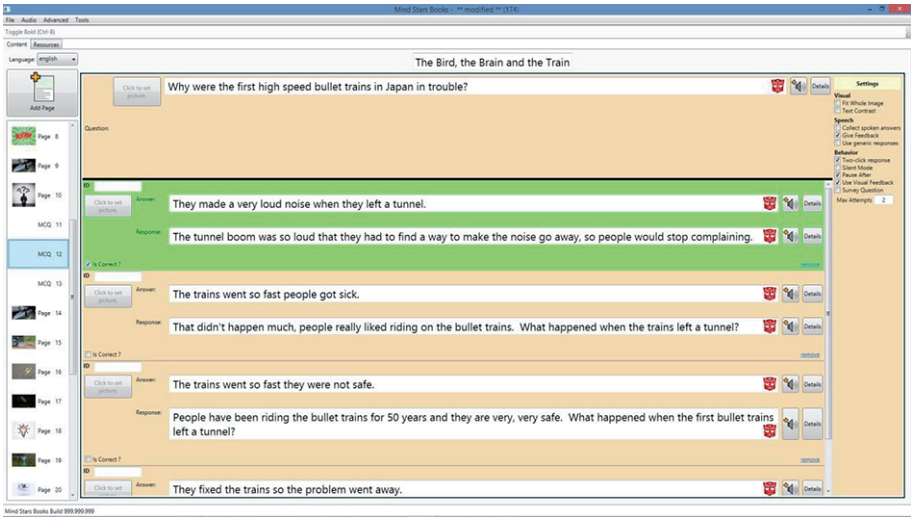
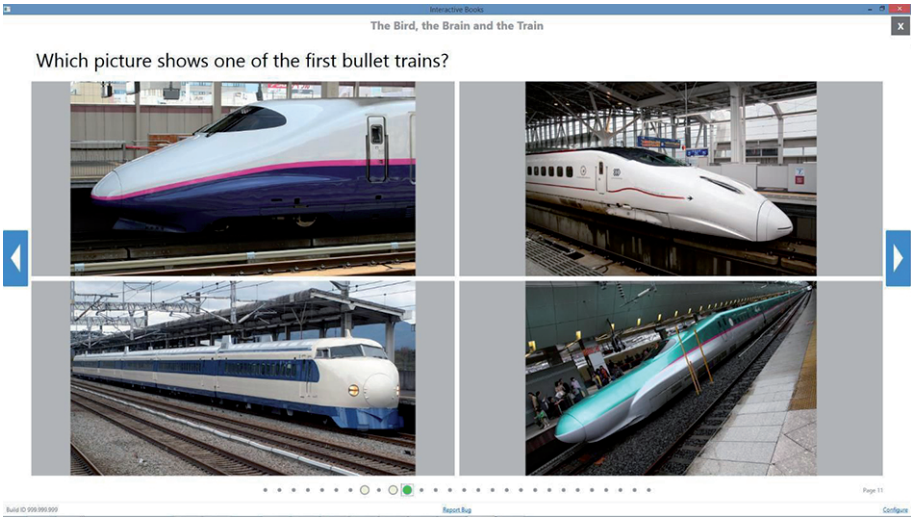


Figure 3b - *Screenshot of multiple-choice question followed by four pictures. The train with the flat nose cone is the correct answer. Answer choices are randomly placed on the screen each time a question is presented*



3. *Oral Reading Fluency (ORF)*

ORF practice and training occurs immediately after the listening comprehension activities are completed; that is, after all pages of the science text have been narrated to the student and MC questions have completed.

The goal of the ORF training is to help students learn to read grade level science texts accurately and fluently; oral reading fluency has been demonstrated to be a strong predictor of reading comprehension and later reading proficiency (LaBerge, Samuels, 1974; Perfetti, 1985; Ward, Cole, Bolaños, Buchenroth-Martin, Svirsky, Van Vuuren, Weston, Zheng & Becker, 2011).

Fluency training occurs through repeated reading of each page of the science text. The student is presented with the first page of the text, with each sentence displayed on the page. The student can choose to practice reading the text, with eventually support from a virtual tutor, before reading it independently. During practice, the student can listen an entire sentence to be read, or individual words in a sentence to be pronounced. The students can record themselves reading these sentences or words and play back their recordings to compare their reading with the correct one, eventually spoken by a virtual tutor. During playback of their recordings, each word is highlighted on the page as it spoken by the student. English learners can listen to MSB read a translation of the sentence in Italian or Spanish. When the student has finished practicing, they click an icon to read the page independently. Immediately after reading the page, the student receives feedback on the number of words they read correctly (out of the total number of words on the page), and their reading rate (relative to MSB's natural reading rate). The MSBs highlight words

that the speech recognizer scored as misread or skipped, so the student can practice reading these words and sentences. Repeating readings of the page, with practice before each reading and feedback on the student's reading performance immediately after independent reading, continues until the student achieves a criterion level of oral reading performance (90% word reading accuracy, reading speed within 10% of MSB's) or after three independent readings. Repeated reading of texts with feedback and practice following each reading has been shown to be a powerful tool to improve reading fluency, which correlates highly with reading comprehension (Baker et al., 2008; Fuchs, 2001; LaBerge, Samuels, 1974; Perfetti, 1985; Reynolds, 2000; Samuels, 1997; Stanovich, 2000).

4. *Final Considerations*

English, Italian and Spanish speech recognizers needed for ORF practice are already well developed and at an advanced level. Development and evaluation of MindStars Books is already on a good stage for English and Spanish and will be hopefully extended to Italian in the next year, depending on the effective funding of various submitted research projects and grants.

Bibliography

Baker, L. (2003). Computer-assisted vocabulary acquisition: The cslu vocabulary tutor in oral-deaf education. In *Journal of Deaf Studies and Deaf Education*, 8(2), 187-198.

Baker, S.K., Smolkowski, K., Katz, R., Fien, H., Seeley, J.R., Kame'enui, E.J. & Beck, C.T. (2008). Reading fluency as a predictor of reading proficiency in low-performing high poverty schools. In *School Psychology Review*, 37, 18-37.

Cole, R., Van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D., Ward, W. & Yan, J. (2003). Perceptive animated inter-faces: First steps toward a new paradigm for human-computer interaction. In *Proceedings of the IEEE*, 91(9), 1391-1405.

Cole, R., Halpern, A., Ramig, L., Van Vuuren, S., Ngampatipatpong, N. & Yan, J. (2007). A virtual speech therapist for individuals with parkinson disease. In *Educational Technology*, 47(1), 51-55.

Cole, R., Wise, B. & Van Vuuren, S. (2007). How Marni teachers children to read. In *Educational Technology*, 24(1), 14-18.

Fuchs, L.S., Fuchs, D., Hosp, M.K. & Jenkins, J.R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. In *Scientific Studies of Reading*, 5, 239-256.

Gough, P.B., Hoover, W.A. & Peterson, C.L. (1996). Some Observations on a Simple View of Reading. In Cornoldi, C., Oakhill, J.V. (Eds.), *Reading comprehension difficulties: Processes and intervention*. Mahway, New Jersey: Lawrence Erlbaum Associates, 1-13.

Gough, P.B., Tunmer, W.E. (1986). Decoding, reading, and reading disability. In *Remedial and Special Education*, 7, 6-10.

Hall, K., Goswami, U., Harrison, C., Ellis, S. & Soler, J. (Eds.) (2010). *Interdisciplinary Per-spectives on Learning to Read. Culture, cognition and pedagogy*. Routledge: London.

Hoover, W.A., Gough, P.B. (1990). The Simple View of Reading. In *Reading and Writing: An Interdisciplinary Journal*, 2, 127-160.

Joshi, R.M., Aaron, P.G. (2000). The component model of reading: Simple view of reading made a little more complex. In *Reading Psychology*, 21(2), 85-97.

Laberge, D., Samuels, S. (1974). Toward a theory of automatic information processing in reading. In *Cognitive Psychology*, 6, 293-323.

Perfetti, C. (1985). *Reading ability*. Oxford, England: Oxford University Press.

Pressley, M. (2000). What should comprehension instruction be the instruction of?. In Kamil, M.L., Mosenthal, P.B., Pearson, P.D. & Barr, R. (Eds.), *Handbook of Reading Research: Volume III.* Mahwah, NJ: Erlbaum, 545-561.

Reynolds, R.E. (2000). Attentional resource emancipation: Toward understanding the interaction of word identification and comprehension processes in reading. In *Scientific Studies of Reading*, 4, 169-195.

Samuels, S. (1997). The importance of automaticity for developing expertise in reading. In *Reading and Writing Quarterly*, 13, 107-122.

Stanovich, K.E. (2000). The interactive-compensatory model of reading: A confluence of developmental, experimental, and educational psychology. In Stanovich, K.E. (Ed.), *Progress in understanding reading: Scientific foundations and new frontiers*. New York, NY: Guilford Press, 44-54.

Uppstad, P.H., Solheim, O.J. (2011). Code and Comprehension in Written Language – Considering Limitations to the Simple View of Reading. In *L1-Educational Studies in Language and Literature*, 11, 159-174.

Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., Van Vuuren, S., Weston, T., Zheng, J. & Becker, L. (2011). My science tutor: A conversational multimedia virtual tutor for elementary school science. In *ACM Transaction on Speech and Language Processing*, 7(4), 18.

Wolf, M., Bowers, P. (1999). The "double-deficit hypothesis" for the developmental dyslexias. In *Journal of Educational Psychology*, 91(3), 415-438.

DOI: 10.17469/O2102AISV000014

LIDIA CALABRÒ

PhonetIC(T)s: teaching and learning geminates in Italian SL through body movement, cooperative learning and mobile apps – an experience

In this contribution the author addresses the use of geminates in Italian L2 pronunciation classes. She presents here an innovative and dynamic way to teach pronunciation. The combination of different teaching practices has been developed with the aim to support learners in pronunciation classes and to provide teachers with a multimodal approach: cooperative learning, body movement, mobile apps, reflection on and discovery of sounds. She describes how lexicon and morpho-syntactic aspects have been combined to phonetics teaching, how the activities have been implemented, and she shows sample activities where students are asked to perform pronunciation. The activities prepared are based on the proficiency level of the students. This approach reveals that pronunciation teaching and learning is far from being a worthless practice. She argues that teachers of Italian L2 should pay attention to perceptive and productive skills as they enhance listening, reading, writing, and speaking.

Key words: geminates, pronunciation learning, pronunciation teaching, ICT, Italian SL.

Introduction

By playing with the language and combining together the word 'Phonetics' and the acronym 'I.C.T.' (*Information Computing Technology*), the first word of the title of this contribution puts in foreground two aspects that will be discussed in the following paragraphs: the teaching of Phonetics in Italian as a SL/FL and the usage of mobile apps as technological instruments to teach and learn pronunciation.

Phonetic and phonological aspects in the teaching of Italian as a second language (SL) are not often taken into consideration compared to the teaching of grammar and vocabulary. They are often considered useless and thus bypassed, but working on segmental and suprasegmental aspects of a language is fundamental to improve students' phonetic and phonological competence and their comprehension and production of the second language. In teachers' training not enough time is dedicated to theories and techniques to stimulate learners to develop this competence. Teachers too, not only learners, need to raise awareness about learners perceptive and productive difficulties.

This contribution aims at showing how some practical activities, realized during an Italian pronunciation workshop, help foreign students to make a re-

flection on segmental and prosodic tracts of the SL. Furthermore, the teacher helps them in practicing by also using body movement, cooperative learning and mobile applications (*QR-Code Reader and Kahoot!*).

The author does specify that the kind of contribution is not a research neither quantitative nor qualitative but it is a teaching experience. It can be used as a basis for further research aimed at seeing if it is possible for learners to improve in pronunciation and, if so, aimed at evaluating the degree of improvement. This contribution is a good starting point to rethink about methods and practices in the teaching of a SL/L2 as it helps to raise awareness in: 1) teachers about the importance of teaching pronunciation and 2) students about the processes involved in sounds production because it also helps self-correction.

1. *The Experience at CLA – Roma Tre University*

1.1 The activities

The activities presented in this contribution took place at *CLA – University of Roma Tre* (*Università degli Studi Roma Tre*) and have been realized with 10 Chinese students at A1/A2 level of the *C.E.F.R.* (*Common European Framework of Reference* – Council of Europe, 2001). Learners were aged between 18 and 25. The topic of the lesson related to the vocabulary and to some expressions used at the supermarket or when people do food shopping. The activities have been prepared and realized *ad hoc* after taking a lesson about the supermarket which was part of a phonetic workshop (see *1.2*). Thus, the vocabulary and the expressions or sentences used for the workshop have been chosen by taking into consideration the words and the texts the students had learnt during the previous lesson 'At the supermarket'. The activities respected the learning process and the students' linguistic competence[1]. According to the Lexical Approach (Lewis, 1993; 1997) and to the *Profilo della Lingua Italiana* (Italian Language Profile) suggestions (Costamagna, 2010a), lexicon has been considered in context. In this particular case, vocabulary has been taught in 'at the supermarket' context on which the phonetic tracts were based. Thus, the author has considered Krashen's 'i+1' theory (1981; 1985) where the 'i' was related to the lexicon and the '+1' to the phonetic aspects. This specific work has been realized by taking into consideration the interferences coming from the learners' mother tongue (L1) as geminates, which are always difficult to perceive and, as a consequence, to produce by foreign students of Italian as a SL. Eckman's Markedness (1977) is a fundamental concept to bear in mind while preparing *ad hoc* activities. As for the theory an asymmetry exists between two phonemes, in a couple of pho-

[1] The author decided not to consider either specific word constructions or stress position as learners were exposed to the vocabulary they had learnt before the workshop. Moreover, the level of acquisition was very low (A1/A2) and, in general, for Chinese students the difficulty to learn Italian words is quite high due to the typological distance between L1 and L2.

nemes the marked tract is more complex and less natural than the other; for example, the voiced /g/ of 'gatto' (cat) is more marked than the voiceless /k/ as the former contains the [+voiced] tract together with the cords vibration (Chini, 2010). The idea of 'naturalness' has been associated to this theory (Dressler, Mayerthaler, Panagl & Wurzel, 1987). Thus a more natural linguistic element is less marked and easier to learn. Markedness is also a reflection of the structure of the human cognition (Ekman, 1977), of language perception and 'processability' modality (Pienemann, 1998) or of cognitive, articulatory and perceptive factors that interact among them (Ferguson, 1984). The knowledge of previous linguistic acquisition, related to the L1, interferes with the acquisitional process of the SL by slowering or fastening it. This is the case of the interference or 'transfer' from L1 or from other L2 previously learnt (Gass, Selinker, 1983; Cook, 2001). In this specific case, Chinese learners show difficulties in learning distinctive consonant duration also at advanced levels because of fossilization (Costamagna, 2010b). In their L1 they also 'tend to produce, within a syllable, a longer duration of stressed and unstressed vowels, reducing the consonant duration in stressed and unstressed syllables'. Thus 'stressed and unstressed syllables have a similar duration' (Romito, Tarasi, 2012). This phenomenon is opposite to Italian where the geminate always has a longer duration than a singleton as shown in Celata, Costamagna (2012). Plosives /p, t, k/ and /b, d, g/ are complex both at a perceptive and productive level as they substitute the voiceless phoneme with the voiced counterpart or *viceversa*. In Chinese language the distinctive tract of sonority is only allophonic, so [b, d, g] are used in non-stressed syllables. A further problem is connected to the graphic Pynin system where the use of graphemes is not clear (Costamagna, 2010b; Dal Maso, 2003).

The activities aimed at improving: a) perception of singletons *vs* geminates through listening to non-words; b) pronunciation of geminates; c) spelling of words with singletons and geminates (food and drinks vocabulary); d) prosodic features such as syllabic stress and length to give fluency to the reading and to spontaneous speech (Celata, Costamagna, 2014; Costamagna, 1996; 2000; D'Annunzio, 2009; Mastrantuono, 2010).

1.2 Phonetic workshop in Italian as a SL/FL

The Phonetic Workshop moves from the studies of Wrembel (2007; 2011) and her suggestions to improve Phonetics teaching practices through body movement and the connections between itself and the acquisitional processes. The author of the present contribution has launched the Phonetic Workshop, which body movement is inserted in, as a way to renew methods in Phonetics teaching (Calabrò, 2015; 2016a; 2016b; Luchini, 2005; Underhill, 2005). The lessons usually take place in a classroom as it conceives the combination of body movement, cooperative learning, mobile apps, reflection and discovery of segmental and suprasegmental tracts of Italian as a SL. This kind of activities give dynamicity to the lesson and Phonetics is not perceived anymore as a boring activity.

The Workshop is not in contrast with the work done in a language lab and it can also be done, and suggested, in all situations in which a language lab does not exist.

Mobile apps give a new teaching perspective while body movement involves the whole person in a physical way but it helps the learner to deeply think about the sounds perception and production and becoming aware of them (Costamagna, Marotta, 2008). Working in a room is a good chance to consider multiple intelligences and learning styles and to also involve students in cooperative learning to negotiate meanings and to make a good reflection on their sounds perception and on how to realize sounds which are different from their L1. This is a first step for the learners to self-correct during their learning processes (Gardner, 1983; 1993; 1999; Kowal, Swain, 1994; Silver, Strong & Perini, 1997; Cook, 2001; Kagan, 2007).

1.3 I.C.T. and its usefulness

Students have been asked to download the mobile apps *QR-Code Reader*[2] and *Kahoot!*[3] on their own mobile phones to perform the tasks assigned. *QR-Codes* can combine images, writing and web links to search for material and in this specific case to discover words. The usage of this mobile app gives dynamicity to the lesson and breaks its monotony. *Kahoot!* is a virtual, interactive and multimedia web and mobile application that allows you to prepare online tests and administer them in a way that the students do not think they are being evaluated. This game has interested the learners because of its gaming characteristics but it is also very interesting for teachers as it helps to consider the students' learning process and their mistakes thanks to the Excel file downloadable soon after the game has been administered. In this way it is possible to monitor their progress and their main difficulties (see 2. and Figure 5 for an example of the results obtained).

The use of mobile apps targets the double of the audience: on the one hand, the author believes it can be useful to teachers and researchers working in experimental phonetics, phonology and psycholinguistics. It can also be beneficial for preparing innovative lessons with a research perspective as Excel files can be easily downloadable and can be used to test the student's progress even if the learners do not realize they are being tested. On the other hand, it can be useful to learners of pronunciation classes.

[2] QR Code Reader is a mobile app that can be downloaded for free on Google Play (https://play.google.com/store/apps/details?id=me.scan.android.client&hl=it) or on Apple itunes (https://itunes.apple.com/it/app/qr-code-reader-and-scanner/id388175979?mt=8).

[3] Kahoot! is another mobile app that can also be set up on teacher's pc to prepare the activities before playing with the learners. The website where it can be downloaded for free is https://getkahoot.com.

2. *Techniques and class material*

Before taking the pronunciation workshop, the students sat a pre-test. Learners had to listen to three sentences and fill in the gaps with geminates. The sentences related to asking for food and buying it at the supermarket (see Figure 1). They had been prepared *ad hoc* and recorded before being administered to the learners. The totality of the students inserted a singleton or a singleton with a different tract (e.g. ro**ss**e → ro**s**e* (*red*), tutte → tute* (*all*), co**tt**o → co**d**o* (*cooked*), gia**ll**e → gia**r**e* (*yellow*), vo**rr**ei → vo**l**ei*(*I would like*). This is a clear example of how difficult it is to perceive a SL sound that has different tracts from those of the L1.

Figure 1 - *Ad hoc created activities: pre-test on geminates*

5. Ascoltate i dialoghi e completate le parole con le lettere mancanti. __/14

1.
A: Vo___ei un e___o di prosciu___o!
B: Lo vuole co___o o crudo?
A: Crudo. Ha anche que___o di Parma?

2.
A: Mi dà due e___i di parmigiano, per favore?
B: Certo! Intero o gra___ugiato?
A: Gra___ugiato, grazie!

3.
A: Mi dà un chilo di mele?
B: Quali vuole? Le ro___e o le gia___e?
A: Tu___e gia___e, grazie. Le ro___e non mi pia____ono!

The techniques to improve perception and production relate to 5 different kinds of activities.

The first type of activities is made of four steps (a-d) as it follows: a) listening to six non-words and sign with an "X" whether the students hear a singleton or a geminate; b) comparing the answers with a class-mate; c) checking the answers with the teacher and say what happens when a geminate sound appears in a word (to make a reflection on the phonetic process); d) reading and pronouncing the six minimal pairs made of non-words (from point 'a') at first on mute, then without reading and by releasing sounds and imitating teacher's pronunciation (see Figure 2).

Figure 2 - *Ad hoc created activities for perception and production of minimal pairs*

1. Ascoltate le seguenti parole e mettete una "X" se sentite una consonante o "XX" se sentite due consonanti. Le parole non esistono in italiano.

X 1 consonante	XX 2 consonanti
1.	1.
2.	2.
3.	3.
4.	4.
5.	5.
6.	6.

2. Confrontate le vostre risposte con quelle di un compagno.

3. Correggete con l'insegnante e dite che cosa succede quando una consonante è doppia.

_____

_____

4. Leggete e pronunciate le non-parole: prima senza voce, poi con la voce.

X 1 consonante	XX 2 consonanti
1. gagola	1. gagolla
2. isoto	2. isotto
3. esole	3. essole
4. diti	4. ditti
5. ado	5. addo
6. popo	6. poppo

The second kind of activities relates to *QR-Codes*. Learners have been asked to search for words, through mobile phones applications, read them, and then write them in a table according to the singleton or geminate sound heard. 20 *QR-Codes* (10 for singletons and 10 for geminate sounds) have been spread around the room and students had to read them first through the *QR-Code Reader* mobile app and consequently write them on the right column of the handout received. This activity, far from being static, is a good way for learners to read and write words. Thus, it develops reading and writing (spelling) skills. As a discovery activity and as a task to accomplish, students do not even realize they are already learning or revising useful vocabulary (see Figure 3).

Figure 3 - *Ad hoc created QR-Codes activities*

QR Code **Parole con una sola consonante**	**QRCode** **Parole con la consonante doppia**
1. _____	1. _____
2. _____	2. _____
3. _____	3. _____
4. _____	4. _____
5. _____	5. _____
6. _____	6. _____
7. _____	7. _____
8. _____	8. _____
9. _____	9. _____
10. _____	10. _____

Soluzioni

QRCode **Parole con una sola consonante**	**QRCode** **Parole con la consonante doppia**
1. ____patate_____	1. _____risotto_____
2. ____basilico_____	2. _____burro_____
3. ____miele_____	3. _____mozzarella_____
4. ____pepe_____	4. _____frutta_____
5. ____mango_____	5. _____pacco_____
6. ____anguria_____	6. _____etto_____
7. ____funghi_____	7. _____zucchero_____
8. ____melanzane_____	8. _____prosciutto_____
9. ____fragola_____	9. _____formaggio grattugiato_
10.____carne_____	10._____cipolla_____

As for third activity, the body movement technique has been used to read words and concentrate on syllable stress and length. The two movements are: a) opening and closing the hands when pronouncing stressed syllables; b) punching when pronouncing geminates to visualize the duration and the length of the sounds. Moving hands and arms allows learners to create a strict connection between length and duration of both sounds and rhythm in order to acquire the prosodic tracts of the new language. The movement is a mean to concentrate on the length of words and sentences.

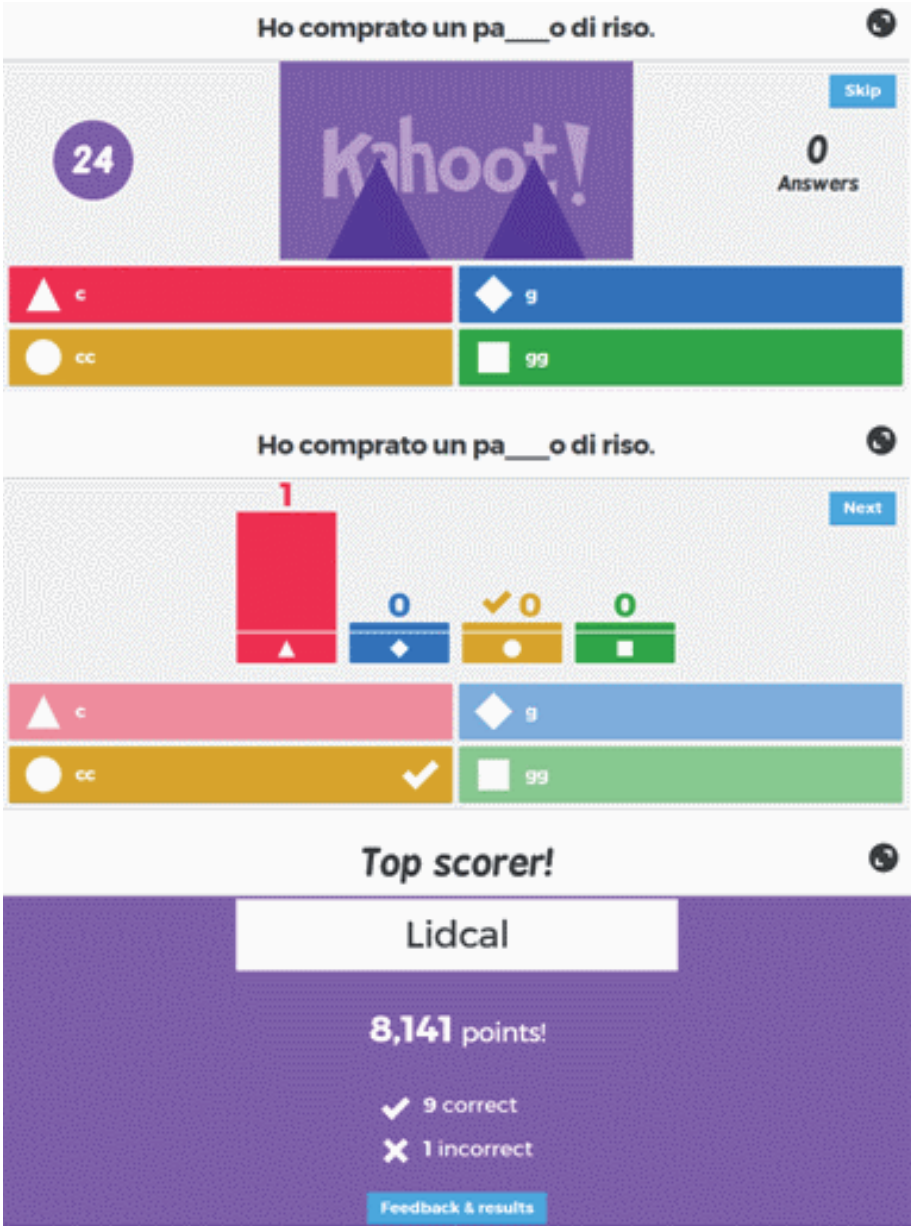
The fourth activity relates to matching cards containing expressions of food, drinks and packaging [e.g. *vorrei un etto di prosciutto (I would like 100 grams of ham); vorrei un pacco di riso (I would like a packet of rice*)] (see Figure 4). Through cooperative learning students revise vocabulary related to collocations, read the short sentences by using body movement to help fluency and sentences rhythm.

Figure 4 - *Ad hoc created collocations activities*

Vorrei	**un pacco**	**di riso**
	una bottiglia	**d'acqua**
	un chilo	**di zucchero**
	un barattolo	**di marmellata**
	una scatoletta	**di tonno**
	un etto	**di prosciutto**
	un litro	**di birra**
	un vasetto	**di funghi**
	una busta	**di patatine fritte**
	un tubetto	**di ketchup**

The fifth activity is a game made up with *Kahoot!* computer and mobile app to improve words spelling and to involve learners in a more competitive, innovative and funnier way. Figure 5 is an example of the activity as it appears when launching the game. The teacher reads quickly the sentence projected on the wall and each student, through his/her mobile phone, pushes on one of the four colorful buttons to answer. At the end of each question the top scorer shows the number of right and wrong answers together with the rank reached by each student.

Figure 5 - *Ad hoc created Kahoot! activities (an example)*



The geminates involved in the *Kahoot!* test where /tt/, /kk/ and /rr/ as the most frequent ones among the vocabulary used for the workshop. Each question, as shown in Figure 5, had four options. The choice was made by considering the two most difficult tracts for Chinese students to learn: a) singleton *vs* geminate; b) voiceless *vs* voiced. The answers to the 10 questions are illustrated as follows (see Table 1).

Table 1 - *Possible answers to Kahoot! activity*

Kahoot! Questions	*Answer a.*	*Answer b.*	*Answer c.*	*Answer d.*
1. Ho comprato un pa___ di riso	c	cc	g	gg
2. La maionese è nel tube__o	d	t	dd	tt
3. Mi prendi un chilo di zu___ine, per favore?	ch	cc	cch	ggh
4. Ho bisogno dell'acqua, del vino e della bi__a.	ll	r	rr	ll
5. Devo comprare la fru__a e la verdura	t	dd	tt	d
6. Mi porta anche le patatine fri__e, per favore?	dd	d	t	tt
7. Preferisco il prosciu__o crudo.	t	d	dd	tt
8. Nella torta c'è tanto zu___ero.	cch	ggh	c	ch
9. Il riso__ è buonissimo.	tt	d	dd	t
10. Ecco a Lei due e__i di salame.	dd	t	tt	d

Table 2 - *Kahoot! activities Excel Results*

n° student	**cc**	**tt**	**cch**	**rr**	**tt**	**tt**	**tt**	**cch**	**tt**	**tt**	**tot**
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**10**
2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**10**
3	✓	✓	× cc	✓	✓	✓	✓	✓	✓	✓	**9**
4	× gg	✓	✓	✓	✓	✓	✓	✓	✓	✓	**9**
5	✓	✓	× cc	✓	✓	✓	✓	✓	✓	✓	**9**
6	✓	✓	✓	✓	✓	✓	empty	✓	✓	✓	**9**
7	✓	empty	✓	✓	✓	× t	✓	✓	✓	P	**8**
8	✓	✓	× ch	✓	✓	× t	✓	✓	✓	✓	**8**
9	empty	✓	× cc	✓	✓	✓	× dd	✓	✓	✓	**7**
10	empty	✓	✓	✓	✓	✓	× d	✓	× d	✓	**7**

The results coming from Excel file (see Table 2) show that only two students answered correctly (10/10), four made a mistake and ranked 9/10, three 2 mistakes (8/10) and only two of them made 3 mistakes (7/10). Based on these results we can also make some assumptions in relation to the type of mistake. In fact, we can see who made the mistake and why[4]. For example, the students n° 4, 9 and 10 had problems with the tract of sonority (that can be explained considering that the phoneme appears in an unstressed syllable) and n° 10 also with duration. The third question

[4] In Excel Result the names of the students also appear as they use their name or nickname to sign in the app.

created a problem to the association between sound and grapheme, problem solved in the eight questions as students were able to recognize the sound above all after having understood the previous mistake. Some answers were left empty. We can assume that students n° 7, 9 and 10 were short in time and had problems in answering quickly, while student n° 6 did not know the answer to the question 7. On the whole we can say that the activity has been successful both for the results obtained and for the personal involvement and reflection on sound perception.

3. *Considerations of the experience*

The activities have helped the students to raise personal awareness about their linguistic gap as they have started, during classes, to make questions to the teacher about how to improve the pronunciation in a way which can be considered similar to that of a native speaker. The learners have been totally involved in a more personal, physical, technological, collaborative and reflective way. The lesson has become dynamic and working on pronunciation has become a pleasant activity for both teacher and students. Some difficulties remain because of: a) learners personal abilities; b) L2 level of knowledge; c) lacking of time to dedicate to pronunciation during classes; d) time students spend with their class mates *vs* native speakers. This work has helped learners: a) to face with difficulties related to their L1; b) to focus on spelling and perception of geminates and singletons; c) to raise awareness about their difficulties and about the importance of practicing with pronunciation improving their L2.

3.1 Students' opinions about the phonetic workshop

The results of a questionnaire, submitted to Chinese (*Ch*) students at the end of the phonetic workshop, show the importance of focusing on perceptive and productive aspects of a language. Although no measurement of improvement in SL has been done, the author considers learners' answers relevant to phonetics teaching methodology.

The questions are intended to investigate on the personal perception of improvement and on the utility of the activities experienced. The questions 1) 'Did you like taking part at the workshop?' and 5) 'Do you think you have improved your pronunciation?' ranked both 'Yes' *Ch*. (10/10) and 'Yes' *Ch* (10/10). To the question 2) 'Why?', improvement has been attributed to: the lessons dynamicity (multimodal approach); the self-perception of enhancing in reading, listening and speaking. To the questions 3) 'What did you find more useful?' and 4) 'What did you find useless?', working on pronunciation has been perceived as: *piacevole* (enjoyable)[5] due to the lesson *atmosfera* (atmosphere), *interessante* (interesting) and *utile* (useful) to correct mistakes. The activities are presented in a very *sfidante* (engaging) way. Nothing has been considered useless. To the question 6) 'If so, in which

[5] The words in Italics are students' expressions.

aspects of pronunciation? If not, why?', the answers show that awareness is raised in relation to: articulation of words and *l'apertura della bocca* (the opening of the mouth), knowledge of phonemes and phones perception, above all those which are different from their mother tongue.

3.2 Strengths and weaknesses

While considering the entire work some weaknesses and strengths come to light. On the one hand, even if some listening and repetition activities are inserted into Italian language books, more hours are needed to plan and realize a phonetic workshop. We also do need to start a proper research with a group of control as well, in order to test if and how this work can be effective. On the other hand, some positive aspects of this contribution can be identified in the following characteristics: the activities described here can also be carried out in places where a language lab is missing; they stimulate personal reflection and raise awareness; they strengthen aspects of the language which are usually left out because of lacking of time; they provide for total and dynamic involvement (all the students, the whole person and his/her body); various activities consider phonetic and phonological difficulties related to the learners' L1; the entire work can be considered as a first step for further research.

3.3 Final remarks

The activities have helped the students to raise personal awareness about the linguistic gap as they have started, during classes, to ask questions to the teacher about how to improve the pronunciation in a way that can get closer to the one of a native speaker. The lesson has become dynamic. The learners have been totally involved in a more personal, physical, technological, collaborative and reflective way. Students do become aware of what is difficult for them and why.

To conclude, working on sounds perception and production during a phonetic workshop, in addition or alternatively to the language laboratory activities, could be considered a regular teaching practice together with the language course program, by further supporting the four communicative skills: listening, reading, writing and speaking.

Bibliography

Calabrò, L. (2015). Il workshop di fonetica in italiano L2/LS. In *Italiano LinguaDue*, 1, 40-49. http://riviste.unimi.it/index.php/promoitals/article/view/5011/Accessed 31.07.15.

Calabrò, L. (2016a). Phone-tic: esperienza pratica e tecnologie per sensibilizzare gli apprendenti stranieri alla riflessione sugli aspetti fonetico-fonologici dell'italiano L2. In Balboni, P., Argondizzo, C. (Eds.), *I 'territori' dei Centri Linguistici Universitari: le azioni di oggi, I progetti per il fututo*. Torino: Utet Università, 456-465.

Calabrò, L. (2016b). Il workshop di italiano L2/LS: accento di parola e sillaba accentata. In *Italiano LinguaDue*, 1, 322-327. http://riviste.unimi.it/index.php/promoitals/article/view/5011/Accessed 07.10.16.

Celata, C., Costamagna, L. (2012). Geminate Timing in the Speech of Estonian L2 Learners of Italian. In De Meo, A., Pettorino, M. (Eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition: The Case of Italian*. Newcastle upon Tyne: Cambridge Scholars Publishing, 115-136.

Celata, C., Costamagna, L. (Eds.) (2014). *Consonant gemination in first and second language acquisition*. Pisa: Pacini.

Chini, M. (2010). Concetti, fenomeni e fattori relativi all'acquisizione di lingue seconde. In Rastelli, S. (Ed.), *Italiano di Cinesi, Italiano per Cinesi. Dalla prospettiva della didattica acquisizionale*. Perugia: Guerra, 23-43.

Cook, V. (2001). *Second language learning and language teaching*. London: Arnold.

Costamagna, L. (1991). Correzione fonetica: utilità del laboratorio linguistico. In *Rassegna di Linguistica Applicata*, 1, 151-176.

Costamagna, L. (1996). *Pronunciare l'italiano*. Perugia: Guerra.

Costamagna, L. (2000). *Insegnare e imparare la fonetica*. Perugia: Guerra.

Costamagna, L. (2010a). I livelli di riferimento e l'insegnamento della fonetica e della fonologia. In Spinelli, B., Parizzi, F. (Eds.), *Profilo della Lingua Italiana. Livelli di riferimento del QCER A1, A2, B1, B2*. Firenze: La Nuova Italia, 75-86.

Costamagna, L. (2010b). L'apprendimento della fonologia dell'italiano da parte di apprendenti sinofoni: capacità e strategie. In Rastelli, S., Bonvino, E. (Eds.), *La didattica dell'italiano a studenti cinesi e il progetto Marco Polo*. Atti del XV seminario AICLU. Pavia: Pavia University Press, 49-62.

Costamagna, L., Marotta, G. (Eds.) (2008). *Processi fonetici e categorie fonologiche nell'acquisizione dell'italiano*. Pisa: Pacini.

Council of Europe (Ed.) (2001). *A Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: CUP.

D'Annunzio, B. (2009). *Lo studente di origine cinese. Risorse per docenti di italiano come L2 e LS*. Perugia: Guerra.

Dal Maso, S. (2003). *Processi di semplificazione della forma delle parole nell'italiano di cinesi in relazione alla struttura e allo statuto della sillaba*. PRIN 2003: Verona.

Dressler, W.U., Mayerthaler, W., Panagl, O. & Wurzel, W. (1987). *Leitmotifs in Natural Morphology*. Amsterdam: Benjamins.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. In *Language Learning*, 27, 315-330.

Ferguson, C. (1984). Repertoire universals, markedness, and second language acquisition. In Rutherford, W. (Ed.), *Language Universals and Second Language Acquisition*. Amsterdam: Benjamin, 247-258.

Gardner, H. (1983). *Frames of Mind: the Theory of Multiple Intelligences*. New York: Basic Books.

Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice*. New York: Basic Books.

Gardner, H. (1999). *Intelligences Reframed: Multiple Intelligences in the 21st Century*. New York: Basic Books.

Gass, S., Selinker, L. (Eds.) (1983). *Language transfer in language learning*. Rowley Mass: Newbury House.

Kagan, S. (2007). *L'apprendimento cooperativo: l'approccio strutturale*. Roma: Edizioni Lavoro.

Kowal, M., Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. In *Language Awareness*, 3, 73-93.

Krashen, S.D. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

Krashen, S.D. (1985). *The Input Hypothesis*. London: Longman.

Lewis, M. (1993). *The Lexical Approach*. Hove: Language Teaching Publications.

Lewis, M. (1997). *Implementing the Lexical Approach*. Hove: Language Teaching Publications.

Luchini, P.L. (2005). A New Approach to Teaching Pronunciation: An Exploratory Case Study. In *The Journal of Asia TEFL*, 2/2, 35-62.

Mastrantuono, E. (2010). Considerazioni teoriche e proposte applicative sull'acquisizione della fonologia nell'insegnamento/apprendimento dell'italiano l2. In *Italiano LinguaDue*, 2/1, 52-65. http://riviste.unimi.it/index.php/promoitals/article/view/630/844/Accessed 20.07.15.

Pienemann, M. (1998). *Language processing and second language development: processability theory*. Amsterdam: Benjamins.

Romito, L., Tarasi, A. (2012). A rhythmic-prosodic analysis of L1 and L2 Italian. In De Meo, A., Pettorino, M. (Eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition: The Case of Italian*. Newcastle upon Tyne: Cambridge Scholars Publishing, 136-152.

Silver, H., Strong, R. & Perini, M. (1997). Integrating learning styles and multiple intelligences. In *Teaching for multiple intelligences*, 55/1, 22-27.

Underhill, A. (2005). *Sound Foundations. Learning and teaching pronunciation*. Oxford: Macmillan.

Wrembel, M. (2007). In search of cross-modal reinforcements in the acquisition of L2 practical phonetics. In Wrembel, M. (Ed.), *Speak Out! The Newsletter of the Pron SIG*, 38, 39-43.

Wrembel, M. (2011). Cross-modal reinforcements in phonetics teaching and learning: an overview of innovative trends in pronunciation pedagogy. In *Proceedings of 17th ICPhS*, Hong Kong, 104-107.

DOI: 10.17469/O2102AISV000015

PAOLO MAIRANO, LIDIA CALABRÒ

Are minimal pairs too few to be used in pronunciation classes?

In this contribution we address the usage of minimal pairs in L2 pronunciation classes. An informal survey with FL teachers of Italian and English revealed that minimal pairs are considered to be scant and difficult to find. We present here a tool (Minimal Pair Finder) that has been developed with the aim to support teachers and learners in pronunciation classes by providing quick access to several minimal pairs via a top-down approach. We describe how this tool can be consulted, how it has been implemented, and we show a sample teaching unit where students are asked to make use of it. Minimal Pair Finder reveals that minimal pairs are generally not too few to be used in pronunciation classes; however, we argue that L2 teachers should wisely choose minimal pairs for their classes based on the proficiency level of their students, by paying attention to parameters such as productivity and word frequency.

Key words: minimal pairs, pronunciation, tool, pronunciation teaching.

Introduction

Minimal pairs are pairs of words which differ by just one sound, e.g. pet – bet. Such pairs have been used since the time of classical phonology (Trubetzkoy, 1939) as a proof that two similar sounds have a distinct function in a language and can thereby be considered as phonemes, rather than mere variants of the same abstract entity (allophones). The procedure of replacing one sound with another in a word and checking whether this produces a new word has been called commutation test and has also been used since classical phonology. Minimal pairs are still important in present-day research in phonology, for example in studies measuring functional load (Oh, Coupé, Marsico & Pellegrino, 2013, and Oh, Pellegrino, Coupé & Marsico, 2015). And they have also been widely used outside proper phonology for various purposes. For instance, they are often used in psycholinguistic experiments testing first/second language acquisition issues (e.g., Pallier, Colomé & Sebastián-Gallés, 2001, as well as Lin, Chang & Cheung, 2004).

It has been suggested by numerrous authors (see for instance Breitkreuz, Derwing & Rossiter, 2009) that minimal pairs can also be profitably employed for didactic purposes, notably to illustrate and teach phonological oppositions to learners of foreign languages. In effect, the observation of minimal pairs is a metalinguistic exercise that helps learners understand the importance of pronouncing and perceiving sounds that they may erroneously consider as the same phonological entity based on their native language. For instance, novice Italian learners of

L2 English will have a tendency to perceive/produce *fill* /fɪl/ and *feel* /fiːl/ as the same word, since both /iː/ and /ɪ/ are assimilated to the closest L1 phonological category, namely /i/ (this and similar phenomena are widely described in L2 phonological models such as Best, Tyler, 2007, and Flege, 1995). Many authors (e.g., Renard, 1979, and Celce-Murcia, Brinton & Goodwin, 1996, and Santiago, 2012) suggested that exercises with minimal pairs can be beneficial for developing phonological awareness. Analysing such pairs of sounds, hearing their pronunciation and observing the change in meaning can be an important contribution to help learners acquire this phonological opposition. Explicit exercises using minimal pairs may in effect contribute to improve learners' pronunciation and phonological awareness.

Minimal pairs are frequent in general ESL/EFL textbooks, as reported by Levis, Cortes (2008). However, we find that not many textbooks of Italian as a SL/FL include exercises on or about minimal pairs, and/or exhaustive lists of minimal pairs to support teachers and students. An informal enquiry among teachers of English/Italian as a FL at the Universities of Warwick and Rome 3 revealed that teachers consider minimal pairs to be few or difficult to find. Lists of such word pairs are not easy to find (except Baker, Goldstein & Dolgin, 1990, for English), so we attempt here to provide a solution to this problem.

In the first part of our contribution we shall present a tool called *Minimal Pair Finder* (MPF), which has been developed in order to assist learners, teachers and linguists to search for minimal pairs of English and Italian. It is freely available online at *http://phonetictools.altervista.org/minimalpairfinder/* and more languages will possibly be added in the future. In the second part of this article, we shall illustrate how MPF can be used in L2 classroom activities: we present pronunciation exercises developed with it and we describe how it can be used in a class of Italian as a FL, along the lines of what is being done at the University of Rome 3. In the third and last part of this article, we shall present our considerations about using minimal pairs in L2 pronunciation classes.

1. *Minimal Pair Finder*

1.1 Using Minimal Pair Finder

MPF has a simple HTML/JavaScript interface (see figure 1) that lets the user specify a language and a pair of phonemes. As for the language choice, only American English and Italian are currently implemented, but collaborations have already been set up with other universities to extend the tool to more languages. As for the phoneme choice, the list contains all phonemes traditionally described for each language and is updated automatically whenever the user switches languages.

Figure 1 - *MPF interface for standard queries*

Once the user launches the search, the tool will dynamically look up a lexicon and return all minimal pairs found in it matching the criteria specified by the user; for productive oppositions, the output can contain several hundred pairs. Depending on the language, the results also contain extra information (such as frequency in a reference corpus for each word in the pair, see figure 2 – details in the next section).

Figure 2 - *Output of a query in MPF*

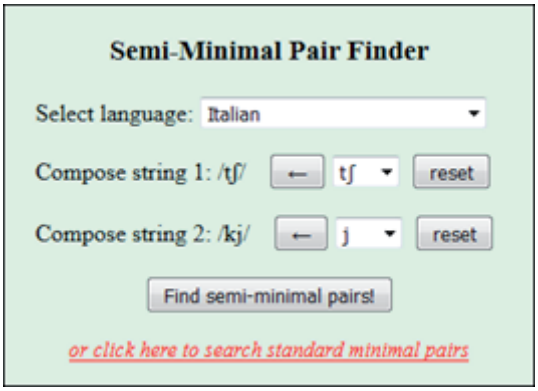
283 minimal pairs for /f/ and /v/ were found in the Italian data.
Words come from text corpora and may contain errors or imprecisions:
the most suspiscious-looking words are marked like this.
Numbers in parenthesis indicate word frequency in the COLFIS corpus.

/f/	/v/
fa (3566) fà (6) /fˈa/	va (1389) vah (0) /vˈa/
fino (1924) /fˈino/	vino (233) /vˈino/
fanno (863) /fˈan:o/	vanno (463) /vˈan:o/
foto (354) /fˈɔto/	voto (369) /vˈɔto/
fede (238) /fˈede/	vede (383) /vˈede/
fu (1450) /fˈu/	v (50) vu (7) vù (2) /vˈu/
fia (11) /fˈia/	via (2563) /vˈia/
inferno (119) /infˈɛrno/	inverno (191) /invˈɛrno/
fai (143) /fˈai/	vai (90) /vˈai/
finto (39) /fˈinto/	vinto (314) /vˈinto/
fan (66) /fˈan/	van (160) /vˈan/
fini (196) /fˈini/	vini (51) /vˈini/

Additionally, MPF also has an advanced feature that makes it possible to search for semi-minimal pairs, i.e. words that are identical except for n>1 phonemes. This feature may be used for various purposes, such as looking for pairs of words opposing /tʃ/ *vs.* /kj/ (e.g. Italian *cedere vs. chiedere*), and once French is ready, /ɑ̃/ *vs.* /an/

and similar, or syllables such as /ma/ *vs.* /no/, or longer segments such as /ˈtart/ *vs.* /ˈtɔrd/, or something totally unrelated and having different lengths such as /ˈtart/ *vs.* /ˈnud/.

Figure 3 - *Advanced MPF interface for searching semi-minimal pairs*



It has to be noted that other software exists to find phonological neighbours for various languages, such as Worden (Origlia, Cangemi & Cutugno, 2015) and the Clearpond database (Marian, Bartolotti, Chabal & Shook, 2012). The distinctive feature of MPF is that it leads the user through a bottom-up (rather than top-down) search. Both Worden and Clearpond let the user input one word or non-word, and they will provide phonological neighbours of various types according to the options specified. Instead, the search in MPF is always bottom-up: the user specifies the terms of the opposition (i.e. the two opposing segments, be they phonemes such as /p/ *vs.* /b/, or longer chunks such as /tʃ/ *vs.* /kj/) and the tool will output corresponding word pairs. This means that MPF is somehow complementary to Worden or Clearpond and, in our view, it responds to the needs of learners and teachers looking for lists of minimal pairs. In effect, we can imagine that teachers are not interested in finding minimal pairs (or other phonological neighbours) of one given word; rather, if they are planning a pronunciation teaching session on the palatal nasal in Italian, they might be interested in finding minimal pairs given the phonemes /n/ *vs.* /ɲ/; or, similarly, if they are planning a teaching session on geminates, they might be interested in finding examples of /m/ *vs.* /mː/, /l/ *vs.* /lː/etc. For this reason, we believe that MPF can profitably be used by learners/teachers of foreign languages.

1.2 The implementation of Minimal Pair Finder

The implementation of MPF is fairly simple and the search engine is written in PHP. It relies on a lexicon with orthographic forms and corresponding phonological transcriptions. For American English, we simply used the *CMU Pronouncing Dictionary*, which is freely available online (*http://www.speech.cs.cmu.edu/cgi-bin/cmudict*). For Italian, we combined two sources: the list of lemmas found in the CoLFIS corpus

(Bertinetto, Burani, Laudanna, Marconi, Ratti, Rolando & Thornton, 2005), and the list of lemmas in Garzanti's Italian dictionary, which can be downloaded from the publisher's website (*http://www.garzantilinguistica.it/lemmario-italiano/*). We combined those two complementary sources in order to have a richer list of entries in the lexicon: Garzanti's list contains only lemmas, while COLFIS contains many word forms that can contribute to the output list of minimal pairs. For example, had we used only lemmas, MPF could not find such productive oppositions as Italian /m/ *vs.* /mː/ in future tense *vs.* conditional (e.g. *andremo* /anˈdrɛmo/ - *andremmo* /anˈdrɛmːo/). Instead, had we used COLFIS only, MPF would not be able to find minimal pairs such as Italian *intorpidire - intorbidire* as neither of these words occur in COLFIS.

Both sources (Garzanti's lemmas and CoLFIS word forms) were transcribed with a component of the Espeak TTS system[1] (*http://espeak.sourceforge.net/*). Transcriptions were mapped to an internal symbol set for programming convenience, but are then further remapped and presented as IPA symbols in the results: this means that the user can comfortably use the IPA alphabet without being aware of the double remapping which happens "under the hood".

For Italian, we also included frequency information for each word form as found in the CoLFIS corpus: this way, the output minimal pairs can be sorted in a tentative *reliability order* with the intention to have "better" (i.e. frequent and native-looking) words high on the list, and foreign or uncommon words further down in the list. This is achieved by a complex set of rules that attribute a score to each word on the basis of

[a] their frequency in CoLFIS and

[b] Italian phonotactic and orthographic restrictions (i.e. giving a certain cost to words ending in one or more consonants, having foreign letters, showing unauthorized consonant clusters, etc.). For instance:
- Exists in Garzanti: +100.
- Occurs in COLFIS: +1 for each occurrence.
- Orthography contains letters (or symbols) that are not included in 'aàbcdeèéfghiìlmnoòópqrstuùvxz' (and corresponding capital letters): -70 for each letter/symbol.
- Contains unauthorized consonant clusters[2] (e.g. 'broadband', 'Burlington', 'Kommerzbank', 'feedback'): -50 for each unauthorized cluster.

[1] Although MPF is already fully functional and available online, we are currently in the process of manually screening output transcriptions and correcting mistakes coming from Espeak's automatic transcriber module.

[2] This check is actually performed by multiple rules. The following regular expression checks for generic unauthorized clusters:
/b[fgpqvxz]|c[bdfgpvx]|d[bfpqtxz]|f[bcdpqvxz]|g[fpqtx]|p[bcdfgvx]|q[^qu]|v[cfpqtxz]|x[bcdfglmnpqrsvxz]|z[bcdfpqvx]/i
Successive rules check for additional constraints on double consonantal graphemes (geminates), making sure they are preceded by a vowel and followed by a vocalic grapheme, or <r>, or <l>.

- Contains typical foreign letter combination (e.g. 'eau', 'ou', 'ées'): -40 for each combination.
- Ends in consonant(s): -30 for each consonant.
- Orthographic form contains double vowel grapheme (e.g. 'scooter'): -25.
- Orthographic form contains >4 consonant graphemes in a row (e.g. 'Gershwin', 'Goldsmith'): -20.
- Etc.

Additionally, we also used a whitelist to account for exception words such as 'per' (*for*) and 'nord' (*north*), which would otherwise be penalized as non-native looking by the rules above. The reliability score of each word in a minimal pair is then combined via another set of rules[3] to get a final score for the pair as a whole, which determines its relative position in the list.

Instead, the ranking feature was not added to English. This has two main reasons: on the one hand, it is not possible to identify non-native English words by relatively simple orthographic rules as we did for Italian. On the other hand, we did not dispose of frequency information for English word forms in a reference corpus. We did implement a ranking of word-forms on the basis of trigram frequencies, but the results are not satisfactory and we are still looking for a better alternative.

Another issue concerns orthographic ambiguity in both Italian and English, and particularly homographs which are not homophones, such as *wind* ([wɪnd] *vs.* [waɪnd]) in English and *pesca* (['pɛska] *vs.* ['peska]) in Italian. These cases are represented in the CMU as two distinct entries; this does not pose any problem to MPF, which also stocks them as two completely separate items, each with its pronunciation. In effect, the automatic transcription coming from a TTS component is deterministic by definition: it only outputs one transcription for such words, meaning that only one of the two (or more) possible pronunciations is represented. The solution has been that of manually building a list of Italian homographs which are not homophones, and appending it to the data. This includes homographs with lexical stress on a different syllable (e.g. *scrivano* ['skrivano] *vs.* [skri'vano]), and minimal pairs opposing /e/-/ɛ/ or /o/-/ɔ/ (the only phonological oppositions of Standard Italian that are not marked by the orthography).

The opposite case concerns homophones that are not homographs, such as *waste vs. waist* ([weɪst]) in English, and *hanno vs. anno* (['anːo]) in Italian. These words have multiple entries in all of the sources we used (CMU, COLFIS, Garzanti). So, in order to overcome this issue, we organized MPF data with phonetic transcrip-

[3] One may think that it is possible to simply add the scores of each word in the minimal pair, but this unfortunately gives odd results as it favours pairs where one word is very frequent, even if the other is not. We found that a slightly more complex algorithm gave cleaner results than a simple score sum: if scores for both words in the pair are > 0, such scores are multiplied; else, they are summed and then divided by 100. This will heavily penalize pairs where one (or both) of the words has a negative score. Moreover, the multiplication of positive scores will heavily favour pairs where both words have high scores, whereas a simple sum would also favour pairs where one word has a high score, and the other does not.

tions (instead of orthographic forms) as the key for retrieval. This means that words like *waste* and *waist* are listed within the same entry (e.g.: [weɪst]#waist#waste) and are both retrieved when the result of a search includes [weɪst].

Many further improvements of MPF are currently being considered, apart from extensions to other languages. From an L2 pronunciation learning/teaching perspective, the most interesting extensions would consist in adding audio to all or some words via speech synthesis, and adding information to each word about its competence level (A1, A2, B1, B2, C1, C2). The latter would make it possible (given 2 phonemes) to exclusively get minimal pairs that are adequate to the learner's level of competence (say C1). It would require a categorization of words into competence levels such as has been done within the *English Vocabulary Profile* project (see Capel, 2012) and by the *Instituto Cervantes* for Spanish (see *http://cvc.cervantes.es/Ensenanza/Biblioteca_Ele/plan_curricular/indice.htm*). Until this feature becomes available, teachers will need to screen all output minimal pairs and select the ones that are appropriate to their students' level.

2. *Minimal Pair Finder in L2 pronunciation classes*

MPF targets a double audience: on the one hand, we believe it can be useful to researchers working in experimental phonetics, phonology and psycholinguistics. The first author started the development of this tool while striving to find minimal pairs matching specific requirements, and then successfully used it for ongoing psycholinguistic experiments. On the other hand, we believe it can be useful to learners and teachers for pronunciation classes.

Teachers can obviously find a wealth of minimal pairs to be used as illustrations in their classes. Learners can also find many examples to learn phonological oppositions; but above all, they can experience first-hand how productive certain oppositions are in the language they are studying. Learners of Italian as an L2 will for instance have the opportunity to see how many minimal pairs can be created by contrasting singletons and corresponding geminates, and thereby (hopefully) realize that this opposition is worth learning. They will also be able to focus on the type of contexts that any given opposition can create: for example, they can observe that /m/ *vs.* /mː/ is a recurring opposition in verbs for future tense *vs.* conditional (e.g. *mangeremo vs. mangeremmo*).

We have in fact prepared some specific activities around this tool that have been tested in July 2016 with 12 Chinese learners of Italian as an L2 at the University of Rome 3 by the second author. After the intervention, students filled in a short questionnaire where, among other things, they were asked to state whether they found MPF to be a useful tool for learning the pronunciation: out of 12 students, 7 gave a positive evaluation, 1 left a negative evaluation, 2 left neutral comments, 1 left no comment, 1 left an unintelligible comment. Full details of this intervention are given in Calabrò, Mairano (in preparation).

The activities (which can be found in the appendix in English translation) use minimal pairs inserted in specific sentence contexts to help learners improve perception and production of the target phonological oppositions. They can be considered as a first attempt to use MPF to create phonetic activities for learners of Italian as a FL/SL: they can be a support for teachers who are not experienced in phonetics but would like to improve their students' production and perception abilities. They can certainly be a useful illustration for teachers on how to create their own activities with MPF, once they have familiarized with it. Learners can do the activities on their own and then discuss them with the teacher/classmates; or else, they can do them during a phonetic workshop.

The aim of the lesson is to practice the opposition /m/ *vs.* /mː/ in Italian. The activities have been prepared for learners at a final A2 or initial B1 level of the *C.E.F.R.* (*Common European Framework of Reference*). The whole lesson is divided in two main parts: two pre-activities (a-b), and six activities proper (c-h).

2.1 Pre-activities (a-b)

The pre-activities have been conceived to awaken the students' previous awareness regarding the /m/ *vs.* /mː/ phonological opposition in production and perception. In (a) learners are asked to start thinking about words that contain the two sounds, and to write them down. If they are working in classroom or in a lab, they can compare their own answers with their classmates' answers. Once they have found an acceptable number of words for their level, they should pass on to the second step and (b) listen to sentences containing one of two words composing a minimal pair. They are asked to mark with an X whether they hear one or the other word, then they can discuss with the teacher and classmates what they have perceived. In this part of the activity the teacher does not give the solution as this should come as a discovery.

2.2 Activities (c-e)

In the real activities students proceed to really use MPF. In (c) they are asked to look at the web page and search for words containing the sounds /m/ *vs.* /mː/ and to write down the five words with the highest frequency for both sounds. The teacher should make sure they all understand that the frequency number is in rounded brackets. This activity can be followed by a comparison of the results with classmates. In (d) learners are asked if they know the words they have found, and what the differences in meaning and pronunciation are. For this part, a plenary discussion with classmates and the teacher is fundamental as it helps to strengthen their awareness and/or correct wrong ideas. In (e) students are invited to start a metalinguistic analysis based on the minimal pairs they have found, specifically about pronunciation. They are asked to observe the sentence context in which the words appear and to think about a rule relating words in these minimal pairs. Of course, the aim of this activity is to let students make a connection between the geminate and the conditional tense *vs.* the singleton and the future tense.

2.3 Post-activities (f-h)

To conclude the work package, we propose 3 post-activities meant to consolidate what students have learnt. In (f) learners can find six sentences taken from COLFIS to be read aloud using word linkers (which are supposed to improve fluency). After practicing, they are subsequently asked to listen to the previous sentences in (g) and to discuss with the teacher about the difference of intonation. Finally, they are asked to listen again to the recorded sentences and repeat what they hear: the aim here is to consolidate what they have learnt and at this stage the teacher should correct the pronunciation of /m/ and /m:/ if necessary.

3. *Are minimal pairs too few to be used in pronunciation classes?*

In the last part of this paper, we shall try to address the question that originally brought us to develop MPF (see the introduction): are minimal pairs too few to be used in pronunciation classes? The answer is not trivial, as much depends on at least 3 factors: (1) the target phonological opposition, (2) the language in question, (3) the level of learners.

3.1 The target phonological opposition

Obviously, not all phonological oppositions are equally productive, as is widely known from the literature since Trubetzkoy (1939). Some phonological oppositions may be attested by a large number of minimal pairs, while others are only attested by a few. The productivity of a phonological pair is in effect linked to the function load of each phoneme. The function load can be defined as the capability of a certain phoneme to create minimal pairs: in fact, recent studies (e.g., Oh et al., 2013; 2015) measure the functional load of a phoneme by calculating the number of minimal pairs that would be neutralized if a phoneme were deleted from a language. Also, some oppositions are only active in certain contexts and are neutralized in other contexts: for instance, the singleton *vs.* geminate opposition in Italian is only active word-internally[4] and in specific phonotactic conditions[5]. Such oppositions may as a very general rule be considered as less productive, but remarkable exceptions exist (Italian gemination being one of them).

MPF shows that the most productive oppositions can create hundreds or even thousands of minimal pairs. For example, the /p/ *vs.* /b/ oppositions outputs 488 minimal pairs in the Italian data, /t/ *vs.* /d/ outputs 422 minimal pairs, /o/ *vs.* /a/ outputs 4034 minimal pairs[6], and /m/ *vs.* /m:/ outputs 251 minimal pairs. By con-

[4] Except for *raddoppiamento fonositattico*, which we will ignore here for the sake of simplicity and because it is not relevant with the subject of this paper.

[5] Namely between two vowels, or preceded by a vowel and followed by one of /l/, /r/, /j/, /w/.

[6] MPF's output is limited by the memory allocated by the server, so the tool cannot screen all data in cases of such productive oppositions. The output for /a/ *vs.* /o/ would be even higher if all data could be screened.

trast, other oppositions can be far less productive. This is the case of, for example, /n/ *vs.* /ɲ/ which yields 64 minimal pairs in our Italian data, /j/ *vs.* /ʎ/ which yields 41 minimal pairs, or /dʒ/ *vs.* /ʒ/ which yields only 32 minimal pairs in our English data[7].

So, we can provisionally claim that many phonological oppositions make it possible to find a wealth of minimal pairs. In the case of less productive oppositions, it may be more difficult to find minimal pairs, but usually they will still come in acceptable numbers for a pronunciation class.

3.2 The language in question

Of course it is no surprise that the language in question also is an important factor for the productivity of a given phonological contrast. The same pair of phonemes (or similar phonological entities existing in two languages) can create many oppositions in one language, and just a few in another language. We can illustrate this by the /tʃ/ *vs.* /ʃ/ opposition, which yields 394 minimal pairs in our English data, but only 99 minimal pairs in our Italian data[8]. This distinction is not really relevant for L2 pronunciation classes because they usually focus on one target language. However, on some occasions, foreign language teachers may still want to keep in mind that phonological oppositions in their students L1 may have a different weight in the L2, or vice versa.

3.3 The proficiency level of learners

The proficiency level is certainly another relevant factor at play when looking for minimal pairs to be used in pronunciation classes. Obviously, it is preferable and advisable to use words with which learners are familiar with, and for two reasons. Firstly, presenting unfamiliar words will concentrate the students' effort and attention to learning such words, rather than learning their correct pronunciation. In fact, according to VanPatten (1996), teachers should focus learners' attention to one aspect at a time (see Akerberg, Espinosa & Santiago, 2016, for an application on L2 pronunciation). Secondly, the reason for using minimal pairs is to make the student realize that the target phonological opposition is important and can create differences in meaning: if only obscure words are used, learners may not be able to grasp the change in meaning and, as a consequence, they would not be motivated to learn the target phonological contrast.

So, we propose that teachers focus on productive or fairly productive oppositions – at least with elementary or intermediate students, so that a wealth of examples is available where at least one of the two words is familiar to the students. This idea is

[7] These results have been obtained with MPF version online on 29th June 2016; the implementation of the tool or the data may be modified in the future, and these numbers may change.

[8] We have to recognize that it can be dangerously misleading to compare output from our Italian and English data, because they are very different. The English data (coming from the CMU pronunciation dictionary) is composed of many more entries and includes many proper names. The Italian data is less rich, so the output minimal pairs tend to be fewer as a consequence.

not new: other authors have suggested that teachers evaluate the functional load of a phonological opposition in order to decide whether or not it is worth teaching it (see Brown, 1988, as well as Levis, Cortes, 2008). Finding suitable words should not be too difficult with MPF, which also outputs word frequency information (from COLFIS): the higher the frequency, the more relevant the word and therefore the more useful the minimal pair. Only with advanced students can teachers dare to use minimal pairs made up of more infrequent words. In fact, at higher proficiency levels, it is even possible to combine a pronunciation lesson with a vocabulary learning lesson using minimal pairs from lesser productive oppositions.

4. *Final remarks*

On the whole, we think that the use of minimal pairs is a viable way to teach students the importance of correctly pronouncing and perceiving sounds that contrast phonologically. The use of minimal pairs may be more appropriate for some oppositions than others, notably for the most productive and yet challenging ones, such as singletons vs. geminates in Italian. This is in line with what has been suggested by recent literature in L2 pronunciation: Munro, Derwing (2006) found that pronunciation errors involving oppositions with high functional load had a heavy impact on ratings of accentedness and comprehensibility. The authors suggest that the functional load principle should therefore guide pronunciation instruction. We also propose that the use of minimal pairs with high vs low functional load needs to be tuned to the learner's proficiency level. Most productive oppositions and more frequent words are to be favoured for students at lower and intermediate proficiency levels, whereas less productive oppositions and less frequent words can be introduced in high proficiency levels (see also Brown, 1988).

We provide a tool (*Minimal Pair Finder*, MPF) that can (a) support teachers in the search for adequate minimal pairs to be used in their classes, and (b) support learners in observing the productivity of certain phonological oppositions in the target language. We are profitably[9] using this tool in L2 pronunciation classes and we provide an illustration of how to do so in the appendix.

Bibliography

Akerberg, M., Espinosa, A. & Santiago, F. (2016). *Manual de pronunciación para profesores de L2*. Mexico: UNAM.

Baker, A., Goldstein, S. & Dolgin, P. (1990). *Pronunciation Pairs: An Introductory Course for Students of English. Student's Book*. Cambridge University Press.

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C. & Thornton, A.M. (2005). Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS). http://linguistica.sns.it/CoLFIS/Home.htm/Accessed 29.06.16.

[9] Cfr. the evaluations given by students after the teaching intervention and briefly reported in section 2.

Bertinetto, P.M., Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. In *Journal of the International Phonetic Association*, 35/2, 131-151.

Best, C.T., Tyler, M.D. (2007). "Nonnative and Second Language Speech Perception: Commonalities and Complementarities." In *Language Experience in Second Language Speech Learning*, in honor of James Emil Flege. Amsterdam: John Benjamins Publishing Company.

Breitkreutz, J., Derwing, T.M. & Rossiter, M.J. (2009). Pronunciation teaching practices in Canada. In *TESL Canada Journal*, 19/1, 51-61.

Brown, A. (1988). Functional load and the teaching of pronunciation. In *Tesol Quarterly*, 22/4, 593-606.

Calabrò, L., Mairano, P. (in prep.). Teaching pronunciation to Chinese learners of Italian with Minimal Pair Finder.

Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. In *English Profile Journal*, 3/1, 1-14.

Celce-Murcia, M., Brinton, D. & Goodwin, J. (1996). *Teaching pronunciation: a reference for teachers of English to speakers of other languages*. New York: Cambridge University Press.

The CMU Pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict/ Accessed 05.10.15.

Grande Dizionario Italiano Garzanti. List of lemmas retrieved from http://www.garzanti-linguistica.it/lemmario-italiano/Accessed 27.10.16.

Espeak TTS system. http://espeak.sourceforge.net/Accessed 04.09.15.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Levis, J., Cortes, V. (2008). Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching. In Chapelle, C.A., Chung, Y.-R. & Xu, J. (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment*. Ames, IA: Iowa State University, 197-208.

Lin, H.L., Chang, H.W. & Cheung, H. (2004). The effects of early English learning on auditory perception of English minimal pairs by Taiwan university students. In *Journal of psycholinguistic research*, 33(1), 25-49.

Marian, V., Bartolotti, J., Chabal, S. & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. In *PLoS ONE* 7(8): e43230.

Munro, M., Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. In *System*, 34/4, 520-531.

Oh, Y., Coupé, C., Marsico, E. & Pellegrino, F. (2015). "Bridging Phonological System and Lexicon: Insights from a Corpus Study of Functional Load". In *Journal of Phonetics*, 53, 153-176.

Oh, Y., Pellegrino, F., Coupé, C. & Marsico, E. (2013). "Cross-language Comparison of Functional Load for Vowels, Consonants, and Tones". In *Proc. of Interspeech 2013*, Lyon, France, 25-29 August, 3032-3036.

Origlia, A., Cangemi, F. & Cutugno, F. (2015). *WORDEN:* a PYTHON interface to automatic (non)word generation. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Language acquisition and Language Loss: acquisition, change and disorders of the language sound structure*, *Studi AISV*, 1, 459-470.

Pallier, C., Colomé, A. & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. In *Psychological Science*, 12(6), 445-449.

Renard, R. (1979). *Introduction à la méthode verbo-tonale de correction phonétique*. Brussels: Didier.

Santiago, F. (2012). La didactique de la phonétique de L2 et la perception auditive: vers une nouvelle approche. In *Synergies Mexique*, 2, 57-70.

Trubetzkoy, N. (1939). Grundzüge der Phonologie. In *Travaux du Cercle de Linguistique de Prague 7*. [English translation: Baltaxe, C.A.M. (trans.) (1969). *Principles of Phonology*. Berkley, CA: University of California Press].

VanPatten, B. (1996). *Input processing and grammar instruction: theory and practice*. New Jersey: Ablex Publishing Corporation.

Appendix

The appendix contains a work package using MPF for classes of Italian as a FL/SL. Instructions have been translated into English for the comfort of readers.

- **Type of activity**: discover minimal pairs opposing /m/ *vs.* /m:/
- **Level of students**: A2-B1 of C.E.F.R.
- **Time**: approx. 75 minutes

Pre-activities

a) Work alone if you are at home or in pairs if you are in the classroom. Think about some words that contain the sounds /m/ and /m:/, and write them down. Then, compare your answers with classmates.
b) Listen to the sentences: mark with an X what you hear. Then discuss with your teacher and classmates (the teacher will not give the solution).

1	a. Tra dieci anni avremo una casa tutta nostra. b. Tra dieci anni avremmo una casa tutta nostra.
2	a. Forse potremo cambiare macchina. b. Forse potremmo cambiare macchina.
3	a. Domani saremo tanto stanchi. b. Domani saremmo tanto stanchi.
4	a. Il camino è piccolo. b. Il cammino è piccolo.
5	a. Dovremo comunicare di più e meglio. b. Dovremmo comunicare di più e meglio.

Activities

a) Look at MPF web page and search for minimal pairs opposing the sounds /m/ and /m:/. What results can you find? Write down the five words with the highest frequency for both sounds (you can find the frequency in rounded brackets next to each word). Then compare your results with your classmates.
b) Do you know these words? What is the difference in meaning within each pair? And what is the difference in pronunciation? Discuss you answer with the teacher and classmates.
c) Work in pairs. Go back to the previous words and observe them in their sentences. Answer the questions below and then discuss them with your teacher and classmates.
   – Where do /m/ and /m:/ appear most often?
   – Can you find a rule for when either sound is used?
   – Which is the difference in pronunciation?

Post-activities

a) Read the following sentences from COLFIS and use word linkers to get a fluent reading.
   1. *Mi dicono: avremo un campionato di calcio eccellente!*
   2. *Noi non avremmo problemi di sorta.*
   3. *Sono certo che dovremo affrontare tutti qualche sacrificio.*
   4. *Dovremmo proprio prendere un tappeto!*
   5. *Vicino al camino, trovato acceso, ai piedi del letto, è stata trovata una tavola imbandita con cibi cucinati.*
   6. *Quando cammino per strada in America vado tranquillo perché non mi riconosce nessuno.*
b) Listen to the sentences above and discuss with the teacher about the difference in intonation.
c) Listen again and repeat the sentences.

DOI: 10.17469/O2102AISV000016

PIERO COSI, GIULIO PACI, GIACOMO SOMMAVILLA, FABIO TESSER

CHILDIT2 – A New Children Read Speech Corpus

One of the main achievement of the recently concluded European FP7 project ALIZ-E ("Adaptive Strategies for Sustainable Long-Term Social Inter-action") has been the collection of various new Italian children's speech annotated corpora. From some of this speech material the CHILDIT2 corpus has been created and this paper describes in detail its design, building and development.

Key words: children, speech, corpus.

Introduction

The Padova Institute of Cognitive Sciences and Technologies (ISTC) of the National Research Council (CNR) has been the partner of the ALIZ-E ("Adaptive Strategies for Sustainable Long-Term Social Interaction") project (Belpaeme, Baxter, Read, Wood, Cuayahuitl, Kiefer, Racioppa, Kruijff-Korbayová, Athanasopoulos, Enescu, Looije, Neerincx, Demiris, Ros-Espinoza, Beck, Cañamero, Hiolle, Lewis, Baroni, Nalin, Cosi, Paci, Tesser, Sommavilla & Humbert, 2013) responsible of carrying out studies in the field of speech technologies, as described in (Tesser, Paci, Sommavilla & Cosi, 2013) and (Paci, Sommavilla, Tesser & Cosi, 2013).

One of its main achievements has been the collection of various new Italian children's speech annotated corpora (Cosi, Paci, Sommavilla & Tesser, 2015) and in this paper the design, building and development of CHILDIT2, a new read children's speech corpus, is described in detail.

1. *Data Collection*

CHILDIT2 is made up by sentences read by young children, and prompts from the FBK CHILDIT corpus (Gerosa, Giuliani & Brugnara, 2007) have been used. They are phonetically balanced sentences, selected from children's literature.

In the original recording set-up, as illustrated in Figure 1, during each session the input coming from the four microphones of Nao (a robot used in the ALIZ-E project), a close-talk microphone and a panoramic one has been recorded, and for CHILDIT2, only the close talk microphone has been taken into consideration.

Figure 1 - *Data Collection framework: A,B,C,D - 4 microphones of Nao (the robot used in the ALIZ-E project); E - 1 close-talk microphone; F - 1 panoramic microphone*

Four main recording sessions in normal silent rooms have been performed during the ALIZ-E project. In July 2011, 31 children (age 6-10) have been recorded at a Summer school at Limena (PD, Italy); in August 2012, at a Summer school for children with diabetes, recordings from 5 children (age 9-14) have been collected. In 2013 two final sessions have been carried out: the first one (March-April 2013, at Istituto Comprensivo "Gianni Rodari", Rossano Veneto) involved 52 young users aged between 11 years to 14 years; in the second one (August 2013), eight children aged between 11 and 13 years have been recorded at the Summer school for children with diabetes at Misano Adriatico. All recording sessions consist of data from 96 Italian young speakers, for a total amount of 4875 utterances, resulting in more than eight and a half hours of children's speech.

For all recording sessions, an external Zoom H4N device connected to a laptop computer's USB port has been used (see Fig. 1). A Shure WH20QTR Dynamic Headset or a Proel RM300 close talk microphone, plugged into the Zoom's input, has been indifferently chosen for recording, depending on the different sessions and the audio format is characterized by the following set: Channels: 1, Sample Rate: 16000 (originally 48000), Precision: 16-bit and Sample Encoding: 16-bit Signed Integer PCM.

2. *Final Considerations*

Free available speech data are essential for small labs to build and develop new ASR systems and to improve their knowledge on speech of specific group of people, such as the children one.

As illustrated in previous papers (Cosi, Nicolao, Paci, Sommavilla & Tesser, 2014; Cosi, 2015) the original CHILDIT corpus was quite useful in the past to build children speech ASR systems, and it was extensively tested with various open-source ASR systems producing very good PER (phoneme-error-recognition) results (see Table 1).

Table 1 - *PER (phoneme-error-recognition) for various open-source systems tested on CHILDIT*[1]

CHILDIT	SPHINX	BAVIECA	SONIC	KALDI	KALDI (DNN)
Applied Adaptation Methods	VTLN+MLLR (5 Loops)	MLLR (5 Loops)	VTLN + SMAPLR (5 Loops)	LDA+MLLT SGMM+MMI (4 Loops)	DNN+ SMBR
Baseline	18.7 %	16.9 %	15.03 %	13.8 %	8.5 %
Best Score	17.3 %	14.7 %	12.4 %	8.6 %	8.1 %

In a set of recent and still not published experiments, KALDI (Povey, Ghoshal, Boulianne, Burget, Glembek, Goel, Hannemann, Motlicek, Qian, Schwarz, Silovsky, Stemmer & Vesely, 2011; Kaldi ASR-web) was tested on CHILDIT+CHILDIT2. Results, shown in Table 2, are quite better than those obtained with the previous experiments where only CHILDIT was used, showing both the importance of using more data to improve recognition performance and also that the quality of the data in the newly created CHILDIT2 corpus is the same as that of CHILDIT.

Table 2 - *PER (phoneme-error-recognition) for KALDI ASR system tested on CHILDIT+CHILDIT2*

CHILDIT + CHILDIT2	KALDI	KALDI (DNN)
	LDA+MLLT SGMM+MMI (4 Loops)	DNN+ SMBR
	12.5 %	7.9 %
	7.9 %	7.3 %

[1] VTLN,: Vocal Tract Length Normalization; MLLR: Maximum Likelihood Linear Regression SMAPLR: Structural Maximum A Posteriori Linear Regression; LDA: Linear Discriminant Analysis; MLLT: Maximum Likelihood Linear Transform; SGMM: Subspace Gaussian Mixture Models; MMI: Maximum Mutual Information; DNN: Deep Neural Network; SMBR: State-level Minimum Bayes Risk.

CHILDIT2 is freely available to the research community[2] and it is licensed by FBK and ISTC CNR, UOS Padova, under a Creative Commons Attribution-Non-Commercial-Share-Alike 4.0 International License.

Acknowledgements

We acknowledge FBK and in particular, Diego Giuliani, for inspiring and guiding the development of the whole CHILDIT2 project. This work was partially supported by the EU FP7 "ALIZ-E" project (grant number 248116).

Bibliography

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayahuitl, H., Kiefer, B., Racioppa, S., Kruijff-Korbayová, I., Athanasopoulos, G., Enescu, V., Looije, R., Neerincx, M., Demiris, Y., Ros-Espinoza, R., Beck, A., Cañamero, L., Hiolle, A., Lewis, M., Baroni, I., Nalin, M., Cosi, P., Paci, G., Tesser, F., Sommavilla, G. & Humbert, R. (2013). Multimodal Child-Robot Interaction: Building Social Bonds. In *Journal of Human-Robot Interacion*, vol. 1, 2, 33-53.

Cosi, P., Nicolao, M., Paci, G., Sommavilla, G. & Tesser, F. (2014). Comparing Open Source ASR Toolkits on Italian Children Speech. In online proceedings of *4th Workshop on Child Computer Interaction (WOCCI 2014)*, Satellite Event of Interspeech 2014, Singapore, 19 September 2014.

Cosi, P., Paci, G., Sommavilla, G. & Tesser, F. (2015). Building Resources for Verbal Interaction – Production and Comprehension within the ALIZ-E Project. In *Atti AISV 2015, XI Convegno Nazionale dell'Associazione Italiana di Scienze della Voce. "Il farsi e il disfarsi del linguaggio. L'emergere, il mutamento e la patologia della struttura sonora del linguaggio"*, Alma Mater Studiorum, Università di Bologna, 28-30 gennaio 2015.

Cosi, P. (2015). A KALDI-DNN-Based ASR System for Italian Experiments on Children Speech. In CD-Rom *Proceedings of IJCNN 2015*, Killarney, Ireland, 12-17 July 2015, CD-paper 15079.

Gerosa, M., Giuliani, D. & Brugnara, F. (2007). Acoustic variability and automatic recognition of children's speech. In *Speech Communication*, 49, 847-860.

KALDI ASR. http://kaldi-asr.org.

Paci, G., Sommavilla, G., Tesser, F. & Cosi, P. (2013). Julius ASR for Italian children speech. In *Proceedings of the 9th national congress, AISV (Associazione Italiana di Scienze della Voce)*, Venice, Italy.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G. & Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. In *Proceedings of IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, IEEE Signal Processing Society, Hilton Waikoloa Village, Big Island, Hawaii, US, December 2011.

[2] For further info mail-to: piero.cosi@pd.istc.cnr.it.

Tesser, F., Paci, G., Sommavilla, G. & Cosi, P. (2013). A new language and a new voice for MARY-TTS. In *Proceedings of the 9th national congress, AISV (Associazione Italiana di Scienze della Voce)*, Venice, Italy, 2013.

DOI: 10.17469/O2102AISV000017

MIRCO RAVANELLI, LUCA CRISTOFORETTI, ROBERTO GRETTER,
MARCO PELLIN, ALESSANDRO SOSI, MAURIZIO OMOLOGO

Il corpus DIRHA-ENGLISH ed i relativi task per il riconoscimento vocale a distanza in ambienti domestici

This paper addresses the contents and the possible usage of the DIRHA-ENGLISH multi-microphone corpus, realized under the EC DIRHA project. The reference scenario is a domestic environment equipped with a large number of microphones distributed in space.
The corpus is composed of both real and simulated material, and it includes 12 US and 12 UK English native speakers' utterances. Each speaker uttered different sets of phonetically-rich sentences, newspaper articles, conversational speech, keywords, and commands. From this material, a large set of 1-minute sequences was generated, which also includes typical domestic background noise and inter/intra-room reverberation effects. Development and test sets were derived.
The paper reports a first set of baseline results obtained using different techniques, including Deep Neural Networks (DNN), aligned with the state-of-the-art at international level. Various tasks and Kaldi recipes have already been developed.

Key words: distant speech recognition, microphone arrays, corpora, Kaldi, DNN.

Introduzione

Il riconoscimento vocale è stato oggetto di molta attenzione negli ultimi anni (Yu, Deng, 2015). Come risultato, ha trovato applicazione in vari campi, come i sistemi di dettatura, il controllo in automobile, la ricerca su web. Nonostante gli sforzi, molte soluzioni sono ancora basate su un'interazione closetalk, usando un microfono vicino al parlatore, e costringendo quindi l'utente ad avvicinarsi al dispositivo. Questa limitazione non è però accettabile nel caso in cui l'utente voglia un maggiore grado di libertà nell'utilizzo di queste tecnologie. È dunque facilmente prevedibile che il riconoscimento vocale a distanza (Wölfel, McDonough, 2009) rivestirà un ruolo di primaria importanza nello sviluppo delle future interfacce uomo-macchina. Un esempio applicativo particolarmente rilevante è l'ambiente domestico, che è stato oggetto del progetto europeo DIRHA. Lo scopo era quello di sviluppare dei servizi automatizzati in ambiente domestico, controllati da un sistema con riconoscimento vocale a distanza funzionante in più lingue.

Nonostante i progressi in questo campo, le tecnologie attuali mostrano ancora una mancanza di robustezza e flessibilità, dovuta alla presenza di rumori ambientali non stazionari e alla presenza di riverbero (Hänsler, Smith, 2008). Per colmare la differenza di prestazioni rispetto ad un sistema close-talk sono necessarie

notevoli quantità di dati, registrati in condizioni adatte e trascritti appositamente. Date le innumerevoli variabili in gioco nell'ambiente domestico, si tratta di un'attività altamente complicata ed impegnativa; diventa quindi indispensabile disporre di corpora multimicrofonici realistici e di alta qualità, finalizzati all'addestramento del riconoscitore. Nonostante la disponibilità di alcuni corpora, la necessità di materiale specifico ci ha portato alla creazione di vari corpora multimicrofonici registrati in ambiente domestico.

Il corpus multi-microfonico DIRHA-ENGLISH è stato realizzato in lingua inglese assieme ad altri corpora (raccolti in quattro lingue: italiano, greco, tedesco e portoghese) nell'ambito del progetto DIRHA (Cristoforetti, Ravanelli, Omologo, Sosi, Abad, Hagmüller & Maragos, 2014). Lo scenario di riferimento è un appartamento equipaggiato con un elevato numero di microfoni, distribuiti nelle varie stanze. La scelta della lingua inglese è dettata dal fatto che questa rappresenta la lingua di riferimento nella comunità internazionale. Il corpus è composto sia da materiale simulato sia da materiale reale registrato nell'appartamento, per permettere di testare le prestazioni del riconoscimento vocale in condizioni reali.

Lo scopo di questo articolo è di descrivere il contenuto del corpus DIRHA-ENGLISH e di fornire alcuni risultati preliminari su frasi foneticamente ricche, ottenuti utilizzando il framework Kaldi (Povey, Ghoshal, Boulianne, Burget, Glembek, Goel, Hannemann, Motlicek, Quian, Schwarz, Silovsky, Stemmer & Vesely, 2011). Il risultante task di tipo TIMIT può essere visto come complementare a task tipo riconoscimento WSJ o di parlato conversazionale.

L'articolo è suddiviso nel seguente modo. La Sezione 1 descrive il progetto europeo DIRHA mentre la Sezione 2 si focalizza sul contenuto e le caratteristiche del corpus DIRHA-ENGLISH. La Sezione 3 riporta una descrizione dei task definiti ed i risultati preliminari corrispondenti. La Sezione 4 fornisce alcune conclusioni.

1. *Il progetto europeo DIRHA*

Il progetto europeo DIRHA, iniziato nel gennaio 2012 e durato tre anni, aveva come obiettivi l'analisi della scena acustica e l'interazione vocale a distanza in un ambiente domestico. Seguono ora una descrizione degli obiettivi, dei task ed dei corpora raccolti.

1.1 Obiettivi e task

Lo scenario applicativo affrontato nell'ambito del progetto è caratterizzato da un sistema vocale interattivo che permette l'interazione da qualsiasi stanza e senza vincoli di posizione. Sfruttando una rete di microfoni distribuiti in varie stanze, il sistema DIRHA reagisce prontamente ai comandi impartiti da un utente. Il sistema è sempre in ascolto, aspettando una specifica parola d'ordine per iniziare un nuovo dialogo con l'utente. Il dialogo permette poi all'utente di accedere a

dispositivi e servizi, tipo l'apertura di porte e finestre, accendere o spegnere luci, controllare la temperatura o ascoltare della musica. Il sistema è inoltre caratterizzato dalla possibilità di gestire più dialoghi in parallelo in stanze diverse e dalla possibilità di fare barge-in (cioè poter interagire anche in presenza di musica o prompt acustici emessi da parte del sistema). Un'altra caratteristica molto importante è la capacità di limitare i falsi allarmi, dovuti alla non corretta interpretazione di suoni ambientali o normali dialoghi tra utenti.

Partendo da queste funzionalità, vari task sperimentali sono stati definiti in combinazione con algoritmi di front-end processing e riconoscimento vocale nelle varie lingue. La maggior parte di questi task si riferiscono ad acquisizioni vocali effettuate nell'appartamento ITEA di Trento.

1.2 I corpora DIRHA

I corpora vocali DIRHA sono stati progettati per mettere a disposizione raccolte multi-microfoniche atte ad essere utilizzate in un ampio numero di task, come menzionato sopra. Alcune raccolte sono basate su simulazioni ottenute tramite contaminazione (Matassoni, Omologo, Giuliani & Svaizer, 2002; Couvreur, Couvreur & Ris, 2000; Haderlein, Nöth, Herbordt, Kellermann & Niemann, 2005), combinando registrazioni close-talk con risposte impulsive stimate e sequenze reali di rumore di fondo (Cristoforetti et al., 2014). Altre raccolte sono invece state registrate in condizioni reali.

A parte il corpus DIRHA-ENGLISH che verrà descritto nella prossima sottosezione, gli altri corpora raccolti sono i seguenti:

- Il corpus DIRHA Sim (30 parlatori per quattro lingue) (Cristoforetti et al., 2014), che consiste in sequenze multi-canale della durata di un minuto, comprendenti vari eventi acustici e frasi;
- Una raccolta basata sul Mago di OZ (WOZ) (Brutti, Ravanelli, Svaizer & Omologo, 2014) per valutare le componenti di speech-activity-detection e di localizzazione del parlatore;
- Il corpus DIRHA AEC (Zwyssig, Ravanelli, Svaizer & Omologo, 2015) che include dati specificatamente raccolti per studiare la cancellazione dell'eco, per rimuovere interferenze acustiche note, diffuse nell'ambiente;
- Il corpus DIRHA-GRID (Matassoni, Astudillo, Katsamanis & Ravanelli, 2014) che include una raccolta multi-canale e multi-stanza di dati simulati, derivanti dalla contaminazione del corpus GRID (Cooke, Barker, Cunninghan & Shao, 2006), composto da brevi comandi in lingua inglese.

2. *Il corpus DIRHA-ENGLISH*

Come per gli altri corpora, anche il corpus DIRHA-ENGLISH è composto da una parte di dati reali ed una parte di dati simulati, questi ultimi ottenuti tramite contaminazione di parlato clean che viene descritto in seguito.

2.1 Il parlato clean

Il materiale clean è stato acquisito in una sala di registrazione in FBK, tramite un microfono di alta qualità (Neumann TLM 103) a 96kHz 24 bit. Sono stati registrati 12 parlatori nativi inglesi e 12 parlatori nativi americani, suddivisi in egual numero tra maschi e femmine. Ognuno ha letto il seguente materiale:

- 15 comandi domestici letti;
- 15 comandi domestici spontanei;
- 13 parole d'ordine (keyword);
- 48 frasi foneticamente ricche (dal corpus Harvard);
- 66/67 frasi dal WSJ-5k;
- 66/67 frasi dal WSJ-20k;
- Circa 10 minuti di parlato conversazionale (ad esempio, il parlatore doveva descrivere un film).

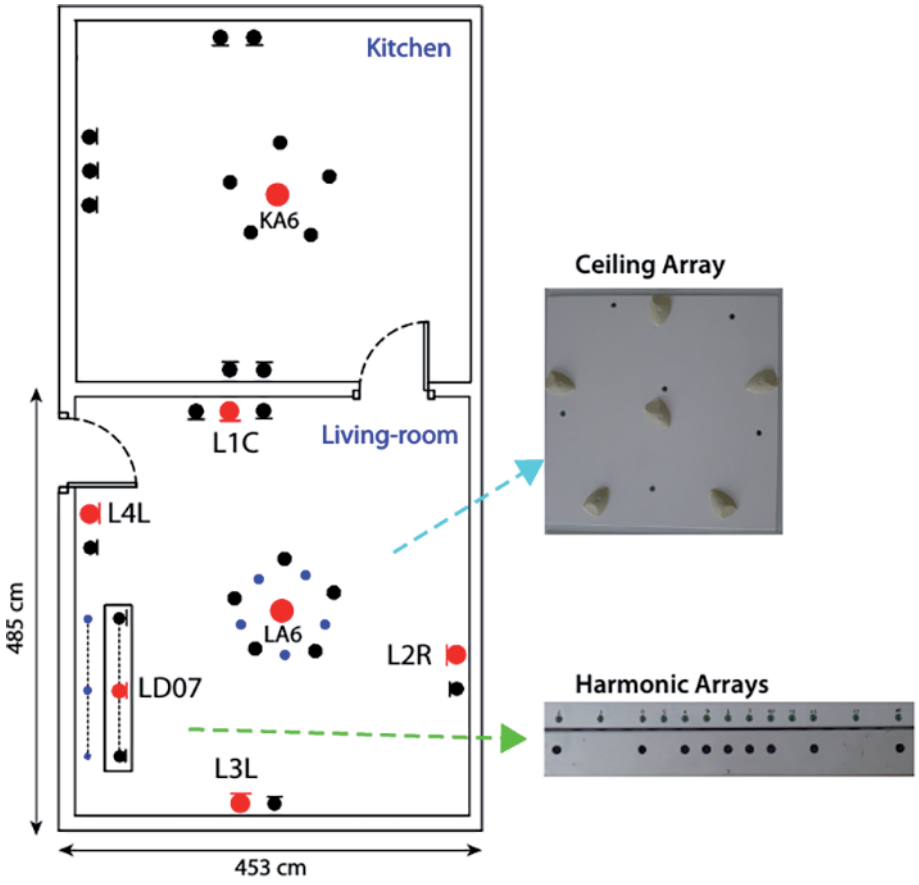
In totale sono state registrate circa 11 ore di materiale vocale e tutte le frasi sono state annotate manualmente. Le frasi foneticamente ricche sono state segmentate a livello fonetico da una procedura automatica (Brugnara, Falavigna & Omologo, 1993); un esperto ha poi controllato le trascrizioni e l'allineamento temporale.

Sei parlatori nativi inglesi e sei parlatori nativi americani sono stati assegnati al development set, mentre gli altri sono stati assegnati al test set. Le assegnazioni sono state effettuate in modo da distribuire le frasi WSJ come nel task originale (Paul, Baker, 1992). Entrambi i set sono compatibili con le specifiche TIMIT.

2.2 La rete microfonica

L'appartamento ITEA è l'appartamento di riferimento che è stato reso disponibile durante il progetto DIRHA per la raccolta di dati e lo sviluppo di prototipi. È composto da cinque stanze che sono state equipaggiate con una rete di diversi microfoni. I microfoni sono in maggior parte degli SHURE MX391 con un pattern omnidirezionale, collegati a delle schede di acquisizione RME Octamic II, campionati a 48kHz - 16 bit in maniera sincrona. Il bagno e altre due stanze sono equipaggiati con un numero limitato di microfoni organizzati in coppie o terne (in totale 12 microfoni), mentre cucina e soggiorno comprendono un numero più elevato di microfoni. Come si può vedere in Figura 1, il soggiorno include tre coppie di microfoni, una terna, due array a soffitto da sei microfoni ognuno (di cui uno composto da microfoni digitali MEMS) e due array armonici (composti rispettivamente da 15 microfoni electret e 15 microfoni digitali MEMS).

Figura 1 - *Schema che rappresenta la distribuzione dei microfoni nel corpus DIRHA-ENGLISH. I punti blu rappresentano i microfoni digitali MEMS, i punti rossi indicano i microfoni utilizzati negli esperimenti mentre i punti neri rappresentano tutti i microfoni disponibili. Le immagini di destra mostrano l'array di microfoni sul soffitto e gli array armonici del soggiorno*



Un'attività particolarmente dispendiosa è stata dedicata a caratterizzare l'ambiente a livello acustico, attraverso più raccolte dati per stimare le risposte impulsive. In totale sono state calcolate più di 10000 risposte impulsive che descrivono come si propaga il suono da vari punti nello spazio ad ognuno dei microfoni presenti. Il metodo adottato per calcolare le risposte impulsive si basa sulla diffusione di uno sweep di frequenze esponenziale (Exponential Sine Sweep, ESS) (Farina, 2000). La rete di microfoni considerata nel DIRHA-ENGLISH (rappresentata in Figura 1) comprende solo soggiorno e cucina, ma include anche gli array armonici e gli array di microfoni MEMS che non sono disponibili negli altri corpora raccolti.

2.3 I data-set simulati

I data-set simulati derivano dal parlato clean descritto nella Sezione 2.1 e dai metodi di contaminazione descritti in (Matassoni et al., 2002; Ravanelli, Sosi, Svaizer & Omologo, 2012). Il corpus risultante consiste in un grande numero di sequenze lunghe un minuto, ognuna comprendente un numero variabile di frasi pronunciate nel soggiorno con differenti livello di rumore di fondo. Sono stati creati quattro tipi di sequenze, corrispondenti ai seguenti task:

- Frasi foneticamente ricche;
- Frasi dal WSJ-5k;
- Frasi dal WSJ-20k;
- Parlato conversazionale (comprendente anche parole d'ordine e comandi).

Sono disponibili le registrazioni di 62 microfoni per ogni sequenza, come descritto nella Sezione 2.2.

2.4 Il data-set reale

Per quello che riguarda le registrazioni di materiale dal vivo, ogni utente ha letto il materiale da un tablet, stando in piedi o seduto in soggiorno. Dopo ogni set di frasi è stato chiesto al parlatore di spostarsi in una nuova posizione con un differente orientamento. Ogni utente ha letto lo stesso materiale che aveva letto in sala di registrazione, descritto nella Sezione 2.1. Le registrazioni dei microfoni MEMS sono state allineate temporalmente con gli altri microfoni in una seconda fase, non essendo stato possibile utilizzare lo stesso clock per sincronizzare le acquisizioni.

Una volta raccolto il materiale sono state derivate sequenze da un minuto in modo da rimanere coerenti con i dati simulati.

3. *Esperimenti e risultati*

Questa sezione descrive i task sperimentali proposti ed i relativi risultati preliminari ottenuti utilizzando la parte US delle frasi foneticamente ricche del corpus DIRHA-ENGLISH.

3.1 Descrizione del contesto sperimentale

3.1.1 Corpora per test e training

In questo lavoro la fase di training è ottenuta impiegando la porzione di training del corpus TIMIT (Garofolo, Lamel, Fisher, Fiscus, Pallett & Dahlgren, 1993). Per gli esperimenti di riconoscimento a distanza il corpus originale TIMIT è stato inoltre riverberato utilizzando tre risposte impulsive misurate nel soggiorno. Inoltre sono state aggiunte alcune sequenze di rumore multi-canale, per simulare condizioni reali. Sia le risposte impulsive che le sequenze di rumore sono differenti da quelle utilizzate per generare il corpus DIRHA-ENGLISH.

La fase di test invece è stata effettuata utilizzando le frasi foneticamente ricche del corpus DIRHA-ENGLISH, sia simulate che reali. In entrambi i casi alle sequenze di un minuto è stato applicato un VAD (Voice Activity Detector) e le sequenze sono state poi sotto-campionate da 48 kHz a 16 kHz.

3.1.2 Estrazione delle feature

Alle frasi è stata effettuata un'estrazione delle feature basata su MFCCs. In particolare, il segnale è stato suddiviso in frame da 25 ms con un overlap di 10 ms; per ogni frame sono state estratte 13 feature MFCCs. Le feature sono state poi raggruppate in un vettore di 39 componenti, assieme alle loro derivate prime e seconde.

3.1.3 Training dei modelli acustici

Negli esperimenti descritti sono stati considerati tre modelli acustici differenti, con una complessità sempre maggiore. La procedura adattata per addestrare i modelli è la stessa utilizzata per la recipe Kaldi di TIMIT s5 (Povey et al., 2011). La prima baseline (mono) si riferisce ad un semplice sistema caratterizzato da 48 fonemi della lingua inglese, indipendenti dal contesto, ognuno modellato con una rete HMM (Hidden Markov Model) a tre stati (in totale usando 1000 gaussiane). La seconda baseline (tri) è basata su una modellizzazione dei fonemi che dipende dal contesto e da un addestramento che si adatta al parlatore. In totale sono utilizzati 2500 stati con 15000 gaussiane.

Per ultima, la terza baseline basata su reti neurali (Deep Neural Networks, DNN) è addestrata con la recipe Karel (Ghoshal, Povey, 2013), composta da sei stati nascosti e 1024 neuroni, su una finestra di 11 frame e un learning rate di 0,008.

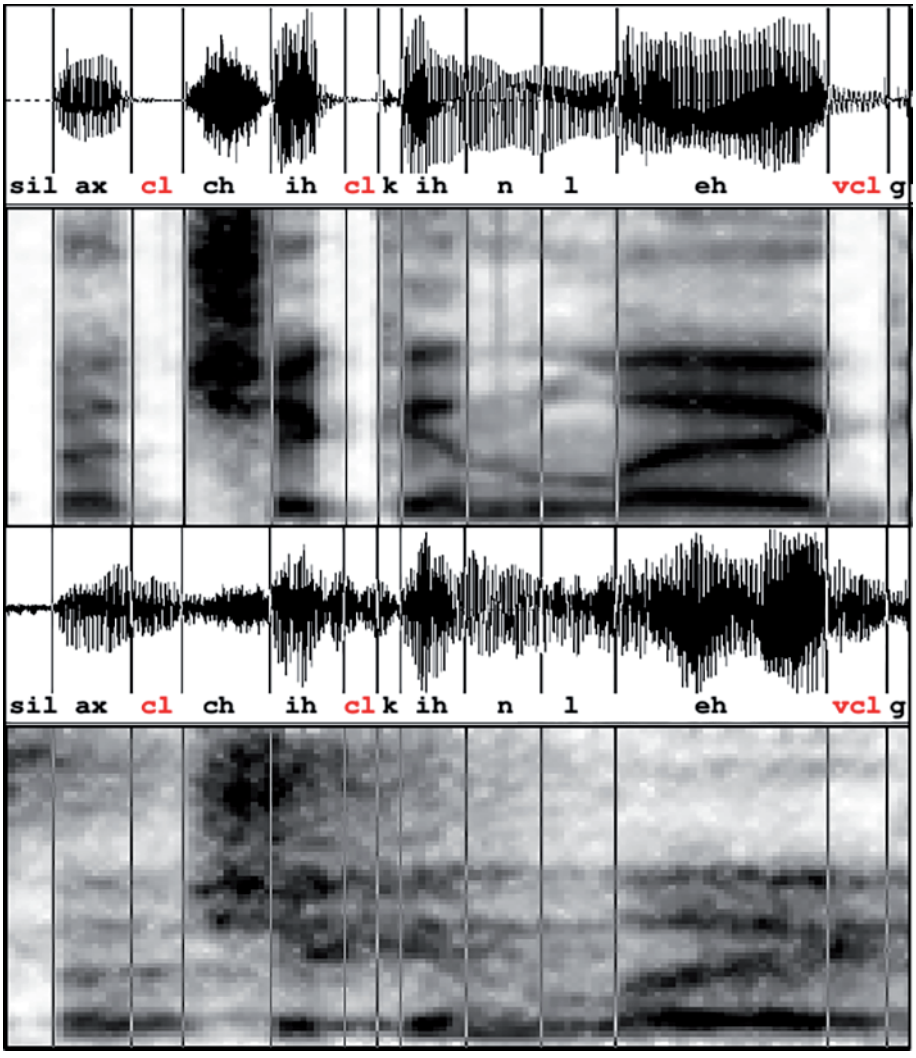
3.1.4 Task proposto e valutazione

La recipe Kaldi si basa sull'impiego di un modello del linguaggio a bigrammi, stimato utilizzando le trascrizioni fonetiche disponibili con il training set. Al contrario, qui proponiamo l'adozione di un puro task phone-loop (zero-grammi) in modo da evitare qualsiasi possibile alterazione dovuta a non-linearità o ad artefatti causati dal modello del linguaggio. I nostri lavori precedenti (Ravanelli et al., 2012; Ravanelli, Omologo, 2014; Ravanelli, Omologo, 2015), infatti, suggeriscono che, sebbene l'impiego di modelli del linguaggio sia certamente utile al fine di aumentare le prestazioni di riconoscimento, l'utilizzo di un semplice task phone-loop è più adatto nel caso di esperimenti che focalizzano l'attenzione sull'informazione acustica.

Un'altra differenza rispetto alla recipe originale Kaldi riguarda la valutazione di silenzi e closures. In fase di valutazione, la recipe standard Kaldi (basata su sclite) mappa le 48 unità fonetiche inglesi in un set ridotto a 39 unità, come originariamente fatto in (Lee, Hon, 1989). In particolare, sei closures (*bcl,dcl,gcl,kcl,pcl,tcl*) vengono mappate come "silenzio opzionale", e possibili cancellazioni di queste unità non vengono considerate come errori in fase di valutazione. Errori relativi a tali unità sono molto frequenti, ed il loro contributo al calcolo complessivo del tasso di errore può introdurre un "bias". Questo aspetto risulta rilevante soprattutto nel caso del riconoscimento vocale a distanza dai microfoni, in cui le code del riverbero ren-

dono l'individuazione di tali unità praticamente impossibile, come evidenziato in Figura 2. Per questa ragione, in questo lavoro proponiamo semplicemente di eliminare silenzi e closures sia dalla sequenza di riferimento che dalla sequenza fonetica prodotta dal sistema di riconoscimento. Questa scelta porta ad un peggioramento nelle prestazioni del sistema, in quanto tutti i silenzi opzionali inclusi aggiunti nel caso della recipe originale non vengono più considerati. Allo stesso tempo, si ottiene una stima più coerente delle prestazioni del sistema, per quel che riguarda unità fonetiche di maggiore importanza, quali ad esempio vocali e burst di occlusive.

Figura 2 - *La frase "a chicken leg" registrata tramite close-talk in studio (in alto) e con microfono distante nell'ambiente reale (in basso). Le closures (in rosso) sono coperte dalla coda di riverbero nella registrazione ambientale*



3.2 Risultati sperimentali

Questa sezione fornisce alcuni risultati *baseline*, che possono risultare utili come riferimento per altri ricercatori che intendessero utilizzare questo corpus. Nelle prossime sezioni vengono presentati risultati ottenuti sia nel caso di input close-talk che di input da microfono posto a distanza dal parlatore.

3.2.1 Performance nel caso di input close-talk

Come riportato in Tabella 1, le prestazioni ottenute decodificando sequenze vocali clean (ovvero acquisite nello studio di registrazione di FBK) attraverso l'impiego di un modello del linguaggio a bigrammi di fonemi o di un semplice loop di unità fonetiche (phone loop).

I risultati sono stati ottenuti attraverso la recipe standard Kaldi s5 e attraverso l'alternativa recipe basata sulla nostra proposta di valutazione degli errori, in modo da poter evidenziare le discrepanze nei risultati fra le due diverse condizioni sperimentali.

Tabella 1 - *Phone Error Rate (PER%) ottenuta applicando differenti recipes Kaldi alle frasi fonetiche acquisite nella sala registrazioni di FBK*

Recipe	LM type	Mono	Tri	DNN
Standard Kaldi s5	Bigram LM	36.4	23.2	20.1
Standard Kaldi s5	Phone-loop	39.4	26.3	22.4
Proposed Evaluation	Bigram LM	42.7	28.6	24.6
Proposed Evaluation	Phone-loop	46.7	32.5	27.5

Come prevedibile, i risultati evidenziano che le prestazioni cambiano significativamente quando si passa dal semplice caso di GMM che modellano unità indipendenti (monofoni) dal contesto al caso di DNN. Inoltre, come evidenziato in Sezione 3.1.4, applicando la recipe Kaldi originale si osserva una riduzione relativa prossima al 20% del tasso di errore rispetto alla procedura da noi proposta, che, di fatto, non corrisponde ad alcun miglioramento nelle capacità del sistema, ma esclusivamente al metodo di valutazione.

Gli esperimenti di riconoscimento a distanza dal microfono che vengono descritti nelle prossime sezioni si riferiscono al solo caso di phone loop e di valutazione basata sulla procedura da noi proposta.

3.2.2 Prestazioni del sistema nel caso di singolo microfono distante

In questa sezione vengono riportati e discussi i risultati che sono stati ottenuti nel caso in cui l'input del riconoscitore corrisponde ad uno fra quattro possibili microfoni (LA6, L1C, LD07, KA6) posti a distanza dal parlatore. La Tabella 2 riporta la lista completa di questi risultati.

Tabella 2 - *Phone Error Rate (PER%) ottenuta con un singolo microfono a distanza dal parlatore. "Sim" si riferisce ai dati simulati, mentre "Real" si riferisce ai dati reali*

	Sim Mono	Sim Tri	Sim DNN	Real Mono	Real Tri	Real DNN
LA6	67.0	57.7	51.6	70.5	60.9	55.1
L1C	67.4	58.5	52.4	70.3	61.7	55.6
LD07	67.5	58.1	53.2	71.5	62.6	57.3
KA6	76.7	67.3	64.0	80.5	73.6	70.3

Come evidenziato in tabella, nel caso di microfono a distanza dal parlatore le prestazioni risultano nettamente peggiori rispetto al caso di input close-talk. Come già osservato nel caso di input close-talk, si osserva inoltre che l'impiego della DNN migliora le prestazioni in modo significativo rispetto a quanto è possibile ottenere con gli altri modelli acustici di riferimento. Questa evidenza sperimentale è confermata con tutti e quattro gli input microfonici che sono stati considerati, sia in caso di segnali simulati che di segnali reali. In realtà, le prestazioni nel caso di segnali reali sono leggermente peggiori rispetto al caso di dati simulati, a causa di un inferiore rapporto segnale rumore che caratterizza le registrazioni in ambiente reale.

È altrettanto importante rilevare che il trend generale delle prestazioni risulta sufficientemente coerente al variare dei modelli acustici esaminati. Solo il caso del microfono installato in cucina (KA6) è caratterizzato da un netto peggioramento delle prestazioni, causato da una generale riduzione del rapporto segnale rumore, dovuta al fatto che tutte le frasi sono state lette in salotto.

3.2.3 Prestazioni nel caso di delay-and-sum beamforming

In questa sezione viene esaminato l'andamento delle prestazioni nel caso di impiego della tecnica di delay-and-sum beamforming (Brandstein, Ward, 2000) per la combinazione di segnali acquisiti rispettivamente dall'array di microfoni posto nel soffitto e dall'array armonico, entrambi installati nel salotto.

Tabella 3 - *Phone Error Rate (PER%) ottenuta applicando il delay-and-sum beamforming agli array di microfoni presenti nel soggiorno dell'appartamento ITEA*

	Sim Mono	Sim Tri	Sim DNN	Real Mono	Real Tri	Real DNN
Array a soffitto	66.2	55.9	50.4	65.9	55.9	50.6
Array armonico	66.2	56.0	51.8	66.2	56.2	51.5

I risultati, riportati in Tabella 3, dimostrano che il beamforming risulta utile nel migliorare le prestazioni del sistema, per es. dal 55.1% PER osservato nel caso di singolo microfono al 50.6% PER che si ottiene applicando questa tecnica combinata con DNN, ai segnali acquisiti attraverso l'array installato nel soffitto.

Sebbene quest'ultimo array consista di soli sei microfoni, le prestazioni che esso offre risultano migliori rispetto al caso di array armonico, comprendente 13 microfoni. Questo risultato sperimentale potrebbe essere dovuto al posizionamento del primo dei due array, il quale spesso acquisisce un maggiore contributo in termini di propagazione diretta del suono rispetto all'array armonico. Una seconda motivazione per questo miglioramento di prestazioni è legata alla migliore qualità dei microfoni.

Si osserva inoltre che il miglioramento di prestazioni introdotto dalla tecnica di delay-and-sum beamforming risulta più evidente nel caso di segnali reali. Ciò conferma il fatto che il filtraggio spaziale risulta particolarmente utile nei casi in cui le condizioni acustiche sono meno stazionarie e quindi meno predicibili.

3.2.4 Prestazioni basate su selezione automatica del microfono

Il corpus DIRHA-ENGLISH può essere utilizzato anche in esperimenti basati sull'impiego di tecniche di selezione automatica del microfono. A questo proposito, risulta quindi interessante esaminare alcune prestazioni di riferimento che possono costituire un *upper-bound* per tali tecniche. La Tabella 4 fornisce un confronto tra i risultati ottenuti nel caso di selezione casuale del microfono e quelli ottenuti nel caso (Oracle) in cui per ciascuna frase viene selezionato come input il microfono che assicura il minimo tasso di errore. L'esperimento è stato condotto utilizzando i sei microfoni del salotto indicati in rosso in Figura 1.

Tabella 4 - *Phone Error Rate (PER%) ottenuta applicando una selezione casuale del microfono (random) oppure attraverso una selezione operata con modalità oracolo (Oracle)*

	Sim Mono	Sim Tri	Sim DNN	Real Mono	Real Tri	Real DNN
Random	67.6	57.7	52.4	70.3	61.0	55.4
Oracle	56.6	47.1	42.0	60.3	49.6	44.0

I risultati dimostrano che una opportuna selezione automatica dinamica del microfono può risultare determinante nel miglioramento delle prestazioni di un sistema DSR. Si può osservare una significativa differenza fra l'upper bound indicato nella riga *Oracle* e il lower bound basato su una selezione random del microfono. Ciò conferma l'importanza da attribuire alla tematica di ricerca riguardante la selezione automatica del microfono, la quale è potenzialmente in grado di fornire risultati migliori rispetto all'impiego di delay-and-sum beamforming. Per esempio, un 50.6% PER ottenuto applicando il beamforming all'array del soffitto va confrontato con il corrispondente 44.0% ottenibile nel caso di selezione automatica del microfono ideale.

4. *Conclusioni e lavori futuri*

Questo articolo descrive il corpus multi-microfonico DIRHA-ENGLISH ed alcuni esperimenti preliminari relativi all'utilizzo delle frasi foneticamente ricche. In generale i risultati sperimentali mostrano le prestazioni che ci si aspettava, allineate con altri lavori in questo campo.

Nella ricerca sul riconoscimento del parlato a distanza ci sono vari vantaggi nell'utilizzo di materiale foneticamente ricco con un così vasto numero di microfoni.

Il corpus contiene anche parti del WSJ e parlato conversazionale che potrebbero essere oggetto di una distribuzione pubblica ed oggetto di future competizioni relative al riconoscimento vocale a distanza. Il parlato conversazionale, in particolare, potrebbe essere utile per investigare altri aspetti chiave, come ad esempio la combinazione di ipotesi multi-microfoniche basate su reti di confusione, reticoli multipli e rescoring.

I futuri lavori prevedono lo sviluppo di baseline e relative recipe per l'utilizzo dei microfoni MEMS digitali, per frasi del WSJ e conversazionali, e per l'inglese britannico.

5. *Rilascio del corpus*

Alcune sequenze di un minuto di durata l'una sono disponibili a questo indirizzo: http://dirha.fbk.eu/DIRHA_English. L'accesso ai dati utilizzati in questo articolo ed i relativi documenti saranno possibili tramite i server FBK, con le modalità che saranno riportate nel sito http://dirha.fbk.eu. In futuro altri dati saranno resi disponibili, corredati da documentazione e recipe, ed istruzioni per poter effettuare dei paragoni tra sistemi differenti.

Ringraziamenti

Il lavoro presentato è stato parzialmente finanziato dalla Comunità Europea nell'ambito del Settimo Programma Quadro (FP7/2007-2013), con il contratto 288121-DIRHA.

Riferimenti bibliografici

Brandstein, M., Ward, D. (2000). *Microphone arrays*. Berlin: Springer.

Brugnara, F., Falavigna, D. & Omologo, M. (1993). Automatic segmentation and labeling of speech based on hidden markov models. In *Speech Communication*, 12, 4, 357-370.

Brutti, A., Ravanelli, M., Svaizer, P. & Omologo, M. (2014). A speech event detection/localization task for multi-room environments. In *Proc. of HSCMA*, 157-161.

Couvreur, L., Couvreur, C. & Ris, C. (2000). A corpus-based approach for robust ASR in reverberant environments. In *Proc. of INTERSPEECH*, 397-400.

Cooke, M., Barker, J., Cunningham, S. & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. In *Journal of the Acoustical Society of America*, 120, 5, 2421-2424.

Cristoforetti, L., Ravanelli, M., Omologo, M., Sosi, A., Abad, A., Hagmüller, M. & Maragos, P. (2014). *The DIRHA simulated corpus.* In *Proc. of LREC*, 2629-2634.

Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Proc. of the 108th AES Convention*, 18-22.

Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S. & Dahlgren, N.L. (1993). *DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM*.

Ghoshal, A., Povey, D. (2013). Sequence discriminative training of deep neural networks. In *Proc. of INTERSPEECH*.

Haderlein, T., Nöth, E., Herbordt, W., Kellermann, W. & Niemann, H. (2005). Using Artificially Reverberated Training Data in Distant-Talking ASR. In *Lecture Notes in Computer Science*, 3658, 226-233. Springer.

Hänsler, E., Schmidt, G. (2008). *Speech and Audio Processing in Adverse Environments*. Springer.

Lee, K.F., Hon, H.W. (1989). Speaker-independent phone recognition using hidden markov models. In *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37, 11, 1641-1648.

Matassoni, M., Omologo, M., Giuliani, D. & Svaizer, P. (2002). Hidden Markov model training with contaminated speech material for distant-talking speech recognition. In *Computer Speech & Language*, 16, 2, 205-223.

Matassoni, M., Astudillo, R., Katsamanis, A. & Ravanelli, M. (2014). The DIRHA-GRID corpus: baseline and tools for multi-room distant speech recognition using distributed microphones. In *Proc. of INTERSPEECH*, 1616-1617.

Paul, D.B., Baker, J.M. (1992). The design for the wall street journal-based csr corpus. In *Proc. of the Workshop on Speech and Natural Language*, 357-362.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G. & Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. In *Proc. of ASRU*.

Ravanelli, M., Omologo, M. (2014). On the selection of the impulse responses for distant-speech recognition based on contaminated speech training. In *Proc. of INTERSPEECH*, 1028-1032.

Ravanelli, M., Omologo, M. (2015). Contaminated speech training methods for robust DNN-HMM distant speech recognition. In *Proc. of INTERSPEECH*.

Ravanelli, M., Sosi, A., Svaizer, P. & Omologo, M. (2012). Impulse response estimation for robust speech recognition in a reverberant environment. In *Proc. of EUSIPCO*, 1668-1672.

Yu, D., Deng, L. (2015). *Automatic Speech Recognition - A Deep Learning Approach*. Springer.

Wölfel, M., McDonough, J. (2009). *Distant Speech Recognition*. Wiley.

Zwyssig, E., Ravanelli, M., Svaizer, P. & Omologo, M. (2015). A multi-channel corpus for distant-speech interaction in presence of known interferences. In *Proc. of ICASSP*, 4480-4485.

DOI: 10.17469/O2102AISV000018

FABIO TESSER, GIACOMO SOMMAVILLA, GIULIO PACI, PIERO COSI

Automatic creation of tts intelligibility tests

This work presents a new method to automatize the creation and the analysis of subjective intelligibility tests for Text To Speech systems. In order to reach this goal, two main ideas have been adopted: the employment of multiple-choice answers and the use of algorithms able to automatically create the appropriate test set for intelligibility tasks. The reasons that led to the design of this new kind of intelligibility tests and the choices regarding the design of the algorithms are described. A practical example of the application of this new methodology has been experimented in the context of an online subjective test for intelligibility evaluation of two TTS voices.

Key words: TTS, Intelligibility.

Introduction

In speech communication, intelligibility is a measure of how comprehensible is speech under particular conditions. In the context of Text To Speech (TTS) Synthesis, the goal is to evaluate how much the speech signal generated from text is comprehensible. This measure can give an indication about the good-ness of the synthesizer, its voices and its modules. In addition, the results can help pointing out how to improve the quality of a TTS system.

The conditions under which intelligibility is evaluated are generally environ-mental acoustic conditions, and then they can be specified by the particular audio device used or the environment in which the test is carried out.

Looking to the literature, different speech intelligibility measures can be defined. It can be useful to measure the segmental intelligibility of a TTS system (Venkatagiri, 2003) in order to understand if the system is less intelligible in correspondence of particular phonemes; in other cases, it can be sufficient the use of the intelligibility at word level (Yu, Yue, Zu & Chen, 2010).

To simulate real life conditions, some methods add noise to the speech signal in order to measure intelligibility depending on the level of added noise (Venkatagiri, 2005). In this way, it is possible to test which TTS systems or technologies are more resistant to noise. Moreover, researchers started investigating on strategies to make the speech signal more intelligible if disturbed by a particular kind of noise (e.g. babble noise) for example using particular speech enhancement techniques (Zorila, Kandia & Stylianou, 2012) or by formants shifting (Godoy, Koutsogiannaki & Stylianou, 2013).

As a matter of fact it was noticed that the intelligibility of a TTS system immersed in noise is lower than the natural voice in the same conditions (King, Karaiskos, 2010).

The scientific and technological communities are aware of this phenomenon, so that, in the context of a speech-in-noise intelligibility challenge, a dedicated task was reserved only for TTS systems (Cooke, Mayo & Valentini-Botinhao, 2013) and researchers in the field of speech synthesis have proposed many ideas to increase the intelligibility of TTS systems (Nicolao, Tesser & Moore, 2013; Valentini-Botinhao, Yamagishi, King & Maia, 2014; Erro, Zorila & Stylianou, 2014).

All these examples suggest that intelligibility is a very important area for both speech technology and voice science.

Objective measures of intelligibility has been proposed, but some of them do not correlate well to subjective intelligibility scores. Moreover, most of them mainly measure the audibility of a signal without taking into account the actual phonetic content of the signal (Valentini-Botinhao, Yamagishi & King, 2011).

Although the use of an objective measure for intelligibility would enable to fully automatize the procedure, because of the limitations mentioned above, currently the most reliable method to evaluate intelligibility are subjective tests.

However, organizing and performing a subjective intelligibility test is a task that needs many resources. In fact traditional subjective intelligibility tests ask the subject to listen to a sentence and then to transcribe what she/he understood. Drawbacks of this method are: 1) it requires a lot of effort from the user; 2) the results are difficult to be analysed due to user typos and different ways of writing the same word.

The new method presented here is a proposal to eliminate the aforementioned issues. The paper is organized as follow: Sections 1 and 2 present the proposed method, Section 3 illustrates the web-oriented intelligibility tool de-signed to dispense the test, Section 4 describes the possible data analysis and Section 5 shows an experiment of intelligibility executed with this method; finally, Section 6 concludes the paper.

1. *Method*

We designed a novel type of subjective intelligibility test, based on multiple-choice answers. The benefits of this methodology with respect to the traditional ones are:

1. less effort is required for the subject;
2. the results are easy to be analysed (also in an automatic way);
3. all the process, from the data creation to the data analysis can be automatized.

The first two advantages are obtained thanks to the multiple-choice answers, while the last one, the automation, is explained in Section 2.

Multiple-choice answers intelligibility tests are not a novelty; Modified Rhyme Test (MRT) (Logan et al., 1989; Goldstein, 1995) was used in the past to measure

intelligibility of speech synthesis systems. However, that method was designed to present to the listener only isolated words and thus it is not suit-able for any kind of overall sentence evaluation, and it is focused only on consonants.

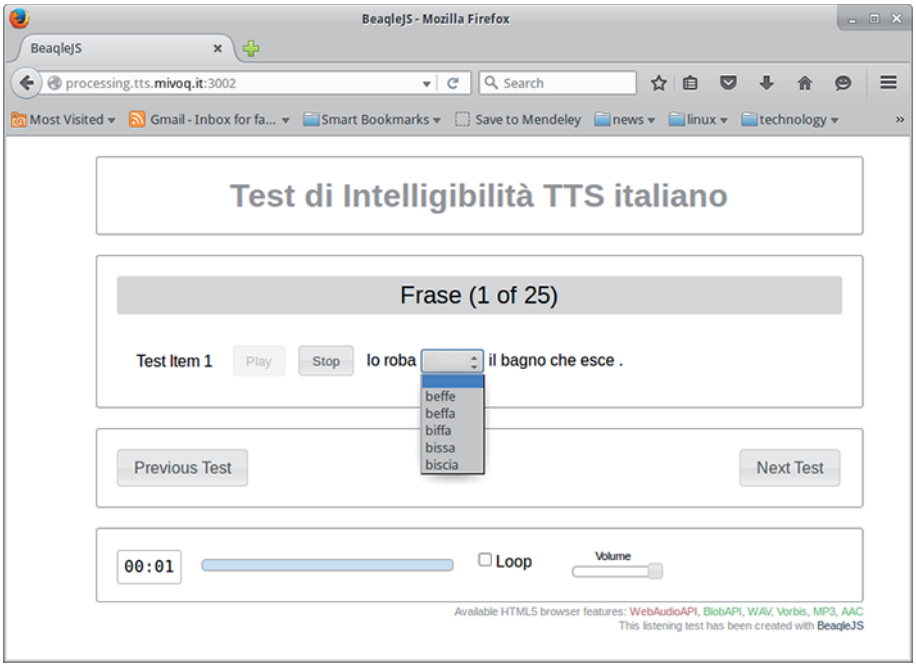
Our method does not suffer from these limitations because it considers all the phonemes and it makes use of full sentences instead of isolated words.

The procedure of the test is as follows:

a. the system presents a sentence to the subject and the user can play the sentence as many time as she/he likes;
b. when the subject feels ready to give an answer, the sentence is dis-played on the screen with the exception of one word called the miss-ing word;
c. the subject is asked to select the missing word she/he heard, by choosing from a list of acoustically-similar words.

To make a visual example, Figure 1 shows a screenshot of our intelligibility web tool taken when the subject is selecting the missing word. It is worth noting that the resulting sentence is semantically nonsense; the reason will be explained in Section 2.3.

Figure 1 - *Intelligibility web evaluation tool: selection of the missing word*



2. *Automatic creation of the test's data set*

Automatic creation of test's data set is a very useful feature if the experiment has to be replicated for different speakers, systems or even for different languages. When

designing the test, in fact, it is considered a constraint the fact of being able to test and prepare the data for other languages with little effort.

In this paper, the term automatic means that there is a procedure able to compute the data or to analyse the results for different languages and different subjects, with the only requirement of having the necessary data and linguistic modules for that language.

Going into detail, the automatic creation of the data set for this intelligibility test includes three main points

RF practice and training occurs immediately after the listening comprehension activities are completed; that is, after all pages of the science text have been narrated to the student and MC questions have completed.

The next three paragraphs explain the motivations and methods concerning these three points:

1. the automatic generation of a Phonetically Balanced Word List;
2. the automatic generation of Acoustically Similar Words;
3. the automatic generation of Semantically Unpredictable Sentences.

2.1 Automatic generation of the Phonetically Balanced Word List

In order to evaluate the intelligibility of the entire phonetic inventory of a language, of course it is necessary to prepare a test that contains sentences and words that globally contains the sounds of that language with the same frequency.

Because the test is based on guessing a missing word, our goal is to find a list of words, which satisfies the aforementioned property.

The first Phonetically Balanced Word List was developed during the Second World War (Egan, 1948) and reformulated later (Logan et al., 1989; Goldstein, 1995), precisely for the purpose of measuring intelligibility, and it was a 50 words list manually computed for the English language.

Today it is possible to automatize the computation of such a list, using algorithms able to do extensive search and computations on big textual corpora available for that language (Wikipedia). We developed this procedure using the same algorithms created to select the sentences that maximize the phonetic coverage for the purpose of building a balanced TTS corpus (Pammi et al., 2005), with the difference that in our case the sentences were composed only of a single word.

The procedure to derive the Phonetically Balanced Word List is the following:

- extract the M more frequent words of the considered language from Wikipedia;
- compute the words' phonetic transcriptions, using a Grapheme to Phoneme module;
- use the before-mentioned algorithm to select the set WN of the N words that maximize the phonetic balance.

While the choice of M and N is left to the test's designers and may vary according to the aim of the test, it is critical to choose N so that the set of words includes all the phonemes of the language.

The procedure can be easily replicated for different languages; in fact, the algorithm only needs a grapheme-to-phoneme module and a big textual corpus available for the language taken into consideration.

2.2 Automatic generation of acoustically similar words

During the process of understanding a verbal message, it can happen of not having understood a particular word. In this circumstance, it is very common to try to mentally find the missing word among the words more acoustically similar to the perceived sounds. For this reason, it is desirable that the words the subject has to choose among are similar from the acoustic point of view.

The task that we want to solve here is then: starting from a word wX (taken from the set of Phonetically Balanced Word List WN), to find the R words more acoustically similar to the word wX looking in a dictionary D.

To do this in an automatic way, we need to compute an acoustic distance between two words. Since at this stage of the procedure we do not know the acoustic realization of the words but only their phonetic representation, we can approximate the distance with acoustic phonetic distance.

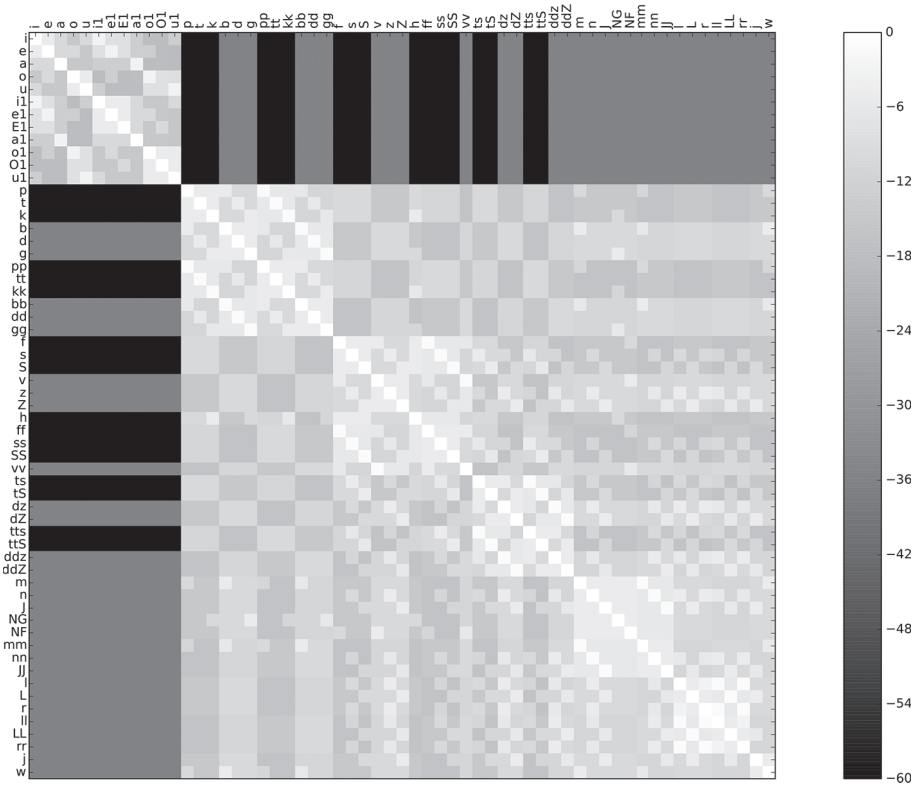
In order to compute the phonetic distance between the pronunciations of two words, we make use of the Needleman-Wunsch algorithm (Needleman, Wunsch, 1970), an algorithm used in bio-informatics to align DNA sequences. The distance between two words is determined by the number of insertions/substitutions/deletions necessary to convert a phonetic sequence into the other one.

The procedure is weighted by a substitution matrix that is responsible of weighting the substitution between phonemes taking into account the similarities between them.

Each element (i, j) of this matrix represents the penalty to be introduced by the algorithm in the replacement of the phoneme i with the phoneme j. Knowing the acoustic-phonetic characteristics of each phoneme (e.g.: vowel, consonant type, place of articulation, ...) of the language it is possible to automatically build this matrix for every language.

An example of a generated substitution matrix for the Italian language is shown in Figure 2. As an example, looking at the figure, it is possible to verify that the penalty for a substitution of a vowel with a consonant is high (darker color) and, on the contrary, the diagonal elements have a penalty equal to zero (white color).

Figure 2 - *Example of the substitution matrix used in the Needleman-Wunsch algorithm for the Italian language. Rows and columns are indexed by Italian phonemes (coded with SAMPA). Dark-gray elements represents a high penalty, light-gray elements represent low penalty*



Also for this task, the proposed procedure can be easily ported to different languages. In fact for each language, the algorithm needs: the definition of the phonetic characteristics (for the generation of the substitution matrix), the Grapheme to Phoneme module and the dictionary D.

2.3 Automatic generation of SUS sentences

SUS (Semantically Unpredictable Sentences) sentences (Benôit et al., 1996) are generated randomly from permissible grammatical structures by a NLG (Natural Language Generation) program. In order to run an experiment for a particular language, these programs must be designed to use the grammatical structures and lexical data peculiar of that language.

Since they are randomly generated, most SUS sentences are semantically anomalous: they can have no meaning or be perfectly plausible (from a semantical point of view).

For this reason the user does not have prior knowledge about the semantical correctness of the sentence he is about to listen, so they are called "semantically unpredict-

able". In the context of intelligibility tests, SUS sentences are used to prevent the subject of the test to be able to "guess" the missing word thanks to the semantic context.

The procedure adopted to generate the sentences of the test is described as follows.

1. A SUS sentence is randomly generated; it presents a correct grammatical structure (e.g.: Subject - verb - object - ...).

   It is worth noting that sentence generation is designed in a way that there can be no agreement in gender and number between nouns, modifiers and verbs.

   For example, consider the following SUS Italian sentence

   – "Lo roba beffa il bagno che esce"

   that has the following literal English translation

   – "The stuff jokes the bathroom that exits".

   In this example, "Lo" is a masculine article, while "roba" is a feminine noun.

2. A word (called "missing word") is chosen randomly among the sentence's words. In addition, the four words most similar to the missing words are calculated.

   In the previous example, "beffa" is chosen as the missing word. The similar words are "beffe | biffa | bissa | biscia".

   We have ensured that the missing word is not always in the same position. In a sentence, it can be at the beginning, in other sentences it can be in the middle or at the end of the sentence.

3. An audio file (containing the speech synthesis of the sentence to be presented to the test user) is generated, by randomly choosing a word, among the similar words, replacing the missing word.

   The word chosen may not agree in gender and number or have a different POS tag with reference to the missing word, thus breaking the correctness of the original grammatical structure.

   With reference to the previous example, the synthesized sentence can be

   – "Lo roba beffe il bagno che esce."

   In this case, "beffe" is a plural noun, while the missing word "beffa" is a verb.

4. The user will listen to the synthesized sentence. Also, the user will be presented with the sentence without the missing word and a menu with the similar words, where she/he will have to choose.

   For example:

   – "Lo roba wx il bagno che esce."

   where wx can be selected in the following menu

   – wx (beffa | beffe | biffa | bissa | biscia).

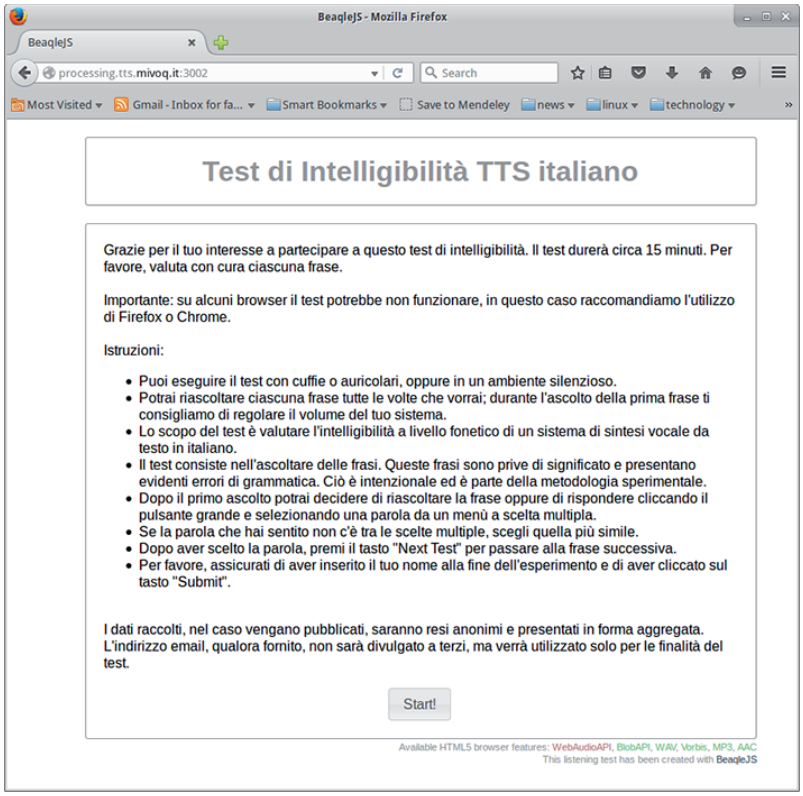
3. *Web based intelligibility evaluation tool*

Another issue related to the organization of subjective tests is recruiting of the adequate number of subjects to reach a statistically valuable result. Nowadays, thanks to the massive diffusion of Internet, it is possible to get in touch with a lot potential subjects and allowing them to perform the test online. Compared to the experiments carried out in the laboratory, the online test is deprived from the opportunity of supervising the subjects when they are performing the experiment, but it has the great advantage of being able to reach more people, leaving them the freedom to carry out the test when and where they want only using a web browser.

For these reasons, we spent efforts on the development of a web-based evaluation tool: we implemented additional features into an open source web application for listening tests called BeaqleJS (Kraft, Zölzer, 2014).

Figure 3 shows the instructions for the experiment available when the subject begins the test.

Figure 3 - *Intelligibility web evaluation tool: initial page (instructions)*



The instructions are configurable, and they should be decided depending on the particular context. An example of instructions for the subjects can be:

– listen to the sentence as many times as you want;

– when you are ready to give the answer, the text of the sentence will be displayed on the screen with the exception of one word: the missing word;
– with reference to the missing word, please select the one that you have understood by choosing from a list of five words in a multiple choice menu.

Figures 4 and 5 show some screen-shots of the web tool during different phases of line test.

Figure 4 - *Intelligibility web evaluation tool: click to play*

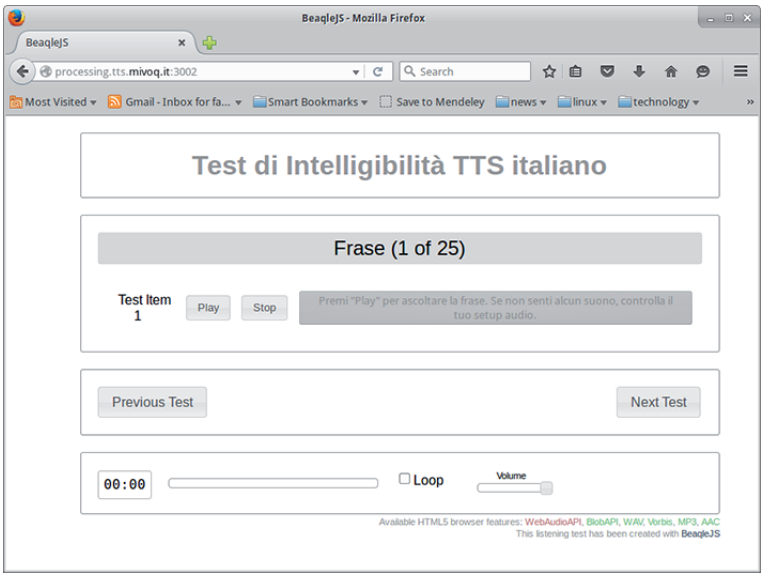
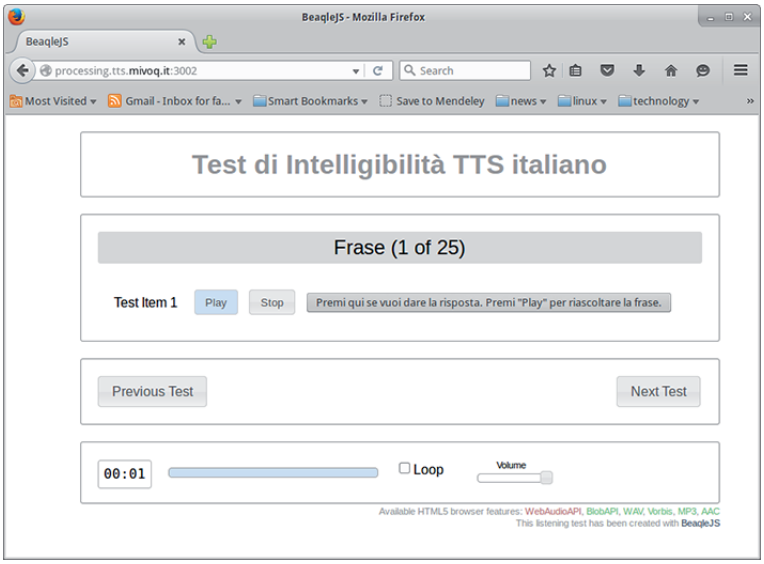


Figure 5 - *Intelligibility web evaluation tool: click to answer*



4. *Data Analysis*

The web application has been designed to store the responses of each subject and also additional information such as the number of listenings (i.e., the number of clicks on the "play" button) and the time the user took to give each answer.

The data are stored in JSON format and collected by a dedicated server. It is easy to write some scripts to analyse the data and compute indexes related to intelligibility.

The typical index that measures intelligibility is the Word Correct Rate (WCR), i.e. the ratio between the number of words correctly identified and the total number of words:

$$WCR = \frac{\text{Number of Correct Word}}{\text{Total Number of Word}} \tag{1}$$

Other interesting indexes that could be computed with this method are:
- number of listenings for the correct words;
- number of listenings for the wrong words;
- time spent on correct words;
- time spent on wrong words.

Finally, a phonetic analysis of the errors is possible with this method. In fact, with the phonetic transcriptions of the reference word and the chosen word it is possible to analyse which phonemes have caused more misunderstandings. This is possible because when a user chooses a word different from the reference word, it is possible to see how the transcripts of the two words differ. In particular, it is possible to investigate the frequency of events such as: phoneme replaced with another, phoneme inserted or phoneme deleted.

5. *Experiment*

In order to prove the whole methodology, we ran an intelligibility experiment to evaluate two Italian voices offered by FA-TTS[1], an open TTS system released in the context of the European project FI-Content2[2]. The two voices, a female one (istc-lucia) and a male one (istc-speaker internazionale) were built with Statistical Parametric Speech Synthesis technology (Zen et al., 2009).

Since in this experiment we evaluated Italian voices, we have sent requests to several Italian mailing lists to recruit subjects. The only required constraint for the target user was being Italian mother tongue. We did not make difference between people expert on speech technology or phonetics/phonology. The targeted end users were Italian people familiar with email and internet.

[1] http://lab.mediafi.org/discover-flexibleandadaptivetexttospeech-overview.html.
[2] http://mediafi.org.

Regarding the conditions upon which our intelligibility test was based, we were interested in ensuring that the sentences were understandable with every device (headphones, headsets, desktop speakers, hi-fi systems); for this reason we did not give any indications about what device to use.

In addition, in order to simulate real conditions, it was decided to leave the user free to listen to the test where and how she/he preferred. In this way, we can receive feedback from people who are running the test in front of a Desktop PC or using a smartphone and earphones. Everything necessary to run the test was: a device with a browser and audio output and a working network connection.

For this experiment, the following design choices were carried out:

– we made use of the freely available Wikipedia corpus to extract the words of the Italian language;
– starting from these, we selected the N = 87 words that maximize the phonetic coverage, using the procedure described in section 2.1;
– for each of these words we computed the R = 5 more similar words, using the algorithm described in section 2.2;
– the SUS generation algorithm, described in section 2.3, was used to generate 5 grammatically different kinds of sentences for each voice and taking into consideration the 87 missing words: a total of $87 \cdot 5 \cdot 2 = 870$ sentence was generated;

For each of these sentences it was generated the corresponding audio file using the related TTS voice. Regarding the missing word, the text passed to the synthesizer was randomly selected between the 5 acoustically similar words.

The web service for delivering this experiment and collecting the data has been running from October 9th to October 13th 2015. For each user, the test session consisted of a set of 25 randomly selected sentences among the 870 ones that have been automatically generated by the system described above. The estimated time necessary to execute the experiment was 10-15 minutes. A total number of 146 subjects participated to the experiment.

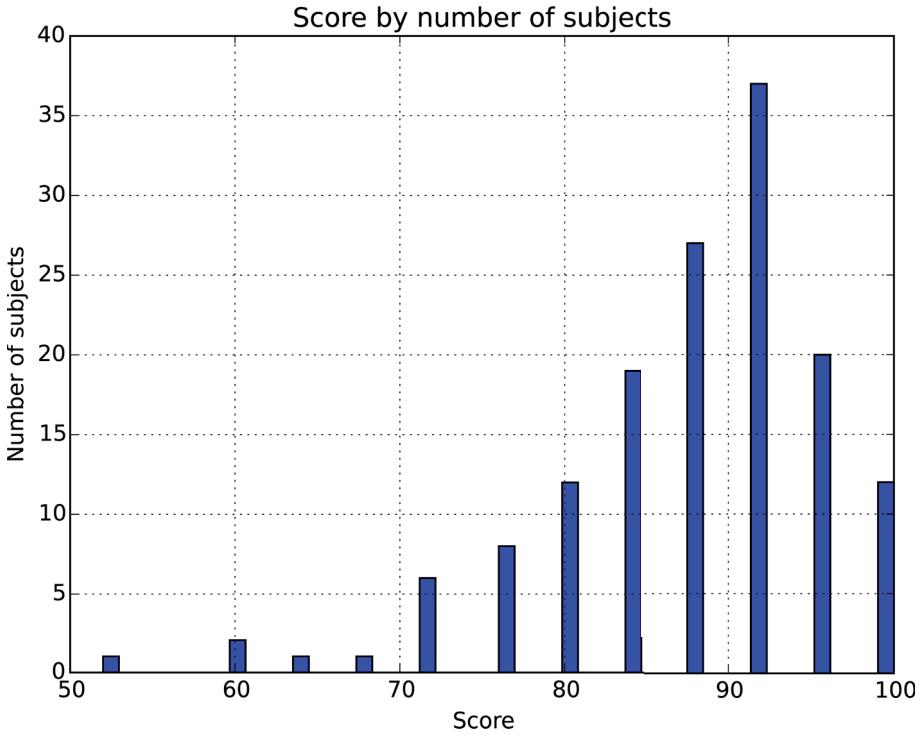
Table 1 shows the Word Correct Rate computed from the analysed data; the global WCR of the Italian voices is 87.7%, and there is not a significant difference between the female voice and the male one.

Table 1 - *Word Correct Rate of the two FA-TTS voices, istc-lucia is the female voice, and istc-speaker internazionale is the male voice*

FA-TTS voices	*WCR (%)*
istc-lucia	87.2
istc-speaker internazionale	88.1
aggregate	87.7

Figure 6 shows the Word Correct Rate by the number of subjects. From this figure we can see that 12 subjects out of 146 have correctly recognized all the missing words (WCR = 100%). Most participants have reached a WCR of 92 %.

Figure 6 - *Word Correct Rate by number of subjects*



Other interesting figures are shown on Tables 2 and 3. Table 2 shows that the average number of listenings when the user guessed right the missing word is not so different from the number of listenings when the answer was wrong. On the contrary, Table 3 shows that the time spent on wrong words is, in average, almost 5 seconds longer than the time spent on correctly guessed words.

Table 2 - *Average number of listenings for the correct and wrong answers*

	Number
Average number of listenings for the correct words	1.89
Average number of listenings of wrong words	1.66

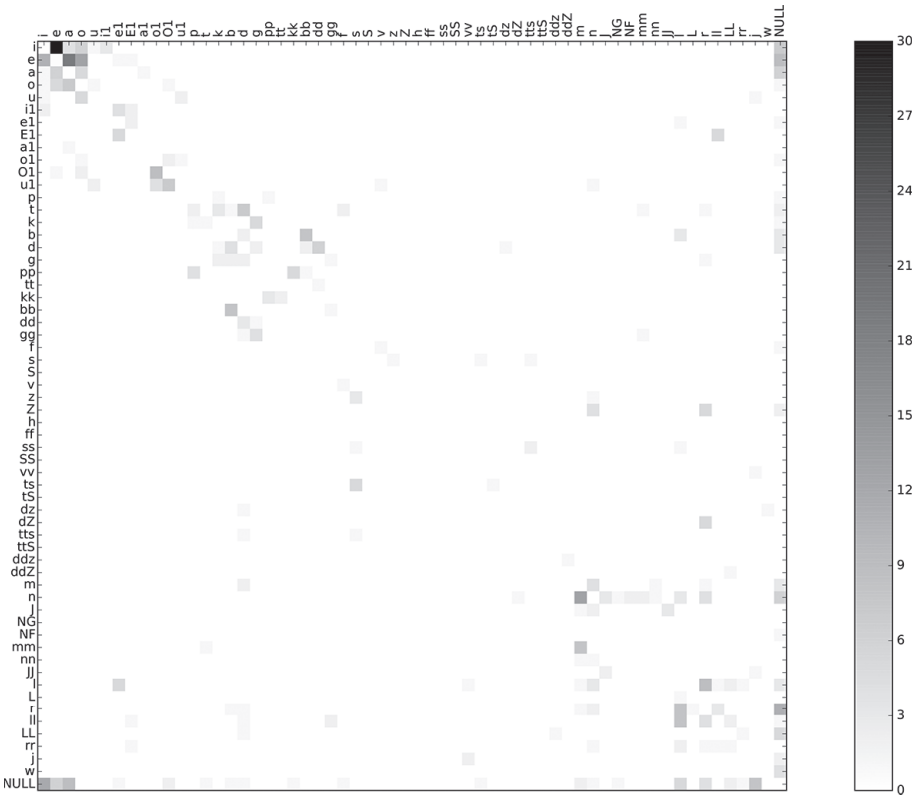
To analyse more in detail the intelligibility of the TTS system, the analysis of phonetic errors done during the experiment was carried out. The error matrix shown in Figure 7 illustrates the phoneme errors: the colour of each element of the matrix

represents the frequency of the event "phoneme in the abscissa was mistaken for the one in the ordinate".

Table 3 - *Average time spent on correct and wrong answers*

	Time (s)
Average time spent on correct words	17.72
Average time spent on wrong words	22.55

Figure 7 - *Error matrix of intelligibility for the two Italian FA-TTS voices: the colour of each element of the matrix represents the frequency of the event "phoneme in the abscissa was mistaken for the one in the ordinate"; white colour means zero errors, black colour means 30 errors. NULL is used to represent the case when a phoneme has been inserted or deleted*



From the error matrix, it is possible to notice that the most common mistake has been to misunderstand very similar vowels: /i/ and /e/ or /e/ and /a/. Another interesting and frequent error was confusing the geminates with its non-geminate counterpart. Also liquid consonants /l/ and /r/ have been misunderstood quite frequently.

6. *Conclusions and future works*

A new method to design and perform subjective tests of intelligibility has been presented and described. This method has the advantage of being easily replicable for each language as long as phonetic information, a Grapheme to Phoneme module, a SUS NLG tool and a big textual database are available for the concerned language.

Analysis of the data does not require manual effort to transcribe or correct user answers because the test is based on multiple-choice answers.

An experiment on the Italian language to test the intelligibility of two TTS voices was designed, executed and tested with success. We aim to investigate this methodology further. The next steps to take, in order to validate deeply this new method comprises:

a. the addition of sentences from a stable reference such as a real voice or a different TTS with a known intelligibility;
b. adding different levels and kind noise to the speech signal, in order to test controlled speech-in-noise intelligibility;
c. make a comparison with the traditional intelligibility test methodology.

Finally, we are evaluating the possibility of using also non-words, to check if this could improve the phonetic similarity among the words in the multiple-choice menu.

Acknowledgment

The authors want to thank all the participants in the subjective test. This work was supported by the EU FP7 "FI-Content2" project (grant number 603662).

Bibliography

Benôit, C., Benôit, C., Grice, M. & Hazan, V. (1996). The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences. In *Speech Communication*, 18, 4, 381-392.

Cooke, M., Mayo, C. & Valentini-Botinhao, C. (2013). Intelligibility-enhancing speech modifications: the Hurricane Challenge. In *Proceedings of Interspeech 2013*, 3552-3556.

Egan, J.P. (1948). Articulation testing methods. In *The Laryngoscope*, 58, 9, 955-991.

Erro, D., Zorila, T.C. & Stylianou, Y. (2014). Enhancing the Intelligibility of Statistically Generated Synthetic Speech by Means of Noise-Independent Modifications. In *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22, 12, 2101-2111.

Godoy, E., Koutsogiannaki, M. & Stylianou, Y. (2013). Assessing the Intelligibility Impact of Vowel Space Expansion via Clear Speech-Inspired Frequency Warping. In *Proceedings of Interspeech 2013*, 8, 1169-1173.

Goldstein, M. (1995). Classification of methods used for assessment of text-to-speech systems according to the demands placed on the listener. In *Speech Communication*, 16, 3, 225-244.

King, S., Karaiskos, V. (2010). The Blizzard Challenge 2010. In *Proceedings of Blizzard Challenge Workshop*, Kyoto, Japan.

Kraft, S., Zölzer, U. (2014). BeaqleJS: HTML5 and JavaScript based Framework for the Subjective Evaluation of Audio Quality. In *Linux Audio Conference* (LAC-2014).

Logan, J.S., Greene, B.G. & Pisoni, D.B. (1989). Segmental intelligibility of synthetic speech produced by rule. In *Journal of the Acoustical Society of America*, 86, 2, 566.

Needleman, S.B., Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. In *Journal of Molecular Biology*, 48, 3, 443-453.

Nicolao, M., Tesser, F. & Moore, R.K. (2013). A phonetic-contrast motivated adaptation to control the degree-of-articulation on Italian HMM-based synthetic voices. In *Proceedings of 8th ISCA Workshop on Speech Synthesis*, Barcelona, Spain, 107-112.

Pammi, S., Charfuelan, M. & Schröder, M. (2005). Multilingual voice creation toolkit for the MARY TTS platform. In *Proceedings of International Conference on Language Resources and Evaluation (LREC 2005)*, Valleta, Malta.

Valentini-Botinhao, C., Yamagishi, J. & King, S. (2011). Evaluation of objective measures for intelligibility prediction of HMM-based synthetic speech in noise. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, IEEE, 5112-5115.

Valentini-Botinhao, C., Yamagishi, J., King, S. & Maia, R. (2014). Intelligibility enhancement of HMM-generated speech in additive noise by modifying Mel cepstral coefficients to increase the glimpse proportion. In *Computer Speech & Language*, 28, 2, 665-686.

Venkatagiri, H.S. (2003). Segmental intelligibility of four currently used text-to-speech synthesis methods. In *Journal of the Acoustical Society of America*, 113, 4, 2095.

Venkatagiri, H.S. (2005). Phoneme Intelligibility of Four Text-to-Speech Products to Non-native Speakers of English in Noise. In *International Journal of Speech Technology*, 8, 4, 313-321.

Yu, Z., Yue, D., Zu, Y. & Chen, G. (2010). Word intelligibility testing and TTS system improvement. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010)*, IEEE, 593-596.

Zen, H., Tokuda, K. & Black, A.W. (2009). Statistical parametric speech synthesis. In *Speech Communication*, 51, 11, 1039-1064.

Zorila, T., Kandia, V. & Stylianou, Y. (2012). Speech-in-noise intelligibility improvement based on power recovery and dynamic range compression. In *Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012)*, Portland, USA, September 2012, 2075-2079.

DOI: 10.17469/O2102AISV000019

ANNA DORA MANCA, GIORGIO DE NUNZIO, MIRKO GRIMALDI

EEG-Based Recognition of Silent and Imagined Vowels

This work proposes a framework for future Silent Speech Interfaces (SSI) based on non-invasive EEG recordings. Specifically, the information embedded in the brain signals related to the production – overt, covert and imagined production – of the Italian vowels /a/ and /i/ allowed to distinguish the vowels relying on discriminative features calculated by the Ambiguity Function in the context of time-frequency analysis, and ranked by the Fisher contrast. The vowels were classified by using a multilayer feed-forward ANN. Overall, intra-subject classification accuracies, as measured by the area under the ROC curve, ranged from 0.84 to 0.96 for overt production, from 0.83 to 0.96 for covert production, and from 0.89 to 0.98 for imagined vowels. Results indicate significant potential for the use of speech prosthesis controllers for clinical and military applications.

Key words: EEG, speech, neural network, vowels, ambiguity function.

Introduction

Several electrocorticographic, electric and magnetic investigations showed that from the brain signals important information for the discrimination of spoken and perceived speech sounds can be extracted (Bouchard, Mesgarani, Johnson & Chan, 2013; Pei, Barbour, Leuthardt & Schalk, 2011; Wang, Perreau-Guimaraes, Carvalhaes & Suppes, 2012; Obleser, Lahiri & Eulitz, 2004; Obleser, Scott & Eulitz, 2006; Scharinger, Idsardi & Poe, 2011; Luo, Poeppel, 2012). It seems also feasible to recognize neuronal traces of non-audible speech sounds evoked during imagined and mouthed (covert) speech processes; the idea is that the mechanisms underlying such operations rely on the same neuronal substrates involved in the processes of overt speech production, thus, tracing and detecting the related cortical signals seems actually plausible (Tian, Poeppel, 2010).

In the last decades, different attempts have been made to decoding the EEG signals associated to non-audible speech mainly with the interest of testing new methodologies for speech recognition systems such as Silent Speech Interfaces (SSIs). These systems acquire data from brain activity associated with overt and covert speech performance and synthesize information by reproducing a digital representation of the signals necessary for their functioning (for a detailed description of SSIs see Denby, Shultz, Honda, Hueber & Gilbert, 2010). The potential usability of these applications is enormous – from medical to military environments – and leads to explore methodological approaches in support of new portable and user friendly EEG-based SSIs. For researchers, this means resolving some critical steps such as the extraction of the most discriminative features of the brain signals

associated with speech sounds and the choice of accurate classification procedures (Šťastný, Sovka & Stančák, 2003). To date, several approaches have been proposed.

First works go back to the end of 90s when Suppes and colleagues (Suppes, Lu & Han, 1997) succeed in decoding electric and magnetic brain signals recorded during imagined words; some years later however, Porbadnigk and colleagues showed that the classification rates were biased for the effects of the temporal artifacts caused by the experimental protocol (Porbadnigk, Wester & Calliess, 2009). Subsequent studies focused mostly on decoding imagined phonemes. For example, D'Zmura and colleagues (D'Zmura, Deng, Lappas, Thorpe & Srinivasan, 2009) showed with spectral analysis techniques that the brain frequency bands were informative for non-audible sounds classification. They recorded the EEG activity of four subjects performing two imagined syllables /ba/ and /ku/ with three different rhythms and achieved a classification accuracy of 87% only for one of the four subjects included in the experiment. Working on the same data set, Brigham and Kumar extended the result demonstrating that classification rates remarkably improved after an intensive technique of artifacts rejection. Here, the features were extracted by autoregressive coefficients and the classification was done with a k-Nearest Neighbor classifier (Brigham, Kumar, 2010). Meanwhile, DaSalla et al. classified the neuronal activity of three healthy subjects associated to the imagined vocalization of the English vowels /a/ and /u/ as compared to a no-state control condition where subjects were at rest (DaSalla, Kambara, Sato & Koike, 2009). The authors applied spatial filters to the EEG time series and tested a support vector machine (SVM) for the classification of the tasks achieving overall good accuracies (/a/ vs. rest: 68%; /u/ vs. rest: 78%). The same EEG dataset was tested with other kinds of classification algorithms (see Santana, 2015; Iqbal, Shanir, Khan & Farooq, 2016) reporting similar accuracy percentages. Yet, the EEG activity evoked during the imagined production of couples of phonemes differing in patterns of vocal articulation was successfully classified (classification rate above 70%) by exploiting information embedded in spectrogram samples at specific brain frequencies (Chia, Hagedorna, Schoonovera & D'Zmura, 2011). Here, classification was done with a Naive Bayes and Linear Discriminant Analysis (LDA) classifiers. Similar results were found in pairwise classification using SVM of the Japanese vowels /a/ and /u/ (Matsumoto, Hori, 2014). To conclude, to the best of our knowledge, only Riaz and colleagues have discriminated EEG data of three subjects performing mouthing tasks of five different vowels, (a, e, i, o, and u). Results of the pairwise comparisons showed an average accuracy of around 75% with the best separation between vowels /a/ vs. /i/ and /e/ vs. /u/ (Riaz, Akhtar, Iftikhar, Khan & Salman, 2014).

In the present work, we explored a procedure for decoding brain signals associated to different experimental speech tasks: overt production (OP), covert production (CP or mouthing) and imagined production (IP) of the Italian vowels [a] and [i]. These two vowels are suitable for our explorative purposes as they are realized by maximally contrastive tongue gestures: /a/ is pronounced by lowering the tongue body and /i/ by raising the tongue body and advancing the tongue root. Our aims

were (i) to determine whether the patterns elicited by OP were elicited even in absence of audible speech signals (i.e., in CP and IP) and (ii) whether the EEG waves contained discriminant information for vowel classification. To do this, spectral analysis and the (symmetric) Ambiguity Function (AF) were used to represent the EEG signals, and a feed-forward Artificial Neural Network (ANN) was tested for vowel classification in each task. If covert and imagined speech conditions reveal as useful levels of investigation, then the framework may be implemented in new methodological approaches for the development of non-invasive SSIs.

1. *Methods*

1.1 Subjects

Twelve students of the University of Salento (Lecce, Italy) (7 males and 5 females, 25±3 years) participated in the experiment after providing a written informed consent. They were right-handed according to Handedness Edinburgh Questionnaire and none of them had any known neurological disorder or other significant health problem. The ethical committee of the local health authority of Lecce approved the study.

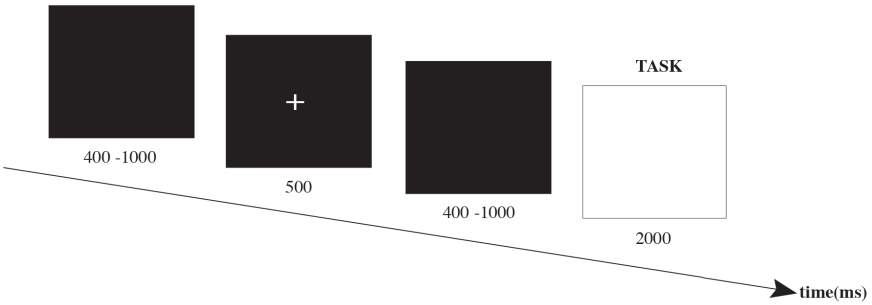
1.2 Experimental procedure

In separate runs, the participants performed three tasks: OP, CP and IP of the vowel /a/ and then of the vowel /i/. In the OP task, the subjects pronounced aloud the vowel, in the CP task, they mouthed the vowel without any emission of sound, and during IP, they had to imagine to produce the vowel without using articulatory muscles (i.e., inertial tongue and mandibular movements) and without uttering any audible sound. The order of the tasks was counterbalanced across participants; the order of the vowels was established before each task.

Each trial began with a black screen displayed for a random time (400-1000 ms) followed by a small white cross (500 ms) in the center of the computer monitor, used to suggest subjects to concentrate and prepare for the task. Another randomized time interval (400-1000 ms) preceded a white screen (2000 ms) which triggered the onset of the task. Each session consisted of 80 visual cues (white screen) and each trial had an average duration of about 3850 ms (Figure 1).

The subjects were instructed to perform as best as possible the experimental task while remaining completely still during imagined phoneme production. All participants took part in a training phase, which was identical to the experimental procedure to ensure an accurate task performance.

Figure 1 - *Schematic illustration of the paradigm employed in the present experiment*



1.3 Data acquisition

Continuous EEG was recorded with a 64-channel actiCAP (10-20 system), a sampling rate of 250 Hz and a band pass filter of 0.1-70 Hz (BrainProducts GmbH, Germany). The vertical Electro-oculogram (EOG) was recorded by means of two electrodes (same type as EEG) just above and below the right eye, and the horizontal EOG was recorded with the FT9 and FT10 electrodes. The online reference was at FCz and impedance was kept under 5 kΩ by electrogel conductant.

2. *Data pre-processing*

2.1 EEG analysis

Off-line signal processing was carried out with MATLAB and the software package EEGLab. Data were digitally filtered at 2–30 Hz, they were re-referenced to the right and left mastoids TP9-T910 and down-sampled to 100 Hz according to EEG studies on the classification of speech stimuli (Wang et al., 2012; Suppes, Han, Epelboim & Lu, 1999). Independent component analysis (ICA) was computed as a pre-processing step to remove muscular and ocular artifacts. A script was written to identify artefactual independent components (ICs) by exploiting their power spectral density (PSD) properties (Vos, Riès, Vanderperren, Vanrumste, Alario, Huffel & Burle, 2010). The basic assumption is that brain EEG signals have lower power at high frequencies whereas muscular EEG signals have higher power at high frequencies. Accordingly, we have considered as potential muscular artifacts the ICs whose average power between 15–30 Hz was at least twice as great as the one between 2-15 Hz. Similarly, keeping in mind that the ocular EEG signal power has a very narrow peak between 0–4 Hz, we considered possible ocular artifacts those ICs whose average power between 2–4 Hz was at least half of that between 4–30 Hz. Finally, we visually inspected the highlighted ICs. Components actually identified as artifacts were rejected, and the original EEG time-courses were reconstructed, using only the preserved ICA components. EEG epochs were extracted with reference to the white screen onset. Each epoch had a duration of 400 ms, 100 ms of pre-stimulus and 300 ms of stimulus.

2.2 Feature extraction

For an accurate characterization of non-stationary signals such as EEG data, Time-Frequency Representations (TFRs) are required. TFRs (Kozek, Hlawatsch, Kirchauer & Trautwein, 1994) are data processing methods in which signals are analyzed simultaneously in the time and frequency domains, in a 2D representation. The rationale for TFRs is that conventional methods as Fourier Transform assume the signals to be periodic or infinite in time, while many real-life signals, such as EEG time series, vary considerably. Known TFRs are the Short-Time Fourier Transform (STFT) and wavelet analysis. In this work, the (symmetric) Ambiguity Function (AF), i.e., the inverse Fourier transform of the Wigner-Ville distribution was proposed to represent EEG signals (Kozek et al., 1994) and an ANN (Haykin, 2008) was used for the vowel classification in the different tasks. EEG epochs for each subject were initially processed by time-frequency analysis in the doppler-delay ambiguity plane. Values of the ambiguity function in the plane were chosen as features for vowel recognition. The most discriminant points in the plane were identified by maximizing the Fisher contrast of the two classes, and the ambiguity values in those points formed the feature vector. A 2-layer, 5-hidden-neuron feedforward ANN was trained and validated for the recognition of the vowels, independently on each subject. ROC (Receiver Operating Characteristic) curves were calculated and the AUC (Area Under the Curve) values were derived as a classification accuracy measure.

The AF of signal x(u), denoted by Ax, is defined as:

$$A_x(\tau,\nu)=\int_{\Re} x\left(u+\frac{\tau}{2}\right)x^*\left(u-\frac{\tau}{2}\right)e^{-2\pi i\nu u}\,du \tag{1}$$

Where t is the time delay, n is the doppler frequency shift, and x* is the complex conjugate of x. The AF can be considered as an autocorrelation function in joint time-frequency domain, which transforms a signal to time delay and frequency shift plane (Ambiguity Plane). Its most useful properties are:

i. The AF modulus is independent of time and frequency shift, that is, if y is a time- and frequency-shifted copy of x: $y(t)=x(t-t_1)e^{2\pi i f_1 t}$, then $A_y(\tau,\nu)=A_x(\tau,\nu)\,e^{2\pi i(f_1\tau-t_1\nu)}$, so that $|A_y(\tau,\nu)|=|A_x(\tau,\nu)|$;
ii. The AF modulus is symmetric with respect to the origin: $|A_x(t,\nu)|=|A_x(-\tau,-\nu)|$. If x is real, then: $|A_x(\tau,\nu)|=|A_x(\tau,-\nu)|$ and $|A_x(-\tau,\nu)|=|A_x(-\tau,-\nu)|$, $|A_x(\mathrm{t},\nu)|=|A_x(-\tau,\nu)|$, $|A_x(\tau,-\nu)|=|A_x(-\tau,-\nu)|$;
iii. The largest AF value is in the axes origin, and equals signal energy: $\forall\tau,\nu$: $|A_x(\tau,\nu)|\le|A_x(0,0)|=\int|x(t)|^2\,\mathrm{dt}$;

Time shift and frequency shift invariance (property (i)) indicates that even if the arriving times and center frequencies of the signal vary from each other, the moduli of their AFs are the same. Therefore, extracting features from the ambiguity plane does not require time alignment and frequency transform. The symmetry properties of AF with real signals (property (ii)) allowed considering only a quarter of the

ambiguity plane without information loss. Some literature exists on the subject of AF applications to pattern recognition and signal classification, where discriminant features are taken from the Ambiguity plane. If the length of a signal (number of samples) is Ns, the AF of the signal is a Ns×Ns matrix, which is generally large. Therefore, it is convenient to project the ambiguity function to a lower-dimensional space. In some studies, (McLaughlin, Droppo & Atlas, 1997; Atlas, Droppo & McLaughlin, 1997; Gillespie, Atlas, 2001), kernel function methods were proposed that extracted features from the ambiguity plane by designing time-frequency kernel functions, which preserved the location of the ambiguity plane that maximized class separability. In Garcia et al. (Garcia, Ebrahimi & Vesin, 2003) and in Ebrahimi et al. (Ebrahimi, Vesin & Garcia, 2003), this method was applied to Brain-Computer Interfacing. As a means of reducing feature-space dimensionality (Garcia, Ebrahimi & Vesin, 2002) used the Fisher Contrast (or Fisher's discriminant ratio, FDR) to locate the N most discriminant locations on the ambiguity plane. Thus, N locations from the ambiguity plane are chosen, in such a way that the values in these locations are very similar for signals from the same class, but they vary significantly for signals from different classes. For our two-class classification of vowels, we followed this methodology, which proved simple but effective. The procedure consisted in determining the coordinates of a number of highest contrast points between two given TFRs in the ambiguity planes (representing the two classes), then using the values of the AF in those points as features for classification. Steps are as follows. Firstly, calculate the FDR for the training sets of the two classes (/a/ and /i/ vowels) in the ambiguity plane (doppler ν, delay τ), for each rebuilt EEG channel c:

$$K_{Fisher}(c,\tau,\nu)=\frac{\left|\overline{A}_{1,c}(\tau,\nu)-\overline{A}_{2,c}(\tau,\nu)\right|^2}{\overline{\overline{A}}^2_{1,c}(\tau,\nu)+\overline{\overline{A}}^2_{2,c}(\tau,\nu)} \tag{2}$$

In the above expression:

$$\overline{A}_{i,c}(\tau,\nu)=\frac{1}{n_i}\sum_{j=1}^{n_i}A_{x_j^{i,c}}(\tau,\nu) \tag{3}$$

$$\overline{\overline{A}}_{i,c}(\tau,\nu)=\frac{1}{n_i}\sum_{j=1}^{n_i}\left(A_{x_j^{i,c}}(\tau,\nu)-\overline{A}_{i,c}(\tau,\nu)\right)^2 \tag{4}$$

are respectively the mean and the variance of the AFs of all epochs (belonging to the training set), calculated for each class and for each channel: i indexes the two classes, n_i is the total number of signals for class i, $A_{x_j^{i,c}}$ is the AF of the j-th epoch of class i, in channel c. The rationale of using the FDR is to optimize the representation space by maximizing the value of $K_{Fisher}(c, \tau, \nu)$, which means increasing the distance between the mean of the two classes, while reducing intra-class dispersion. Secondly, chose a number of points Np in the AF planes, according to a criterion of maximum

discriminant power as measured by the KFisher contrast between the average AF planes for the two classes.

After some tests with a variable number of features (from 10 to some hundreds) we concluded that no important gain could be obtained by using too large Np values, so we set Np = 100 which appeared as a good compromise between accuracy and feature space dimensionality. Each point had coordinates $F_m(c_m, \tau_m, \nu_m)$, m = 1, ..., Np. Discriminant features were chosen as the AF calculated in $\{F_m\}$. Then, by considering the channels in which most frequently the features were chosen by the FDR, information about the most discriminant EEG electrodes was collected. The analysis of the EEG channels more frequently chosen by the FDR value put in evidence a noticeable variability between the subjects, but it allowed, anyway, to derive some interesting common information. In particular, the most discriminative sites for vowel discrimination were Cz, CP2, CP4, CP6 for the OP task, FT7-FT8-T7-T8-C6 for CP, and CP1-CP2-CP3 for the IP task.

2.3 Classification method

Each validation EEG epoch *x* was classified by a supervised classifier as class i, according to the set of features $AFx(c_m, \tau_m, \nu_m)$, m = 1, ..., Np. Supposing that the overall number of training trials for a subject is NT = NT1 + NT2 (the summation of trials for class 1 and 2 respectively, 160 per subject in our experiments), the classifier input is a NT×Np matrix (i.e. 160×100). Finally, a feed-forward ANN was chosen as the classifier (1 hidden layer with 5 hidden neurons, HNs). In intrasubject experiments, the training set was randomly divided into two subsets of equal cardinality, and the training-validation process was repeated 50 times, each time calculating the ROC curve and its AUC as a measure of accuracy. The result was a mean AUC with an associated error (the standard deviation). This step was repeated for each subject for the OP, CP, and IP tasks. Vowel classification was also performed intersubject in LOSO (Leave One Subject Out) cross validation.

3. *Results*

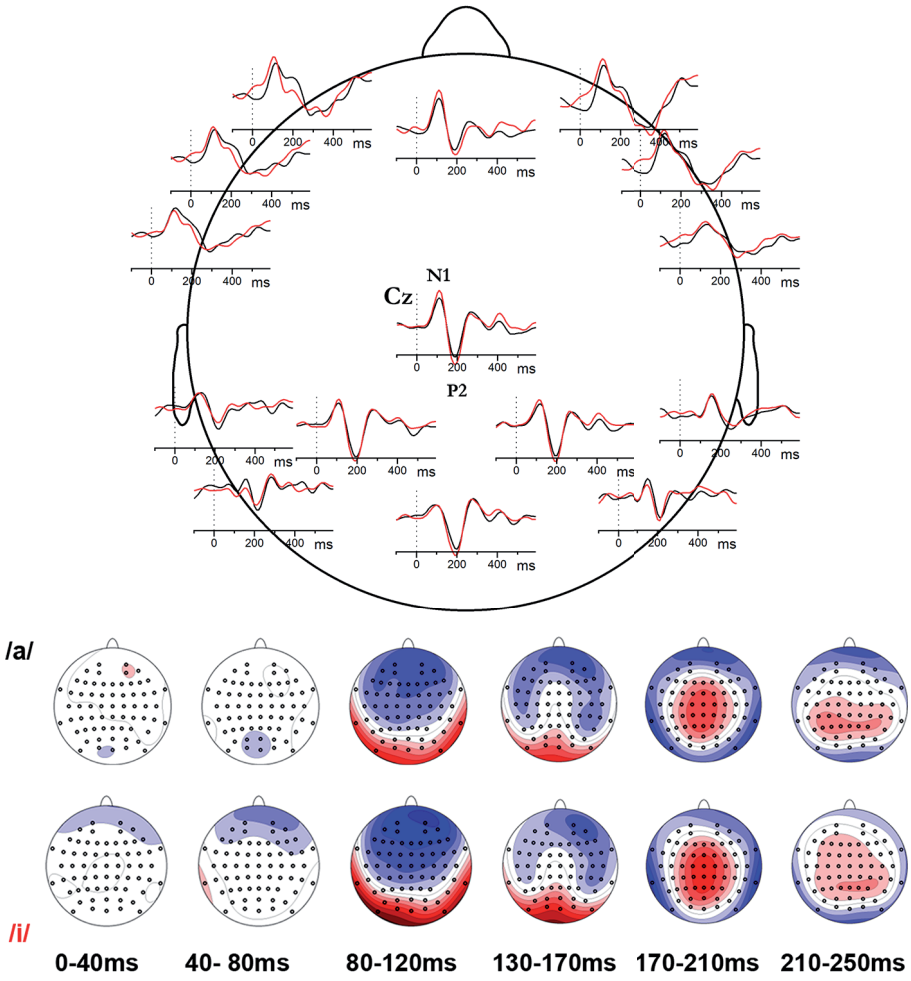
3.1 EEG patterns

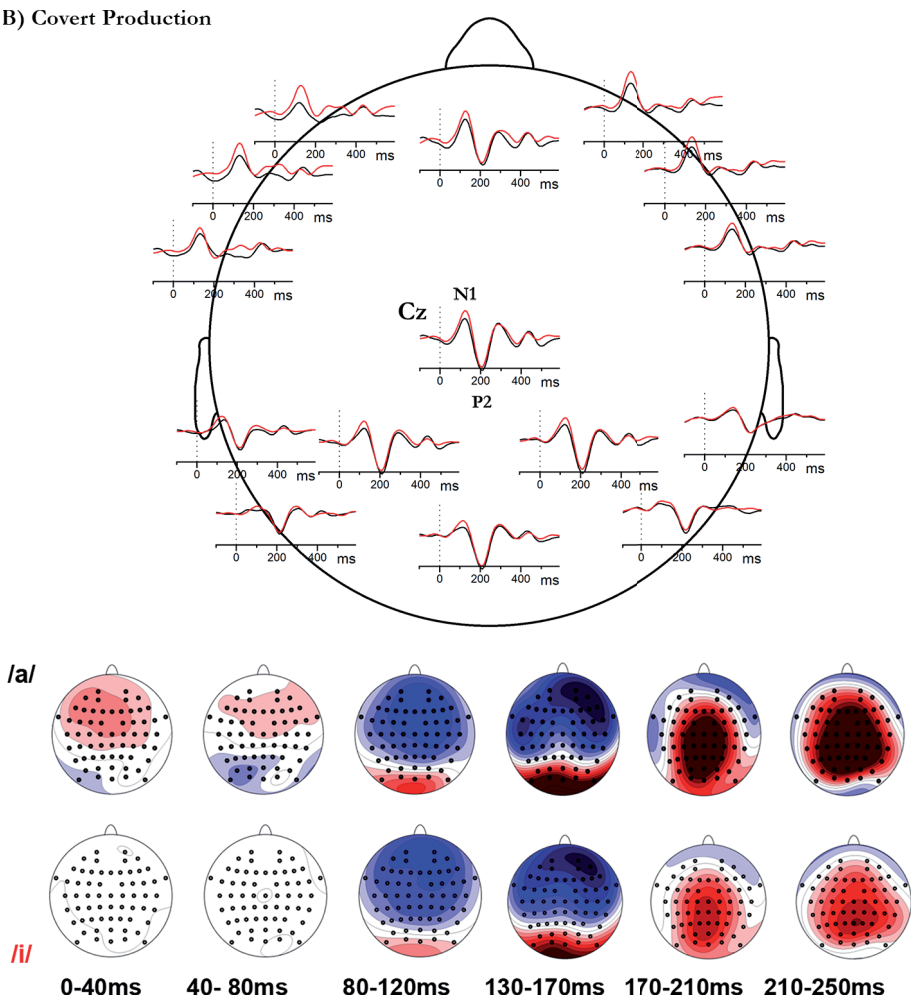
To visualize the responses to each vowel we calculated the average of all epochs in the three tasks separately. Figure 2 (A, B, C) shows the grand averages at the most important electrodes for vowel classification (Section 2.2); time 0 ms coincides with the appearance of the visual cue triggering the onset of the task execution. In each experimental condition, we recognized a negative trend reaching the most negative peak between 80-120 ms over the fronto-central electrodes, and a late positive shift peaking between 170-200 ms at central and parietal sites. These peaks resemble the typical neuronal audito-

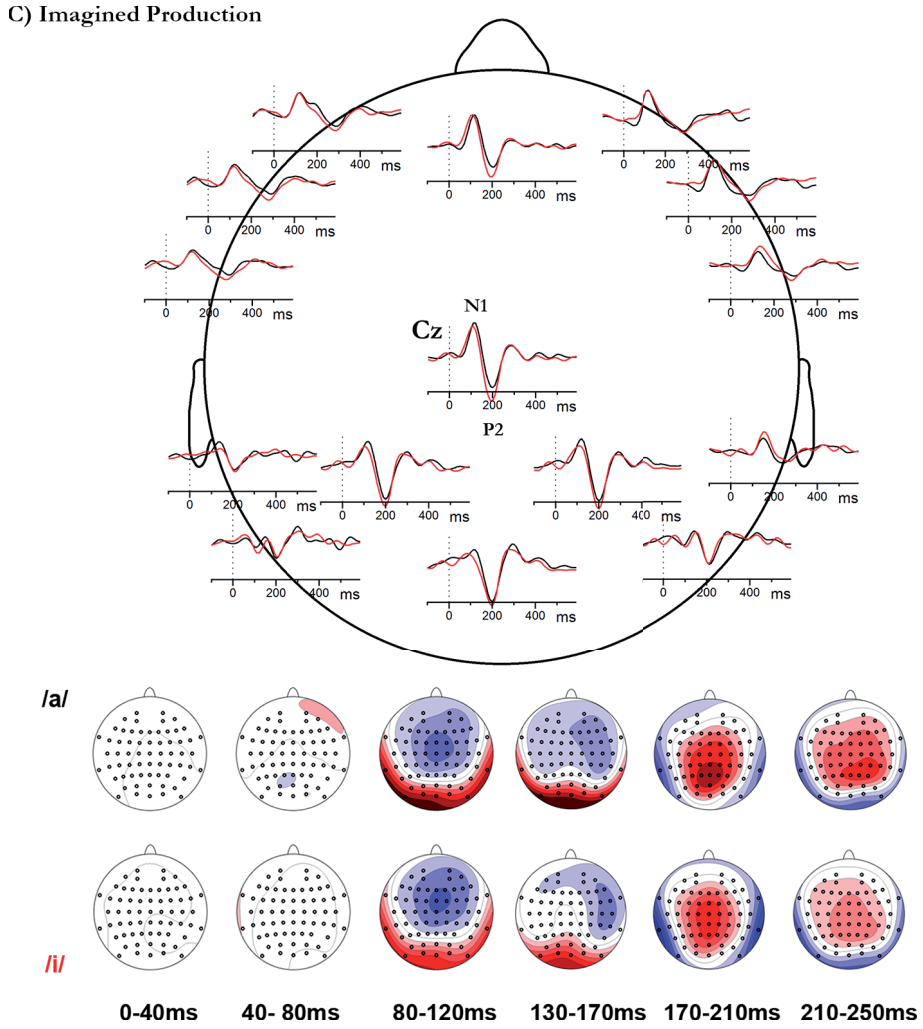
ry N1/P2 pattern (e.g., in Manca, Grimaldi, 2013). The mean amplitude of these peaks was calculated considering an interval of 60 ms centered at the maximum peak.

Figure 2 - *Grand average waveforms (N=12 subjects) at fronto-central and centro-parietal electrodes and topographic maps for the vowel /a/ (in black) and /i/ (in red)*

A) Overt Production




B) Covert Production
Cz
N1
P2
0 200 400 ms
/a/
/i/
0-40ms
40- 80ms
80-120ms
130-170ms
170-210ms
210-250ms



The presence of the N1 and P2 pattern of response was verified by a series of t-tests against zero at the midline electrodes FCz, Fz, and Cz ($p < 0.05$). Data were normality distributed ($p > 0.5$) as evaluated by a series of Shapiro-Wilk tests on the N1 and P2 amplitude and latency values at Cz electrode (n=12) where the components had the maximum distribution (Table 1).

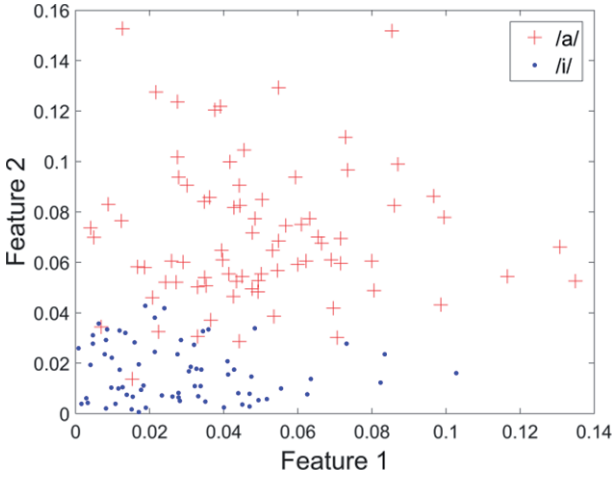
Table 1 - *Normality tests on N1 and P2 values for each experimental tasks*

	Overt Production		*Covert Production*		*Imagined Production*	
	Statistic	*Sig.*	*Statistic*	*Sig.*	*Statistic*	*Sig.*
N1 Amplitude /a/	,930	,376	,662	,089	,662	,089
N1 Latency /a/	,982	,990	,893	,129	,893	,129
P2 Amplitude /a/	,972	,926	,700	,098	,700	,098
P2 Latency /a/	,920	,282	,928	,355	,928	,355
N1 Amplitude /i/	,911	,221	,927	,347	,927	,347
N1 Latency /i/	,986	,998	,960	,785	,960	,785
P2 Amplitude /i/	,942	,520	,920	,286	,920	,286
P2 Latency /i/	,884	,098	,930	,375	,930	,375

3.2 Vowel classification

Vowel classification was obtained by a 1-hidden-layer feed-forward ANN with back-propagation (5 HNs). Tests with less or no HNs (in the hypothesis that the discrimination problem might be liner) gave poor results. Increasing the number of HNs gave no relevant accuracy improvement. As an example of the discriminant power of the chosen features, Figure 3 shows the scatter plot of the two classes (vowel /a/ and vowel /i/) in the plane of the two best features (named Feature 1 and Feature 2 in the graph), for the IP task of one of the subjects. The good class separation is evident.

Figure 3 - *Scatter plot of the two vowel classes in the plane of the two best features, for the IP task of one of the subjects*



AUC values for vowel classification, for each task and for each subject are reported in Table 2. AUC were averaged on 50 iterations of the ANN training/validation process, and standard deviations are reported as the uncertainties. The CP and IP

tasks showed an average classification accuracy (as measured by the ROC AUC) of 0.91 and 0.93 respectively, which suggests slightly better performance compared to the OP task (i.e., 0.89). This was statistically significant for OP vs IP comparison (according to a paired t-test applied to the mean values in Table 2, giving $p = 10^{-3}$). This finding needs anyway deeper investigation and confirmation.

Intersubject classification in LOSO cross validation was finally tested with very poor results (AUC about 0.50-0.60).

Table 2 - *Average AUC values and standard deviations calculated with 50 runs of the training/validation process, for each subject (S) and each task*

S	AUC – OP	AUC - CP	AUC – IP
1	0.85±0.03	0.95±0.02	0.93±0.03
2	0.85±0.04	0.96±0.02	0.89±0.03
3	0.88±0.03	0.85±0.04	0.92±0.03
4	0.88±0.03	0.95±0.02	0.94±0.02
5	0.87±0.05	0.93±0.03	0.91±0.03
6	0.95±0.02	0.84±0.05	0.94±0.02
7	0.83±0.05	0.90±0.03	0.93±0.02
8	0.94±0.02	0.95±0.02	0.96±0.02
9	0.96±0.02	0.91±0.03	0.96±0.02
10	0.93±0.03	0.95±0.02	0.97±0.02
11	0.93±0.03	0.86±0.05	0.98±0.02
12	0.87±0.03	0.90±0.03	0.89±0.03
Mean	0.89	0.91	0.93

4. *Discussion*

The current study shows that is feasible to recognize the features distinguishing the vowels /a/ and /i/ from information embedded in the EEG signals generated during covert and imagined production. Two main conclusions derived from the results.

First, the CP and IP tasks elicit similar neural responses to the OP showing the typical auditory N1/P2 responses to speech sounds (Näätänen, Picton, 1987) as in previous studies on Italian vowels. In particular, an event–related study on the perception and production processes of the same vowel pairwise (Manca, Grimaldi, 2013) suggested the N1/P2 pattern as index of the activation of auditory neurons to the linguistically relevant properties of sounds; yet, the generators of the auditory activity to the perceived vowels was localized in the supratemporal auditory cortex of both hemispheres (Manca, Di Russo & Grimaldi, 2015). Furthermore, models of speech production (Guenther, Hampson & Johnson, 1998; Tian, Poeppel, 2010) have established that during speech production two efferent copies – auditory and motor – are created from stored models of previous speech motor acts; when the speech command is executed, auditory feedback of the spoken sound is heard at the level of

the peripheral auditory system and processed through the central auditory pathway to the bilateral auditory temporal lobe (Tourville, Reilly & Guenther, 2008). Yes, further MEG studies revealed that the auditory cortical potentials at a latency of approximately 100 ms are modulated during speech execution (Gunji, Hoshiyama & Kakigi, 2000; Heinks-Maldonado, Nagarajan & Houde, 2006). In the present work, the topography of our waves resembles the same neuronal pattern revealing some hints of auditory activation also in the tasks where there exists no auditory feedback (Figure 2). That is, since the subjects in the present experiment were instructed to generate different forms of speech production, we can speculate that the early activity elicited during the CP and IP tasks (as described by N1/P2 pattern) represents the output of sensory-motor circuits along which the auditory system is activated even when motor activity is inhibited or absent. Anyway, the role played by the motor areas in speaking cannot be ruled out: studies on movement-related cortical potentials (Deecke, Engel, Lang & Kornhuber, 1986) reported negatives potentials with symmetric activities at 100 ms post vocalization that overlap the auditory activity (Gunji et al., 2000). In our study, the early frontal activity (from 80 to 170 ms) may be actually related to the motor act of speaking (as in OP and CP) that requires interconnection among the frontal, temporal, and parietal lobes of the brain (Guenther, 2007). Further investigations are needed to improve the understanding of the activities that are recruited in the speech production tasks and, in this perspective, it will be necessary to take into account recent intracranial investigations finding no activation of the Broca's area during actual articulation (Flinker, Korzeniewska, Shestyuk, Franaszczuk, Dronkers, Knight & Crone, 2015).

The second finding is that information extracted by the early dynamics contains sufficient discriminative features for the vowel cortical classification. As to features, D'Zmura (2009) used matched filters, DaSalla et al. (2009) and Matsumoto and Hori (2014) applied the Common Spatial Patterns (CSP) methods that exalts more discriminative EEG channels, taking variance as the discriminating feature. We used the Ambiguity Function that gave us a large number of features (the values of the Ambiguity Function at each point of the ambiguity planes) then reducing the feature space dimensionality by FDR.

As for the choice of the classifier, we used a feed-forward ANN to recognize the features distinguishing the vowels /a/ and /i/ and showed that a 2-layer, 5-HNs feed-forward ANN can be successfully trained for the intrasubject recognition of overt, covert and imagined vowel production, in line with previous studies using different approaches. Other studies preferred the use of SVMs as classifiers (DaSalla et al., 2009; Matsumoto, Hori, 2014; Riaz et al., 2014) that has some advantages: e.g., the guaranty of finding the global minimum during training, but we preferred ANNs because they are naturally fit for problems with nonlinear decision hyperplanes, while SVMs require the selection of appropriate kernels and parameters. Surprisingly, we found that the pairwise comparison performed slightly better in the CP and IP tasks (i.e., 0.91 and 0.93 respectively) than in the OP task (i.e., 0.89), and the difference between OP and IP was judged as statistically significant by the t-test. This was tenta-

tively explained by the reduced presence of motor artifacts in the IP task where motor activity should be absent. It is likely that the CP and IP trials contain a large number of useful information for vowel cortical distinction since they are less affected by muscular activity as compared to OP signals. Furthermore, the location of the most discriminative electrodes of the scalp showed that the most informative sites are placed over both sensorimotor areas in CP, very close to the motor cortex (Riaz et al., 2014), and over posterior regions for IP task suggesting that the classification of these signals may be based on mostly on the imagined speech muscle movements as shown in other speech imagery studies (DaSalla et al., 2009). Further works are needed to provide additional validation to out hypothesis.

5. *Conclusions and Limitations*

The current work proposes a method that may be used for classifying speech sounds from brain signals and suggests the EEG technique as a pursuable and a necessary approach for developing frameworks for EEG-based SSIs systems. Other techniques result limiting in that perspective: functional magnetic resonance imaging (fMRI) has reduced temporal resolution, magnetoencephalography (MEG) is not sensitive to all the currents generated by the brain and ECoG requires the implantation of electrodes in the brain during neurosurgical operations.

However, in this study a series of limitations has to be taken into account, at least because finding highly significant results in such an experiment is new. For example, there is good reason to believe that the fixed order of vowel affected the accuracy performance as suggested by Porbadnigk and colleagues (2009) although, a more recent study has also revealed no significant difference between the imagined vocalization of vowels presented in fixed and in random order (Matsumoto, Hori, 2013). In future works, we are going to extend our probes on the activities involved during the motor preparation and to select small temporal windows (shorter than 300 ms) in order to provide a more fine-grained picture of the phenomenon under investigation.

Yet, data filtering needs to be much intensive: for example, building a composite system in which CSP is used as a preprocessing step before time-frequency analysis, may be a good solution; the reduction of the number of prominent electrodes as well as their physical significance for classification, remain subjects for future studies. To conclude, it will be also important to compare the performance of well-known procedures (e.g. SVM *vs*. ANN) and to test other efficient classification techniques moving beyond the pairwise classification of vowels. Currently, we are working to examine all these points in an EEG study with all Italian vowels.

Knowledgements

This work is supported by the "Programma Operativo Nazionale (PON) 254/Ric – Ricerca e competitività 2007/2013" of the Italian Ministry of Education, University,

and Research (upgrading of the "Centro ricerche per la salute dell'uomo e dell'ambiente" PONa3_00334)."

Bibliography

Atlas, L., Droppo, J. & McLaughlin, J. (1997). Optimizing time-frequency distributions via operator theory. In *Proceeding of SPIE*, 3162, 161-171.

Bouchard, K., Mesgarani, N., Johnson, K. & Chan, E. (2013). Functional organization of human sensorimotor cortex for speech articulation. In *Nature*, 495, 327-332.

Brigham, K., Kumar, B.V. (2010). Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In *Bioinformatics and Biomedical Engineering (iCBBE)*, 4th International Conference on. IEEE, 1-4.

Chia, X., Hagedorna, J.B., Schoonovera, D. & D'zmura, M. (2011). EEG-based discrimination of imagined speech phonemes. In *International Journal of Bioelectromagnetism*, 13, 201-206.

DaSalla, C.S., Kambara, H., Sato, M. & Koike, Y. (2009). Single-trial classification of vowel speech imagery using common spatial patterns. In *Neural Networks*, 22, 1334-1339.

Deecke, L., Engel, M., Lang, W. & Kornhuber, H.H. (1986). Bereitschaftspotential preceding speech after holding breath. In *Experimental Brain Research*, 65(1), 219-223.

Denby, J.B., Shultz, T., Honda, K., Hueber, T. & Gilbert, J.M. (2010). Silent speech interfaces. In *Speech Communication*, 52, 270-287.

D'Zmura, M., Deng, S., Lappas, T., Thorpe, S. & Srinivasan, R. (2009). Toward EEG sensing of imagined speech. In *International Conference on Human-Computer Interaction*. Springer: Berlin Heidelberg, 40-48.

Ebrahimi, T., Vesin, J. & Garcia, G. (2003). Brain-Computer Interface in Multimedia Communication. In *IEEE Signal Processing Magazine*, 20, 14-24.

Flinker, A., Korzeniewska, A., Shestyuk, A.Y., Franaszczuk, P.J., Dronkers, N.F., Knight, R.T. & Crone, N.E. (2015). Redefining the role of Broca's area in speech. In *Proceedings of the National Academy of Sciences*, 112(9), 2871-2875.

Garcia, G., Ebrahimi, T. & Vesin, J.M. (2002). Classification of EEG signals in the ambiguity domain for brain-computer interface applications. In *14th International Conference on Digital Signal Processing (DSP2002)*, 301-305.

Garcia, G., Ebrahimi, T. & Vesin, J. (2003). Joint Time-Frequency-Space Classification of EEG in a Brain Computer Interface Application. In *Eurasip Journal on Applied Signal Processing – Special issue on Neuromorphical Signal Processing*, 7, 713-729.

Gillespie, B., Atlas, L. (2001). Optimizing time-frequency kernels for classification. In *IEEE Transactions on Signal Processing*, 49, 485-496.

Guenther, F.H. (2007). Neuroimaging of normal speech production. In Ingham, R.J. (Ed.), *Neuroimaging in Communication Sciences and Disorders*. San Diego: Plural Publishing Inc., 1-51.

Guenther, F.H., Hampson, M. & Johnson, D.A. (1998). Theoretical investigation of reference frames for the planning of speech movements. In *Psychological review*, 105(4), 611.

Gunji, A., Hoshiyama, M. & Kakigi, R. (2000). Identification of auditory evoked potentials of one's own voice. In *Clinical Neurophysiology*, 111(2), 214-219.

Haykin, S.O. (2008). In Haykin, S.O. (Ed.). *Neural Networks and Learning Machines.* USA: Pearson.

Heinks-Maldonado, T.H., Nagarajan, S.S. & Houde, J.F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. In *Neuroreport*, 17(13), 1375. https://www.ncbi.nlm.nih.gov/pubmed/16932142.

Iqbal, S., Shanir, P.M., Khan, Y.U. & Farooq, O. (2016). Time Domain Analysis of EEG to Classify Imagined Speech. In *Proceedings of the Second International Conference on Computer and Communication Technologies*. Springer India, 793-800.

Kozek, W., Hlawatsch, F., Kirchauer, H. & Trautwein, U. (1994). Correlative time-frequency analysis and classification of nonstationary random processes. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 417-420.

Luo, H., Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. In *Frontiers in Psychology*, 3(1), 170. https://www.ncbi.nlm.nih.gov/pubmed/22666214.

Manca, A.D., Grimaldi, M. (2013). Perception and production of Italian vowels: an ERP study. In *Proceedings of INTERSPEECH*, 916-920.

Manca, A.D., Di Russo, F. & Grimaldi, M. (2015). Orderly organization of vowels in the auditory brain: the neuronal correlates of the Italian vowels. In Vayra, M., Avesani, C. & Tamborini, F. (Eds.), *Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio/Language acquisition and language loss. Acquisition, change and disorders of the language sound structure*. Milano: AISV, 357-368.

Matsumoto, M., Hori, J. (2013). Classification of silent speech using adaptive collection. In *Computational Intelligence in Rehabilitation and Assistive Technologies (CIRAT)*, 5-12.

Matsumoto, M., Hori, J. (2014). Classification of silent speech using support vector machine and relevance vector machine. In *Applied Soft Computing*, 20, 95-102.

McLaughlin, L., Droppo, J. & Atlas, L. (1997). Class-dependent time-frequency distributions via operator theory. In *Proceeding of ICASSP*, 3, 2045-2048.

Näätänen, R., Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. In *Psychophysiology*, 24(4), 375-425.

Obleser, J., Lahiri, A. & Eulitz, C. (2004). Magnetic Brain response mirrors extraction of phonological features from speakers vowels. In *Journal of Cognitive Neuroscience*, 16, 31-39.

Obleser, J., Scott, S.K. & Eulitz, C. (2006). Now you hear it, now you don't: Transient traces of consonants and their unintelligible analogues in the human brain. In *Cerebral Cortex*, 16, 1069-1076.

Pei, X., Barbour, D.L., Leuthardt, E.C. & Schalk, G. (2011). Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. In *Journal of neural engineering*, 8(4). https://www.ncbi.nlm.nih.gov/pubmed/21750369.

Porbadnigk, A., Wester, M. & Jan-p Calliess, T.S. (2009). EEG-based speech recognition impact of temporal effects. In *Biosignals 2009*, 376-381.

Riaz, A., Akhtar, S., Iftikhar, S., Khan, A.A. & Salman, A. (2014). Inter comparison of classification techniques for vowel speech imagery using EEG sensors. In *Systems and Informatics (ICSAI), 2014 2nd International Conference*, 712-717.

Santana, R. (2015). Supervised classification of vowel speech imagery. In *Actas de la XVI Conferencia CAEPIA*, Albacete, 951-961.

Scharinger, M., Idsardi, W.J. & Poe, S. (2011). A Comprehensive Three-dimensional Cortical Map of Vowel Space. In *Journal of Cognitive Neuroscience*, 23, 3972-3982.

Šťastný, J., Sovka, P. & Stančák, A. (2003). EEG signal classification: introduction to the problem. In *Radioengineering*, 12(3), 51-55.

Suppes, P., Lu, Z.L. & Han, B. (1997). Brain wave recognition of words. In *Proceedings of the National Academy of Sciences*, 94(26), 14965-14969.

Suppes, P., Han, B., Epelboim, J. & Lu, Z.L. (1999). Invariance between subjects of brain wave representations of language. In *Proceedings of the National Academy of Sciences*, 96(22), 12953-12958.

Tourville, J.A., Reilly, K.J. & Guenther, F.H. (2008). Neural mechanisms underlying auditory feedback control of speech. In *Neuroimage*, 39(3), 1429-1443.

Tian, X., Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. In *Frontiers in Psychology*, 1, 166.

Vos, D.M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.X., Huffel, V.S. & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. In *Neuroinformatics*, 8(2), 135-150.

Wang, R., Perreau-Guimaraes, M., Carvalhaes, C. & Suppes, P. (2012). Using phase to recognize English phonemes and their distinctive features in the brain. In *Proceedings of the National Academy of Sciences*, 109(50), 20685-20690.

PARTE III

QUESTIONI DI FONETICA ARTICOLATORIA, ACUSTICA E PERCETTIVA

DOI: 10.17469/O2102AISV000020

LORENZO CIAURELLI, ARAVIND NAMASIVAYAM, GRAZIANO TISATO,
PASCAL VAN LIESHOUT, CLAUDIO ZMARICH

Consonantal and vocalic gestures in the articulation of the Italian glides /j/ and /w/ at different syllable positions

From a phonological point of view, four glides (or approximants) exist in Italian: /j/, /w/, [i̯] and [u̯]. Glides still raise a lot of questions, from the definition of the necessary and sufficient features for their identification (Chitoran, Nevins, 2008), to their characterization at the acoustic and articulatory levels of speech production (Gick, 2003). In this paper, in order to describe the articulatory features of Italian glides, we analyzed the kinematics of both consonantal and vocalic gestures involved in the production of /j/ and /w/, by using 3D electromagnetic articulography (EMA; Carstens Medizinelektronik GmbH). The results show similar articulatory features for both glides in the way they differentiate themselves from corresponding vowels [i] and [u].

Key words: glides, kinematics, articulatory features, diphthong.

Introduction

Italian glides present a number of questions, ranging from the definition of their phonological nature (phonemic vs. allophonic status), to their phonetic and articulatory features (vocalic vs. consonantal, greater vs. lesser degree of constriction than corresponding vowels). In the phonological literature (Marotta, 1988; Nespor, 1993; Schmid, 1999; Bertinetto, Loporcaro, 2005) [j] and [w] are considered phonemes (/j/ and /w/) when placed in a word's initial position (onglides) preceding a vowel (V) as in "iodio" and "uomo". In contrast [i̯] and [u̯] are treated as positional, non-syllabic, allophones (offglides) of the corresponding vowels /i/ and /u/ as in "daino" and "auto". Onglides, also known as "semiconsonanti", are the non-nuclear elements of rising diphthongs, whereas offglides, also known as "semivocali", are the non-nuclear elements of falling diphthongs.

From an articulatory perspective, one can distinguish three different theoretical approaches which have tried to deal with the problem of glides characterization. The "Featural Hypothesis" and the "Structural Hypothesis" are the best known. Proponents of the first view (Nevins, Chitoran, 2008) assume that glides like /j/ and possibly /w/ are less vocalic than vowels like /i/ and /u/ because of their greater constriction degree. Proponents of the second view (Gick, 2003) assume that the glides /j/ and /w/ and their allophones can be characterized by the timing relationship between their gestural components, due to the position they occupy in the

word (or syllable), without having to specify their features directly. The third view (Maddieson, 2008), states that glides are characterized by the absence of a stable acoustic or articulatory target position, although they cannot be considered intrinsically transitional.

According to the "Structural Hypothesis", Browman and Goldstein (1992; 1995) have identified two gestural syllable position effects through which one can identify the properties of allophones in final or initial syllable positions: a) syllable position-specific timing between different tautosegmental[1] gestures (a property of gestural configuration); b) final reduction (a property of gestural scaling). Assuming that glides consist of two gestures (Sproat, Fujimura, 1993), namely a C-gesture (consonantal in nature) and a V-gesture (vocalic in nature), they can be distinguished by analyzing the behavior of their component gestures in different syllable positions. As to the English glides /j/ and /w/, empirical studies have shown that the C-gesture of initial allophones is greater in magnitude than the C-gesture in final allophones, and it temporally precedes the V-gesture, whereas in final allophones, C- and V-gestures are phased more closely together. Ambisyllabic allophones behave somewhat in between the characteristics found for initial and final allophones. In other words, final allophones are more vowel-like and initial allophones are more consonant-like (Gick, 2003).

In past years, Italian glides have mostly been studied by means of acoustic analysis, which does not always provide clear information on the actual gestural configurations of their production. Further, acoustic analysis was found to be unsuitable to identify some constituent differences between glides and vowels. For example, Salza, Marotta & Ricca (1987) showed that onglides can be distinguished from corresponding vowels by mean of acoustic duration (/j/ and /w/ are shorter than /i/ and /u/ respectively), whereas offglides ([i̯] and [u̯]) were similar in duration to unstressed vowels.

There is only one preliminary articulatory study on Italian glides using the Reading EPG system (Calamai, Bertinetto, 2006). The authors found that global tongue-palate contacts tends to be more extended in /i/ compared to /j/ and in /u/ compared to /w/. These results, albeit quite unexpected and at odds with findings from English glide productions (Nevins, Chitoran, 2008), are not unrealistic. Indeed, as Maddieson and Emmeroy (1985) have demonstrated, there is a wide cross-linguistic variability in the production of glides, due to underlying differences between glides and homorganic vowels. However, the EPG methodology used in the study of Calamai and Bertinetto (2006) is not particularly suited for studying articulatory behaviors in glides because EPG does not indicate which part of the tongue contacts the palate nor does it record lips movements, which is a constituent gesture of /w/ and /u/. Moreover, EPG cannot track the transition from syllable nucleus (in this case /a/) to glides. In contrast, 3D electro-magnetic articulography (EMA) is a more reliable instrument for this type of research (van Lieshout,

[1] In the Articulatory Phonology framework *tautosegmental gestures* refer to the internal organization of segment, that is to those overlapping gestures which characterize the segment.

Merrick & Goldstein, 2008) since it can track the movements of multiple articulators in a 3D space and with a higher temporal resolution (200 Hz against the 100 Hz of EPG).

In this paper we will try to shed light on the different hypotheses put forward in the previous works in this area. As stated by Featural Hypothesis, and at odds with Calamai and Bertinetto (2006), we expect to find a greater constriction degree for glides (Marotta, 1988; Bertinetto, Loporcaro, 2005). Moreover, by analyzing the steady-state and transitional portions of glides and vowels, we want to verify whether glides lack a stable target position (Maddieson, 2008). Finally, we want to study whether there is a cross-linguistic variability in the production of glides (Stone, Lundberg, 1996; Maddieson, Emmeroy, 1985).

In order to verify whether Italian glides are more vowel- or consonant-like we studied the movements of specific articulators used in the production of /j/ and /w/ by means of three parameters: a) extent of constriction degree; b) duration of steady-state portion of articulation; c) duration of transition from glide to syllable nucleus. We analyzed a wide set of articulators possibly involved in the production of glides, then, based on preliminary analysis, we discarded those that were found unsuitable for characterizing allophonic variations of glides (e.g. front-back movements of tongue). The articulators studied in this paper were tongue body for /j/ and tongue back with lips (upper & lower) for /w/, similar to previous work in this area (Gick, 2003).

In this paper, we will focus on a selection of Ciaurelli's (2015) data; a preliminary analysis of the production of one participant from the same dataset was presented in Zmarich, van Lieshout, Namasivayam, Limanni, Galatà & Tisato (2011).

1. *Methods and Materials*

1.1 *Participants*

The ten participants (8 females and 2 males, average age 32 years) involved in the experiment were all fluent speakers of Italian as their first language. They were all Italian students living in Toronto for a short period of time and they were recruited by flyers and word-of-mouth and paid for their participation. To avoid a bias due to the influence of regional Italian dialects, we were careful not to include people coming from Campania and Emilia-Romagna. This was done because there is a tendency for extreme diphthongization and for producing actuals diphthongs as hiatuses (['pje.de] vs ['pi.e.de]) in Campania and for spirantization of the /w/ in words like "auto" and "attuale" in Emilia-Romagna (Telmon, 1997).

In this paper, we only present data from 5 female participants (average age 27 years). The five participants (subj2, subj4, subj5, subj6 and subj8) were chosen for the completeness of their data (there was no problem during recording sessions) and the clarity of their pronunciation. We will only refer to words containing onglides, offglides and vowel targets as I-words and U-words. As we already presented data

for subj2 on U-words previously (Zmarich et al., 2011), we will not present these data here.

All participants were asked to complete a short questionnaire about general demographic data and to sign a consent form. All of them have normal vision and no history of hearing or speaking difficulties. The study was approved by the Health Sciences Research Ethics Board at the University of Toronto.

1.2 Instrumentation

We used an AG500 articulograph (Carstens Medizinelektronik, GmbH) setup at the Oral Dynamics Lab (ODL) in the Department of Speech-Language Pathology at the University of Toronto to record kinematic and acoustic data from the participants. The AG500 allows for 3D recordings of articulatory movements inside the vocal tract by tracking the movement of transducer coils placed on the articulators in the following manner: 2 coils on the vermillion borders of the upper and lower lip respectively, 1 on the tongue tip (1 cm behind the actual tip of the tongue), 1 on the tongue body (2 cm behind the tongue tip coil location), 1 on the tongue back (at least 1 cm behind tongue body location), 1 on the lower incisors of the lower jaw. We also recorded head motion by placing additional coils on subject's forehead, bridge of the nose and left and right skin covering the mastoid. This allowed us to afterwards correct the movement of articulators for head motion. Acoustic recordings were made with a 44 kHz sampling rate at 16 bits using a supplementary headset microphone connected to a solid-state audio recorder, which was synchronized with the kinematic signals. These are standard procedures developed at the ODL (Henriques, van Lieshout, 2013).

Two different sets of measures were obtained using the INTERFACE program (Tisato, Cosi, Drioli & Tesser, 2005). The movement patterns used for analyzing /i/, /j/ and [i̲] were:

– Tongue body vertical (TB_VERT, i.e. the position in high-low dimension of the coil on the tongue body).

For /u/, /w/ and [u̲] we used:

– Tongue back vertical (TBACK_VERT, i.e. the position in high-low dimension of the coil on the tongue back).
– Lip opening (LIP_OPEN, i.e. the vertical distance between the coils on the vermillion border of upper and lower lip).

For each of these component gestures associated with glides and vowels we calculated mean and standard deviation (SD) values for the following parameters:

– extent of constriction degree (i.e. the spatial value – in millimeters – of the position of the TB_VERT or TBACK_VERT or LIP_OPEN coils, represented in Figure 1 by green and red triangles).

– Duration of steady-state portion of articulation (i.e. the value of the temporal interval representing the difference between the two red or green arrows in Figure 1; these arrows were automatically detected by INTERFACE as temporal locations where the velocity of the movement under examination reached a threshold of 15% of the maximal velocity); in other words the arrows enclose only the portions of the articulatory trajectory characterized by a zero or very low velocity (i.e., the steady state).
– Duration of transition from glide to syllable nucleus (i.e. the value of the temporal interval representing the difference between the second arrow of a steady-state portion and the first arrow of the following steady-state portion); in other words, the arrows enclose only the portions of the articulatory trajectory characterized by a non-zero or low velocity.

Finally, after manually segmenting the speech signal (Salza, 1991), we made acoustic measurements of segment duration by using Praat software (Boersma, Weenink, 2009).

1.3 Stimuli

In order to elicit the production of the target glides and vowels by the participants we set up a series of short sentences. The target segment "I" and "U" (we will use the capital letters to refer to both glides and vowels) were added to vowels [e], [ɛ], [o] and [a] in order to produce hiatuses and diphthongs. All targets (hiatuses and diphthongs) were inserted in the carrier phrase "Ha detto X chiaramente" ("he said X clearly"), where X is the word containing the targets. Each sentence was repeated twice over one session, and there were 3 sessions using a normal, habitual rate and 3 other sessions using formal, slow rate, for a total of 408 sentences. All sequences were presented in random order to the participants on a computer screen using *Direct RT* (Jarvis, 2008), a stimulus presentation program. In order to obtain the stimuli at slow rate, we made participants listening to questions prompting for an answer that would put contrastive focus on the target word. The Table 1 shows the words contained in the carrier phrase.

Due to the small sample of participants statistical analysis was performed on each participant separately.

In order to better compare the kinematic values for I and U segments, we normalized the kinematics values for each articulator by subtracting the peak value achieved for each parameter for the vocalic targets /i/ and /u/ produced in isolation ("Ha detto i/u chiaramente") from the peak values achieved for the I and U targets in the carrier. Then, we performed a statistical analysis separately for I-words ("mia" and "maiale) and U-words ("tua", "attuale", "auto" and "baule").

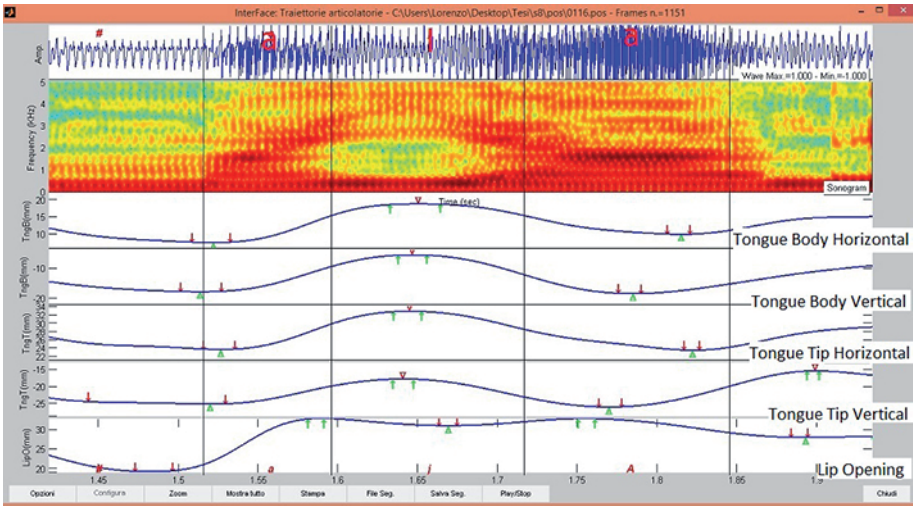
Table 1 - *Words contained in carrier phrase ("Ha detto X chiaramente") sorted by target segment, position and syllabic status*

Segments	Contextual vowels	Hiatus (V.V)	Onglide (GV)	Offglide (VG)	Hiatus (V.V)
u/w	a	tua /'tu.a/	attuale /at.'twa.le/	auto /'aw.to/	baule /ba.'u.le/
i/j	a	mia /'mi.a/	maiale /ma.'ja.le/	mai /'mai/	faina /fa.'i.na/

In order to study the differences between glides and homorganic vowels, we made the following comparisons: a) "mia" vs. "maiale"; b) "faina" vs. "mai"; c) "tua" vs. "attuale"; d) "baule" vs. "auto". In this way, we were able to analyze the transition from glides to vowel nucleus (and vice versa) as well.

Regrettably, we were unable to analyze the offglide-vowel opposition for I-words ("mai" vs "faina"), as the preliminary analyses showed that vertical movement of the tongue dorsum of [i̯] in "mai" was strongly influenced by the production of the velar stop /k/ which followed in the sentence carrier ("Ha detto mai chiaramente").

Figure 1 - *Screenshot of APmanager tool of INTERFACE showing the /aIa/ segment of the word "maiale". From top to bottom: 1) waveform; 2) Spectrogram; 3) Trajectories of selected articulators, from top to bottom: a) Tongue Body Horizontal; b) Tongue Body Vertical; c) Tongue Tip Horizontal; d) Tongue Tip Vertical; e) Lip Opening. The red and green triangles represent maximum and minimum positions in the trajectory of articulators respectively. The flanking red and green arrows locate the point where velocity of the movement reaches a threshold of 15% of maximal velocity (i.e. a relatively stable portion of a trajectory, referred to as a steady-state portion)*



2. *Results*

Table 2 shows the acoustic durations of glides and vowels. A Student's t-test was performed on acoustic durations, with I-words as factor. No statistical significance was found among I-words. A one-way ANOVA was also performed on acoustic durations, with U-words as factor. The difference among U-words was significant for all participants (subj4: f-ratio = 181.447, p-value < 0.001; subj5: f-ratio = 84.815, p-value < 0.001; subj6: f-ratio = 27.853, p-value < 0.001; subj8: f-ratio = 147.716, p-value < 0.001). A Bonferroni post-hoc pairwise analysis showed that for all subjects the acoustic duration for /u/ in "tua" was significantly greater than the duration for /w/ glide in "attuale" (subj4: p-value < 0.001; subj5: p-value < 0.001; subj6: p-value = 0.002; subj8: p-value < 0.001). We also found that the duration of /u/ in "baule" was significantly longer than [u̯] in "auto" for all subjects (subj4: p-value < 0.001; subj5: p-value < 0.001; subj6: p-value < 0.001; subj8: p-value < 0.001).

Table 2 - *Range, mean (s) and standard deviation of acoustic durations of I- and U- targets*

Subject	*Word*	*Target*	*Range*	*Mean*	*Std*
SUBJ_2	mia	/i/	0.097	0.162	0.038
SUBJ_2	maiale	/j/	0.093	0.160	0.034
SUBJ_4	mia	/i/	0.120	0.145	0.033
SUBJ_4	maiale	/j/	0.064	0.122	0.019
SUBJ_4	tua	/u/	0.065	0.131	0.016
SUBJ_4	attuale	/w/	0.038	0.084	0.013
SUBJ_4	baule	/u/	0.045	0.185	0.012
SUBJ_4	auto	[u̯]	0.029	0.073	0.010
SUBJ_5	mia	/i/	0.112	0.124	0.034
SUBJ_5	maiale	/j/	0.054	0.133	0.018
SUBJ_5	tua	/u/	0.086	0.137	0.026
SUBJ_5	attuale	/w/	0.044	0.046	0.012
SUBJ_5	baule	/u/	0.069	0.161	0.021
SUBJ_5	auto	[u̯]	0.060	0.077	0.018
SUBJ_6	mia	/i/	0.169	0.175	0.069
SUBJ_6	maiale	/j/	0.147	0.150	0.046
SUBJ_6	tua	/u/	0.128	0.150	0.047
SUBJ_6	attuale	/w/	0.097	0.088	0.034
SUBJ_6	baule	/u/	0.116	0.220	0.041
SUBJ_6	auto	[u̯]	0.090	0.106	0.028
SUBJ_8	mia	/i/	0.042	0.141	0.010
SUBJ_8	maiale	/j/	0.037	0.139	0.012
SUBJ_8	tua	/u/	0.068	0.149	0.019
SUBJ_8	attuale	/w/	0.029	0.076	0.009
SUBJ_8	baule	/u/	0.069	0.191	0.024
SUBJ_8	auto	[u̯]	0.030	0.074	0.009

Figure 2 shows the /i/-normalized Tongue Body Vertical (TB_VERT) values for the I-targets in the I-words. A Student's t-test was performed on TB_VERT, with I-words as factor. The normalized value for /i/ in "mia" was significantly greater than the value for /j/ in "maiale" for three subjects (subj4: t = -3.729, df = 21.192, p-value = 0.001; subj6: t = -3.420, df = 17.830, p-value = 0.003; subj8: t = -2.630, df = 18.318, p-value = 0.017).

Figure 2 - */i/-normalized Tongue Body Vertical (TB_VERT) values (mm) for I-words "maiale" and "mia"*

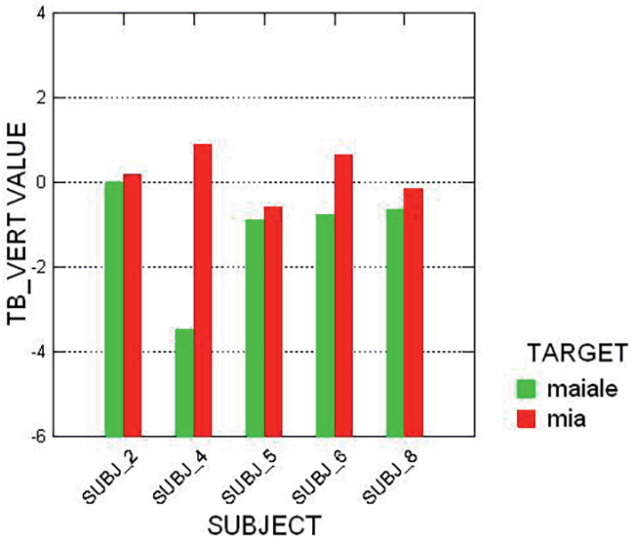


Figure 3 shows the /u/-normalized Tongue Back Vertical (TBACK_VERT) values for the U-targets in "attuale" and "tua". A one-way ANOVA was performed on TBACK_VERT with U-words as a factor. The difference for the U-words was significant for all participants (subj4: f-ratio = 30.680, p-value < 0.001; subj5: f-ratio = 29.085, p-value < 0.001; subj6: f-ratio = 28.808, p-value < 0.001; subj8: f-ratio = 7.313, p-value = 0.001). A Bonferroni post-hoc pairwise analysis showed that for 3 subjects the normalized value for /u/ in "tua" was significantly greater than the value for /w/ glide in "attuale" (subj4: p-value = 0.005; subj6: p-value = 0.001; subj8: p-value = 0.003). We also found that as for subj4 the normalized value for /u/ in "baule" was greater than the value for the [u̯] glide in "auto" (p-vaue < 0.001), whereas for subj6 (p-value < 0.001) that value was significantly greater for the[u̯] glide than for /u/.

A one-way ANOVA was performed on LIP_OPEN with U-words as a factor, but no significant differences were found.

We also studied the durations of the steady-state intervals of articulatory movements, that is the temporal interval where the velocity of articulator movement is lower than 15% of the maximal velocity.

Figure 4 shows the durations of TB_VERT steady-state for I-targets in "mia" and "maiale". A Student's t-test was performed on TB_VERT with I-words as factor. All

participants produced a significantly longer steady-state for /i/ in "mia" than for /j/ in "maiale" (subj2: t = -3.946, df = 13.251, p-value = 0.002; subj4: t = -4.181, df = 11.074, p-value 0.002; subj5: t = -2.940, df = 11.065, p -value = 0.013; subj6: t = -3.251, df = 10.231, p-value = 0.008; subj8: t = -2.630, df = 12.823, p-value < 0.001).

Figure 3 - */u/-normalized Tongue Back Vertical (TBACK_VERT) values (mm) for U-words "attuale" and "tua"*

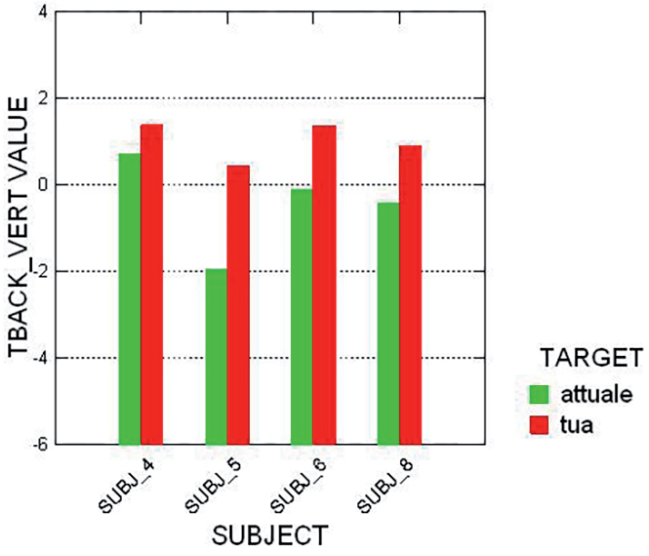


Figure 4 - *steady-state duration (ms) for TB_VERT for I-targets in "mia" e "maiale"*

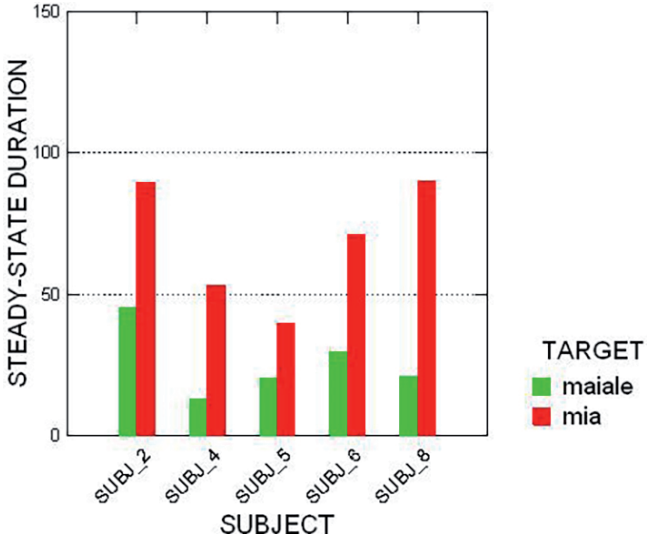
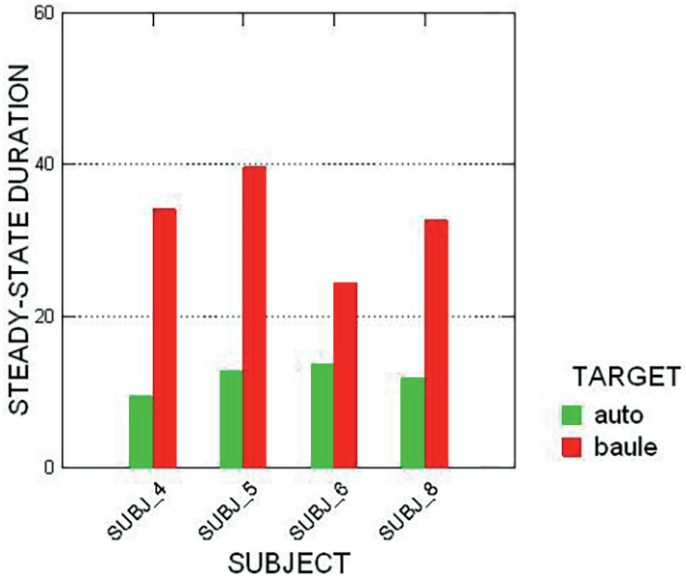


Figure 5 shows the durations of TBACK_VERT steady-state for U-targets in "baule" and "auto". A one-way ANOVA was performed on TBACK_VERT with U-words as a factor. The difference in U-words was significant for all participants (subj4: f-ratio = 17.363, p-value < 0.001; subj5: f-ratio = 54.434, p-value < 0.001; subj6: f-ratio = 20.874, p-value < 0.001; subj8: f-ratio = 10.573, p-value < 0.001). A Bonferroni post-hoc pairwise analysis showed that for all participants the steady-state portion for /u/ in "baule" was significantly longer than for [u̯] in "auto" (subj4 and subj8: p-value = 0.001; subj5 and subj6: p-value < 0.001). As for the onglide-vowel contrast, the steady-state portion for /u/ in "tua" was significantly longer than for [u̯] in "attuale" only for subj8 (p-value = 0.006).

Further a one-way ANOVA was performed on LIP_OPEN with U-words as a factor, but no significant differences were found.

Figure 5 - *steady-state duration (ms) for TBACK_VERT for U-targets in "auto" e "baule"*

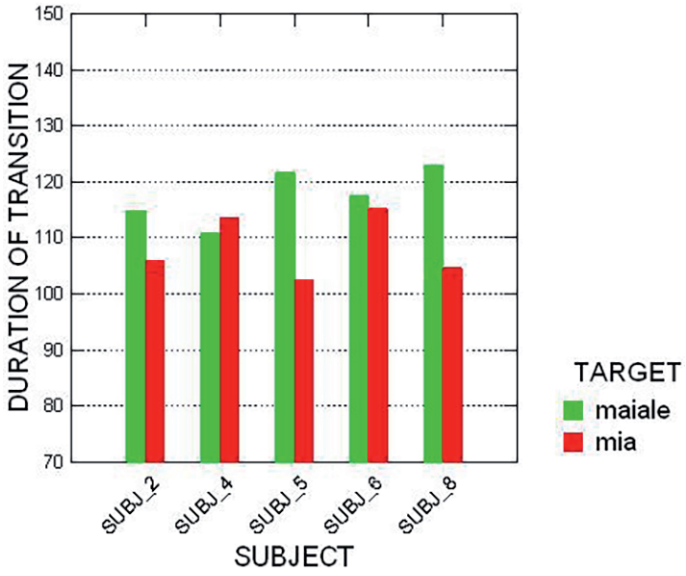


We analyzed the duration of the transitions from vowels to glides (or vice versa) as the distance between two contiguous steady-states (that is from I to /a/ and for U to the following and preceding /a/; see Figure 1).

Figure 6 shows the values of duration of transitions for I-words. A Student's t-test was performed on TB_VERT with I-words as factor. For 3 out of 5 subjects the transition from /j/ to /a/ in "maiale" was significantly longer than the transition from /i/ to /a/ in "mia" (subj2: t = 2.079, df = 21.877, p-value = 0.05; subj5: t = 2.563, df = 21.823, p-value = 0.018; subj8: t = 5.418, df = 14.768, p-value < 0.001). A one-way ANOVA was performed on TBACK_VERT and LIP_OPEN with U-words as factor. The difference among U-words was significant for just one subject and only for TBACK_VERT (subj4: f-ratio = 72.440, p-value < 0.001). A

Bonferroni post-hoc pairwise analysis showed that for this subject the transition from /w/ to /a/ in "attuale" was longer than the transition from /u/ to /a/ in "tua".

Figure 6 - *duration of transition (ms) from I-target to /a/ for TB_VERT in I-words*



3. *Discussion and Conclusion*

The aim of this study was to gain a better understanding of the nature of Italian glides and how they contrast with homorganic vowels. Specifically, we wanted to verify, by means of a kinematic analysis, whether glides are more vowel- or consonant-like and whether there is any difference between I and U glides with respect to their component gestures (in this case, only for U-targets, because they are constituted by both tongue and lips gestures). To this end, we analyzed the behavior of specific articulators involved in glides and vowel productions using 3 parameters: a) constriction degree; b) duration of steady-state portion; c) duration of transition from glides to syllable nucleus. The main findings show that for both onglides, the vertical position of the tongue (tongue body for I, and tongue back for U) distinguishes onglides from the corresponding vowels, in the sense that both vowels /i/ and /u/ show a greater degree of constriction than /j/ and /w/ onglides respectively. These results are at odds with data from English (Ladefoged, Maddieson, 1996; Stone, Lundberg, 1996;) but comparable with the findings of Calamai and Bertinetto (2006) in Italian speakers who showed smaller tongue-palate contact for the onglides when compared to the homorganic vowels.

We also found that the offglide [u̯] is distinguished from its corresponding vowel predominantly by a difference in the duration of the steady-state portion. In fact, all participants produced the /u/ segment with longer duration of the articulatory

steady-state portion than the [u̯] segment, whereas only two out of four participants distinguished /u/ from [u̯] by showing a different value for constriction degree (for one of the subjects, constriction degree was greater in the vowel than in the glide, whereas for the other subject constriction degree was greater in the glide than in the vowel).

The duration of transition was found to be longer for /j/ than for /i/ for three subjects out of five and for /w/ than for /u/ for just one subject. No significant difference was found with regard to offglide-vowel opposition ([u̯] vs /u/).

Finally, no significant results were found for lip gesture movements (LIP_OPEN) involved in the production of U-targets.

In trying to interpret these results, we have to take into account some limitations of the current experimental design. The greater constriction degree found for /i/ and /u/ vowels with respect to /j/ and /w/ glides could be due to the fact that full vowels analyzed here carried lexical stress. However, one could also interpret this result as a reflection of the assumed hypo-articulated nature of glides. Furthermore, as Maddieson and Emmeroy (1985) have demonstrated, there is a wide variability in production of glides across the languages, which could perhaps account for the difference found between Italian and English glides.

Although these limitations prevent us from making some general statements about the difference between I and U offglides, the behaviors of /w/ and [u̯] show a clear differentiation in the way they contrast with homorganic vowels. In fact, the onglide-vowel contrast is triggered by a difference in constriction degree, whereas the offglide-vowel contrast is based on a difference in steady-state duration. Following Salza, Marotta & Ricca (1987) and Marotta (1988), one could consider the [u̯] offglide as a non-stressed vowel as the only significant difference with respect to the /u/ vowel is in duration, both articulatory (steady-state) and acoustically. This finding, together with the finding on transition duration between glides and vowels, might suggest that glides are not inherently transitional (Maddieson, 2008), and it could depend on underlying differences between glides and homorganic vowels (Maddieson, Emmeroy, 1985).

Although the comparison between acoustic and articulatory analyses is beyond the aims of this study, the results from acoustic analysis seem to confirm the previous statement. In fact, although both onglides (/j/ and /w/) can be differentiated from the homorganic vowels (/i/ and /u/) by means of articulatory steady-state duration, only /w/ can be differentiated from the homorganic vowel by means of overall acoustic duration as well. So it could be accounted for by a difference in shift from vowel to glide for /i/ and /u/.

Finally, the fact that no significant results were found for lip opening in differentiating /w/ and [u̯] from vowels seems to suggest that for U-glides, the primary articulator is the tongue (i.e. the tongue back movement in high-low dimension). However, a preliminary analysis on horizontal lip movements (not included in this paper) seems to reveal an important role of lip protrusion in U-glides production that needs further exploration.

4. *Future Perspectives*

In this paper we have presented only a selection of the data that were collected originally. A next step will be to investigate further claims made by Gick (2003). To this end, we will analyze the onset of glides' constituent gestures in order to verify the hypothesis stated by Browman and Goldstein (1992; 1995) about gestural syllable-position effects.

Further analyses would be necessary to determine the syllabic role of Italian glides. Following Hsieh and Goldstein (2015) one could analyze the temporal behaviour of gestures in glides to determine the gestural organization of complex onset and complex coda sequences considering that:

- onset consonants are hypothesized to be coupled in-phase to the following vowel and anti-phase to each other (Fowler, 2015), whereas;
- coda consonants are hypothesized to be sequential, with the first coda consonant coupled in anti-phase mode to the preceding vowel and following consonants coupled anti-phase to preceding consonant.

Given these assumptions, we will be able to directly compare onglides and offglides and to gain some insight into their gestural organization.

Bibliography

Bertinetto, P.M., Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. In *Journal of the International Phonetic Association*, 35(2), 131-151.

Boersma, P., Weenink, D. (2009). Praat: Doing phonetics by computer. Version 5.1.20 [Computer program]. Retrieved 31.10.09.

Browman, C.P., Goldstein, L. (1992). Articulatory phonology: an overview. In *Phonetica*, 49(3-4), 155-180.

Browman, C.P., Goldstein, L. (1995). Gestural syllable position effects in American English. In Bell-Berti, F., Raphael L.J. (Eds.), *Producing Speech: Contemporary Issues*. For Katherine Safford Harris. New York: AIP Press.

Calamai, S., Bertinetto, P.M. (2006). Per uno studio articolatorio dei legamenti palatale, labio-velare e labio-palatale dell'italiano. In Giordani, V., Bruseghini, V. & Cosi, P. (Eds.), *Atti del III Convegno Nazionale*, Trento, 29-30 novembre - 1 dicembre 2006. Torriana (RN): EDK Editore, 43-56.

Chitoran, I., Nevins, A. (2008). Introduction. In *Lingua*, 118, 1900-1905.

Ciaurelli, L. (2015). Gesti consonantici e vocalici nell'articolazione dei glide /j/ e /w/ dell'italiano in posizioni sillabiche diverse. Tesi di Laurea Magistrale in Linguistica, Università degli Studi di Roma "La Sapienza", Roma.

Fowler, C.A. (2015). The segment in Articulatory Phonology. In Raimy, E., Cairns, C.E. (Eds.), *The segment in phonetics and phonology*. Oxford: Blackwell, 25-43.

Gick, B. (2003). Articulatory correlates of ambisillabicity in English glides and liquids. In Local, J., Ogden, R. & Temple, R. (Eds.), *Phonetic interpretation. Papers in Laboratory Phonology VI.* Cambridge (UK): Cambridge University Press, 222-236.

Henriques, R.N., van Lieshout, P.H.H.M. (2013). A comparison of methods for decoupling tongue and lower lip from jaw movements in 3D articulography. In *Journal of Speech, Language and Hearing Research*, 56, 1503-1516.

Hsieh, F.Y., Goldstein, L. (2015). Temporal organization of off-glides in American English. In *Proceedings of International Congress of Phonetic Sciences (ICPhS 2015)*, Glasgow, UK, 2015.

Jarvis, B.G. (2008). DirectRT. Version 2008.1.0.13 [Computer Software]. New York, NY: Empirisoft Corporation.

Ladefoged, P., Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford, UK: Blackwell Publishers.

Maddieson, I. (2008). Glides and gemination. In *Lingua*, 118, 1926-1936

Maddieson, I., Emmorey, K. (1985). Relationship between semivowels and vowels: Cross-Linguistics investigations of acoustic difference and coarticulation. In *Phonetica*, 42, 163-174.

Marotta, G. (1988). The Italian dipthongs and the autosegmental framework. In Bertinetto, P.M., Loporcaro, M. (Eds.), *CertamenPhonologicum, I°*. Torino: Rosenberg & Sellier, 389-420.

Nespor, M. (1993). *Fonologia*. Bologna: Il Mulino.

Nevins, A., Chitoran, I. (2008). Phonological representations and the variable patterning of glides. In *Lingua*, 118, 1979-1997.

Salza, P.L. (1991). La problematica della segmentazione del segnale vocale. In Magno Caldognetto, E., Ferrero, F. (Eds.), *Trattamento del segnale vocale ed elaborazione statistica dei dati. Atti delle I Giornate di Studio del GFS*, Padova, 3-6 novembre 1990, 23-48.

Salza, P.L., Marotta, G. & Ricca, D. (1987). Duration and formant frequencies of Italian bivocal sequences. In *Proceedings of the Eleventh international congress of phonetic sciences (ICPhS)*, 1-7 August 1987, Tallinn, Estonia. Tallinn: Academy of Sciences of the Estonian SSR, 6 voll., vol. 3, 113-116.

Schmid, S. (1999). *Fonetica e Fonologia dell'Italiano*. Torino: Paravia.

Sproat, R., Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. In *Journal of Phonetics*, 21, 291-311.

Stone, M., Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. In *Journal of the Acoustical Society of America*, 99(6), 3728-3737.

Telmon, T. (1997). Varietà regionali. In Sobrero, A.A. (Ed.), *Introduzione all'italiano contemporaneo*, vol. II. *La variazione e gli usi*. Bari: Laterza, 93-149.

Tisato, G., Cosi, P., Drioli, C. & Tesser, F. (2005). InterFace: New Tool for Building Emotive/Expressive Talking Heads. In *Proceedings of INTERSPEECH 2005*, Lisbon, Portugal, 2005, 781-784. http://www2.pd.istc.cnr.it/INTERFACE.

van Lieshout, P.H.H.M., Merrick, G. & Goldstein, L. (2008). An articulatory phonology perspective on rhotic articulation problems: A descriptive case study. In *Asia Pacific Journal of Speech, Language, and Hearing*, 11(4), 283-303.

Zmarich, C., van Lieshout, P., Namasivayam, A., Limanni, A., Galatà, V. & Tisato, G. (2011). Consonantal and vocalic gestures in the articulation of Italian glide /w/ at different syllable positions. In Gili Fivela, B., Stella, A., Garrapa, L. & Grimaldi, M. (Eds.), *Contesto comunicativo e variabilità nella produzione e percezione della lingua, Atti del 7° Convegno Nazionale dell'Associazione Italiana Scienze della Voce*, Lecce, 26-28 gennaio 2011. Roma: Bulzoni editore, 9-24.

DOI: 10.17469/O2102AISV000021

ROSALBA NODARI

Occlusive sorde aspirate e modalità di fonazione: prime ricognizioni acustiche

This paper investigates the phonation of the vowel /a/ when preceded by a voiceless aspirated stop in the Regional Italian from central Calabria. The results showed that variation in Voice Onset Time duration has a strong correlation with the phonation of the following vowel. The findings showed that, after a voiceless aspirated stop, the onset of the vowel is realized with higher values of Open Quotient (OQ) and Spectral Tilt, suggesting that the phonation type is different from a vowel after a voiceless short-lag VOT stop.

Key words: phonation, aspiration, VOT, coarticulation, Calabrian Italian.

Introduzione

Ciò che ci si propone nel seguente lavoro è fornire dei primi dati di natura acustica relativi alla modalità di fonazione delle vocali seguenti una occlusiva sorda aspirata. Numerosi studi, condotti su diverse lingue, hanno infatti dimostrato come sia possibile trovare delle vocali articolate con fonazione non modale qualora esse siano precedute da una occlusiva aspirata. Un'analisi accurata del fenomeno per l'ambito italo-romanzo è ancora mancante. Gli studi a disposizione sull'aspirazione delle occlusive sorde presenti nei dialetti italiani si concentrano infatti principalmente sul parametro del VOT: esso spesso viene considerato l'unico fattore utile a discriminare il grado di aspirazione delle occlusive, per quanto i lavori condotti su altre lingue abbiano dimostrato come esso non sia l'unico. In questo studio ci si occuperà in particolare di sillabe con occlusiva aspirata nell'italiano regionale calabrese, a partire da un corpus di dati raccolto per la varietà parlata da giovani di Lamezia Terme (Nodari, 2016a).

Il lavoro è strutturato nel seguente modo: il par. 1 sarà dedicato al problema della realizzazione delle occlusive sorde come aspirate in Calabria: si tratterà del problema dell'aspirazione così come riportato negli studi di impronta dialettologia e classificatoria, e si passeranno in rassegna gli studi di natura acustica dedicati al fenomeno; nel par. 2 si passeranno in rassegna gli studi che si sono dedicati alla modalità di fonazione delle vocali seguenti una occlusiva sorda aspirata; nel par. 3 si analizzeranno gli indici selezionati per indagare la presenza di una eventuale modalità di fonazione sospirata, e cioè l'indice relativo al quoziente di apertura (H1-H2) e un indice di *Spectral Tilt* (H1-A3) relativo alla repentinità del movimento di chiusura delle pliche vocali; nel par. 4 si descriverà il corpus sul quale è stata condotta l'analisi, mentre il par. 5 sarà dedicato all'esposizione dei risultati

relativi al VOT e alla modalità di fonazione; il par. 6 sarà dedicato alla discussione dei risultati, i quali confermano che anche nella varietà calabrese di italiano, la vocale /a/ seguente una occlusiva sorda aspirata è realizzata con modalità di fonazione sospirata, mostrando valori più alti di H1-H1 e H1-A3; infine, nel par. 7 si tireranno le conclusioni e si accennerà a eventuali linee di ricerca future.

1. *L'aspirazione delle occlusive sorde nei dialetti e nell'italiano della Calabria*

Diversi studi hanno attestato la presenza di occlusive sorde aspirate in alcuni dialetti meridionali e nei corrispettivi italiani regionali, soprattutto nella penisola salentina e in Calabria (Canepari, 1986; Telmon, 1993; Fanciullo, Librandi, 2002), dove l'aspirazione è di natura allofonica. Per quanto riguarda la dialettologia, i primi accenni al fenomeno sono già in Rohlfs (1966: 277), il quale notava che "in alcuni dialetti della provincia di Cosenza [Spezzano Grande, Aprigliano, San Giovanni in Fiore] si osserva una chiara aspirazione della -t- [...], fenomeno non limitato però alla posizione intervocalica (per esempio anche fatthu, vienthu 'vento')".

Altri lavori di ambito dialettologico (Falcone, 1976) e di ambito fonetico (Sorianello, 1996; Stevens, Hajek, 2010; Nodari, 2016a) confermano come, almeno in alcune aree della Calabria (sono state indagate le province di Cosenza, Catanzaro, Reggio Calabria) sia aspirata l'intera serie delle occlusive sorde in particolari contesti fonotattici. Secondo Falcone (1976: 42), a Reggio Calabria l'aspirazione riguarderebbe l'intera classe delle occlusive sorde (bilabiali, dentali, palatali e alveolari) se precedute da una nasale, da una vibrante o se geminate.

Menzioni all'aspirazione compaiono anche nelle rassegne sull'italiano regionale calabrese. Gli autori sono concordi nel riportare il fenomeno come presente nella regione, ma non sono tutti concordi in merito alle consonanti implicate e ai contesti. Per Fanciullo, Librandi (2002) e Telmon (1993) sono aspirate le occlusive sorde se geminate o precedute da nasale e vibrante, mentre per De Blasi (2014) il tratto di "maggiore evidenza e di più netta delimitazione regionale è l'aspirazione dell'occlusiva sorda dentale intensa (-tt-) o collocata dopo –r-, -n-, che evidentemente ne provocano una pronuncia rafforzata".

I dati di Sorianello (1996) riguardanti l'aspirazione delle occlusive sorde in dialetto cosentino e nel corrispettivo italiano regionale[1] mostrano, come nelle due varietà indagate la durata del VOT, in generale, sia minore nelle occlusive sorde scempie, maggiore nelle postvibranti e ancor più nelle geminate, sia per l'italiano

[1] L'autrice prende in considerazione un corpus di 44 frasi lette per tre volte da tre parlanti cosentini (due uomini e una donna) sia in italiano regionale, sia nella varietà dialettale cosentina (per la lettura in dialetto veniva chiesto ai soggetti di tradurre la lista di frasi fornita in italiano). Le frasi del corpus contenevano parole di uso comune nel quale comparissero le occlusive sorde /p t k/ e [c] in contesto di geminazione, quando precedute da suono rotico e quando intervocaliche scempie (quest'ultimo contesto è stato inserito come contesto di controllo).

regionale sia per il dialetto. Sia in dialetto sia in italiano regionale cosentino l'autrice conferma quindi la presenza dell'aspirazione per l'intera classe delle occlusive sorde nei contesti di geminazione e di postvibrante.

Stevens, Hajek (2010) conducono analisi sulla durata delle occlusive sorde /p t k/ così come realizzate in diverse varietà regionali di italiano in contesto di geminazione[2]: gli autori dimostrano che a Catanzaro le occlusive sorde sono effettivamente realizzate con un VOT più lungo, e soprattutto i valori del VOT dell'alveolare sono significativamente più lunghi rispetto a quelli delle altre città italiane, avvicinandosi alle durate che caratterizzano la velare.

Per l'area cosentina l'osservazione di Rohlfs (1966) in merito all'aspirazione dell'alveolare è parzialmente smentita da Mele (2009: 83) in uno studio dedicato alla fonetica e alla fonologia del dialetto di San Giovanni in Fiore, in provincia di Cosenza. Mele nota come proprio a San Giovanni in Fiore è sì aspirata l'occlusiva sorda alveolare, ma solo quando essa è geminata o preceduta da nasale o vibrante. La regola non è limitata alla sola alveolare ma riguarda in genere le occlusive /p t k/, che sono realizzate come aspirate quando geminate o precedute dalle sonoranti /r/ o /N/ (indicando l'autore con N le nasali [n ŋ m]), sia all'interno di parola sia in contesto di frase. L'aspirazione non si attiva invece quando l'occlusiva ricorre in contesto prepausale, a causa della desonorizzazione della vocale finale (es. [ˈtsap:ḁ]).

Da ultimo, Nodari (2016a) riporta come nell'italiano regionale lametino sono aspirate le occlusive sorde /p t k/ quando geminate, precedute da suono rotico, nasale o laterale. Si può così generalizzare affermando che nell'italiano regionale calabrese l'aspirazione delle occlusive sorde riguarda l'intera serie delle occlusive sorde quando geminate o precedute da sonorante[3].

[2] Gli autori si avvalgono dei materiali disponibili nel corpus CLIPS (Corpora e Lessici di Italiano Parlato e Scritto), che raccoglie 100 ore di parlato, equamente ripartito tra voci maschili e femminili, e strutturato in modo da rappresentare la stratificazione sia diatopica, sia diastratica. Viene analizzata la produzione di otto parole contenenti uno o più segmenti /p: t: k:/ e lette in isolamento da locutori provenienti da 15 città italiane del nord (Bergamo, Genova, Milano, Parma, Torino, Venezia), del centro (Firenze, Perugia, Roma) e del sud (Bari, Cagliari, Catanzaro, Lecce, Napoli, Palermo).

[3] Come suggerito da un revisore, si potrebbe offrire un'interpretazione fonologica dell'aspirazione e considerarla come un processo di fortizione che riguarda la consonante in posizione di *onset* sillabico. Già Sorianello (1996) fa però notare che un'interpretazione fonologica dell'aspirazione è più complessa e non può limitarsi alla sola osservazione di un processo di rafforzamento. L'aspirazione è infatti sensibilmente più lunga nelle sillabe atone, alla fine di un gruppo di respiro. Secondo Sorianello (1996: 147) questa osservazione farebbe propendere per un fenomeno non tanto di fortizione, quanto di indebolimento: l'autrice conclude rimarcando perciò la complessità del fenomeno dell'aspirazione, il quale è caratterizzato tanto da processi di fortizione (assenza di aspirazione nel contesto intervocalico) quanto da processi di indebolimento (attivazione del processo in posizione prepausale, maggior coinvolgimento delle coronali).

2. *Occlusive sorde aspirate e modalità di fonazione della vocale seguente*

La speciale coordinazione temporale richiesta nell'adduzione delle pliche vocali per produrre una consonante aspirata può interferire con la modalità di fonazione della vocale seguente. Secondo Löfqvist (1980) e Goldstein, Browman (1986), la differenza principale che intercorre tra occlusive aspirate e non aspirate è da riscontrarsi, infatti, nel diverso coordinamento temporale dei gesti laringei relativi alla chiusura orale. Come riportato in Gobl, Ní Chasaide (1999: 123), nell'articolazione di una occlusiva sorda l'abduzione delle pliche vocali avviene quasi simultaneamente al gesto della chiusura orale. Nell'articolazione di un'occlusiva sorda non aspirata, il momento di massima apertura glottidale viene raggiunto all'incirca a metà della fase di chiusura, e ciò permette quindi un movimento adduttore adeguato per l'iniziazione della vibrazione delle pliche vocali subito dopo il rilascio dell'occlusione. Nelle occlusive sorde aspirate, invece, il gesto di abduzione delle pliche si prolunga durante la fase di chiusura e, di conseguenza, il momento di massima apertura glottidale si raggiunge all'incirca durante la fase di rilascio dell'occlusione, causando quindi un ritardo nel raggiungimento della posizione adeguata per la vibrazione delle pliche vocali. In Stevens (2000) si nota come nei primi cicli glottici di una vocale preceduta da un'occlusiva sorda aspirata le armoniche superiori alla prima hanno un'ampiezza ridotta, dando luogo ad una vocale caratterizzata da una modalità di fonazione sospirata (*breathy voice*).

Gobl, Ní Chasaide (1999) conducono una dettagliata analisi acustica delle vocali precedute da /p b/ nella produzione di parlanti di diverse lingue: in inglese, svedese e tedesco si realizza un contrasto tra occlusiva sonora /b/ e occlusiva sorda aspirata [p^h], mentre in francese e italiano si realizza un contrasto tra occlusiva sonora /b/ e occlusiva sorda /p/. I risultati mostrano che in italiano e francese non ci sono effetti di coarticolazione progressiva, né per l'occlusiva sorda né per la sonora: in entrambi i casi, le vocali seguenti raggiungono infatti valori di fonazione modale subito dopo il rilascio della consonante. Il tedesco invece mostra delle differenze: in particolare, i primi cicli glottici della vocale che segue l'occlusiva sorda aspirata [p^h] mostrano una modalità di fonazione di tipo sospirato mentre i cicli successivi raggiungono valori di fonazione modale simili a quelli di una vocale che segue l'occlusiva sonora [b]. Lo stesso effetto non è stato invece trovato in inglese e svedese, dove gli attacchi vocalici non si differenziano in base alla presenza o assenza di aspirazione nell'occlusiva precedente. Inoltre, in inglese e svedese, anche dopo la [p^h] si può avere un'iniziazione immediata della vibrazione delle pliche secondo una fonazione tipicamente modale. Gli autori ipotizzano quindi che anche la tensione delle pliche vocali possa giocare un ruolo nel differenziare i due tipi di produzione.

Effetti simili a quelli riscontrati per il tedesco sono stati riscontrati da altri autori anche in altre lingue; così, ad esempio, in georgiano il tipo di occlusiva influenza la modalità di fonazione della vocale seguente: le vocali seguenti un'eiettiva possiedono un attacco caratterizzato da fonazione cricchiata (*creaky voice*),

mentre le vocali seguenti una occlusiva sorda aspirata sono articolate con modalità di fonazione sospirata (Vicenik, 2010: 80)

3. *Gli indici acustici per la misurazione della fonazione*

In questo studio sull'italiano nella varietà calabrese di Lamezia Terme, si è misurato in primo luogo la durata del VOT, per verificare la realizzazione del fono occlusivo come aspirato; in seconda battuta, per verificare la modalità di fonazione della vocale seguente si sono analizzati il quoziente di apertura (o *Open Quotient*, d'ora in poi OQ) e lo *Spectral Tilt*.

L'OQ fa riferimento alla fase del ciclo glottico nella quale le corde vocali vanno da un massimo di chiusura a un massimo di apertura: esso può essere calcolato a partire dallo spettrogramma a banda larga, sottraendo l'ampiezza della seconda armonica all'ampiezza della prima armonica (H1-H2) (Klatt, Klatt, 1990). Secondo Hanson et al. (2001) un maggiore quoziente di apertura fa sì che la forma d'onda glottidale sia più vicina a una sinusoide, per cui nel dominio delle frequenze l'ampiezza della prima armonica aumenta in relazione alle altre frequenze. L'indice ha una lunga tradizione negli studi fonetici: secondo Holmberg et al. (1995) la differenza H1-H2 è correlata "with the proportion of the glottal cycle during which the glottis is open (the open quotient)", e viene comunemente definita come una delle misure più valide per l'analisi delle modalità di fonazione (Esposito, 2010). Normalmente ci si aspetta che una misura positiva di H1-H2 sia da rapportare a un maggiore quoziente di apertura e, di conseguenza, le vocali con modalità di fonazione sospirata tendono ad avere un quoziente di apertura maggiore e uno spettro dominato dalla prima armonica H1. Dunque, la nostra ipotesi è che, nella varietà calabrese di italiano, le vocali che seguono ad una occlusiva sorda aspirata mostrino valori di OQ più alti rispetto alle vocali che seguono ad una occlusiva sorda non aspirata.

Accanto alla differenza in ampiezza tra la prima e la seconda armonica, si è scelto anche di considerare H1-A3, e cioè la differenza tra l'ampiezza della prima armonica e l'ampiezza dell'armonica più prominente nella regione di F3 ("*Spectral Tilt*"). È stato dimostrato che questa misura è in correlazione con la velocità di chiusura delle pliche vocali (Stevens, Hanson, 1994; Esposito, 2004; Esposito, 2012). La vibrazione delle pliche vocali può essere possibile grazie a una ferma adduzione delle aritenoidi, che permette quindi la vibrazione nella parte anteriore: come notato da Stevens (1977: 174), "[w]ithin this region, the mechanical properties of the folds are more uniform, and an abrupt closure along the length of the vibrating portion can be expected. A more rapid rate of closure is also expected, since the inward (adducting) force on the folds is greater when the arytenoids are tightly adducted". Questa configurazione influenza direttamente la forma d'onda, per cui una vocale di tipo sospirato mostrerà una riduzione d'energia nelle armoniche alle più alte frequenze, con conseguenti valori più alti di *Spectral Tilt*. In conseguenza di tutto ciò, la nostra ipotesi è che, nella varietà ca-

labrese di italiano, le vocali che seguono ad una occlusiva sorda aspirata mostrino valori di *Spectral Tilt* più alti rispetto alle vocali che seguono ad una occlusiva sorda non aspirata.

4. *Metodologia*

4.1 Materiali

I dati analizzati in questo lavoro derivano da una serie di registrazioni effettuate fra il febbraio e il maggio 2014 presso due scuole superiori di Lamezia Terme (CZ). Il corpus fa parte di uno studio sociofonetico sulla correlazione tra persistenza di tratti locali di pronuncia e variabili sociali e individuali in diversi gruppi di parlanti giovani di Lamezia Terme (Nodari, 2016a). I dati qui riportati provengono dalla lista di frasi in italiano che i soggetti dovevano leggere ad alta voce, cercando di mantenere una velocità di eloquio costante, nella maniera più naturale possibile. La lista comprende 156 frasi di identica lunghezza (8 sillabe fonetiche) e struttura intonativa uniforme. Ogni frase contiene al proprio interno una parola *target* la quale a sua volta contiene, in una posizione determinata, una delle occlusive sorde dei tre diversi luoghi di costrizione (/p t k/). L'occlusiva può essere una geminata (GEMINATA), oppure una scempia preceduta da suono rotico (POSTR), nasale (POSTN) o laterale (POSTL). Si è incluso anche il contesto di occlusiva sorda scempia intervocalica (SCEMPIA) come contesto di controllo. Le parole target possono essere parossitone (con occlusiva sorda in posizione pretonica o postonica, es. [ka'tːʰura] e ['saltʰa]), o proparossitone (con occlusiva sorda in posizione postonica, es. ['saltʰano]). Per questo studio ci si è limitati alle sole sillabe con vocale /a/. Come riportato in Esposito (2010: 186), numerosi studi sulla qualità della voce si sono concentrati sulla /a/ perché il valore elevato di F1 minimizza gli effetti della prima e della seconda armonica, a differenza di quanto avviene nelle vocali alte.

Il dataset finale comprende 420 parole, prodotte da 8 soggetti (4 di sesso maschile e 4 di sesso femminile) e di età compresa tra i 15 e 18 anni.

4.2 Il setting sperimentale

Le registrazioni sono state condotte all'interno della scuola, in un'aula vuota messa a disposizione per gli esperimenti. Il materiale è stato raccolto utilizzando un microfono ad archetto Sennheiser collegato a un registratore Edirol R-09HR portatile, registrando direttamente in formato .WAV (44.1kHz / 16-bit).

4.3 Analisi

Le frasi sono state segmentate e annotate su Praat (Boersma, Weenink, 2015) in modo da avere informazioni sulla durata della fase di occlusione, durata del VOT e durata della vocale seguente. Gli intervalli sono stati annotati osservando sia la forma d'onda sia lo spettrogramma. Per l'annotazione del VOT si è scelto di se-

gnalare l'inizio includendo il *burst*; nei casi di *burst* multipli si è adottato il criterio di Cho, Ladefoged (1999), misurando cioè il VOT dall'ultimo *burst* presente sullo spettrogramma. I casi in cui il *burst* non era chiaramente visibile sono stati scartati. Per definire la fine dell'intervallo del VOT si è fatto riferimento a Lisker, Abramson (1964), i quali intendono il VOT come l'intervallo che intercorre tra il rilascio dell'occlusione e l'attacco delle vibrazioni glottidali. Di conseguenza, l'attacco della vocale seguente è stato identificato nel momento in cui si riscontrava sulla forma d'onda l'attacco delle vibrazioni della vocale. Questo criterio è risultato migliore rispetto a quello suggerito da Ladefoged, Maddieson (1996: 70), secondo i quali "aspiration is a period after the release of a stricture and before the start of regular voicing [...] in which the vocal folds are markedly further apart than they are in modally voiced sounds". Secondo questa ultima definizione infatti, ciò che si ottiene, più che un valore effettivo di VOT, è piuttosto un tempo di attacco per la sonorità di tipo modale (o *Modal Voice Onset Time*, come riportato da Helgason, Ringen, 2008), che dunque include anche eventuali fasi di vibrazione glottica non modale.

Per ottenere informazioni sugli indici scelti per l'analisi si sono utilizzati due script di Praat diversi:

1. uno script per estrarre le durate in millisecondi di tutti i segmenti etichettati;
2. uno script che fornisse dei dati sulla modalità di fonazione della vocale tonica seguente. Per questo indice ci si è affidati a uno script accessibile online, messo a punto da Chad Vicenik, basato sul programma VoiceSauce. Lo script estrae diversi indici, tutti correlati alla modalità di fonazione della vocale, e cioè H1-H2, H1-A1, H1-A2 e H1-A3; i valori sono calcolati per ogni terzo della durata della vocale, ottenendo quindi valori di fonazione all'*onset*, allo *steady state* e all'*offset* della vocale. In questo lavoro si sono tenuti in considerazione i parametri relativi ad H1-H2 e ad H1-A3 nel primo terzo della vocale, nella sua parte centrale e nella parte finale, per verificarne l'andamento nel tempo. Specificamente, l'ipotesi è che i valori delle due variabili dipendenti siano significativamente diversi nel confronto tra vocali precedute o non precedute da occlusiva aspirata, relativamente al primo terzo; negli altri due campionamenti (parte centrale e finale), ci si aspetta che le differenze non siano altrettanto marcate.

5. *Risultati*

5.1 Il Voice Onset Time

Un'analisi preliminare sulla durata del VOT nel campione di dati selezionato è servita a confermare che le occlusive sorde sono realizzate come aspirate, con l'eccezione delle occlusive scempie intervocaliche, che sono state incluse come contesto di controllo. I valori del VOT per i cinque diversi sottogruppi in base al contesto fonotattico sono riportati nella Tabella 1.

Tabella 1 - *Durata del VOT nei cinque contesti fonotattici*

	VOTms	st-dev.
SCEMPIA	18,3	7,1
GEMINATA	45,1	16,8
POSTR	45,6	17,2
POSTN	44,8	20,6
POSTL	45,3	16,3

5.2 Il modello statistico

Per l'analisi statistica della modalità di fonazione della vocale seguente sono stati realizzati modelli misti di regressione lineare, utilizzando il pacchetto lme4 in Rstudio (Bates, Mächler, Bolker & Walker, 2014). Le variabili dipendenti, cioè i valori di OQ (H1-H2) e dello *Spectral Tilt* (H1-A3), estratti in corrispondenza delle tre diverse porzioni della vocale, sono state modellate in funzione di diverse variabili indipendenti, elencate di seguito.

Fattori random:
1. PARLANTE
2. PAROLA

Fattori fissi:
1. LUOGO di articolazione dell'occlusiva sorda (3 livelli): bilabiale /p/ (BIL), alveolare /t/ (ALV), velare /k/ (VEL);
2. CONTESTO fonetico (5 livelli): occlusiva sorda intervocalica (SCE) (contesto non soggetto ad aspirazione), occlusiva sorda geminata (GEM), preceduta da suono rotico (POSTR), nasale (POSTN), laterale (POSTL);
3. ACCENTO lessicale (2 livelli): vocale tonica preceduta da occlusiva sorda in parola parossitona (es. [toˈkːare]) (TONICA), vocale atona preceduta da occlusiva sorda in parola parossitona (es. [ˈfatta]) o in parola proparossitona (es. [ˈbutːano]) (ATONA);
4. SESSO dei parlanti: maschi (M) vs femmine (F).

È stato anche valutato il grado di correlazione tra ciascuna delle variabili dipendenti (OQ e *Spectral Tilt*) ed i valori del VOT per le medesime sillabe, sulla base di una correlazione di Pearson a due code.

5.3 Risultati relativi ad OQ (H1-H2)

I dati mostrano come, in generale, il valore di H1-H2 sia modellato dalle variabili ACCENTO e CONTESTO; non sono risultati significativi il LUOGO e il SESSO.

Tabella 2 - *Risultati relativi ad H1-H2 per il primo terzo della vocale*

Coefficients	*Estimate*	*Std. Error*	*T Value*
(Intercept)	-3.1254	0.8583	-3.642
LUOGOp	-0.3785	0.6227	-0.608
LUOGOt	-0.2692	0.6156	-0.437
ACCENTO_TONICA	-1.8208	0.5101	-3.570
CONTESTO_GEM	5.4935	0.8046	6.828
CONTESTO_POSTR	4.9123	0.8376	5.865
CONTESTO_POSTN	5.5211	0.8247	6.695
CONTESTO_POSTL	4.8185	0.8289	5.813
SESSO_F	0.6022	0.6540	0.921

Tabella 3 - *Risultati relativi ad H1-H2 per il secondo terzo della vocale*

Coefficients	*Estimate*	*Std. Error*	*T Value*
(Intercept)	-3.3327	0.6569	-5.073
LUOGOp	0.5560	0.5246	1.060
LUOGOt	0.3193	0.4885	0.654
ACCENTO_TONICA	-1.5995	0.4297	-3.723
CONTESTO_GEM	2.3022	0.6721	3.425
CONTESTO_POSTR	1.8763	0.7035	2.667
CONTESTO_POSTN	2.5507	0.6912	3.690
CONTESTO_POSTL	1.6088	0.6956	2.313
SESSO_F	0.6749	0.3705	1.822

Tabella 4 - *Risultati relativi ad H1-H2 per il terzo terzo della vocale*

Coefficients	*Estimate*	*Std. Error*	*T Value*
(Intercept)	-2.6315	0.6901	-3.813
LUOGOp	-0.6438	0.5128	-1.256
LUOGOt	-0.2119	0.4982	-0.425
ACCENTO_TONICA	-1.5765	0.4159	-3.791
CONTESTO_GEM	1.0113	0.6505	1.555
CONTESTO_POSTR	0.7639	0.6805	1.122
CONTESTO_POSTN	0.8977	0.6700	1.340
CONTESTO_POSTL	0.7371	0.6742	1.093
SESSO_F	0.3061	0.5058	0.605

Per quanto riguarda ACCENTO, non sorprende che vi sia una significativa differenza tra vocali atone e vocali toniche. In particolare, i valori della dipendente sono più bassi per le vocali toniche, indicando cioè che tali vocali sono prodotte con una

modalità di fonazione modale. Questo effetto è forte e presente in tutti e tre i campionamenti temporali della vocale.

Ugualmente significativa è la differenza nei valori di H1-H2 se si confrontano i CONTESTI soggetti ad aspirazione dell'occlusiva sorda (GEM, POSTR, POSN, POSTL) con il contesto non soggetto ad aspirazione (SCE). In tutti i casi, il valore di H1-H2 è maggiore nei contesti con aspirazione, indicando dunque che la vocale che segue ad una occlusiva aspirata è realizzata con una modalità di fonazione diversa rispetto alla vocale preceduta da una occlusiva non aspirata. L'effetto di CONTESTO è però significativo solo nel primo e nel secondo terzo della vocale; nella parte finale, invece, le differenze si annullano. Ciò indica che l'attacco della vocale è massimamente influenzato dal modo in cui viene articolata l'occlusiva precedente, mentre lo stadio intermedio e lo stadio finale della vocale non risentono degli effetti di coarticolazione.

I valori di OQ non sono influenzati dal punto di articolazione delle occlusive precedenti, a conferma del fatto che gli effetti riscontrati per le variabili ACCENTO e CONTESTO valgono per sillabe contenenti tutte e tre le occlusive sorde /p t k /. Similmente, sulla modalità di fonazione delle vocali non gioca nessun ruolo il fatto che il parlante sia di sesso maschile o femminile.

Tabella 5 – *Valori di H1-H2 nei cinque diversi contesti fonotattici*

	Media di H1-H2_ PRIMO TERZO	Media di H1-H2_ SECONDO TERZO	Media di H1-H2_ TERZO TERZO
SCEMPIA	-3,68	-3,34	-3,33
GEMINATA	1,54	-1,20	-2,50
POSTR	1,26	-1,41	-2,55
POSTN	1,92	-0,77	-2,38
POSTL	1,24	-1,59	-2,38

Per verificare se il grado di aspirazione dell'occlusiva sorda influenza i valori dell'OQ, si è valutato il grado di correlazione tra le due variabili. Quando i valori del VOT sono messi in correlazione con i valori di OQ relativi al primo e al secondo terzo della vocale, la correlazione è significativa (primo terzo: $t = 4.4147$, $df = 418$, $p<0.001$, $cor = 0.21$; secondo terzo: $t = 2.0991$, $df = 418$, p-value = 0.03, $cor = 0.10$). Non risulta invece significativa la correlazione tra OQ e VOT quando si considerano i valori di H1-H2 nell'ultimo terzo della vocale ($t = 1.5458$, $df = 418$, p-value = 0.12, $cor = 0.07$).

5.4 Risultati relativi allo Spectral Tilt (H1-A3)

I dati mostrano come H1-A3 sia modellato dalle variabili LUOGO, ACCENTO e CONTESTO; la variabile SESSO non è risultata significativa.

Tabella 6 - *Risultati relativi ad H1-A3 per il primo terzo della vocale*

Coefficients	*Estimate*	*Std. Error*	*T Value*
(Intercept)	2.4095	3.1196	0.772
LUOGOp	2.0478	1.4712	1.392
LUOGOt	5.4126	1.5475	3.498
LUOGO2_t(vsP)	3.3648	1.5582	2.159
ACCENTO_TONICA	-3.4019	1.1894	-2.860
CONTESTO_GEM	10.5749	1.8754	5.639
CONTESTO_POSTR	9.7089	1.9525	4.973
CONTESTO_POSTN	10.6574	1.9260	5.533
CONTESTO_POSTL	11.7254	1.9353	6.059
SESSO_F	-0.6085	3.6843	-0.165

Tabella 7 - *Risultati relativi ad H1-A3 per il secondo terzo della vocale*

Coefficients	*Estimate*	*Std. Error*	*T Value*
(Intercept)	0.7379	2.8715	0.257
LUOGOp	1.9161	1.3737	1.395
LUOGOt	5.3668	1.4441	3.716
LUOGO2_t(vsP)	3.4507	1.4545	2.372
ACCENTO_TONICA	-4.9050	1.1107	-4.416
CONTESTO_GEM	7.4359	1.7513	4.246
CONTESTO_POSTR	7.1433	1.8233	3.918
CONTESTO_POSTN	8.5007	1.7986	4.726
CONTESTO_POSTL	7.9049	1.8073	4.374
SESSO_F	0.7235	3.3700	0.215

Tabella 8 - *Risultati relativi ad H1-A3 per il terzo terzo della vocale*

Coefficients	*Estimate*	*Std. Error*	*T Value*
(Intercept)	3.8977	3.2612	1.195
LUOGOp	0.3916	1.5195	0.258
LUOGOt	4.0202	1.5991	2.514
LUOGO2_t(vsP)	3.6285	1.6096	2.254
ACCENTO_TONICA	-2.8157	1.2283	-2.292
CONTESTO_GEM	4.9410	1.9368	2.551
CONTESTO_POSTR	5.5951	2.0163	2.775
CONTESTO_POSTN	5.9067	1.9891	2.970
CONTESTO_POSTL	5.3058	1.9987	2.655
SESSO_F	0.4346	3.8714	0.112

La variabile LUOGO è risultata significativa nel modulare la variazione di H1-A3. La vocale /a/, infatti, quando preceduta da /t/ mostra valori di *Spectral Tilt* più alti rispetto quando è preceduta dalle altre occlusive sorde (/p/ e /k/). Ciò indica che la vocale se preceduta da /t/ è realizzata con modalità di fonazione più sospirata. Tale effetto è generalizzato in tutti e tre i campionamenti, cosa che suggerisce che tale proprietà è diffusa sulla durata vocalica in tutta la sua interezza.

Anche la variabile CONTESTO contribuisce significativamente a modellare i valori di H1-A3; i dati mostrano infatti che la vocale nei contesti di aspirazione è significativamente diversa dalla vocale nel contesto di occlusiva scempia. In particolare, i quattro contesti con occlusiva sorda aspirata mostrano valori maggiori di H1-A3. Tale effetto è significativo per l'intera durata della vocale, come dimostrano i risultati dei tre punti di campionamento della vocale. Cionostante, vi è un decremento sistematico nei valori di *Spectral Tilt*, man mano che dall'*onset* della vocale si procede verso il suo *offset*, come mostrano i valori contenuti nella Tabella 10.

Tabella 9 - *Valori di H1-A3 nei cinque diversi contesti fonotattici*

	Media di H1-A3_ PRIMO TERZO	Media di H1-A3_2 TERZO	Media di H1-A3_3 TERZO
SCEMPIA	3,4	1,8	4,6
GEMINATA	14,1	9,0	9,6
POSTR	12,2	8,1	9,5
POSTN	14,2	10,3	10,6
POSTL	15,0	9,8	9,9

Per quanto riguarda la variabile ACCENTO, le vocali atone mostrano valori maggiori di *Spectral Tilt* rispetto alle vocali toniche, in tutti e 3 i punti di campionamento. Ciò indica quindi che le vocali atone sono prodotte con modalità di fonazione più sospirata rispetto alle vocali toniche, sia all'inizio che nelle fasi successive dell'articolazione.

Per verificare se il grado di aspirazione dell'occlusiva sorda influenza i valori dello *Spectral Tilt*, si è valutato il grado di correlazione tra le due variabili. Quando i valori del VOT sono messi in correlazione con i valori dello *Spectral Tilt* relativi al primo e al secondo terzo della vocale, la correlazione è significativa (primo terzo: $t = 3.3902$, $df = 418$, $p < 0.001$, $cor = 0.16$; secondo terzo: $t = 2.5106$, $df = 418$, p-value = 0.01, $cor = 0.12$). Non risulta invece significativa la correlazione tra *Spectral Tilt* e VOT quando si considerano i valori di H1-A3 nell'ultimo terzo della vocale ($t = 1.6317$, $df = 418$, $p = 0.10$, $cor = 0.07$).

6. *Discussione*

In questo studio, si sono analizzati i valori del quoziente di apertura (OQ) e lo *Spectral Tilt* di /a/ tonica e atona dopo occlusiva sorda aspirata dell'italiano calabrese di 8 giovani studenti di Lamezia terme. Il quoziente di apertura e lo *Spectral Tilt* sono indici generalmente utilizzati per valutare il modo di fonazione dei segmenti vocalici, con particolare riferimento alla differenza tra fonazione modale e fonazione sospirata. Date queste premesse, ci aspettavamo di trovare, per le vocali che seguono a una occlusiva sorda aspirata, valori di OQ e di *Spectral Tilt* più alti rispetto a quelli di vocali che seguono a una occlusiva non aspirata. Inoltre, ci aspettavamo di trovare che le maggiori differenze si concentrano in corrispondenza del primo terzo della vocale, a causa del fatto che l'influsso coarticolatorio dell'aspirazione consonantica sulla modalità di fonazione della vocale successiva è generalmente attribuito alla fase di attacco della vocale stessa.

I risultati delle analisi fin qui condotte sono parzialmente coerenti con le ipotesi iniziali.

Innanzitutto, per le occlusive sorde aspirate del calabrese come per altre lingue studiate in letteratura, si riscontrano degli effetti coarticolatori della consonante che vanno a influenzare la modalità di fonazione della vocale seguente: sono state infatti trovate delle differenze significative che governano la variazione sia dell'OQ, sia dello *Spectral Tilt*, in rapporto al tipo di occlusiva precedente (aspirata o non aspirata). Il contrasto laringeo sembra perciò avere un diretto effetto sulla modalità di fonazione della vocale, per entrambe le misure che si sono prese in considerazione. Esse, peraltro, appaiono strettamente correlate ai valori di durata del VOT delle occlusive sorde precedenti, a riprova del fatto che le variazioni dello stato laringeo durante la produzione della vocale sono tanto maggiori quanto maggiore è il grado di aspirazione delle occlusive precedenti

Le vocali precedute da una consonante in uno dei quattro contesti di aspirazione (occlusiva geminata o inclusa in un nesso con /r/, /l/ o /n/) mostrano infatti valori più alti di H1-H2 e H1-A3, almeno nel primo terzo della vocale. Il risultato suggerisce che, in effetti, la vocale viene articolata con una modalità di fonazione diversa, che potrebbe essere definita di tipo sospirato, ossia con una fase di abduzione delle corde vocali relativamente più lunga, rispetto alla norma della fonazione modale Il risultato relativo alla *Spectral Tilt* suggerisce in particolare che le vocali che seguono a una consonante aspirata sono prodotte con un movimento più lento di adduzione/abduzione delle pliche vocali.

Inoltre, i valori dell'OQ suggeriscono che la differenza è maggiore nelle fasi di attacco della vocale rispetto alle fasi della sua conclusione. Ciò è coerente con le aspettative rispetto al fatto che l'influsso dell'aspirazione consonantica sulla vocale è un influsso coarticolatorio di tipo perseverativo, dunque maggiormente presente quanto più ci si avvicina all'*offset* della consonante. Per quanto riguarda lo *Spectral Tilt*, invece, l'effetto dell'aspirazione sembra riguardare tutta la produzione della vocale nel suo complesso, senza differenziazione tra parte iniziale e parte finale. Da questo risultato si può dunque concludere che, mentre l'effetto dell'aspirazione

sulla durata relativa della fase di apertura delle pliche vocali è presente solo nella fasi iniziali della produzione della vocale, l'effetto dell'aspirazione sulla velocità del movimento di adduzione/abduzione persiste per tutta la durata della produzione della vocale. In entrambi i casi, gli indici suggeriscono una modalità di fonazione di tipo sospirato. Solo l'indice relativo all'OQ, però, fa riferimento ad una modifica temporanea dell'attacco vocalico; l'indice relativo allo *Spectral Tilt* suggerisce che tale proprietà si estende durante tutta la produzione del fono vocalico interessato.

Le nostre analisi hanno però anche messo in luce la presenza di alcuni effetti ulteriori.

I dati riguardanti l'accento mostrano che le vocali toniche sono significativamente diverse dalle atone, per entrambi i parametri considerati (OQ e *Spectral Tilt*). In particolare, le atone avrebbero valori di *Spectral Tilt* e di OQ maggiori rispetto alle vocali toniche, cosa che suggerisce cioè che siano prodotte con una più lunga fase di apertura glottidale e una minore velocità di adduzione, rispetto alle toniche. Da un lato, questo risultato appare in linea con quanto riscontrato in altre lingue: come riportato nella rassegna di Uguzzoni (2006), valori più bassi di H1-A3 per le vocali toniche rispetto alla loro controparte atona sono stati trovati in olandese, tedesco e inglese americano, tanto da far ipotizzare che, per queste lingue, un'enfasi spettrale nelle medie e alte frequenze sia da considerare come un correlato acustico dell'accento lessicale. I nostri dati sull'italiano regionale calabrese confermano dunque che i due parametri acustici considerati possono variare anche in dipendenza di fattori indipendenti dall'aspirazione dell'occlusiva precedente.

Per quanto riguarda il luogo di articolazione dell'occlusiva, si è riscontrato che la vocale che segue ad una /t/ ha un valore più alto di *Spectral Tilt* rispetto alle vocali che seguono a /p/ e /k/. L'effetto potrebbe essere messo in relazione con una maggiore durata del VOT della consonante alveolare, e quindi con un suo maggior grado di aspirazione: l'esplorazione dei dati ha però portato a scartare quest'ipotesi, poiché la durata del VOT nei tre luoghi di articolazione è in linea con i risultati attestati dalla bibliografia (/p/ 28,9 ms., /t/ 32,5 ms., /k/ 42,5 ms.) (Sorianello, 1996). Quest'effetto inatteso necessita quindi di ulteriori analisi per poter essere interpretato.

Da ultimo, è inoltre importante notare come i dati non abbiano mostrato un effetto della variabile del sesso dei parlanti: indipendentemente dal tipo di occlusive esaminato, studi condotti su altre lingue hanno riportato spesso una maggiore incidenza tra i soggetti di sesso femminile di modalità di fonazione di tipo sospirato (su tutti, si veda Gordon, Ladefoged, 2001 per il San Lucas Quiaviní Zapotec). Il risultato assume maggiore interesse se si considera come la stessa differenza non significativa tra maschi e femmine sia stata riscontrata anche per la durata del VOT delle occlusive sorde prodotte da un campione più ampio di soggetti (nel quale sono inclusi anche gli otto soggetti del seguente studio, Nodari, 2016b). Locutori di entrambi i sessi parlanti italiano regionale calabrese sembrano quindi mostrare lo stesso comportamento nella realizzazione di una occlusiva sorda aspirata, sia per quanto

riguarda la durata del VOT, sia per quanto riguarda la modalità di fonazione della vocale seguente.

7. *Conclusioni e prospettive future*

Per quanto condotto su un numero limitato di parlanti, questo studio contribuisce a dettagliare maggiormente il processo di aspirazione delle occlusive sorde presente nell'italiano regionale calabrese. Infatti, oltre alle classiche analisi del parametro del VOT, il quale risulta in stretta correlazione con gli indici analizzati, anche le caratteristiche della fonazione della vocale immediatamente successiva suggeriscono che l'aspirazione delle occlusive sorde è il risultato di una mancata sincronia tra gesto di rilascio dell'occlusione orale e gesto di iniziazione della vibrazione cordale.

Inoltre, la verifica della modalità di fonazione di tipo sospirato nella vocale che segue a una occlusiva aspirata, oltre a offrirci un quadro più chiaro sulla realizzazione effettiva di un'occlusiva sorda aspirata, potrebbe risultare di notevole interesse per verificare le dinamiche della percezione: è possibile cioè che i parlanti siano in grado di percepire la distinzione tra consonanti aspirate e non aspirate anche a partire dalla realizzazione sospirata o non sospirata della vocale (Esposito, 2006), e che la modalità di fonazione costituisca una spia percettiva secondaria utile per la discriminazione, all'interno di una teoria della percezione come elaborazione interattiva di molteplice spie (*cue-weighting*; cfr. Kingston, Diehl, Kirk & Castleman, 2008; Llanos, Dmitrieva, Shultz & Francis, 2013 per le distinzioni di sonorità nelle ostruenti).

Riferimenti bibliografici

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2014). Package lme4: linear mixed-effects models using Eigen and S4. In *Journal of Statistical Software*, 67, 1, 1-48.

Boersma, P., Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.09. http://www.praat.org/Accessed 01.06.15.

Canepari, L. (1986). *Italiano standard e pronunce regionali*. Padova: CLEUP.

Cho, T., Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. In *Journal of Phonetics*, 27, 207-229.

De Blasi, N. (2014). *Geografia e storia dell'italiano regionale*. Bologna: Il Mulino.

Esposito, C.M. (2004). Santa Ana del Valle Zapotec phonation. M.A. Thesis, University of California, Los Angeles.

Esposito, C.M. (2006). The effects of linguistic experience on the perception of phonation. PhD Thesis, UCLA.

Esposito, C.M. (2010). Variation in contrastive phonation in Santa Ana Del Valle Zapotec. In *Journal of The International Phonetic Association*, 40, 2, 181-198.

Esposito, C.M. (2012). An acoustic and electroglottographic study of White Hmong phonation. In *Journal of Phonetics*, 40, 3, 466-476.

Falcone, G. (1976). *Calabria*. Pisa: Pacini.

Fanciullo, F., Librandi, R. (2002). La Calabria. In Cortelazzo, M., Marcato, C., De Blasi, N. & Clivio, G. (Eds.), *I dialetti italiani. Storia, struttura, uso*. Torino: Utet, 793-833.

Gobl, C., Ní Chasaide, A. (1999). Voice source variation in the vowel as a function of consonantal context. In Hardcastle, W.J., Hewlett, N. (Eds.), *Coarticulation: theory, data and techniques*. Cambridge: University Press, 122-143.

Goldstein, L.M., Browman, C.P. (1986). Representation of voicing contrasts using articulatory gestures. In *Journal of Phonetics*, 14, 339-42.

Gordon, M., Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. In *Journal of Phonetics*, 29, 383-40.

Hanson, H., Stevens, K., Kuo, H., Chen, M.Y. & Slifka, J. (2001). Towards models of phonation. In *Journal of Phonetics*, 29, 451-480.

Helgason, P., Ringen, C. (2008). Voicing and aspiration in Swedish stops. In *Journal of Phonetics*, 36, 607-628.

Holmberg, E., Hillman, R., Perkell, J., Guiod, P. & Goldman, S. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. In *Journal of Speech and Hearing Research*, 38, 1212-1223.

Kingston, J., Diehl, R.L., Kirk, C.J. & Castleman, W.A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. In *Journal of Phonetics*, 36, 28-54.

Klatt, D., Klatt, L. (1990). Analysis, synthesis and perception of voice quality variations among male and female talkers. In *Journal of the Acoustical Society of America*, 87, 820-85.

Ladefoged, P., Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lisker, L., Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. In *Word*, 20, 3, 384-422.

Löfqvist, A. (1980). Interarticulatory programming in stop production. In *Journal of Phonetics*, 8, 475-90.

Llanos, F., Dmitrieva, O., Shultz, A. & Francis, A.L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. In *Journal of the Acoustic Society of America*, 134, 3, 2213-2224.

Mele, B. (2009). *Fonetica e fonologia del dialetti di San Giovanni in Fiore*. Tübingen-Basel: Francke.

Nodari, R. (2016a). Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio/Language acquisition and language loss. Acquisition, change and disorders of the language sound structure*. Milano: AISV.

Nodari, R. (2016b). L'italiano degli adolescenti: aspirazione delle occlusive sorde in Calabria e percezione della varietà locale. PhD Thesis, Scuola Normale Superiore.

Rohlfs, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino: Einaudi.

Sorianello, P. (1996). Indici fonetici delle occlusive sorde nel cosentino. In *Rivista Italiana di Dialettologia*, 20, 123-159.

Stevens, K. (1977). Physics of Laryngeal Behavior and Larynx Modes. In *Phonetica*, 24, 264-279.

Stevens, K. (2000). *Acoustic phonetics.* Cambridge: MIT.

Stevens, K., Hanson, H. (1994). Classification of glottal vibration from acoustic measurements. Paper presentato all'*8 Vocal Fold Physiology Conference*, Kurume, Japan, 7-9 April 1994.

Stevens, M., Hajek, J. (2010). Post-aspiration in standard Italian: some first cross-regional acoustic evidence. In *Proceedings of Interspeech, Makuhari, Japan*, 1557-1560.

Telmon, T. (1993). Varietà regionali. In Sobrero, A.A. (Ed.), *Introduzione all'italiano contemporaneo, la variazione e gli usi*. Roma-Bari: Laterza, 93-149.

Uguzzoni, A. (2006). I valori di H1-A2 e H1-A3 come correlati della intensità "rivisitata". Aspetti e problemi. In *Atti del II Convegno Nazionale AISV, Analisi prosodica: teorie, modelli e sistemi di annotazione*, DVD, 566-592.

Vicenik, C. (2010). An acoustic study of Georgian stop consonants. In *Journal of the International Phonetic Association*, 40, 1, 59-92.

DOI: 10.17469/O2102AISV000022

MASSIMILIANO M. IRACI, MIRKO GRIMALDI, BARBARA GILI FIVELA

Phonology Drives Compensation: bridging linguistic and clinical evaluation for a classification of speech impairment in dysarthria

Kinematic data collection is providing new possibilities to enhance (and objectivise) the evaluation of the impairment in Motor Speech Disorders. Focusing on Hypokinetic Dysarthria in Parkinson's Disease, recent studies reveal that pathological speakers, despite showing deficits in amplitude and coordination of speech gestures, are able to correctly realise kinematic and acoustic correlates of phonological contrast (such as in the alternation of singletons and geminates in Italian) through some compensatory strategies. Our hypothesis is that phonological constraints drive the compensation, but constraints due to the pathology act at the phonetic level, on contiguous gestures. This seems to be the case when analysing speech production. In order to check this hypothesis on listeners' perception of pathological productions, an auditory test aiming to collect both phonological and phonetic information was designed. Furthermore, the information collected were also used in order to more objectively classify pathological speakers' productions. Results seem to confirm our hypothesis and suggest that a phonologically-phonetically based evaluation of the level of Motor Speech Disorders's impairment may correspond to subjective clinical evaluation, and thus can be eligible for objectivising clinical assessment.

Key words: Parkinson's Disease, Dysarthria, Impairment evaluation, Articulatory Phonology, Compensation strategies.

Introduction

In the last decades, the research on Motor Speech Disorders (MSD) benefited of motion tracking instruments, such as Electromagnetic Articulography (EMA), in order to study this class of speech pathologies at the level of motion, i.e. measuring the production of speech sounds directly from the dynamics of articulators. Before researchers had the possibility to exploit motion tracking instruments for these purposes, the standard for the study of MSD was the perceptual evaluation – still in use in the clinical practise. Perceptual evaluation has been crucial for the classification of motor speech disorders as we know them, from the pioneering studies by Darley, Aronson and Brown (cfr. Darley, Aronson & Brown, 1969) to the most recent volume by Duffy (2005).

1. *Hypokinetic Dysarthria in Parkinson's Disease*

1.1 Amplitude and coordination of speech gestures

This study focuses on Hypokinetic Dysarthria (HD): HD is a motor speech disorder typically shown by people affected by Parkinson's Disease (PD) (Duffy, 2005). From the point of view of speech production, it entails disturbances to the execution and control of speech gestures' amplitude and coordination (as for the amplitude, Ackermann, Ziegler, 1991; Gili Fivela, Iraci, Sallustio, Grimaldi, Zmarich & Patrocinio, 2014; Iraci, Zmarich, Grimaldi & Gili Fivela, 2016; Skodda, Gronheit & Schlegel, 2012; Skodda, Visser & Schlegel, 2011; Wong, Murdoch & Whelan, 2010; 2011; as for coordination, Connor, Abbs, Cole & Gracco, 1989; Gili Fivela, Iraci, Grimaldi & Zmarich, 2015; Iraci, Grimaldi & Gili Fivela, in revisione; Tjaden, 2000; 2003; Tjaden, Wilding, 2005; Weismer, Yunusova & Westbury, 2003). Actually, the issue is quite controversial since, on the one hand, speech gestures' amplitude has been found to be both reduced (Skodda et al., 2011; 2012) or increased (Wong et al., 2010; 2011). Moreover, this happened simultaneously in native Italian dysarthric PD speakers, depending on the axis of movement (Gili Fivela et al., 2014; 2015; Iraci et al., 2017b): more in detail, given a mid-sagittal plane of observation, Italian pathological speakers, when compared to control speakers, can show increased tongue gestures' amplitude on the anterior-posterior dimension, while the opposite happens on the vertical dimension (reduced gestures' amplitude). On the other hand, studies on coordination still report uncertain results probably because of methodological differences. While Connor et al. (1989) find that the coordination between lips and jaw fails in the production of bilabial consonants, Tjaden, despite some slight coordination deficit, states that patterns of coordination are mostly preserved (Tjaden, 2000; 2003; Tjaden, Wilding, 2005). Finally, Weismer et al. (2003) report similar considerations, noticing only differences in the timing of lip protusion for the production of /u/. Concerning Italian, also Gili Fivela et al. (2015) can infer some slight differences between PD and control speakers in the coordination between tongue and lip, but the patterns of supposed incoordination remain unclear in their preliminary study.

1.2 Phonology and compensatory strategies

Despite the patterns of misarticulation and/or incoordination that the research on this topic individuates and describes, dysarthric PD speakers appear to be able to realise meaningful differences mainly based on articulatory gestures' amplitude and duration (such as singleton vs. geminate consonants in Italian; Gili Fivela et al., 2014; 2015; Iraci et al., in revisione; 2017b). This can be possible hypothesising that HD does not carry any direct effect at the phonological level, but only indirect effects due to extreme lack of accuracy. In other words, speech alterations (due to misarticulation and/or incoordination) affect the range of phonetic variation without threatening the phonological contrast. This is true at least when the level of impairment is not extremely severe (Iraci et al., in revisione). For instance, the dif-

ference in the acoustic duration, which is one of the main correlates of the singleton vs. geminate contrast, is maintained even though the average geminate duration in PDs' production is similar to the singleton duration in controls' production (Gili Fivela et al., 2015). In fact, dysarthric speakers seem to exploit some compensatory strategies (Schröter-Morasch, Ziegler, 2005; McCabe, 2010) that are not functional to the accuracy of speech but, as we hypothesize, are likely to maintain the phonological plan – or, in other words, they are driven by phonological constraints. For example, in a preliminary study, Iraci et al. (in revisione) individuated and described subject-specific articulatory strategies of dysarthric PD speakers relying on EMA data (AG 501, Carstens GmbH): subjects were all able to realise both acoustic and articulatory correlates of singleton vs. geminate contrast, showing some subjective alterations to contiguous (or co-produced secondary) gestures.

1.3 Bridging phonetic and clinical assessment: a summary of research questions

Our general hypothesis about the conservation of the phonological plan through compensatory strategies affecting the accuracy of speech has been tested in studies on speech production so far. As mentioned above, MSD's impairment evaluation has been based on auditory analysis and this method is the one the clinical practise is mostly relying on. Thus, in order to bridge production and auditory analysis, it seems crucial to test our hypothesis through perceptual data from experimental subjects listening to pathological voices. Especially, we wonder if listeners will recognise those consonants which in our previous studies are said to be consistently realised by pathological speakers as examples of the expected categories.

In addition to this, when consonants are correctly recognised by listeners, we wonder which will be their evaluation about the accuracy of the whole production. This kind of information is useful as to measure the extent to which the pathological realisation is well-suited or at the edges of the admitted range of phonetic variation.

However, the answers to these two questions will offer us an objective evaluation of both the phonological and the phonetic characteristics of the PD productions.

Furthermore, bridging perception and production might also mean bridging linguistic-phonetic and clinical evaluation. Researchers never know whether the clinically established level of impairment corresponds to the phonetically established one. Bridging phonetic (objective) and clinical (subjective) assessment is of crucial importance for the definition of common starting points and aims for both linguistic and medical sciences.

1.4 Bilabials, geminates, voicing and nasality

In order to test our hypothesis, the same items collected for the study of pathological speech production (Gili Fivela et al., 2014; 2015; Iraci et al., in revisione; 2017b) have been administered to non-pathological native Italian listeners coming from the same area of pathological speakers. Items were acoustic recordings of pseudo-words coupled in minimal pairs differing for the medial consonant. Such consonant could

be a bilabial singleton or the corresponding geminate – whose correlates of contrast are also kinematically based (Gili Fivela, Zmarich, 2005; Gili Fivela, Zmarich, Perrier, Savariaux & Tisato, 2007; Zmarich, Gili Fivela, 2005; Zmarich, Gili Fivela, Perrier, Savariaux & Tisato, 2007; 2009; 2011) and have been demonstrated to be realised as such by pathological speakers from a kinematic and acoustic point of view (Gili Fivela et al., 2014; 2015; Iraci et al., in revisione; 2017b).

The contrast between Italian singletons and geminates has been selected as a factor of interest for two reasons. As already stated, it is also based on kinematic correlates such as amplitude and duration of gestures, whose implementation is problematic in HD (see §1.1 for references). Moreover, syllable structure is supposed to switch from CV.CV in the case of singleton, to CVC.CV in the case of geminate (Bertinetto, 1981; Loporcaro, 1996). So, this contrast allows to check whether a switch from a simple to a (more) complex syllable structure influences pathological speakers' performances.

Among potential items to administer, bilabials have been selected for continuity with our production studies. In fact, in Gili Fivela et al. (2014; 2015) and Iraci et al. (2017b), bilabials have been exploited as a baseline case to check for the phasing of gestures, specifically for the purpose of the singleton vs. geminate contrast which the phasing should be relevant for: this choice allowed to exclude cases of shared articulators between consonants and vowels' gestures and concentrate only on productions whose consonants were associated to a bilabial gesture, and vowels to a tongue dorsum gesture.

Moreover, voicing and nasality have been included in this experiment because PD speakers can show alterations to the management of vocal folds and/or the velopharyngeal sphincter. The first case is acknowledged in the classical literature reporting, for example, reduced vibratory intensity, incomplete vocal closure, increased phonation threshold pressure and glottal tremor (cfr. Duffy, 2005, but also Zhang, Jiang & Rahn, 2005). On the contrary, there is a limited amount of studies on the effects of the Velopharyngeal dysfunction (VPD) on speech (e.g. see Hammer, Barlow, Lyonsc & Pahwac, 2011 for the treatment of VPD in PD through Deep Brain Stimulation). VPD consists in inadequate velopharyngeal control, the latter crucial for the realisation of velum opening gestures for nasal consonants. Analysing voicing and nasality can provide new information about articulators which are hard to be instrumentally inspected.

It is worth reminding that voicing and nasality are to be considered not only as physical components which can clinically stress HD speakers. Following the theoretical framework of Articulatory Phonology (cfr. amongst others, Browman, Goldstein, 1989; 1990; cfr. for recent developments, Gafos, Goldstein, 2011), voicing and nasality should be considered as two gestures, the first realised by the vocal folds, the second by the velum[1]. They are produced in coordination with the other oral gestures. For these reasons, items containing voiced and nasal bilabials will be analysed as within an increasing scale of phonetic demand, in comparison with the baseline case (the simplest case) represented by unvoiced segments.

[1] Despite Articulatory Phonology considers voicing to be a default modality (hence, less marked than lack of voicing, i.e. glottal abduction), for the clinical reasons above exposed, this study will consider vocal folds vibration to be more problematic than glottal abduction for unvoiced consonants.

2. *Experiment*

2.1 Aims and hypotheses

Despite HD entails disturbances to the execution and control of speech gestures' amplitude and coordination, dysarthric PD speakers are able to realise meaningful differences mainly based on articulatory gestures' amplitude and duration. For this reason, our main hypothesis, driven by production results is that HD does not carry any direct effect at the phonological level (but only indirect effects): speech alterations are supposed to affect the phonetic characteristics of speech, while clear inconsistencies with the phonological form manifest when execution is too much disrupted. Given those premises, we intend to verify from the aural-perceptual point of view:

i. if minimal pairs realised by pathological speakers, and consistently differentiated on the kinematic and acoustic dimension, will correctly be categorised by non-pathological listeners,
ii. the phonetic accuracy in PD productions, as evaluated by listeners.

In line with our main hypothesis and with the observation of cases in which, given a significant difference between singleton and geminates by PD speakers, the average geminate duration in PDs' production is similar to the singleton duration in control speakers' realizations (see §1.2.), we expect results to correlate with the level of impairment.

Therefore, we would expect that:

1. the higher the impairment, the higher the number of minimal pairs not correctly categorised by listeners,
2. the higher the impairment, the lower the accuracy of pathological productions.

However, our corpus includes bilabial consonants, whose voiced segment, in the area where recordings took place (Lecce), shows *rafforzamento* in intervocalic medial position (Gaillard-Corvaglia, Kamiyama, 2008) and domain initial position (Gili Fivela, d'Apolito, Stella & Sigona, 2008; 2010). For this reason, it may be hard to distinguish between a singleton and the corresponding geminate, which, as will be discussed in the following sections, are our main term of comparison. Thus, for what concerns the voiced bilabial we expect:

1a anomalous results (i.e. not in line with expectations above exposed), in that the geminate should be hardly distinguished from the singleton showing *rafforzamento* (and vice-versa);
2a results in line with previous expectations as for accuracy, in that the phenomenon of *rafforzamento* has no effects on the phonetic accuracy of the consonant realisation.

Nevertheless, as mentioned in §1.2, we cannot take for granted that the clinically established level of impairment corresponds to the phonetically established one. Rather, we assume that measuring the amount of alterations of both the original

phonological plan and the actual phonetic execution (of the expected phonological plan) might provide a phonetically and phonologically-based rating of the impairment, offering a satisfying description (and hopefully classification) of the speech impairment itself.

Interestingly, in the literature on second language acquisition, Guion, Flege, Akahane-Yamada & Pruitt (2000) proposed an index to measure the accurateness in L2 speech, in which no perfect match with the ideal L2 production is expected. In their study on Japanese learners of English, "English consonants with relatively high *fit indexes* would be readily accepted as instances of a Japanese consonant category, whereas those with relatively low fit indexes would be heard either as "foreign" or as distorted instances of a Japanese category" (Guion et al., 2000). In our opinion, a similar index may be useful to measure the accurateness also in pathological speech, since no perfect match with the ideal production is expected in this case as well[2]. Thus, a similar index – intended to correspond to a phonetically and phonologically-based rating of the impairment – is proposed in this paper, and the expectations in 1) and 2) are checked with reference to such index (see §2.4 for this index calculation).

Therefore, we even intend to verify:

iii. if, and to what extent, a phonetically and phonologically-based classification matches with the clinical assessment.

In this respect, as a working hypothesis, we assume that:

3. clinical evaluation resembles the phonologically-phonetically obtained index of classification.

2.2 Corpus and subjects

The corpus is composed of disyllabic pseudo-words inserted in a carrier phrase such as "La CVC(C)V blu" (transl. "The CVC(C)V blue"). Consonants are all bilabials and can be unvoiced, voiced and nasals; the medial consonant can be singleton and geminate; the vocalic context can be aCiCa o iCaCi (corpus: /'pa.pi/, /'pi.pa/, /'ba.bi/, /'bi.ba/, /'mi.ma/ and corresponding geminates in medial position). Speakers repeated the corpus 6/7 times; listeners heard all repetitions for every item/speaker.

Pathological speakers are 5, all affected by PD and having developed a HD. According to the clinical evaluation, their level of impairment can vary from mild-to-moderate to moderate-to-severe (see Tab. 1). All speakers come from Lecce's area and are aged 64 to 81; speakers declared to have not being diagnosed of any other neuro-cognitive impairment or other speech-language-hearing disease.

[2] For a comparable application of the fit index by Guion et al. (2000) to pathological speech, we refer to Iraci, Grimaldi & Gili Fivela (2017) where indexes attributed to pathological productions have been compared to indexes derived from typical speech. In this study, only the creation of a similar index for the purpose of pathological speech study will be illustrated and used for comparison with the clinical assessment.

Listeners are 11, all coming from Lecce's area, aged 22 to 36 and holding a degree (though from high school to university). They all declared to have never reported any neuro-cognitive impairment or speech-language-hearing disease.

All subjects (speakers and listeners) read the informative sheet and signed the consensus module.

Table 1 - *Summary of pathological speakers*

Speaker	*Age*	*Clinical evaluation*
PD-1	65	Moderate-to-severe
PD-2	81	
PD-4	74	
PD-3	75	Mild-to-moderate
PD-5	64	

2.3 Perceptual test

Target sentences produced by pathological speakers have been automatically segmented[3] and inserted in a data base (10 pseudo-word x 6/7 repetitions x 5 speakers = 335 items). The data base has been randomised and then presented to the listeners in the form of a perceptual test (realised on Praat, Boersma, Weenink, 2009) characterised by two tasks: an identification and a goodness rating task. In details, firstly listeners had to reply to a phonemic categorisation test aimed to the recognition of the word-internal target consonant's status (identification test with forced binary choice, singleton vs. geminate); secondly, an evaluation of the whole sentence's accuracy was required (goodness rating, on a 1-5 Likert scale).

The corpus was illustrated to the listeners in a short training phase in which the target consonant(s), target of the first task, was/were explicitly pointed out. During the same phase, listeners were informed that in second task they would have evaluated the entire production and that, the higher the rating, the most fluent, accurate and well-controlled the production. The test's average duration was about 40 minutes. Listeners used headphones at comfortable volume level and were allowed to take a break whenever they liked. None of them benefited from more than a 5 minutes break.

After starting the script on Praat, a list of instructions appeared on the screen: "Dopo aver ascoltato una parola, clicca sul numero 1 se hai sentito una sola consonante, altrimenti clicca sul numero 2 se hai sentito una consonante doppia. Poi valuta se il parlante ha prodotto la parola che hai ascoltato in maniera fluente, accurata e ben controllata in una scala da 1 a 5" (trans.: "after every sentence, click number 1 if you heard a singleton consonant, or click number 2 if the consonant was geminate. Then evaluate if the sentence you heard was accurate, fluent and well-controlled in a scale from 1 to 5"). Then, clicking on any point on the screen, the first audio stimulus was played and the listener could visual-

[3] For details on data collection and post-processing see Gili Fivela et al. (2014; 2015) and Iraci et al. (2017b).

ise a question: "Hai sentito una consonante scempia o una geminata" (trans. "Did you hear a singleton or a geminate consonant?"). Following the question, two buttons were displayed, one showing a number "1", the other a number "2", respectively corresponding to "singleton" and "geminate". Once clicked on the selected button, another question was unlocked in the lower part of the screen: "La frase era accurata, fluente e ben controllata?" (trans. "To what extent was the sentence accurate, fluent and well-controlled?"). Following the question, five buttons reporting number from 1 to 5 were displayed: once listeners expressed their evaluation by clicking on the selected button, the next audio stimulus was automatically launched (unless the listener decided to take a break).

2.4 Measures

In order to verify our hypothesis we measured:

1. Percentages Of Categorization (POC): the number of times the listener choice concerning the medial consonant matched the expected consonantal status – the expected consonantal status corresponded to the form reported in the script the speakers read in the course of kinematic recordings. Such percentages have been calculated on the whole number of listeners' responses as a function of
   a. consonant status: singleton vs. geminate. For example, in order to calculate the POC of all singletons, we used the following formula (1):

(1) $$POC_{singleton} = \frac{\text{number of all listeners' correct match}}{\text{sample of all speakers' singletons}} \times 100^{4};$$

   b. consonant distinctive features: unvoiced vs. voiced vs. nasal. For example, for unvoiced consonants POC, we used formula (2):

(2) $$POC_{unvoiced} = \frac{\text{number of all listeners' correct match}}{\text{sample of all speakers' unvoiced}} \times 100;$$

   c. the single pathological speaker:

(3) $$POC_{PD\text{-}1} = \frac{\text{number of all listeners' correct match}}{\text{sample of all PD1's realisations}} \times 100;$$

2. Goodness Rating's Average (GRA): values contributed to averages only when the item was correctly categorised in the previous task[5]; averages are calculated

[4] In this paper, "correct match" corresponds to the cases in which the listener's choice corresponds to the script the speaker read during the kinematic production recordings, e.g. when /b/ was identified as /b/. When /b/ was identified as /b:/ (or the opposite), "no match" can be used in the text.

[5] When an item was not identified in the first task, its goodness rating did not contribute to the GRA. This choice is due to the impossibility of accepting an evaluation of accuracy on a word that has not been understood. In other words, the phonological plan failed in being implemented (i.e. the produced consonant(s) in the minimal pair could not be successfully identified). Let's hypothesise the extreme case in which only 1 item's repetition out of 7 was identified (1 correct match, 6 no match) by a given listener, and that item was assigned a very high goodness value: GRA will however be equal to that single goodness value because we assume that speaker to show clear problems to the preservation

out of the number of repetitions of a given pseudo-word produced by a given speaker and evaluated by a single listener, i.e. we calculated one value per word/speaker/listener[6] in order to measure differences as a function of

a. consonant status: singleton vs. geminate
b. consonant distinctive features: unvoiced vs. voiced vs. nasal;
c. the single pathological speaker;

The effect of these three factors were calculated through a statistical test (see §2.5).

3. classification index (INDEX): obtained multiplying the index of categorization (the number of times the item was identified divided by the number of repetitions produced by the PD speaker for the given pseudo-word) by the GRA calculated by item/speaker/listener[7]. As already mentioned, this corresponds to the *fit index* proposed by Guion, Flege, Akahane-Yamada & Pruitt (2000) to investigate the relations between cross-language mapping and discrimination. Indeed, in our opinion the index is useful to measure the accurateness in both L2 or pathological speech as in both cases no perfect match with the ideal production is expected. Factors considered for the analysis of INDEXes are
   a. consonant status: singleton vs. geminate
   b. consonant distinctive features: unvoiced vs. voiced vs. nasal;
   c. the single pathological speaker.

To sum up, for the purpose of this study we will consider

– POC to represent the phonological information since it is generated by the amount of times the minimal pair was not correctly identified;
– GRA to represent the phonetic information since it is generated by the listener's perception of speech accuracy.

Thus, our INDEX is defined as phonologically-phonetically obtained since it is a function of

a. speaker's accuracy (when the phonological plan's execution is not altered)
b. the inferred number of times the phonological plan's execution is altered

(4) $$INDEX_{word} = POC_{word} \times GRA_{word}$$

of the phonological planning, but when the latter is preserved, the phonetic form may be not entailed. GRA measures only the phonetic form.

[6] E.g., listener F6's GRA_{mimma} for PD-2 = 2.75. Listener F6 recognised item /'mim.ma/ produced by speaker PD-2 4 times out of 7. So, 3 values were excluded (the 3 previous task's no match); resting values are 3, 3, 2, 3; the formula is (3+3+2+3)/4 = 2.75.

[7] Given the example in footnote 4, the INDEX attributed to listener F6 for the item /'mim.ma/ produced by PD-2 is equal to 1.57. The listener recognised the item 4 times out of 7, so the formula is (4/7)*2.75= 1.57.

2.5 Statistical tests

Data obtained from the categorization test are calculated in percentage out of:
- all observations concerning singletons (1848 observations) and geminates (1837) produced by all pathological speakers (for the analysis of consonant status);
- all observations concerning unvoiced (1485), voiced (1432) and nasal (770) segments produced by all pathological speakers (for the analysis of distinctive features);
- all observations concerning a pseudo-word (from 77 to 154) for inter-speaker differences[8].

As for GRAs and INDEXes, they have been analysed through a linear mixed effects model (Bates, Maechler, Bolker & Walker, 2014), in the R environment (R Core Team, 2015). The model was aimed to evaluate the effect of the following fixed factors: "consonant status" (henceforth *status*: 2 levels, singleton vs. geminate) and "distinctive features" (henceforth *feature*: 3 levels, unvoiced vs. voiced vs. nasal). Moreover, in order to inform the model that items were produced by 5 speakers, and further analyse inter-speaker differences, a fixed effect called *speaker* was included[9]. Finally, the model was attributed two random effects: one to account for listeners', another for items' variability.

(5) *Dependent variable ~ status + features + speaker + (1|listeners) + (1|items)*

Post-hoc tests have been run with package *multcomp* (Hothorn, Bretz & Westfall, 2008); significance threshold was considered <0.05.

3. *Results*

3.1 Percentages of Categorisation

The categorisation test revealed that singleton consonants have been correctly identified 77,92% of times (1440/1848), while geminates 83,66% (1537/1837).

However, the following percentages reveal differences as a function of consonant's distinctive features: listeners recognised the consonant 94,94% of times (1410/1485) when it was an unvoiced obstruent (including both singletons and geminates), 59,77% (856/1432) when it was a voiced one, and 87,01% (670/770) when nasal.

[8] Observations are to be considered always out of the total of all listeners. E.g. 77 observations for PD-3's /'mi.ma/ means that speaker produced 7 repetitions of that pseudo-word but the sample is calculated out of 11 listeners.

[9] Of course, no factors have been included as to account for speech impairment level since we wanted this information to arise from our results.

Table 2 - *Speakers ranked by global POCs*

Speakers	*in %*
PD-5	89,58
PD-3	82,46
PD-4	79,41
PD-2	77,2
PD-1	75,89

Considering the observations per speaker, pathological subjects report POCs between 75% and 89%. In particular, PD-5's POC is 89,58% (611/682), PD-3's 82,46% (635/770), PD-4's 79,41% (594/748), PD-2's 77,2% (586/759), PD-1's 75,89% (551/726).

Concerning PD-1, $POC_{singleton}$ is 96,41% (350/363), $POC_{geminate}$ is 55,37% (201/363): in particular, $POC_{singleton}$ of only unvoiced segments is 99,35% (153/154), $POC_{singleton}$ of voiced segments 99,24% (131/132), $POC_{singleton}$ for nasals is 85,71% (66/77); $POC_{geminate}$ for unvoiced, voiced and nasals is respectively 70,12% (108/154), 27,27% (36/132) and 74,02% (57/77)[10].

PD-2's $POC_{singleton}$ is 71,42% (275/385), while $POC_{geminate}$ is 83,15% (311/374): as for singletons, unvoiced segments were identified 92,85% of times (143/154), voiced segments 38,31% of times (59/154), nasals 94,8% of times (73/77); as for geminates, unvoiced segments were identified 100% of times (154/154), voiced segments 76,22% of times (109/143), nasals 62,33% (48/77).

PD-3 reports $POC_{singleton}$ of 68,05% (262/385), and $POC_{geminate}$ of 96,88% (373/385): concerning singletons, respectively, unvoiced, voiced and nasal segments were identified 99,35% (153/154), 20,77% (32/154), and 100% of times (77/77); concerning geminates, unvoiced, voiced and nasal segments, were respectively identified 99,35% (153/154), 100% (154/154) and 85,71% of times (66/77).

As for PD-4, $POC_{singleton}$ is 70,58% (264/374), $POC_{geminate}$ is 88.23% (330/374): in particular, $POC_{singleton}$ of only unvoiced segments is 94,15% (145/154), $POC_{singleton}$ of voiced segments 31,46% (45/143), $POC_{singleton}$ for nasals is 96,1% (74/77); $POC_{geminate}$ of only unvoiced segments is 99,3% (142/143), $POC_{geminate}$ of voiced segments is 77,27% (119/154), $POC_{geminate}$ of nasals is, 89,61% (69/77).

Concerning PD-5, $POC_{singleton}$ is 84,75% (289/341), $POC_{geminate}$ is 94,42% (322/341): $POC_{singleton}$ for only unvoiced, voiced, and nasal segments respectively is 98,48% (130/132), 72,72% (96/132), and 81,81% (63/77); $POC_{geminate}$ for unvoiced, voiced and nasals is respectively 97,72% (129/132), 87,87% (116/132) and 100% (77/77).

[10] All speakers' results are grouped in Tab. 3 for better reading.

Table 3 - *Summary of by-speaker/by-word POCs*

In %	PD-1		PD-2		PD-3		PD-4		PD-5	
	sing	gem	sing	gem	sing	gem	sing	gem	sing	gem
Unvoic.	99,35	70,12	92,85	100	99,35	99,35	94,15	99,3	98,48	97,72
Voiced	99,24	27,27	38,31	76,22	20,77	100	31.46	77,27	72,72	87,87
Nasal	85,71	74,02	94,8	62,33	100	85,71	96,1	89,61	81,81	100

3.2 Goodness Rating Averages

GRA does not change as a function of *status* ($\chi2(1)=3.02$, $p=0.08$) but significantly varies both as a function of *feature* ($\chi2(2)=13.38$, $p=0.001$) and *speaker* ($\chi2(4)=211.8$, $p<0.0001$), with an interaction between *status* e *feature* ($\chi2(2)=18.86$, $p=0.0008$). GRA is lower when the target consonant is voiced, if compared to nasals and unvoiced respectively. Concerning the factor *speaker*, GRA differs by *speaker* as follows: PD-2 < PD-4, PD-1 < PD-5 < PD-3. The post-hoc on the interaction shows that GRA is definitely lower in case of voiced singleton; voiced and nasal geminates report intermediate values; finally, higher values are reported in case of unvoiced singletons and geminates (the nasal singleton let report intermediate, but not significantly different values between the last two groups).

3.3 Index of classification

INDEX does not change as a function of *status* ($\chi2(1)=1.43$, $p=0.23$) but factors *feature* and *speaker* are both significant (respectively [$\chi2(1)=15.5$, $p=0.0004$] and [$\chi2(4)=151.51$, $p<0.0001$]), as well as the interaction between *status* and *feature* ($\chi2(2)=18.93$, $p=0.0007$). INDEX is lower when *speaker*s are realising voiced consonants; higher values are reported respectively in case of nasals and unvoiced. According to the factor *speaker*, pathological speakers are exactly grouped following the clinically established level of impairment: PD-2, PD-4, PD-1 < PD-5, PD-1. The post-hoc reveals that INDEX values distribute as follows in ascending order: voiced singleton < voiced geminate < nasal geminate (nasal singleton) < unvoiced geminate, unvoiced singleton. Mean INDEXes grouped by pseudo-word and speaker are reported in the following table:

Table 4 - *Summary of by-speaker/by-word mean INDEXes*

	PD-1		PD-2		PD-3		PD-4		PD-5	
	sing	gem	sing	gem	sing	gem	sing	gem	sing	gem
[‘pa.pi]	3.99	3.65	2.42	3.58	4.8	4.72	3.19	3.63	4.20	4.07
[‘pi.pa]	4.11	2.20	2.94	3.54	4.67	4.47	3.52	3.25	4.02	4.31
[‘ba.bi]	3.18	0.55	0.44	2.37	0.82	4.63	0.87	2.12	2.98	2.86
[‘bi.ba]	3.59	1.10	1.04	2.36	0.33	4.73	0.59	2.77	1.99	3.81
[‘mi.ma]	2.72	2.36	2.87	1.51	4.84	4.04	3.18	2.73	2.90	3.75

4. *Discussion*

Overall results from the first test (identification test with forced binary choice, singleton vs. geminate) showed higher rates of correct match in the case of geminates, rather than singletons.

Despite so, analysing this outcome in terms of distinctive features can provide a different point of view. Looking at results for only unvoiced consonants (the least in a scale of phonetic demand), it is possible to notice that speakers' productions are recognised in a higher percentage of cases, with the only exception being PD-1's unvoiced geminates (<75%). In fact, PD-2 and PD-4 show lower percentages in case of unvoiced singletons if compared to other speakers, though these are greater than 90%. The productions by most impaired speakers (PD-1, PD-2 and PD-4) seem to be more easily identified if they were intended to correspond to singletons (PD-1) or geminates (PD-2 and PD-4). Since subjects are said to differently compensate, this result may be seen in terms of subjective "preferences" for two hypothesised distinct patterns of production. The first (related to PD1) maybe a preference for a CV.CV pattern, or a general tendency towards hypoarticulation with no explicit compensation leading to correctly articulating singletons, while showing reduction in case of switch to a different dynamical regime (e.g. in case of geminates, where syllable structure is supposed to change to a structure as CVC.CV). The second (PD-2 and PD-4) maybe a preference for a CVC.CV pattern, where compensation to hypoarticulation leads to lengthen the slots of time useful for articulating sounds (thus showing target reaching only in case of geminates).

When considering nasality, the productions by all speakers generally correspond to comparable or lower POCs if compared to results for unvoiced segments. Looking at plain data, when comparing nasal singletons to unvoiced singletons, no evident distinctions arise: PD-1 and PD-5 report lower proportions in the case of nasals; PD-2, PD-3 and PD-4 nasal singleton > unvoiced singleton but these differences seem to be negligible (respectively in percentage, PD-2, 94,8 vs. 92,85; PD-3, 100 vs. 99,35; PD-4 96,1 vs. 94,15). Turning the comparison to geminates, it is possible to notice that PD-2, PD-3 and PD-4 report lower POCs in the case of nasals, while PD-1 and PD-5 nasal geminate > unvoiced geminate, but proportions are comparable (respectively in percentage, PD-1 74,02 vs. 70,12; PD-5 100 vs. 97,72). This result suggest that nasality can slightly worsen the identification of minimal pairs. Nevertheless, major differences lay in the relationship between singletons and geminates belonging to the same group (i.e. nasal singletons and geminates): all speakers show lower POCs for geminates (with the only slight exception being PD-5) as if, increasing phonetic demand (that is, adding a nasal gesture), there is a general preference for CV.CV patterns, thus showing a general trend towards degemination. It is crucial to consider that in the case of nasal consonants not only a linguistic feature is added, but a specific clinical factor has to be considered as well: a great number of subjects affected by PD can show a VPD (see §1.4; cfr., for example, Hammer et al., 2011 for the treatment of VPD in PD through Deep Brain

Stimulation), consisting in inadequate velopharyngeal control, the latter crucial for the realisation of velum gestures for nasal consonants.

Results for voiced segments are to be considered apart from the rest of data because of the regional variety typical pronunciation of intervocalic voiced bilabial plosives as lengthened. Indeed, in production it is possible to notice extreme values, such as for PD-2, PD-3 and PD-4 that probably always produced lengthened consonants due to *rafforzamento*, and PD-1 that, instead, still confirms his preference for simple syllable structures maybe just due to greater difficulties in producing longer segments (which could be related to the slightly lower POCs obtained by his geminates). On the contrary, PD-5 report uncertain results, as no clear trend is detectable, probably because, in these cases, it was very hard to distinguish between a singleton showing *rafforzamento* and a geminate.

Finally, looking at POCs registered for every speaker out of the total of realisations, speakers seem to follow the clinical evaluation's tendency since most impaired speakers have been attributed to values under 80%, and less impaired ones to higher values: PD-1 (75,89), PD-2 (77,2), PD-4 (79,41), PD-3 (82,46), PD-5 (89,58). However, POCs seem to suggest a scalar difference in the productions of most PD speakers and a quite clear differentiation of PD-5's productions only. It may be the case, then, that POCs allow a more fine grained classification of speech by PD speakers which are borderline at a specific intermediate impairment level.

Results from the second task show that there are no consistent differences in terms of accuracy, depending on the alternation of singleton and geminate consonants. In particular, significant differences are found because of the alternation of distinctive features here considered, and by-word results are even helpful for inferring information on the relationship between singletons and geminates: lowest GRAs are attributed to the voiced singleton, followed by the respective geminate that, in turn, is accompanied by the nasal geminate; higher GRAs are attributed to unvoiced segments, while the nasal singleton lays somewhere in between the last two groups. Though differently arisen, the same picture depicted in discussing results from the first test seems to come out of this test as well. Possibly, there are fewer differences due to the alternation of singletons and geminates when segments are unvoiced (that is, for the lowest phonetic demand). But when segments are nasalised, geminates are regularly more inaccurate. On the contrary, voiced segments are definitely not accurate when they are singletons, probably because they can be hardly distinguishable from geminates, mainly because of *rafforzamento*; otherwise because, if not lengthened, they might even be approximated. By-speaker's result still look very close to both POCs and clinical evaluation's trend: PD-2 < PD-4, PD-1 < PD-5 < PD-3.

The INDEX produced after crossing the two information obtained "classifies" single words realisation very similarly to what seen in terms of GRAs, with the only difference being the nasal geminate, this time differing also from the voiced geminate. Again, it is interesting to notice that no statistical differences are found for the *status* factor, but the nasal singleton seem to be better perceived than the respective

geminate. Looking at Tab. 3 – and bearing in mind Tab. 2 – it is possible to notice that those preferences exhibited in relation with productions by most impaired speakers (in terms of POCs) for singletons (PD-1) or geminates (PD-2 and PD-4), now are slightly more evident, though not completely for PD-4. Nasality still lowers INDEXes and the nasal singleton (rather than the geminate) is again confirmed as the best output, as shown by listeners, together with the only exception of PD-5. Voicing strongly lowers INDEXes for the reasons already exposed and tied to the regional variety. Finally, it is worth to notice that the clinical evaluation matches with the statistically-individuated inter-speaker differences (PD-2, PD-4, PD-1 < PD-5, PD-3).

5. *Summary and Conclusions*

PD speakers are able to realise meaningful differences mainly based on articulatory gestures' amplitude and duration despite Hypokinetic Dysarthria entails disturbances to the execution and control of speech gestures' amplitude and coordination. According to our previous works, in order to do so, PD speakers seem to exploit some compensatory strategies. These strategies are supposed to be aimed to the conservation of the phonological plan, though at the expenses of speech accuracy. Our hypothesis is that phonological constraints drive the compensation, but constraints due to the pathology act on contiguous gestures. Hence, Hypokinetic Dysarthria would not carry any direct effect (but only some indirect effects) at the phonological level. As a consequence, speech alterations would mainly remain within a range of phonetic variation that is consistent with the expected message interpretation, apart from the case of speech produced by high severity level patients. Thus we believe that measuring the amount of alterations at both the phonological and the phonetic level might provide a satisfying and objective description (and hopefully evaluation) of the speech impairment. In order to obtain a measure of this kind we set-up a perceptual experiment that allowed us to extract these two source of linguistic information, and further crossed the two information for the purpose of evaluation.

First, we wondered whether minimal pairs (differing for the medial consonant being singleton or geminates) realised by pathological speakers, and differentiated on the kinematic and acoustic dimension, were correctly categorised by non-pathological listeners. We hypothesised that the higher the speakers' impairment, the higher the number of minimal pairs not correctly categorised by listeners. We considered the impairment level expected on the basis of the clinical evaluation and that calculated on the basis of listeners' judgements and, further, we compared the two. Results have been analysed in order to even check for the influence of consonant status (singleton, geminate), distinctive features (voiced, unvoiced, nasal, plosive), and the single speaker. Less impaired speakers, eventually, showed no relevant alterations at the phonological level. On the contrary, according to the POCs registered as a function of the singleton vs. geminate relation, most impaired speakers

showed some alterations differently distributed for the presence/absence of some specific distinctive features (nasality).

At the lowest level of phonetic demand (unvoiced bilabial plosive segments), most impaired speakers are hypothesised to follow two distinct patterns. When the lowest percentages of match have been found in case of geminates (e.g. PD-1), a preference for a CV.CV pattern was hypothesised. This should be due to a general tendency towards hypoarticulation with no explicit compensation, leading to correctly articulating singletons, but showing reduction in case of switch to a different dynamical regime (e.g. in case of geminates where syllable structure is supposed to change to a structure as CVC.CV). The second hypothesised pattern may be a preference for a CVC.CV pattern (e.g. PD-2 and PD-4). In this case, compensation to hypoarticulation leads to target reaching only in case of geminates, that is when a lengthened slot of time is available for articulating sounds.

Out of these two supposed patterns, independently of the clinically-established level of impairment, only the first was found in case of nasals. This means that nasal geminates were identified less frequently than nasal singletons. We suppose this happened not only because of the increase of phonetic demand, but also because most people affected by PD can show an inadequate control of the velopharyngeal sphincter (VPD) that probably prevent pathological speakers from switching to a different dynamical regime. Indeed, the nasal geminate seem to be generally penalised if compared to the nasal singleton and the unvoiced segments. However, we need further check on nasal segments and on the switch to different dynamical regime in order to confirm what supposed.

Voiced segments represented a special case in that, in the area where recordings took place, the singleton shows *rafforzamento*. For this reason, it was expected to be hardly distinguished from the respective geminate. Anomalous results were expected and soon found: two speakers have been likely considered to always realise lengthened singletons, while two other speakers represented a matter of confusion for our listeners in both cases. One speaker (PD-1) is an exception again, as he mainly produced segments identifiable as singletons and showed clear difficulties with geminates.

Independently of the status of the consonant and/or distinctive features, speakers can be roughly grouped by POCs (i.e. the identification of phonological categories) into two different levels yet at this stage, resembling the clinical assessment, though not precisely.

Secondly, we wondered the extent to which listeners would have evaluated the phonetic accuracy. We hypothesised that the higher the impairment, the lower the accuracy of pathological productions. The status of the consonant seemed to play a clear role on the accuracy. The nasal geminate still have been penalised by listeners, probably because of the reasons exposed before. Surprisingly, even the voiced segments was attributed to low values. Concerning the latter, results in line with our main hypothesis were expected since the phenomenon of *rafforzamento* was hypothesised not to influence the phonetic accuracy of the consonant realisation.

Instead, the lowest GRAs have been reported exactly in case of voiced segments. Listeners referred that it was always very hard to distinguish between a bilabial singleton vs. geminate so, often, they assigned low goodness ratings to items that in the previous task were identified with difficulties (i.e. listeners were not self-confident with what they were evaluating). Moreover, it has to be considered that the singleton showing *rafforzamento* is not a geminate but a sort of halfway. This fact probably generates confusion because listeners are asked to choose between two categories that actually manifest themselves as three ones, and probably not always realised within the ranges of expected (non-pathological) phonetic variation. These three categories are not necessarily distinguishable from each other since, in this area, a trustful contrast between a real singleton vs. a real geminate in the case of voiced bilabial consonants can be found only in very highly-controlled realisations.

At this stage, speakers grouped by GRAs (i.e. the degree of appreciation of the actual phonetic implementation) reflected a trend similar to the clinically-established level of impairment, though slightly different from the trend obtained when speakers have been grouped by POCs (i.e. the phonological information). It is likely that compensatory strategies have an amount of "subjectivity" such as non-pathological speech with subjective idiosyncrasies. On the contrary, linguistic constraints seem to play a stronger role and the phonetic alteration seem to be directly proportional to the phonological one. This is the case of voiced segments that, since they were not easily recognised, they have been even evaluated as inaccurate. Or still, it is the case of the nasal geminate, that in the previous task resulted to be among the most penalised and, when identified, it even resulted to be quite inaccurately produced.

Finally, the phonetically and phonologically-based INDEX obtained showed to be sensitive to differences for the status of the consonant (despite this did not come from statistical tests) and to distinctive features differences already individuated across the two previously reported analysis. Again nasality played a crucial role lowering values, and nasal geminates were finally penalised. Concerning the influence of VPD in the realisation of geminates, it seems plausible that a mechanical/physical (pathological) constraint limit the phonological structure's expected execution. In gesture-based phonologies like Articulatory Phonology, linguistic constraints are supposed to interact with mechanical/physical constraints, finally producing the phonetic execution with its amount of variation. When pathological constraints intervene in this process, variation can go out of what listeners perceive as an admitted range in their mother-tongue. Without phonological constraints (whose absence would be attributed to lesions to speech production/comprehension's areas, rather than to MSD) speech probably would be totally disrupted. In our opinion, in HD related to PD, compensatory strategies are supposed to be driven by phonological constraints. This means that the articulators, hindered by pathological constraints, (i) may tend towards a new resting position, and/or (ii) may reach targets with evident limits in the phonetic parameters (amplitude, duration), but without limits in the phonological/dynamic parameters (target, stiffness, damping

coefficients) of the gesture (cfr. Nam, Saltzman, 2003; Goldstein, Byrd & Saltzman, 2006). Nevertheless this is just a hypothesis that still has to be confirmed crossing this data with previously recorded kinematic data.

Thus, summing up, although with the abovementioned differences related to the variation in articulatory demand required for different distinctive features and despite the anomalous results related to bilabial voiced plosives, our findings confirm that

1. the higher the impairment, the higher the number of minimal pairs not correctly categorised by listeners;
2. the higher the impairment, the lower the accuracy of pathological productions;
3. the phonologically-phonetically obtained index of classification resembles clinical evaluation.

Thus it is supposed that basing the clinical evaluation on some objective indicators of the phonological level alterations and of the phonetic accuracy alterations as well, it is possible to offer an objective evaluation of the speech impairment at least in HD, hopefully to extend to MSDs in general.

6. *Future investigations*

Findings from this study suggest at least two directions for further researches. First, this classification method needs to be tested on a greater number of pathological speakers and possibly listeners as well; further, it needs to be tested on different types of MSDs. Second, we will cross perceptual data with acoustic and kinematic data in order to look at every realisation from all possible points of view: articulation, acoustic and perception. This will allow us to focus on the reciprocal interactions between the three phonetic dimensions.

Bibliography

Ackermann, H., Ziegler, W. (1991). Articulatory deficits in Parkinsonian dysarthria: an acoustic analysis. In *Journal of Neurology, Neurosurgery, and Psychiatry*, 54, 1093-8.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-5. http://CRAN.R-project.org/package=lme4/Accessed 18.05.16.

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

Browman, C., Goldstein, L. (1989). Articulatory gestures as phonological units. In *Phonology*, 6, 201-51.

Browman, C., Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J., Beckman, M. (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge: Cambridge University Press, 341-97.

Connor, N.P., Abbs, J.H., Cole, K.J. & Gracco, V.L. (1989). Parkinsonian deficits in serial multiarticulate movements for speech. In *Brain*, 112, 997-1009.

Darley, F.L., Aronson, A.E. & Brown, J.R. (1969). Differential Diagnostic Patterns of Dysarthria. In *Journal of Speech, Language, and Hearing Research*, June 1969, 12, 246-69.

Duffy, J.R. (2005). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. 2°ed. Elsevier Mosby.

Gafos, A., Goldstein, L. (2011). Articulatory representation and organization. In Cohn, A., Huffman, M. & Fougéron, C. (Eds.), *Handbook of Laboratory Phonology*. Oxford University Press.

Gaillard-Corvaglia, A., Kamiyama, T. (2008). La /b/ "forte" in salentino (Puglia): uno studio acustico, percettivo e fisiologico. In Pettorino, M. (Ed.), *La Comunicazione Parlata: atti del congresso internazionale*, Napoli, 23-25 febbraio 2006. Napoli: Liguori, 87-99.

Gili Fivela, B., d'Apolito, S., Stella, A. & Sigona, F. (2008). Domain Initial Strengthening in sentences and paragraphs: preliminary findings on the production of voiced bilabial plosives in two varieties of Italian. In Sock, R., Fuchs, S., Laprie, Y. (Eds.), *Proceedings of the International Seminar on Speech Production*, Strasbourg, France, 8-12 December, 205-208.

Gili Fivela, B., d'Apolito, S., Stella, A. & Sigona, F. (2010). Domain initial strengthening: dati acustici e articolatori relativi a due varietà di italiano. In Cutugno, F., Maturi, P., Savy, R., Abete, G. & Alfano, I. (Eds.), *Parlare con le persone, parlare alle macchine: la dimensione interazionale della comunicazione verbale, Atti del VI Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*, Università di Napoli, 3-5 febbraio 2010. Torriana (RN): EDK Editore, 173-195.

Gili Fivela, B., Iraci, M.M., Grimaldi, M. & Zmarich, C. (2015). Consonanti scempie e geminate nel Morbo di Parkinson: la produzione di bilabiali. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio/Language acquisition and language loss. Acquisition, change and disorders of the language sound structure*. Milano: AISV, 247-273.

Gili Fivela, B., Iraci, M.M., Sallustio, V., Grimaldi, M., Zmarich, C. & Patrocinio, D. (2014). Italian Vowel and Consonant (co)articulation in Parkinson's Disease: extreme or reduced articulatory variability?. In *Proceedings of the 10th International Seminar on Speech Production (ISSP'14)*, Cologne, Germany, 5-8 May 2014, 146-9.

Gili Fivela, B., Zmarich, C. (2005). Italian Geminates under Speech Rate and Focalization Changes: Kinematic, Acoustic, and Perception Data. In *InterSpeech 2005*, Lisbon, Portugal, 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 *Proceedings of International Conference of Phonetic Sciences (ICPhS'07)*, Saarbrücken, Germany, 6-10 August 2007, 469-472.

Goldstein, L., Byrd, D. & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In Arbib, M.A. (Ed.), *Action to Language via the Mirror Neuron System*. Cambridge: Cambridge University Press, 215-49.

Guion, S.G., Flege, J.E., Akahane-Yamada, R. & Pruitt, J.C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. In *Journal of the Acoustical Society of America*, 107(5), Pt. 1, May 2000, 2711-2724.

Hammer, M.J., Barlow, S.M., Lyonsc, K.E. & Pahwac, R. (2011). Subthalamic nucleus deep brain stimulation changes velopharyngeal control in Parkinson's disease. In *Journal of Communication Disorders*, 44(1), 37-48.

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. In *Biometrical Journal*, 50(3), 346-363.

Iraci, M.M., Grimaldi, M. & Gili Fivela, B. (2017). Dalla L2 al parlato patologico: un indice di categorizzazione fonemica per la valutazione dell'intelligibilità del parlato disartrico. In *XVII Congresso Internazionale dell'Associazione Italiana di Linguistica Applicata (AItLA)*, Università degli Studi di Napoli "L'Orientale" - Università della Campania "Luigi Vanvitelli", S. Maria Capua Vetere, 23-25 febbraio.

Iraci, M.M., Grimaldi, M. & Gili Fivela, B. (in revisione). Il contributo della Fonologia alla riabilitazione logopedica personalizzata di soggetti parkinsoniani disartrici. In Dovetto, F. (Ed.), *Lingua e Patologia: le frontiere interdisciplinari del linguaggio, Linguistica delle differenze*, II incontro di studio tra Medici e Linguisti, Accademia Pontaniana e Societa Napoletana di Scienze, Lettere ed Arti, Napoli, 10-11 dicembre 2015.

Iraci, M.M., Zmarich, C., Grimaldi, M. & Gili Fivela, B. (2017b). Il parlato nel morbo di Parkinson: ampiezza dei gesti articolatori e distintività dei suoni linguistici. In Sorianello, P. (Ed.), *Il linguaggio disturbato. Modelli, strumenti, dati empirici*. Roma: Aracne Editrice.

Loporcaro, M. (1996). On the analysis of geminates in Standard Italian and Italian dialects. In Hurch, B., Rhodes, R. (Eds.), *Natural Phonology: The State of the Art. Papers from the Bern Workshop on Natural Phonology*, September 1989. Berlin-New York-Amsterdam: Mouton de Gruyter, 153-87.

McCabe, P. (2010). Advances in Motor Learning: Emerging evidence and new ideas. In *ACQuiring Knowledge in Speech, Language and Hearing*, 12(1), 3-1.

Nam, H., Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 12th International Congress of Phonetic Sciences*, Barcelona, 2253-6.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/Accessed 18.05.16.

Schröter-Morasch, H., Ziegler, W. (2005). Rehabilitation of impaired speech function (dysarthria, dysglossia). In *GMS current topics in otorhinolaryngology, head and neck surgery*, 4, 1-13.

Skodda, S., Gronheit, W. & Schlegel, U. (2012). Impairment of Vowel Articulation as a Possible Marker of Disease Progression in Parkinson's Disease. In *PLoS ONE*, 7(2), 1-8.

Skodda, S., Visser, W. & Schlegel, U. (2011). Vowel Articulation in Parkinson's Disease. In *Journal of Voice*, 25, 4, 467-72.

Tjaden, K. (2000). An Acoustic Study of Coarticulation in Dysarthric Speakers with Parkinson's Disease. In *Journal of Speech, Language and Hearing Research*, December 2000, 43, 1466-1480.

Tjaden, K. (2003). Anticipatory Coarticulation in Multiple Sclerosis and Parkinson's Disease. In *Journal of Speech, Language and Hearing Research*, August 2003, 46, 990-1008.

Tjaden, K., Wilding, G.E. (2005). Effect of rate reduction and increased loudness on acoustic measures of anticipatory coarticulation in multiple sclerosis and Parkinson's disease. In *Journal of Speech, Language and Hearing Research*, April 2005, 48(2), 261-77.

Weismer, G., Yunusova, Y. & Westbury, J.R. (2003). Interarticulator Coordination in Dysarthria: An X-ray Microbeam Study. In *Journal of Speech, Language and Hearing Research*, 46, October 2003, 1247-1261.

Wong, M.N., Murdoch, B.E. & Whelan, B.M. (2010). Kinematic analysis of lingual function in dysarthric speakers with Parkinson's disease: An electromagnetic articulograph study. In *International Journal of Speech-Language Pathology*, 12, 414-25.

Wong, M.N., Murdoch, B.E. & Whelan, B.M. (2011). Lingual Kinematics in Dysarthric and Nondysarthric Speakers with Parkinson's Disease. In *Parkinson's Disease*, 1-8.

Zhang, Y., Jiang, J. & Rahn, D.A. (2005). Studying vocal fold vibrations in Parkinson's Disease with a nonlinear model. In *Chaos*, 15, 033903, 1-10.

Zmarich, C., Gili Fivela, B. (2005). Consonanti scempie e geminate in italiano: studio cinematico e percettivo dell'articolazione bilabiale e labiodentale. In *Atti del Convegno Nazionale AISV (Associazione Italiana di Scienze della Voce) "Misura dei parametri"*, Padova, 2-4 dicembre 2004. ISBN 88-88974-69-5. Torriana (RN): EDK, 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 (Associazione Italiana di Scienze della Voce) "Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche"*, Povo di Trento, 29-30 November-1 December 2006. Torriana (RN): EDK, 151-163.

Zmarich, C., Gili Fivela, B., Perrier, P., Savariaux, C. & Tisato, G. (2009). L'organizzazione temporale dei gesti vocalici e consonantici nelle consonanti scempie e geminate dell'Italiano. In Romito, L., Galatà, V. (Eds.), *Atti del IV Convegno Nazionale AISV (Associazione Italiana di Scienze della Voce) "La Fonetica Sperimentale: Metodo e Applicazioni"*. Torriana (RN): EDK, 89-104.

Zmarich, C., Gili Fivela, B., Perrier, P., Savariaux, C. & Tisato, G. (2011). Speech timing organization for the phonological length contrast in Italian consonants. In Cosi, P., De Mori, R., Di Fabbrizio, G. & Pieraccini, R. (Eds.), *Proceedings of InterSpeech 2011*, Florence, 28-31 August 2011, 401-4.

DOI: 10.17469/O2102AISV000023

CLAUDIO ZMARICH, SIMONA BERNARDINI, GIOVANNA LENOCI,
GIULIA NATARELLI, CATERINA PISCIOTTA

Could the frequency of Stuttering-Like-Disfluencies predict persistent stuttering in children who have just started to stutter?

Stuttering onset occurs for 95% of people who begin to stutter before the age of 4 years, tipically in the third year of life. *Spontaneous recovery* during childhood is common, with recovery rates estimated at 68-96%, usually no later than the fourth year post-onset. If symptoms persist beyond this time, the efficacy of treatment might result more problematic. As a result it is important to refer the subjects that tend to persist to early treatment. The CNR project "Phonetic indexes predictive of chronic stuttering in preschool children" started in 2008, and aimed to identify, among different behavioural indexes, the ones which are able to predict stuttering persistence at early ages, in order to assure to more at-risk subjects the best therapeutic interventions. The aim of the current study is to evaluate the clinical efficacy of the *Disfluency Profile* (i.e. the percentage of Stuttering-Like Disfluencies over 100 spoken syllables) in identifying children at greater risk of persistence. Results of the study suggest that the predictive power of the *Disfluency Profile* at the session of 9-15 months post-onset is low, according to clinical standards of sensitivity and specificity, but it increases over the next six months, albeit not to the standard minimum of 80%. However, the use of the Disfluency Profile is preferable to the *Stuttering Severity Instrument* (third edition), which has been proposed as a predictive tool by some researchers.

Key words: disfluencies, children, stuttering persistence, stuttering recovery, developmental stuttering.

Introduction

It is well known that stuttering is characterized by involuntary part-word and monosyllabic word repetitions, as well as disrhythmic phonations which consist of prolongations of sounds and/or arrests of speech. Yairi, Ambrose (2005) referred to these as 'Stuttering-Like Disfluencies' (SLD) to distinguish them from 'Other Disfluencies' (OD). OD typify the speech of people who do not stutter, and include multisyllable/phrase repetitions, interjections, revisions and incomplete utterances. According to Yairi, Ambrose (2005), for 95% of people who begin to stutter, the onset occurs before 4 years of age (see also Reilly, Onslow, Packman, Cini, Ukoumune, Bavin, Prior, Eadie, Block & Wake, 2013). Among preschool children who experience stuttering, approximately 90% will recover spontaneously, usually within 4 years post-onset (Yairi, Ambrose, 2013). If symptoms persist beyond this time, then natural recovery rarely occurs without any intervention, and treatment itself risks not to be successful. Furthermore, it is well known that stuttering is characterized by a strong hereditary component (Kraft, Yairi, 2011) and is often associated with negative attitudes towards communicative situations (Clark,

Conture, Frankel & Walden, 2012). Because of these evidences, the clinical importance of early treatment for children at high risk of persistent stuttering becomes apparent.

The purpose of the research project "Phonetic indexes predictive of persistent stuttering in preschool children" (CNR-RSTL n. 995, granted in 2007 to the first author) is to identify some behavioural indexes that might serve as predictors of persistence at early ages.

Specifically, several clinical factors in previous researches have been proved quite able to differentiate children at high risk of chronic stuttering from those who are likely to recover (Yairi, Ambrose, 2005: 348). Apart from biological factors, like to be male or to have one or both the parents who still stutter, we choose to verify the following statements in the second year after the onset there is a risk of persistence if: 1) the SLD percentage (%) remains relatively high (Yairi, Ambrose, 2005); 2) there is a high degree of Consonant-Vowel coarticulation in perceptually fluent speech (Subramanian, Yairi & Amir, 2003); 3) there is a mal-attitude toward speech (negative attitude towards communication is high). This contribution only deals with the first assumption, and it is based on the SLD's developmental paths of the persistent and recovered stutterers as described by Yairi, Ambrose in the *Illinois Longitudinal Study* (2005, Chapt. 5, see table 1). Our choice of this measure was driven by current best research evidence and by the criterion of clinical suitability, and for the last reason we did not consider other not frankly clinical measures (such as the acoustic ones).

Table 1 - *Mean (standard deviation) of SLD and OD for Persistent, Recovered and Control Groups of Children (modified from Yairi, Ambrose, 2005: 173)*

Subjects	0-6 months	7-12 m.	13-18 m.	19-24 m.	25-36 m.	37-48 m.	49-60 m.
Persistent							
SLD	11.31 (6.12)	9.76 (6.32)	7.82 (5.31)	7.34 (6.75)	7.93 (6.40)	5.85 (8.37)	3.61 (4.42)
OD	11.03 (6.74)	5.41 (2.09)	5.42 (2.48)	5.75 (3.57)	7.49 (4.20)	5.54 (1.67)	6.38 (3.01)
Recovered							
SLD	11.03 (6.74)	5.38 (4.37)	3.01 (2.65)	1.99 (1.51)	1.62 (1.56)	1.10 (0.81)	0.91 (0.64)
OD	5.85 (3.00)	5.21 (2.34)	5.13 (2.92)	4.80 (2.25)	4.93 (2.89)	5.07 (2.06)	5.75 (2.51)
Controls							
SLD	1.42 (1.01)		1.11 (0.79)		1.08 (0.97)	0.93 (0.89)	
OD	4.42 (2.27)		4.39 (1.60)		4.67 (2.17)	5.42 (2.02)	

As shown in table 1, the SLD percentage of the recovered stutterers was diminished by nearly half between the first (11.03%) and the second semester post-onset

(5.38%), and it was systematically reduced in the following semesters. Conversely, the SLDs rate reduction (if any) of the persistent stutterers was very low over the same period (11.31% in the first and 9.76% in the second semester).

This kind of results could be discovered only by research projects that are longitudinal in design. In fact, the emerging view on stuttering considers it as a neurodevelopmental disorder involving multiple variables, including motor, language and emotional factors (Smith, 2016). Only projects designed to follow children who stutter from the onset to the final remission or persistence of the disorder (at least 5 years long), like the Illinois Longitudinal Project (Yairi, Ambrose, 2005) or the Purdue Stuttering Project (Smith, 2016), not to say of the prospective "Victoria Study" (see Reilly et al., 2013) are equipped to track these factors.

We started from the results of our recent report on 10 Italian children who stutter (Zmarich, 2015), whose persistence was better predicted by the "16-22 months" scores of the *Disfluency Profile* than the "9-15 months" scores. However, the accuracy of prediction of disfluency profile at 16-22 months was not much better than the prediction based on the severity scores (SSI-3, Riley, 1994). As a consequence, the application of these instruments seemed result in similar outcomes.

Here we present the data regarding 13 children (they constitute the whole sample at the last session of the project), ten of which were already participants of the previous report (Zmarich, 2015). The present update is justified by two main reasons: (1) the increase of the sample by three subjects, whose analyses have been completed after Zmarich (2015); (2) the increased post-onset interval (almost two years more), which adds reliability to the assessment of the final clinical outcome. The experimental purpose is to evaluate which of the following scores could best predict persistence at different stages and with different tools: the *Disfluency Profile* at 9-15 months post-onset, the *Disfluency Profile* at 16-22 months post-onset, or the SSI-3 scores at 16-22 months?

1. *The CNR Project*

The CNR project started in 2008, and aimed to identify some behavioural indexes to predict persistence or recovery from stuttering.

Forty families were enrolled in the study when the following criteria/conditions were satisfied: at least one member who stuttered at that time (or had already stuttered) and a child aged 12-23 months in the same family. In this way we could be sure to maximize the probability to collect stuttering cases (see Kloth, Kraaimaat, Janssen & Brutten, 1999, for a similar design).

All children were first audio and video recorded when they were 24-months-old, and as soon as a child showed the first symptoms of stuttering, he/she was addressed to the "Centro Medico di Foniatria" (CMF) in Padua in order to receive a formal diagnosis, and to be evaluated for speech and language abilities in the attempt to rule out other main diseases (Zmarich, 2015). At the same time, the child began to be audio and video recorded at home, every 3 months up to 16-22

months post-onset (for a total of 6 recordings), in order to collect data on the phonetic development (TFPI, Zmarich, Fava, Del Monego & Bonifacio, 2012; "routines" for verbal play, Stoel-Gammon, 1989), lexical development (MacArthur-Bates CDI, Caselli, Pasqualetti & Stefanini, 2007), severity of stuttering (SSI-3, Riley, 1994), and communication attitude (KiddyCAT, Vanryckeghem, Brutten, 2007). If, at the end of that period, the child was still stuttering and the parents requested it, a treatment was initiated at the CMF and the experimental observation consequently ceased.

2. *Materials and Methods*

2.1 Subjects

The subjects of the current study are 13 children of the 14 who began to stutter after the recording at 24 months of age (table 2).

Table 2 - *Subjects (*treated subjects), gender, age at the stuttering onset, months elapsed from stuttering onset to first recording, months elapsed from stuttering onset to final clinical outcome*

subjects	*gender*	*age onset*	*months post onset at the first recording*	*months post onset at the final evaluation*
BM	F	51	2	42
CA	F	30	0	68
CG	M	30	1	38
FM	M	29	0	46
FMd	F	36	3	37
GS	F	30	6	31
MG	M	40	0	67
*ML	M	30	6	39
*RF	M	36	3	41
SC	M	31	2	24
*SL	M	21	5	33
*TA	M	31	3	40
VL	F	33	4	33
Mean		32,27(m,d)	2,21(m,d)	41,15 (m,d)

According to parents' reports, children started to stutter between 21 and 51 months of age (mean: 33 months) and they were first recorded from 0 to 6 months after the onset (mean: 2,21; months, days). Four of the subjects at the end of the observation period were treated.

In order to determine the final clinical outcome of each child, a structured telephone interview to parents was made, after an average of 41 months and 15 days from stuttering onset (range: 24-68 months). Based on these interviews, 3 children

had become persistent stutterers (S) and 10 had recovered spontaneously (NS). S children were among the children who attended a therapy.

2.2 Instruments

Disfluency Profile: it is an index drawn from an original method for counting disfluencies put forward by E. Yairi and N. Ambrose in several studies, collected in Yairi, Ambrose (2005). It is based on the calculation of the percentage of SLD out of 400 target syllables;

Stuttering Severity Instrument – Third Edition (SSI–3; Riley, 1994): it is a reliable and valid norm-referenced stuttering assessment tool that can be used for clinical and research purposes. It has been developed to evaluate stuttering severity both in adults and children. Three features have to be obtained to calculate an SSI-3 score: (1) Frequency of stuttering, (2) Duration of the three longest stuttering moments, and (3) Physical Concomitants. Frequency is expressed in percent syllables stuttered and it is converted to scale scores of 2-18. Duration is timed to the nearest one tenth of a second and it is converted to scale scores of 2-18. The four types of Physical Concomitants are summed and converted to scale scores of 0-20;

Structured telephone interview: parents had to respond to a series of 9 YES or NO questions (answering 'yes' to at least three of them qualified the child as stutterer), and they also had to place the level of stuttering severity along a 7-point Likert scale, where 1 represents the absence of stuttering and 7 the maximum of severity.

2.3 Procedure

As regards the experimental session, the child was audio and video recorded in a play activity where she/he was manipulating objects belonging to three different structured daily activities (from Stoel-Gammon, 1989), and while she/he was describing a picture from SSI-3. The recordings lasted around an hour, and allowed the collection of a sample of at least 500 syllables. Once transcribed, the same syllables were selected for calculating the *Disfluency Profile* as well as the SSI-3 score.

The experimental design consists in predicting the persistence of the disorder from the *Disfluency Profile* scores, calculated at 9-15 months or at 16-22 months post-onset, and from the SSI-3 scores (at 16-22 months post-onset).

To this end, on the basis of the *Illinois Longitudinal Study*, as described by Yairi, Ambrose (2005, Chapter 5th), we tried to formulate some operating criteria for adapting the SLD's developmental paths of the persistent and recovered stutterers to prognostic purposes. Based on these criteria, the child candidate to recover spontaneously would be characterized by the following conditions, to be applied in sequence:

– at *n* semester from the stuttering onset, the reduction of the SLD % should be stronger (or equal) than the reduction of SLD % of the recovered subjects

reported in tab. 5.5 by Yairi, Ambrose (2005: 173). This reduction could be quantified by the ratio of the SLD percentage of the equivalent *n* semester divided by the SLD percentage at the onset;
– if the reduction rate does not decrease, the SLD percentage should be closer to the average SLD % of the recovered subjects, as reported in tab. 5.5 by Yairi, Ambrose (2005: 173) at the equivalent *n* semester from the onset, than to the average SLD % of persistent stutterers in the same table.

The use of SSI-3 as predictor of persistent stuttering, however, is more problematic, because the scores for the preschool children do not present any cut-off value able to distinguish the non stuttering children from the children with *Very Mild* stuttering (Riley, 1994). Hence, we decided to refer to the proposal of Howell, Davis (2011) which assume that, in order to be considered recovered, a child who stutters should 1) score less than 24 at the SSI-3 and 2) present a reduction of at least 2 points compared to the evaluation made at stuttering onset. To evaluate the effectiveness of the prediction, we referred to clinical standards (Meisels, 1988), according to which the instrument under evaluation must simultaneously exhibit a level of sensitivity (measurement accuracy in detecting S) and specificity (measurement accuracy in detecting NS) of at least 80%, although we are aware that such a small sample size would not formally allow this kind of analysis (Jones, Gebski, Onslow & Packman, 2002). We calculated an a priori simulation about the size of the sample needed to reach statistical power for a sensitivity analysis (Jones, Carley & Harrison, 2003), obtaining that, with a likely recovery rate of 88% (Yairi, Ambrose, 2013), a sample of size 40 participants was needed. Regrettably, this size was very far from the possibilities of our low-funded project, but nevertheless we wanted to use this analysis in order to find a clue that we were on the right way.

3. *Results*

Table 3 shows the scores of the *Disfluency Profile* (SLD) and the SSI-3 for the subjects in three different sessions (0-6 months, 9-15 months, 16-22 months post-onset). The predictions of final clinical outcome, S (stutters) or NS (non stutterers), are based on the scores of the last two sessions and were checked against the results of the structured telephone interview.

The values of the *Disfluency Profile* at the 9-15 months post-onset session are correctly predicting the outcomes for 2 S and 6 NS, while producing 4 false positives and 1 false negative (sensitivity: 66.3%, specificity: 60.0%).

Table 3 - *Individual percentages (%) of SLD at the Disfluency Profile and SSI-3 scores, at three evaluation sessions, and prediction (Pre) of persistence (S) or recovery (NS). The rightmost column shows the final clinical outcome (*treated subjects)*

Subjects	*0-6 months*		*9-15 months*		*16-22 months*				*Clinical outcome*
	SLD	SSI-3	SLD	Pre	SLD	Pre	SSI-3	Pre	
BM	4,7	23,0	4,1	S	2,3	NS	18,0	NS	NS
CA	12,0	20,0	8,8	S	6,0	S	20,0	S	NS
CG	5,5	13,0	1,3	NS	0,7	NS	10,0	NS	NS
FM	3,0	14,0	9,0	S	4,3	NS	17,0	S	NS
FMd	15,1	21,3	7,7	NS	7,9	S	20,0	S	NS
GS	4,4	12,3	2,0	NS	2,4	NS	8,0	NS	NS
MG	5,3	19,0	4,0	NS	3,5	NS	18,0	S	NS
*ML	7,0	21,0	2,8	NS	5,3	S	22,0	S	S
*RF	3,3	32,0	5,3	S	6,0	S	29,0	S	NS
SC	4,0	8,0	3,0	NS	2,0	NS	8,0	S	NS
*SL	1,9	5,2	2,2	S	14,0	S	25,4	S	S
*TA	20,0	26,0	24,0	S	10,8	S	28,0	S	S
VL	2,6	10,0	1,6	NS	0,5	NS	0,8	NS	NS
Mean	6,8	17,3	5,8		5,0		17,2		

Half a year later, the predictive power increases: the values of the Disfluency Profile correctly predict the final outcomes for 3 S and 7 NS, with 3 false positives and 0 false negatives (sensitivity: 100%, specificity: 70.0%). For this session (16-22 months after the onset) there are also the SSI-3 scores: by applying the cut-off criterion (Howell, Davis, 2011), 3 S and 4 NS are correctly predicted, while there are 6 false positives and 0 false negatives (sensitivity: 100%, specificity: 40.0%). Results are summarized in table 4.

Table 4 - *Results for the sensitivity & specificity analysis after an average interval of 41;15 (months;days) from the stuttering onset (final evaluation: September 2015). S=stutterers; NS=non stutterers*

		Actual clinical output by september 2015	
		S (3)	**NS** (10)
Disfluency profile prediction by 9-15 m. post-onset	S (6)	2 (66.6%)	4 false positives
	NS (7)	1 false negative	6 (60%)
Disfluency profile prediction by 16-22 m. post-onset	S (6)	3 (100%)	3 false positives
	NS (7)	0 false negative	7 (70%)
SSI-3 prediction by 16-22 m. post-onset	S (9)	3 (100%)	6 false positives
	NS (4)	0 false negative	4 (40%)

4. *Discussion*

The aim of the present study was to determine whether the *Disfluency Profile* or SSI-3 scores, measured at different age levels, might have potential as a clinical marker of stuttering persistence in Italian children.

The findings of the present study suggest that the predictive power of the *Disfluency Profile* at the session of 9-15 months post-onset is low, according to clinical standards, but it increases in the next six months, albeit not to the standard minimum of 80%. However this index performs better than SSI-3, which has been proposed as a predictive tool by Howell, Davis (2011). To date, there is no validated instruments for clinicians to predict the persistence of stuttering; therefore, the *Disfluency Profile* exhibits, compared to the SSI-3, the advantage of being free, less demanding (easy-to-use), less arbitrary (for instance, no need to evaluate tension on monosyllabic-word repetitions) and faster to administer and process (for instance, no need of video recording).

In conclusion, we are aware of the low statistical power of our results, and of the not "ideal" current process of verification of clinical outcome. Furthermore, for almost all the subjects currently at least 5 years have passed since the onset of stuttering, and this interval is deemed sufficient to exclude possible variations of the clinical status in the future (Yairi, Ambrose, 2005). Indeed, we have already planned to check again the clinical outcomes of the subjects by collecting updated and more reliable information directly at the subjects' home. Even so, larger studies will be essential to verify the extent to which these effects generalize across individuals and the degree to which they are driven by other factors such as the process of verification of clinical outcome and/or the attitude and capability of clinicians to work with these instruments.

Bibliography

Caselli, M.C., Pasqualetti, P. & Stefanini, S. (2007). *Parole e frasi nel "Primo Vocabolario del Bambino"*. Milano: Franco Angeli.

Clark, C., Conture, E., Frankel, C. & Walden, T. (2012). Communicative and psychological dimensions of the KiddyCAT. In *Journal of Communication Disorders*, 45, 223-234.

Howell, P., Davis, S. (2011). Predicting persistence of and recovery from stuttering by the teenage years based on information gathered at age 8 years. In *Journal of Developmental & Behavioral Pediatrics*, 32, 3, 196-205.

Jones, M., Gebski, V., Onslow, M. & Packman, A. (2002). Statistical Power in Stuttering Research: A Tutorial. In *Journal of Speech, Language, and Hearing Research*, 45, 2, 243-255.

Jones, S.R., Carley, S. & Harrison, M. (2003). An introduction to power and sample size estimation. In *Emergency Medicine Journal*, 20, 453-458.

Kloth, S.A.M., Kraaimaat, F.W., Janssen, P. & Brutten, G.J. (1999). Persistence and remission of incipient stuttering among high-risk children. In *Journal of Fluency Disorders*, 24, 253-265.

Kraft, S.J., Yairi, E. (2011). Genetic Bases of Stuttering: The State of the Art 2011. In *Folia Phoniatrica et Logopaedica*, 64, 34-47.

Meisels, S.J. (1988). Developmental screening in early childhood: The interaction of research and social policy. In *American Review of Public Health*, 9, 527-550.

Reilly, S., Onslow, M., Packman, A., Cini, E., Ukoumune, O.C., Bavin, E.L., Prior, M., Eadie, P., Block, S. & Wake, M. (2013). Natural history of stuttering to 4 years of age: A prospective community-based study. In *Pediatrics*, 132, 3, 460-467.

Riley, G.D. (1994). *Stuttering Severity Instrument for children and adults. Third edition - SSI-3.* Austin, Texas: Pro-Ed.

Smith, A. (2016). A multifactorial neurodevelopmental approach to stuttering: (1) Language and motor factors and (2) Pathways to persistence and recovery. In Tomaiuoli, D. (Ed.), *Proceedings of the 2nd international Conference on Stuttering.* Trento: Erickson, 31-40.

Stoel-Gammon, C. (1989). Language Production Scale. In Olswang, L.B., Stoel-Gammon, C., Coggins, T.E. & Carpenter, R. (Eds.), *Assessing prelinguistic and Early Linguistic Behaviors in Developmentally Young Children.* Washington: Washington Press.

Subramanian, A., Yairi, E. & Amir, O. (2003). Second formant transition in fluent speech of persistent and recovered preschool children who stutter. In *Journal of Fluency Disorder*, 36, 59-75.

Vanryckeghem, M., Brutten, A. (2007). *The KiddyCAT: A communication attitude test for preschool and kindergarten children who stutter*. San Diego, CA: Plural Publishing.

Yairi, E., Ambrose, N. (2005). *Early childhood stuttering, for clinicians by clinicians.* Austin, Texas: Pro-Ed.

Yairi, E., Ambrose, N. (2013). Epidemiology of Stuttering: 21st century advances. In *Journal of Fluency Disorders*, 38, 66-87.

Zmarich, C. (2015). Il profilo delle disfluenze come indice predittivo precoce di cronicità in bambini che hanno appena iniziato a balbettare. In Busà, M.G., Gesuato, S. (Eds.), *Lingue e contesti. Studi in onore di Alberto M. Mioni*. Padova: Cleup, 805-818.

Zmarich, C., Fava, I., Del Monego, G. & Bonifacio, S. (2012). Verso un "Test Fonetico per la Prima Infanzia". In Falcone, M., Paoloni, A. (Eds.), *La voce nelle applicazioni*, Atti dell'8° Convegno Nazionale AISV, 25-27 gennaio 2012. Roma: Bulzoni, 51-66.

DOI: 10.17469/O2102AISV000024

CINZIA AVESANI, MARIO VAYRA, VALERIA LONGO

Attrito e transfer tra dialetto e italiano regionale. Quantità e lunghezza vocalica nel parlato intramurario di Bologna[1]

We present the results of three production experiments on stressed vowel lengthening in the linguistic repertoires of Bologna in order to ascertain: 1) if contrastive vowel length (CVL) is maintained in the urban dialect of Bologna as currently spoken within the town walls; 2) if the dialect's contrastive length has been transferred as stressed vowel lengthening to the variety of regional Italian spoken in Bologna due to the close contact of the two reperoires; 3) if stressed vowel lengthening in Bolognese Italian is greater or comparable to that found in other varieties of Italian that miss any relationship with CVL in the related dialect, as Florentine. Results show that CVL is consistently maintained in Bolognese dialect, while the hypothesis of a transfer of CVL as stressed vowel lengthening in Bolognese regional Italian is not supported.

Keywords: contrastive vowel length, prosodic lengthening, Bolognese dialect and regional Italian, L1 transfer, attrition.

1. *Quantità vocalica e allungamenti prosodici nel repertorio linguistico bolognese*

Il repertorio linguistico bolognese, che comprende sia il dialetto che la varietà di italiano regionale parlata a Bologna, è caratterizzato da fenomeni interessanti che colpiscono il sistema vocalico in sede tonica.

1.1 Il dialetto e la quantità vocalica

Il dialetto bolognese è uno dei dialetti emiliani che esibiscono opposizioni di lunghezza vocalica, come nelle coppie minime presentate a titolo di esempio nella Tabella 1. Diversamente da altri dialetti emiliani, in bolognese il fenomeno prosodico della quantità vocalica è presente sia in monosillabi che in bisillabi piani. Opposizioni di lunghezza vocalica per questo dialetto sono state illustrate e discusse ad esempio da Coco (1970), Canepari, Vitali (1995), Filipponio (2012), Foresti (1994; 1988-2005; 2010) e Hajek (1992; 1994; 1995; 1997a; 1997b; 2000); per le opposizioni di quantità presenti in altri dialetti emiliani il riferimento è agli studi di

[1] Il lavoro nasce dalla collaborazione tra i primi due autori che lo hanno ideato e disegnato sperimentalmente. A Valeria Longo si devono la trascrizione fonetica e le analisi acustiche di tutto il corpus. La responsabilità del lavoro è divisa come segue: Avesani § 2, 4, 5; Vayra § 1, 3; Avesani, Vayra e Longo § 6.

Uguzzoni, Busà (1995), Uguzzoni, Azzaro & Schmid (2003), Filipponio (2012) e Bernardasci (2015).

Tabella1 - *Esempi di opposizioni di quantità vocalica el dialetto bolognese*

[sak(:)]	*secco*	vs.	[sa:k]	*sacco*
[frask(:)]	*fresco*	vs.	[fra:sk]	*frasca*
[mel(:)]	*mille*	vs.	[me:l]	*miele*
[fos(:)]	*fosse* (cong., 3a sing)	vs.	[fo:s]	*fosso*
[ˈfat(:)a]	*fetta*	vs.	[ˈfa:ta]	*fatta*
[ˈmet(:)er]	*mettere*	vs.	[ˈme:ter]	*metro*
[ˈmol(:)a]	*mula*	vs.	[ˈmo:la]	*mola*
[ˈtsok(:)a]	*ceppo*	vs.	[ˈtso:ka]	*fiasco, zucca*

Una seconda caratteristica interessante del bolognese nel dominio delle durate segmentali è il rapporto complementare che lega, in superficie, la durata della vocale tonica a quella della consonante immediatamente seguente. Questo fenomeno di natura compensativa, già descritto in Coco e denominato da Hajek *CV complementation,* accomuna il bolognese ad altri dialetti emiliani, in particolare a quelli appennininci: una V lunga è necessariamente seguita da C breve mentre una V breve è seguita da C lunga creando alternativamente sequenze [V:C] o [VC:] (Hajek, 1994; 1995; 1997). Come nota Loporcaro (2015), poiché la lunghezza consonantica, dipendente dal contesto, viene trattata come allofonica, spesso non viene annotata nelle descrizioni sui dialetti emiliani (come, ad esempio, in Coco, 1970): il tratto rilevante è l'opposizione di lunghezza vocalica.

In questo lavoro non ci occuperemo del fenomeno compensativo, che verrà trattato in altra sede.

1.2 Rinascita delle opposizioni di quantità vocalica nel dialetto

Come è noto, nessuna delle varietà linguistiche derivate dal latino ha ereditato per via diretta l'opposizione di quantità vocalica. Tuttavia, un numero consistente di sistemi linguistici nella parte centro-settentrionale della Romània, dagli Appennini al Mare del Nord, ha successivamente sviluppato opposizioni di lunghezza vocalica in stadi diversi della loro storia e in modi differenti (Loporcaro, 2015). Sui processi e sulle fasi storiche che hanno portato allo sviluppo della lunghezza vocalica distintiva in seguito al collasso del sistema fonologico quantitativo latino, moltissimo è stato scritto e moltissimo si è dibattuto e si continua a dibattere. Sulla base dell'amplissima letteratura sull'argomento, la rinascita delle opposizioni di lunghezza vocalica (LV) è considerata il successore di quel processo del tardo latino che ha provocato l'allungamento della vocale tonica in sillaba aperta causando la perdita delle opposizioni di quantità vocalica del latino classico (per un'analisi serrata ed esaustiva dell'argomento rimandiamo a Loporcaro, 2015, con riferimenti)[2]. Dei tre

[2] La caduta dell'opposizione di quantità si completa, in accordo a Loporcaro (2015), solo alla fine dell'impero, anche se precursori del fenomeno appaiono già nel II secolo. La tesi dello studioso, dimostratata attraverso un'intensa analisi critica della amplissima letteratura sull'argomento, è che ad un

tipi di lunghezza vocalica che, in accordo a Loporcaro (2015: 61), interessano in sincronia le lingue romanze, quello del bolognese ricade nel tipo italo-romanzo settentrionale.

Nell'Italia del nord, le opposizioni di LV sono attestate in parecchie varietà dialettali, dagli Appennini alle Alpi, dalla Liguria al Friuli ma non in tutte (cfr. Loporcaro, 2015: 82); manca ad esempio nei dialetti veneti, nel ferrarese, nel piemontese e nel lombardo orientale. Nelle varietà in cui LV è presente, la vocale fonologicamente lunga deriva da una vocale tonica che in protoromanzo si trovava in sillaba aperta, e la vocale breve da una vocale tonica originariamente in sillaba chiusa. In alcuni dialetti la situazione attuale collima fedelmente con tale distribuzione (ad esempio nel cremonese); in altri dialetti, come in quelli emiliani, la situazione attuale non rispecchia direttamente l'applicazione dell'allungamento vocalico in sillaba aperta. Il bolognese, ancor più degli altri dialetti emiliani, ha subito ulteriori trasformazioni, quali una rotazione vocalica (ad es., la /a/ breve del bolognese deriva dalla /e/ in sillaba chiusa del protoromanzo) e un successivo allungamento delle vocali basse e medio-basse (Coco, 1970; Filipponio, 2012; Loporcaro, 2015). Il risultato in sincronia è che non tutti i timbri vocalici hanno una vocale fonologicamente lunga e una breve. Seguendo la trattazione di Coco, la distinzione di quantità riguarda le seguenti coppie di vocali: /e:/ vs /e/, /a:/ vs /a/, /o:/ vs /o/, mentre le vocali alte /i:/ e /u:/ non hanno la controparte breve[3] (Coco, 1970: 112).

1.3 Allungamenti prosodici

Accanto ad opposizioni di quantità vocalica, esistono in bolognese interessanti fenomeni di allungamento vocalico condizionati dalla struttura prosodica a livello frasale. Sia nel dialetto che nell'italiano regionale, gli allungamenti della durata vocalica sono particolarmente evidenti nelle sedi metricamente forti dei sintagmi prosodici (*intermediate* e *intonational phrase*); inoltre, per l'italiano bolognese, questi risultano percettivamente più salienti degli allungamenti normalmente attestati nelle stesse sedi metriche in altre varietà di italiano regionale, in primis il toscano.

Si può apprezzare questo fenomeno ascoltando due clip tratte, per il dialetto, da "Pizunèra" (Piccionaia), cartone animato in dialetto bolognese[4] e, per il parlato regionale, da un servizio del TG3 dell'Emilia Romagna dedicato alle scuole di dialetto sorte a Bologna in questi ultimi anni[5]. Negli esempi seguenti, tratti da questi brani, si noti in (1) come un super-allungamento in dialetto interessi la vocale nucleare [u] di "ranuci" nel primo *intermediate phrase* (ip), ma non la vocale nucleare [o] di "pirocca" nel secondo ip (in questo contesto di parola la vocale è in sillaba chiusa da consonante geminata).

---

certo momento sotto l'impero romano varietà regionali diverse del latino potessero differire tra loro esibendo lunghezza vocalica distintiva, un tratto conservativo, oppure allungamento vocalico in sillaba aperta, un tratto diacronicamente successivo (Loporcaro, 2015: 58). Così, ad opposizioni di quantità vocalica si sostituiscono gradualmente in diverse regioni della Romània distinzioni di lunghezza vocalica condizionate dalla struttura della sillaba: vocali lunghe in sillaba aperta, vocali brevi in sillaba chiusa.

[3] Ma si veda Canepari, Vitali (1995), secondo i quali la distinzione di quantità vale non per tre ma per sette vocali.

[4] L'intero cartone animato è visibile alla pagina web https://www.youtube.com/watch?v=N3DfYF1hmv4.

[5] Il servizio televisivo è disponibile alla pagina web https://www.youtube.com/watch?v=ToIVyfER564.

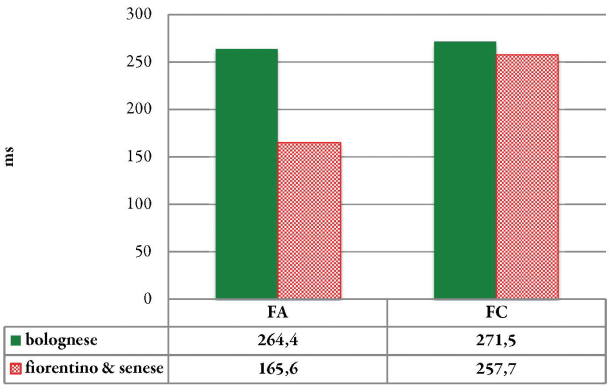
Nell'esempio (2) di italiano regionale bolognese, si noti come il super-allungamento interessi solo il dittongo nucleare del secondo IP [ja] ma non la vocale nucleare degli *intermediate phrases* dell'IP precedente, rispettivamente [ɛ] e [o]:

(1a) [[ˈkwɑ̃ŋd i **rɐˈnu::tɕi**]$_{ip}$ [purˈtɛvɛŋ lɜ piˈrok:ɜ]$_{ip}$]$_{IP}$
(quando i ranocchi portavano la parrucca)

(2) [[koˈr:ɛt:ɜ]$_{ip}$ [sə nɔ noŋ]$_{ip}$]]$_{IP}$ [[noŋ lo **promwoˈʊja::mo** $_{ip}$]]$_{IP}$
(corretto se no non non lo promuoviamo)

Nella letteratura dedicata all'italiano regionale bolognese, peraltro non molto corposa (Rizzi, 1986; 1989; De Dominicis, 2001; Endo, Bertinetto, 1997), tali allungamenti non sono stati oggetto di indagine. Una prima indicazione quantitativa in tal senso proviene da una tesi di laurea magistrale (Università di Bologna) dedicata alla prosodia del parlato bolognese intramurario (Chen, 2013). Tra i vari risultati ottenuti, uno che interessa direttamente il tema di questo lavoro è relativo alla durata sillabica misurata in diverse posizioni metriche e in diverse condizioni di *focus*.

Figura 1 - *Durata della sillaba tonica in bisillabi piani finali di frase in condizione di Focus Ampio (FA) e Focus Correttivo (FC) (tratta da Chen, 2013)*



La Figura 1 riporta le durate sillabiche di trisillabi piani in posizione finale di frase, in condizione di *focus* ampio e *focus* contrastivo-correttivo. I dati sul bolognese, confrontati con dati equivalenti sul fiorentino e sul senese (tratti da Bocci, 2013) mostrano una durata decisamente maggiore per il bolognese rispetto ai dati accorpati del senese e del fiorentino, anche se limitatamente alla condizione di sillaba tonica nucleare in *focus* ampio.

2. *Dal dialetto all'italiano regionale*

Dal quadro di insieme delineato nei paragrafi precedenti scaturiscono i due fulcri di interesse attorno ai quali ruota il presente lavoro.

La prima questione che affrontiamo è il permanere delle opposizioni di lunghezza vocalica nel sistema del vocalismo tonico del dialetto bolognese. Poiché i

dati di Coco fotografano una situazione che riguarda ormai 56 anni fa, è lecito domandarsi se questo tratto sia oggi ancora mantenuto, considerando che, in un contesto sociolinguistico di contatto con l'italiano, si ritiene che il dialetto sia una varietà linguistica in regressione (Badini, 2002), soggetta come altri dialetti a forte attrito da parte dell'italiano (Berruto, 2012). In particolare, abbiamo voluto verificare se la quantità sia ancora mantenuta nel dialetto bolognese studiando la varietà urbana intramuraria, nella quale da una parte "duration levels were found to be extreme in all contexts when compared to the rural variety" (Hajek, 2000: 127) ma che, d'altro canto, è meno conservativa rispetto al dialetto rurale. Se l'opposizione di quantità è tuttora attiva e non soggetta ad erosione da parte dell'italiano, essa dovrebbe essere realizzata da differenze molto significative nella durata di vocali brevi vs. lunghe.

Che il dialetto sia parlato da una minoranza dei bolognesi è un fatto assodato, e lo abbiamo sperimentato in prima persona nella difficoltà di rintracciare parlanti nativi di bolognese-L1: parlanti, cioè, nati a Bologna da genitori bolognesi che parlino il dialetto abitualmente e in modo fluente. Una nota positiva su questo fronte, però, viene dal rinnovato interesse per il dialetto come lingua identitaria; interesse che, motivato anche da ragioni socio-politiche, ha portato alla creazione di scuole di dialetto a Bologna – così come in altre aree dell'Emilia Romagna – nelle quali i cittadini si riappropriano del loro dialetto, come lingua seconda (si vedano sul sito http://www.bulgnais.com/corso.html i documenti relativi ai corsi di bolognese tenuti dal 2002 al 2005 e il servizio del TG3 dell'Emilia Romagna citato in §1.3).

La seconda questione, di tipo sociolinguistico, riguarda l'italiano regionale bolognese e si riaggancia ad una interessante osservazione sugli allungamenti nella durata vocalica offerta da Rizzi (1989: 32). Allungamenti vocalici vengono segnalati solo per alcune condizioni sillabiche: essi sarebbero frequenti nella vocale tonica in sillaba chiusa da nessi consonantici; fenomeno questo non comune né tra le lingue del mondo (in Maddieson, 1985, l'opposta tendenza al *Closed Syllable Vowel Shortening* è esemplificata su un numero molto elevato di lingue e discussa come un "quasi universale" fonetico), né in varietà di italiano vicine allo standard come il toscano (cfr. Vayra, Avesani & Fowler, 1999, con riferimenti)[6]. Il fenomeno non sarebbe tuttavia presente davanti a geminate, secondo una distribuzione che abbiamo visto valere anche nel dialetto (vedi *supra*, esempio 1a).

In questo lavoro ci siamo chiesti se nel processo di convergenza verticale dai dialetti italo-romanzi all'italiano (*advergence*, cfr. Auer, 2005; 2011; Berruto, 1989; Cerruti, Regis, 2011), che ha portato come conseguenza la divergenza delle varietà regionali l'una dall'altra (Cerruti, 2011; Dal Negro, Vietti, 2011), il sostrato dialettale abbia determinato il transfer del tratto di lunghezza vocalica nell'italiano regionale. L'ipotesi forte che vorremmo verificare è se nel passag-

[6] Occorre rilevare che alcuni studi sul vocalismo dell'italiano fondati su parlato "connesso" hanno osservato la mancanza di accorciamento vocalico in sillabe CVC vs. CV (si rinvia per tutti ad Albano Leoni, Caputo, Cerrato, Cutugno, Maturi & Savy, 1995, con riferimenti).

gio dalla varietà dialettale a quella standard regionale la lunghezza vocalica sia stata mantenuta mutando il suo *status* da fonologico a fonetico. Ci siamo chiesti insomma se, in una situazione di contatto linguistico, esistano allungamenti fonetici vocalici di tipo segmentale che interessano gli stessi timbri vocalici che nel dialetto sono quantitativamente distintivi, nell'ipotesi che tali allungamenti vengano poi rinforzati quando la sillaba tonica è associata a posizioni metricamente forti (accento nucleare), dando così origine a quegli allungamenti così marcati diatopicamente che abbiamo illustrato negli esempi precedenti.

Il lavoro procede come segue: nel § 3 presentiamo i risultati di un primo esperimento di produzione volto a verificare la persistenza di opposizioni di quantità nel dialetto intramurario contemporaneo; nel § 4 presentiamo i dati di un secondo esperimento di produzione volto a verificare la presenza, nell'italiano regionale bolognese, di differenze di durata vocalica sistematiche riconducibili alle differenze di quantità vocalica del dialetto; nel § 5 confrontiamo le durate vocaliche dell'italiano regionale bolognese con quelle del fiorentino, in un terzo esperimento di produzione; nel § 6 trarremo le conclusioni di questo studio.

3. *Esperimento 1: dialetto bolognese*

3.1. I parlanti selezionati per il primo e il secondo esperimento sono cinque soggetti bolognesi, quattro uomini e una donna, di età compresa tra i cinquanta e i sessant'anni. A questi abbiamo presentato su uno schermo, in ordine randomizzato, dei brevi brani in italiano composti da due o tre frasi, una delle quali conteneva una delle parole target del nostro corpus. Il parlante doveva leggere mentalmente il brano, tradurlo mentalmente in dialetto, e quindi produrlo in modo naturale.

Le frasi sono state concepite in modo che la parola target compaia nella stessa posizione sintattica, alla fine della prima clausola di un periodo complesso composto da due coordinate. Poiché un confine sintattico forte normalmente fa scattare un confine prosodico forte (di *intonational phrase*), abbiamo la ragionevole certezza che le clausole coordinate vengano realizzate come due sintagmi prosodici distinti, nei quali la parola chiave compare nella posizione nucleare del primo sintagma intonativo. Mantenere la stessa posizione nella struttura prosodica garantisce che la parola avrà lo stesso grado di prominenza prosodica in tutte le frasi del corpus e che la vocale tonica non sarà soggetta a variazioni di durata legate alle diverse posizioni della struttura metrica. Si è cercato inoltre di mantenere costante la lunghezza (in numero di sillabe) del sintagma intonativo entro il quale si colloca la parola target. Un esempio di frasi che ospitano uno dei membri di una coppia minima distinta per quantità vocalica è il seguente:

(3) [[la ˈʤoʌa la tolt ũŋ **sa:k**]$_{ip}$ [e la rĩŋˈpe ed bãˈnã:n]$_{ip}$]$_{IP}$
[[al prɛ: le tot **sak**]$_{ip}$ [e ɐˈdɛs al sjãŋk sãŋs ã:kwa]$_{ip}$]$_{IP}$

[[Giulia ha preso un **sacco**]$_{ip}$ [e lo ha riempito di banane]$_{ip}$]$_{IP}$
[[Il prato è tutto **secco**]$_{ip}$ [e ora siamo anche senz'acqua]$_{ip}$]$_{IP}$

Le 12 parole target e le 6 coppie minime da esse costituite sono elencate nella Tabella 2[7].

Tabella 2 - *Parole target del corpus*

Italiano	*Dialetto*
sacco vs. *secco*	[sa:k] vs. [sak]
frasca vs. *fresco*	[fra:sk] vs. [frask]
fossi (N) vs. *fossi* (V)	[fo:s] ([fu:s]) vs. [fos]
fatta vs. *fetta*	[ˈfa:ta] vs. [ˈfata]
massa vs. *messa*	[ˈma:sa] vs. [ˈmasa]
mola vs. *mula*	[ˈmo:la] vs. [ˈmola]

Ciascuno dei parlanti ha ripetuto il compito dalle tre alle cinque volte. Poiché una condizione fondamentale era che le frasi fossero tutte prodotte con la stessa struttura prosodica, qualora il soggetto avesse avuto esitazioni o prodotto pause in una posizione non prevista, la frase è stata scartata. Il totale di frasi registrate per il dialetto è 195, i casi utilizzati sono 189.

3.2 Analisi

Le frasi sono state segmentate, trascritte e analizzate con il software *Praat,* e i valori di durata (in ms) della vocale tonica target sono stati normalizzati. Per quanto riguarda la normalizzazione, per ciascun locutore e ciascuna coppia di frasi, abbiamo convertito in *z*-scores la durata assoluta di ogni esemplare di vocale target. Il valore *z* è stato ottenuto sottraendo dal valore di ciascun esemplare la media dei valori ottenuti per tutte le ripetizioni delle vocali target in una determinata coppia minima (ad esempio tutti i valori di [a:] e [a] in tutte le ripetizioni di [ˈma:sa] e [ˈmasa] di un determinato locutore). Successivamente, il valore ottenuto è stato diviso per la relativa deviazione standard. I valori *z* così ottenuti saranno positivi se la durata del singolo caso è maggiore rispetto alla media dei valori osservati nella coppia minima, e negativi se la durata del singolo caso è minore della media. Ci aspettiamo quindi che le vocali che sono fonologicamente brevi abbiano valori *z* negativi, mentre vocali fonologicamente lunghe abbiano valori *z* positivi. Sia i valori in ms che i valori *z* sono stati sottoposti ad ANOVA a misure ripetute.

3.3 Risultati

Come mostra la Tabella 2, il corpus elicitato contiene sia parole monosillabiche che parole bisillabiche, così come sono attestate nelle tradizionali fonti dialettologiche (ad es., Coco, 1970). Il primo dato interessante che abbiamo ottenuto è che non tutti i monosillabi attesi sono stati realizzati come tali: nel 29% dei casi le voci target monosillabiche sono state realizzate come bisillabi (30

[7] Le coppie minime riguardano i timbri vocalici /a/ e /o/.

casi su un totale di 103 occorrenze di parole comunemente monosillabiche). Le parole che presentano realizzazione lessicale variabile sono: [sak]/[sa:k] ("secco"/"sacco"), [fos]/[fo:s]-[fu:s] ("fossi"(V)/"fossi"(N), [frask]/[fra:sk] ("fresca"/"frasca")[8]. Si noti come 'fossi' (nome) possa essere realizzato anche con un timbro diverso da quanto più frequentemente attestato, sostituendo alla vocale posteriore medio-alta [o:] una vocale posteriore alta [u:]. In tali casi la coppia è quindi distinta sia dalla quantità che dalla qualità vocalica.

Da un'analisi più dettagliata sulla distribuzione della variazione tra i locutori emerge che la fonte di variazione maggiore è la donna, con 15 casi di realizzazioni bisillabiche su un totale di 17 occorrenze dei 6 monosillabi. La composizione del nosto campione di soggetti non è bilanciata e poiché su 5 locutori uno solo è donna, non siamo in grado di stabilire se ci troviamo in presenza di una variazione correlata a una differenza di genere o se si tratti di una tendenza idiosincratica di questa specifica locutrice. Questo è uno punto che potremo dirimere solo attraverso indagini ulteriori.

In circa la metà dei casi (14/30), e in maniera trasversale tra i diversi locutori, la nuova sillaba atona termina con [ə]; nei casi rimanenti la sillaba termina con una vocale piena.

Gli altri 15 casi di realizzazioni bisillabiche sono distribuite tra i quattro locutori maschili del nostro campione. Per tre di loro l'unica voce monosillabica che viene resa come bisillabica è 'frask' ("frasca"), che può essere realizzata come [ˈfra:ska] o [ˈfra:skə], mentre il quarto locutore realizza come bisillabici sia 'frask' che 'fos' ("fossi": N) (tre casi). Di tutte le occorrenze di "frask" nei cinque locutori (17), solo in un caso la realizzazione rimane monosillabica.

Si può pensare che la vocale paragogica sia qui inserita attraverso un processo di "riparo sillabico" che elimina la presenza di un nesso consonantico strutturalmente "sgradito" in posizione di coda sillabica[9]. Non è un caso infatti che sia proprio l'unica parola target monosillabica con coda bi-consonantica ad essere realizzata sistematicamente dai quattro locutori come un bisillabo.

La semplificazione dei nessi consonantici è un fenomeno presente fra le lingue del mondo (Nespor, 1993), e accomuna i dialetti italiani settentrionali (Repetti, 1992; 2000): esso può avvenire sia grazie alla cancellazione di una consonante, sia grazie all'inserzione di una vocale non etimologica. Nel caso di nessi in coda sillabica, la sillabificazione può avvenire per mezzo di una vocale paragogica o per mezzo dell'inserzione di una vocale epentetica tra le due consonanti (Benincà, Parry & Pescarini, 2016; Repetti, 1995). Quale che sia la strategia usata, la semplificazione dei nessi consonantici produce forme lessicali bisillabiche che presentano una struttura sillabica universalmente meno marcata rispetto ai corrispondenti bisillabi.

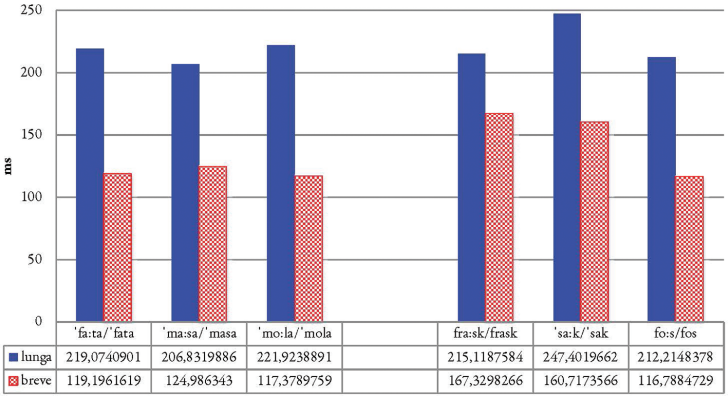
[8] Oltre ai casi di realizzazione bisillabica di parole target monosillabiche, riscontriamo anche un caso di realizzazione monosillabica di "fatta" come [ˈfa:ʈ].

[9] Siamo grati ad un anonimo revisore per questo suggerimento.

L'interpretazione fonologica che si attaglia alla realizzazione bisillabica di 'frask' non può essere estesa però a spiegare le realizzazioni bisillabiche di (quasi) tutte le voci target monosillabiche della nostra locutrice. Essa infatti inserisce una vocale paragogica per semplificare un nesso consonantico finale di parola (['fra:sk]), ma anche in monosillabi con vocale breve seguita da consonante lunga (cfr. [sak:] ('secco') e [fos:] ("fossi", V), e in monosillabi con vocale lunga seguita da consonante breve (cfr. [sa:k] ('secco') e [fo:s] ('fossi', V). Propendiamo quindi per interpretare questi casi come forme dialettali che rivelano un processo di attrito in corso sotto la spinta dell'italiano; se da una parte esse mantengono la distinzione di quantità, dall'altra modificano la struttura sillabica che si avvicina a un tipo fonologicamente meno marcato.

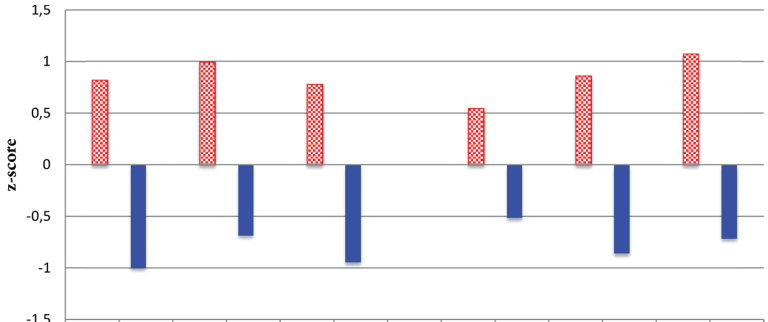
I dati sulla durata assoluta e normalizzata delle vocali toniche sono rappresentati rispettivamente nelle figure 2 e 3. In entrambi, i bisisillabi sono presentati nella parte sinistra e i monosillabi nella parte destra del grafico.

Figura 2 - *Dialetto bolognese. Durata (ms) delle vocali toniche nei bisillabi (a sinistra nel grafico) e nei monosillabi (a destra)*



	'fa:ta/'fata	'ma:sa/'masa	'mo:la/'mola		fra:sk/frask	'sa:k/'sak	fo:s/fos
lunga	219,0740901	206,8319886	221,9238891		215,1187584	247,4019662	212,2148378
breve	119,1961619	124,986343	117,3789759		167,3298266	160,7173566	116,7884729

Figura 3 - *Dialetto bolognese. Durata normalizzata (z-score) delle vocali toniche nel dialetto bolognese. Bisillabi (a sinistra nel grafico) e monosillabi (a destra)*



	'fa:ta	'fata	ma:sa	masa	mo:la	mola		fra:sk	frask	sa:k	sak	fo:s	fos
breve		-0,9986		-0,6833		-0,9452			-0,515		-0,8595		-0,7145
lunga	0,81946		0,99384		0,77837			0,54533		0,85951		1,07495	

Sia per i bisillabi che per i monosillabi la durata assoluta della vocale lunga è maggiore della durata della vocale breve (Figura 2) e il valore *z* normalizzato è sistematicamente negativo per le vocali brevi e positivo per le vocali lunghe. Un'Analisi della Varianza a misure ripetute a una via è stata calcolata separatamente per bisillabi e monosillabi, con variabile dipendente la Durata Vocalica – sia assoluta che normalizzata – e con fattore la Quantità Vocalica (lunga vs. breve). L'analisi indica che la differenza tra vocali brevi e lunghe è significativa sia per i valori assoluti che per quelli normalizzati (durata in ms: monosillabi: $F(1,67) = 101,84$ $p < .0001$; bisillabi: $F(1,111) = 528,97$ $p < .0001$. Durata normalizzata: monosillabi: $F(1,67) = 134,96$ $p < .0001$; bisillabi: $F(1,111) = 378,41$ $p < .0001$).

I risultati indicano quindi con certezza che la quantità vocalica è mantenuta nel dialetto bolognese attuale, e mostrano che la vocale fonologicamente breve ha in media una durata minore del 60% rispetto a quella della vocale fonologicamente lunga.

4. *Esperimento 2: italiano regionale bolognese*

Sia nei dati dell'esperimento 1 che nelle osservazioni riportate in Rizzi (1989) troviamo tracce degli effetti bidirezionali del contatto tra dialetto e italiano regionale. Abbiamo da una parte un processo di attrito del dialetto ad opera dell'italiano, che è evidenziato nei nostri dati dai casi di semplificazione dei nessi consonantici in coda sillabica dei monosillabi; l'inserzione di una vocale paragogica porta, come abbiamo visto sopra, alla creazione di parole bisillabiche che ricalcano la forma italiana e coesistono con le forme tradizionali monosillabiche nel dialetto. Dall'altra parte abbiamo l'attestazione della presenza di allungamenti vocalici in sillaba chiusa in italiano regionale che non si riscontrano nello standard su base toscana. Tali allungamenti, inoltre, sono attestati solo nel caso in cui la sillaba sia chiusa da un nesso consonantico ma non da una geminata, distribuzione che per l'italiano regionale collima con quanto riportato da noi per il dialetto nell'esempio (1). Siamo quindi di fronte a casi che testimoniano la compresenza di fenomeni di italianizzazione del dialetto che di dialettalizzazione dell'italiano (Berruto, 1989; 2005).

Proiettando questa situazione sul caso degli allungamenti vocalici in sede tonica, ci siamo chiesti se la dialettizzazione dell'italiano abbia potuto portare ad un transfer della lunghezza vocalica del dialetto nell'italiano regionale, mutandone nel passaggio la natura, che da fonologica nel dialetto diviene fonetica nell'italiano bolognese. L'ipotesi operativa è che, a parità di condizioni, le vocali che in dialetto presentano vocale breve avranno nell'italiano regionale una durata minore delle vocali che in dialetto presentano vocale lunga.

A questo fine abbiamo predisposto un secondo esperimento sull'italiano regionale, usando lo stesso metodo dell'esperimento 1.

4.1 Metodo

Gli stessi 5 locutori, in una sessione di registrazione diversa, hanno letto gli stessi brani che nell'esperimento 1 erano stati prodotti in dialetto. Le parole target, riportate nella Tabella 1, costituiscono gli equivalenti lessicali delle parole dialettali esaminate nell'esperimento precedente (naturalmente, i monosillabi del dialetto hanno un equivalente italiano bisillabico). Il numero di frasi registrate è pari al numero di casi utilizzati: 234. Un esempio di frasi che contengono, nelle stesse posizioni prosodiche, la coppia di parole che in dialetto era distinta in base alla quantità è presentato in (3):

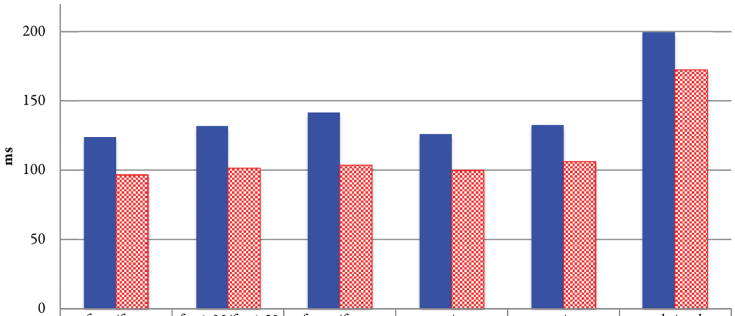
(3a)[[tante persone così fanno una bella **massa**]$_{ip}$ [speriamo che ci ascoltino!]$_{ip}$]$_{IP}$
(3b) [[domani non dobbiamo perdere la **messa**]$_{ip}$ [è domenica!]$_{ip}$]$_{IP}$

4.2 Risultati

I risultati sulla durata assoluta della vocale tonica e sulla durata normalizzata sono riportati rispettivamente nelle figure 4 e 5. Nella legenda, vocale "breve" indica la vocale dell'italiano bolognese che in dialetto è fonologicamente breve, "lunga" la vocale dell'italiano bolognese che in dialetto è fonologicamente lunga. Nella Figura 4, per ciascuna coppia minima, la vocale "breve" è rappresentata nella barra a destra, la vocale "lunga" nella barra a sinistra. Come si vede, tutte le vocali toniche che in dialetto sono brevi hanno durata minore delle vocali toniche che in dialetto sono lunghe. Parimenti, nella Figura 5, la durata normalizzata delle vocali "brevi" ha uno *z-score* negativo mentre quella delle vocali che in dialetto sono fonologicamente lunghe ha uno *z-score* positivo.

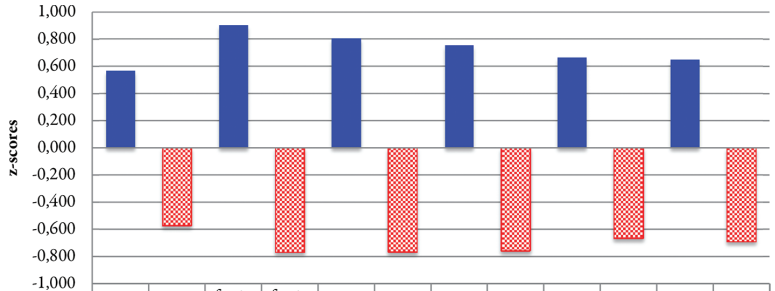
Poiché le coppie minime dell'italiano sono composte da parole che hanno in sede tonica sia sillaba chiusa che sillaba aperta, abbiamo condotto le analisi statistiche sul sottoinsieme omogeneo composto da tutte le parole con sillaba tonica chiusa, scartando di fatto solo una coppia minima (*mola* vs *mula*). L'ANOVA a misure ripetute (con gli stessi fattori usati nell'esperimento 1) conferma la distintività della differenza di durata, sia considerando come variabile dipendente i valori assoluti in ms sia i valori normalizzati come *z-scores* (durata in ms: $F(1,196) = 155{,}64$ $p < .0001$; durata normalizzata: $F(1,335) = 12.169$ $p = 0.0006$).

Figura 4 - *Italiano bolognese. Durata (ms) delle vocali toniche*



	fatta/fetta	fossi_N/fossi_V	frasca/fresco	massa/messa	sacco/secco	mola/mula
lunga	124,0	132,1	141,8	126,2	132,6	199,8
breve	96,6	101,5	103,7	99,9	106,3	172,5

Figura 5 - *Italiano bolognese. Durata normalizzata (z-score) delle vocali toniche*



	fatta	fetta	fossi_ N	fossi_ V	frasca	fresco	massa	messa	sacco	secco	mola	mula
D: lunga	0,570		0,905		0,807		0,758		0,666		0,653	
D: breve		-0,570		-0,769		-0,767		-0,758		-0,666		-0,689

La distribuzione dei dati nelle due figure precedenti rende immediatamente evidente anche un altro fatto: esiste una covariazione tra la quantità delle vocali del dialetto e l'altezza delle vocali dell'italiano. Le parole che in dialetto hanno una vocale fonologicamente lunga hanno in italiano una vocale più bassa rispetto alle parole che in dialetto presentano una vocale fonologicamente breve:

dialetto	*italiano*	
[a:]	[a]	[ˈsak:a]
[a]	[e]	[ˈsek:a]
[o:]	[ɔ]	[ˈfɔs:i]
[o]	[o]	[ˈfos:i]

Poiché sappiamo che, a parità di condizioni, l'altezza vocalica è un fattore fonetico che incide direttamente sulla durata intrinseca della vocale (vocali basse hanno durate maggiori; si veda ad es., Peterson, Lehiste, 1960 e, per l'italiano, Esposito, 2002), è *prima facie* impossibile stabilire se la maggiore durata delle nostre vocali sia il prodotto di una diversa altezza vocalica o sia l'effetto del transfer nell'italiano di differenze di lunghezza delle corrispondenti parole del dialetto.

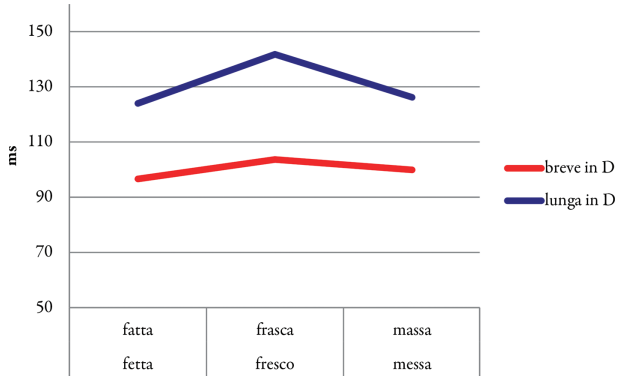
Per poter sciogliere questo fattore di confusione un aiuto proviene dal confronto con i dati di due studi sulla durata vocalica in italiano: quello di Esposito (2002) e quello di Ferrero, Magno Caldognetto, Vagges & Lavagnoli (1978). I valori di durata riportati in questi studi sono relativi ad altre varietà di italiano, nelle quali il sostrato dialettale non è interessato da differenze di quantità vocalica (i dati usati da Esposito sono una media di 7 parlanti provenienti soprattutto da aree centrali e meridionali; i dati di Ferrero et al. (1978) si riferiscono a 10 parlanti fiorentini, figli di genitori fiorentini).

Se dal confronto con dati equivalenti forniti in questi due studi la percentuale di allungamento della vocale bassa rispetto alla vocale alta rilevata nei nostri dati sarà

maggiore di quella attestata in quelle sedi, avremo un indizio, anche se indiretto, a sostegno della nostra ipotesi di transfer.

Abbiamo identificato un sottoinsieme di parole bisillabiche all'interno del nostro corpus compatibili con quelle del corpus utilizzato da Esposito, precisamente quelle che presentano vocali toniche [a] ed [e] in sillaba chiusa. La serie scelta per il confronto è: *fatta-fetta, frasca-fresco, massa-messa*, le cui durate sono rappresentate nella Figura 6.

Figura 6 - *Durata (ms) della vocale tonica nell'italiano bolognese*



In questo sottoinsieme, la vocale breve è pari al 77% di quella lunga, e la percentuale di allungamento è pari all'1,29% (ottenuta dividendo la media di [a] = 131 ms per la media di [e] = 101 ms). In Esposito (2002: 204) [e] è pari al 91,3% della durata di [a] ([a] =150 ms, [e] = 137 ms) con percentuale di allungamento dell'1,09%. I dati di Ferrero et al. (1978), riportati in Esposito (2002), indicano per [a] una media di 200 ms e per [e] una media di 170 ms., con percentuale di allungamento dell'1,17%.

Dal confronto risulta che la percentuale di allungamento di [a] rispetto ad [e] nei nostri dati è maggiore (Tabella 3). Se aggiungiamo a tale confronto anche i dati del dialetto, nel quale la percentuale di allungamento nelle stesse coppie di parole ammonta al 1,55%, emerge una diminuzione progressiva nella percentuale di allungamento di [a] rispetto ad [e] tra il dialetto bolognese, l'italiano bolognese e altre varietà di italiano, con il massimo grado di accorciamento per il dialetto e il minimo accorciamento per l'italiano secondo lo schema seguente:

dialetto bolognese > italiano bolognese > altre varietà di italiano.

Tale progressione può far ritenere che nell'italiano bolognese la durata vocalica rifletta allo stesso tempo sia un fenomeno fonetico intrinseco relativo all'altezza della vocale, legato a condizioni articolatorie e quindi potenzialmente universali, sia un fenomeno linguo-specifico legato all'influenza esercitata dal dialetto. In ogni caso, non escludono quest'ultima possibilità.

Tabella 3 - *Percentuale di allungamento di [a] tonica rispetto a [e] tonica in sillaba chiusa in dialetto bolognese, italiano regionale bolognese (BO) e nei dati di altre varietà di italiano riportati in Esposito (2002) e Ferrero et al. (1978)*

dialetto	*italiano*
BO: 1,55%	BO: 1,29 % Ferrero: 1,17% Esposito: 1,09%

5. *Esperimento 3: italiano regionale bolognese e fiorentino a confronto*

Il modo più diretto per separare il peso relativo dell'influenza (possibile) del dialetto da quella (certa) dell'altezza vocalica è confrontare direttamente la serie di parole pronunciate dai nostri locutori bolognesi con la stessa serie pronunciata da locutori di una varietà di italiano il cui dialetto corrispondente non manifesti fenomeni di quantità vocalica nelle stesso contesto sintattico e prosodico. Il fiorentino è sicuramente una varietà che risponde alle nostre esigenze.

Se la durata delle vocali toniche del bolognese è il risultato combinato dell'influenza del dialetto e della durata intrinseca determinata dall'altezza vocalica, allora ci aspettiamo che le vocali del bolognese abbiano valori di durata maggiori oppure che risultino meno variabili rispetto a quelle del fiorentino, perché sulle variazioni di tipo intrinsecamente fonetico (altezza vocalica) si innesterebbe come un moltiplicatore il coefficiente di variazione indotto dal dialetto, con il risultato di contribuire ad una maggiore uniformità nella distribuzione dei valori.

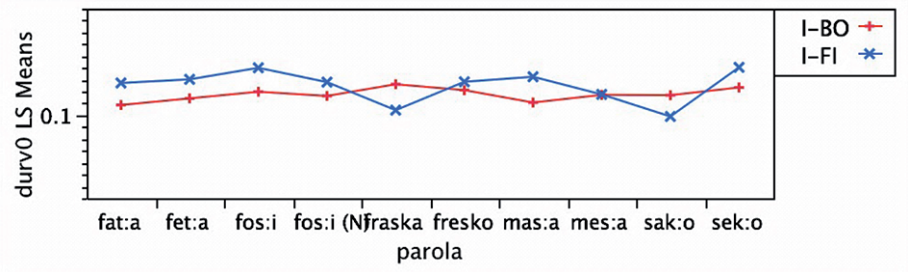
5.1 Risultati

I dati che presentiamo in questa sezione sono il risultato dell'analisi della durata vocalica di cinque ripetizioni dello stesso corpus di frasi pronunciate dai locutori bolognesi prodotte da tre locutrici fiorentine[10] (179 casi). Riportiamo nella Figura 7 il grafico relativo alle durate normalizzate di tutti i bisillabi con sillaba chiusa.

I dati sono stati sottoposti ad un Mixed Model Repeated Measure ANOVA con variabile dipendente la durata in ms o la durata normalizzata; i fattori fissi sono Apertura Vocale, Varietà di italiano (bolognese vs. fiorentino) e Parola, mente i Soggetti costituiscono il fattore random.

[10] Si tratta di tre studentesse universitarie di età compresa tra i 24 e i 26 anni.

Figura 7 - *Durata normalizzata (z-score) della vocale tonica in sillaba chiusa di parole bisillabiche in italiano regionale bolognese e fiorentino*



I risultati indicano che il grado di apertura della vocale, che nell'italiano regionale bolognese covaria con la quantità nel dialetto, è significativo (durata normalizzata: $F(1,1) = 14.948$ $p = 0.0001$), il tipo di parola in cui si trova la vocale è significativo ($F(1,9) = 3.339$ $p = 0.0006$), così come lo è la loro interazione ($F(1,9) = 6.407$ $p < 0.0001$), ma il fattore Varietà di Italiano, non risulta significativo. Gli stessi tipi di significatività valgono anche per il test condotto sui valori di durata assoluta in ms, le cui medie sono riportate nella Tabella 4.

Tabella 4 - *Medie dei valori di durata assoluta delle coppie minime considerate in italiano regionale bolognese e fiorentino*

Bolognese		*Fiorentino*	
+ alta (lunga in dialetto)	- alta (breve in dialetto)	+ alta	- alta
131.33	101.62	138.23	115.72

9. *Discussione e conclusioni*

I diversi italiani regionali si distinguono per caratteristiche fonetiche e fonologiche (segmentali e prosodiche), morfologiche, sintattiche e lessicali; insieme esse rivelano immediatamente all'ascoltatore l'origine diatopica della varietà di italiano del parlante. Negli ultimi anni è stata molto studiata la struttura intonativa (melodica) di parecchie varietà di italiano regionale sia per gli aspetti fonologici (tonali) (ad esempio, Gili Fivela, Avesani, Barone, Bocci, Crocco, D'Imperio, Giordano, Marotta, Savino & Sorianello, 2015), sia per gli aspetti fonetici relativi al *range* dell'estensione melodica (ad esempio Sardelli, Marotta, 2007), sia in prospettiva sociolinguistica, esaminando il rapporto che lega la prosodia dell'italiano regionale a quella del dialetto locale (cfr. ad es., Barone, Gili Fivela & Prieto, 2013, per un esempio di contatto fra italiano regionale e dialetto di Pescara). L'interesse che ha motivato questo nostro lavoro concerne il dominio prosodico delle durate. Ci siamo interrogati sulla natura e sull'origine degli speciali allungamenti vocalici che caratterizzano,

anche percettivamente, la prosodia del bolognese (italiano e dialetto), e più specificamente le sillabe toniche in posizioni metricamente forti della struttura prosodica frasale; ci siamo chiesti, in chiave sociofonetica, se all'origine di questi allungamenti risiedano le distinzioni di quantità vocalica che la letteratura attesta per il dialetto. La nosta ipotesi forte è che le opposizioni di quantità, realizzate foneticamente con differenze molto consistenti della durata vocalica, possano essere state trasferite nell'italiano regionale su base lessicale, cioè mantendo gli allungamenti fonetici della vocale tonica in quelle parole che nel dialetto prevedono una vocale lunga, e che questi allungamenti costituiscano una componente importante degli allungamenti percepiti come particolarmente salienti nelle sedi frasali metricamente forti.

Abbiamo verificato per prima cosa se le opposizioni di quantità siano ancora presenti nel dialetto bolognese intramurario. I risultati del primo esperimento confermano che le opposizioni di quantità sono ad oggi mantenute sistematicamente e che le relative differenze di durata risultano statisticamente significative. Hanno inoltre rivelato un cambiamento fonologico in atto nella struttura sillabica del dialetto che va nella direzione di una sua minore marcatezza: l'inserzione nei monosillabi con nesso consonantico in coda sillabica di una vocale atona paragogica, che in un terzo dei casi casi è uno schewa, comporta da un lato la creazione di una sillaba CV meno marcata e dall'altro il cambiamento della parola da mono- a bisillabica. Tale cambiamento sembra riconducibile all'attrito esercitato dall'italiano sul sistema dialettale.

Nel secondo e nel terzo esperimento abbiamo verificato la presenza di un transfer nell'italiano regionale degli allungamenti vocalici presenti nel dialetto. I risultati hanno posto in luce come le parole che in dialetto hanno una vocale lunga siano realizzati nell'italiano regionale con una vocale di durata significativamente maggiore rispetto alle parole che in dialetto presentano una vocale breve. Ma hanno evidenziato nello stesso tempo la presenza di una covariazione tra questa differenza di durata e una differenza di altezza vocalica: le parole italiane che in dialetto hanno una vocale lunga presentano una vocale più bassa delle parole che in dialetto hanno una vocale breve. Come ci si attendeva sulla base della letteratura sull'argomento, anche i nostri dati mostrano che le vocali relativamente più alte hanno durata minore delle vocali relativamente più basse.

Al fine di separare l'effetto dell'altezza vocalica da un possibile effetto della quantità presente nelle corrispondenti voci del dialetto, abbiamo condotto un terzo esperimento nel quale abbiamo confrontato le vocali toniche prodotte dai locutori bolognesi con le stesse vocali toniche prodotte da locutori fiorentini, ovvero da locutori di una varietà di italiano che non è in alcun modo riconducibile a fenomeni di quantità vocalica. I risultati dell'esperimento non hanno fatto emergere differenze significative nella durata vocalica tra le due varietà. Dobbiamo tuttavia sottolineare che il campione dei parlanti sottoposti al confronto non è bilanciato sociolinguisticamente: il campione dei parlanti fiorentini non corrisponde a quello dei bolognesi né per età (i bolognesi sono più anziani dei soggetti fiorentini) né per genere (quattro unomini e una donna per l'italiano regionale bolognese vs. tre donne per

il fiorentino). Quindi non possiamo escludere la possibilità che un'analisi futura su un campione perfettamente bilanciato ribalti questi risultati.

In conclusione, le nostre indagini hanno confermato il permanere della quantità vocalica nel dialetto bolognese contemporaneo, ma, almeno allo stato attuale dei lavori, non hanno supportato la nostra ipotesi forte di un transfer delle differenze di durata dal dialetto alle corrispondenti parole dell'italiano. La natura di quelli che percettivamente sentiamo come allungamenti particolari del parlato regionale bolognese rimane dunque per il momento irrisolta: i risultati aprono la strada ad una nuova indagine per cercare di svelarne l'origine.

Ringraziamenti

Siamo grati a tutti i soggetti che hanno partecipato a titolo gratuito a questo esperimento e agli studenti Cheng Chen e Arianna Capirossi per il loro aiuto nella registrazione e nella segmentazione di parte dei dati del bolognese. A Giuliano Bocci un ringraziamento speciale per le preziose discussioni sul metodo sperimentale. Desideriamo infine ringraziare i recensori anonimi per i suggerimenti che hanno contribuito a migliorare questo lavoro.

Riferimenti bibliografici

Albano Leoni, F., Caputo, R., Cerrato, L., Cutugno, F., Maturi, P. & Savy, R. (1995). Il vocalismo dell'italiano: analisi di un campione televisivo. In *Studi Italiani di Linguistica Teorica e Applicata*, 24, 405-411.

Auer, P. (2005). Europe's sociolinguistic unity, or: A typology of European Dialect/standard constellations. In Delbeque, N., van der Auwera, J. & Geeraerts, D. (Eds.), *Perspectives on variation. Sociolinguistic, historical, comparative*. Berlin-New York: Mouton de Gruyter, 7-42.

Auer, P. (2011). Dialect vs. standard: a typology of scenarios in Europe. In Kortmann, B., van der Auwera, J. (Eds.), *The languages and linguistics of Europe. A comprehensive guide*. Berlin-New York: de Gruyter, 485-500.

Badini, B. (2002). L'Emilia Romagna. In Cortelazzo, M., Marcato, C., De Blasi, N. & Clivio, G. *I dialetti italiani, Storia, Struttura, Uso*. Torino: UTET.

Barone, M., Gili Fivela, B. & Prieto, P (2013). Intonational change in Pescara: sociolinguistic and linguistic factors. Presentato a *Phonetics and Phonology in Iberia (PaPI)*, 2013.

Benincà, P., Parry, M. & Pescarini, D. (2016). The dialects of northern Italy. In Ledgeway, A., Maiden, M. (Eds.), *The Oxford Guide to the Romance Languages*. Oxford University Press, 185-215.

Bernardasci, C. (2015). Aspetti quantitativi del vocalismo tonico del dialetto di Piandelagotti. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio/ Language acquisition and language loss. Acquisition, change and disorders of the language sound structure.* Milano: AISV, 87-102.

Berruto, G. (1989). On the typology of linguistic repertoires. In Ammon, U. (Ed.), *Status and function of languages and language varieties*. Berlin-New York: de Gruyter, 552-69.

Berruto, G. (2005). Dialect/standard convergence, mixing, and models of language contact: the case of Italy. In Auer, P., Hinskens, F. & Kerswill, P. (Eds.), *Dialect change. Convergence and divergence in European Languages*. Cambridge: Cambridge University Press, 81-97.

Berruto, G. (2012). *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci Editore.

Bocci, G. (2013). *The Syntax-Prosody Interface*. Amsterdam: John Benjamins.

Canepari, L., Vitali, D. (1995). Pronuncia e grafia del bolognese. In *Rivista Italiana di Dialettologia - RID Lingue, dialetti, società*, 19, 119-164.

Cerruti, M. (2011). Regional varieties of Italian in the linguistic repertoire. In *International Journal of the Sociology of Language*, 210, 9-28.

Cerruti, M., Regis, R. (2011). Standardization patterns and Dialect/standard convergence: A northwestern Italian perspective. In *Language in Society*, 43(1), 83-111.

Cheng, C. (2013). Ricerca pilota sull'intonazione dell'italiano regionale bolognese: confronto con quella fiorentina. Tesi di Laurea Magistrale, Università di Bologna.

Coco, F. (1970). *Il dialetto di Bologna. Fonetica storica e analisi strutturale*. Bologna: Forni.

Dal Negro, S., Vietti, A. (2011). Italian and Italo-Romance Dialects. In *International Journal of the Sociology of Language*, 210, 71-92.

De Dominicis, A. (2001). Intonazione assertiva e interrogativa a Bologna. In Magno Caldognetto, E., Cosi, P. (Eds.), *Multimodalità e multimedialità nella comunicazione*. Padova: Unipress, 137-144.

Endo, R., Bertinetto, P.M. (1997). Aspetti dell'intonazione in alcune varietà dell'italiano. In Cutugno, F. (Ed.), *Fonetica e fonologia degli stili dell'italiano parlato, Atti delle VII Giornate di Studio del G.F.S*. Roma: Esagrafica, 27-49.

Esposito, A. (2002). *On Vowel Height and Consonantal Voicing Effects: Data from Italian*. In *Phonetica*, 59, 197-231.

Ferrero, F., Magno Caldognetto, E., Vagges, K. & Lavagnoli, C. (1978). Some acoustic characteristics of Italian vowels. In *Journal of Italian Linguistics*, 3, 87-89.

Filipponio, L. (2012). *La struttura di parola dei dialetti della Valle del Reno*. Bologna: Forni.

Foresti, F. (1988-2005). Italienisch: Arealinguistik V. Emilia Romagna. In Holtus, M., Metzeltin, M. & Schmitt, C. (Eds.), *Lexikon der Romanistischen Linguistik*, vol. IV. Tübingen: Niemeyer, 569-593.

Foresti, F. (1994). Bologna e la Romagna. In Bruni, F. (Ed.), *L'italiano nelle regioni. Testi e documenti*. Torino: UTET, 383-417.

Foresti, F. (2010). *Profilo linguistico dell'Emilia-Romagna*. Roma-Bari, Laterza.

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M. & Sorianello, P. (2015). Varieties of Italian and their intonational phonology. In Frota, S., Prieto, P. (Eds.), *Intonation in Romance*. Oxford: Oxford University Press, 140-197.

Hajek, J. (1992). A Preliminary Investigation of V/C Complementation in Bolognese. In *Progress Reports from Oxford Phonetics*, 5, 25-34.

Hajek, J. (1994). Phonological Length and Phonetic Duration in Bolognese: Are They Related? In Togneri, R. (Ed.), *Proceedings of the Fifth Australian International Conference on Speech Science and Technology*, 2, 662-667.

Hajek, J. (1995). A first acoustic study of the interaction between vowel and consonant duration in bolognese. In *Rivista Italiana di Dialettologia*, 19, 3-10.

Hajek, J. (1997a). Emilia-Romagna. In Parry, M., Maiden, M. (Eds.), *The dialects of Italy*. London: Routledge, 271-278.

Hajek, J. (1997b). Analisi acustica delle quantità segmentali in area bolognese. In *Rivista Italiana di Dialettologia*, 2.

Hajek, J. (2000). How many moras? Overlength and Maximal Moraicity in Standard Italian and Italian Dialects. In Repetti, L. (Ed.), *Phonological Theory and the Dialects of Italy*. Amsterdam: John Benjamins, 111-135.

Loporcaro, M. (2015). *Vowel Length from Latin to Romance.* Oxford: Oxford University Press.

Maddieson, J. (1985). Phonetic cues to syllabification. In Fromkin, V. (Ed.), *Phonetic linguistics.* Orlando: Academic Press, 2013-221.

Nespor, M. (1993). *Fonologia*. Bologna: Il Mulino.

Peterson, G.E., Lehiste, I. (1960). Duration of syllable nuclei in English. In *Journal of the Acoustical Society of America*, 32, 6, 693-703.

Repetti, L. (1992). Vowel Length in Northern Italian Dialects. In *Probus*, 155-182.

Repetti, L. (Ed.) (1995). Epentesi nei dialetti emiliani e romagnoli. In Banfi, E., Bonfadini, G., Cordin, P. & Iliescu, M. (Eds.), *Italia settentrionale: crocevia di idiomi romanzi*. Tübingen: Max Niemeyer Verlag, 181-86.

Repetti, L. (Ed.) (2000). *Phonological Theory and the Dialects of Italy*. Amsterdam-Philadelphia: John Benjamins.

Rizzi, E. (1986). Variabili consonantiche nell'italiano regionale di Bologna. In *Rivista Italiana di Dialettologia. Scuola società territorio*, numero unico, 111-127.

Rizzi, E. (1989). *Italiano regionale e variazione sociale: l'italiano di Bologna.* Bologna: CLUEB.

Sardelli, E., Marotta, G. (2007). Prosodic parameters for the detection of regional varieties of italian. In *Proceedings of ICPhS 2007*, 1281-1284.

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 XVth International Congress of Phonetic Science*, Barcelona, August 2003. Barcelona, vol. III, 2717-2220.

Uguzzoni, A., Busà, M.G. (1995). Corrrelati acustici della opposizione di quantità vocalica in area emiliana. In *Rivista Italiana di Dialettologia*, 19, 7-39.

Vayra, M., Avesani, C. & Fowler, C. (1999). On the phonetic bases of vowel-consonant coordination in Italian: a study of stress and "compensatory shortening". In *Proceedings of the 14th International Congress of Phonetic Sciences*, 1495-498.

Vitali, D., Canepari, L. (1995). Pronuncia e grafia del Bolognese. In *Rivista Italiana di Dialettologia*, 19, 119-164.

Autori

Cinzia Avesani – Istituto di Scienze e Tecnologie della Cognizione del CNR (ISTC-CNR), Padova.
cinzia.avesani@pd.istc.cnr.it

Linda Badan – Department of Translation, Interpreting and Communication, Ghent University, Ghent, Belgium.
Linda.Badan@UGent.be

Linda Barone – Università degli Studi di Salerno.
lbarone@unisa.it

Simona Bernardini – ABC Balbuzie, Padova.
simonabernardini.psic@gmail.com

Lidia Calabrò – CLA, Centro Linguistico di Ateneo, Università degli Studi Roma Tre, Roma.
lidia.calabro@gmail.com

Violetta Cataldo – Università degli Studi di Salerno.
violetta.cataldo@live.it

Maria Assunta Ciardullo – Laboratorio di Fonetica, Dipartimento di Lingue e Scienze dell'Educazione, Università della Calabria, Cosenza.
maria_assunta.ciardullo@unical.it

Lorenzo Ciaurelli – Università degli Studi di Roma "La Sapienza", Roma.
lorenzo.ciaurelli@uniroma1.it

Rone Cole – BLT - Boulder Language Technology, Boulder, Colorado, USA.
rcole@bltek.com

Piero Cosi – ISTC-CNR, UOS Padova, Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Unità Organizzativa di Supporto di Padova, Padova.
piero.cosi@pd.istc.cnr.it

Luca Cristoforetti – Fondazione Bruno Kessler (FBK-irst).
cristofo@fbk.eu

Claudia Crocco – Department of Linguistics, Ghent University, Ghent, Belgium.
Claudia.Crocco@UGent.be

Sonia d'Apolito – Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento & Laboratorio Diffuso di Ricerca Applicata alla Medicina – (DReAM), Lecce.
sonia.dapolito@gmail.com

Anna De Marco – Dipartimento di Studi Umanistici, Università degli Studi della Calabria, Cosenza.
anna.demarco@unical.it

Anna De Meo – Università di Napoli L'Orientale.
ademeo@unior.it

Giorgio De Nunzio – Department of Mathematics and Physics, University of Salento, INFN, and Laboratorio Diffuso di Ricerca Applicata alla Medicina – (DReAM), Lecce.
giorgio.denunzio@unisalento.it

Mariapaola D'Imperio – Laboratoire Parole et Langage, CNRS - Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France.
mariapaola.dimperio@lpl-aix.fr

Mirco Fasolo – Dipartimento di Neuroscienze, Imaging e Scienze Cliniche, Università degli Studi "G. d'Annunzio", Chieti-Pescara.
mirco.fasolo@unich.it

Manuela Frontera – Laboratorio di Fonetica, Dipartimento di Lingue e Scienze dell'Educazione, Università della Calabria, Cosenza.
manuela.frontera@unical.it

Dalia Gamal Abou-El-Enin – Ain Shams University, Il Cairo, Egitto.
daliagamal60@hotmail.com; daliagamal@alsun.asu.edu.eg

Barbara Gili Fivela – Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento & Laboratorio Diffuso di Ricerca Applicata alla Medicina – (DReAM), Lecce.
barbara.gili@unisalento.it

Roberto Gretter – Fondazione Bruno Kessler (FBK-irst).
gretter@itc.it

Mirko Grimaldi – Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento & Laboratorio Diffuso di Ricerca Applicata alla Medicina – (DReAM), Lecce.
mirko.grimaldi@unisalento.it

Massimiliano M. Iraci – Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento & Laboratorio Diffuso di Ricerca Applicata alla Medicina – (DReAM), Lecce.
massimiliano.iraci@unisalento.it

Giovanna Lenoci – Scuola Normale Superiore di Pisa, Pisa.
giovanna.lenoci@sns.it

Valeria Longo – già Alma Mater Studiorum, Università di Bologna.
valerialongom@gmail.com

Paolo Mairano – Centre for Applied Linguistics, Warwick University, Coventry, UK.
p.mairano@warwick.ac.uk

Anna Dora Manca – Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento & Laboratorio Diffuso di Ricerca Applicata alla Medicina – (DReAM), Lecce.
annadora.manca@unisalento.it

Giuliano Mion – Università "G. d'Annunzio" di Chieti, Pescara.
giuliano.mion@unich.it

Aravind Namasivayam – University of Toronto, Oral Dynamics Lab, Toronto, Canada.
a.namasivayam@utoronto.ca

Giulia Natarelli – Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, Padova.
giulia.natarelli@phd.unipd.it

Rosalba Nodari – Scuola Normale Superiore di Pisa, Pisa.
rosalba.nodari@sns.it.

Francesco Olivucci – già Alma Mater Studiorum, Università di Bologna.
francesco.olivucci@studio.unibo.it

Maurizio Omologo – Fondazione Bruno Kessler (FBK), Povo, Trento.
omologo@fbk.eu

Riccardo Orrico – Università degli Studi di Salerno.
rorrico@unisa.it

Giulio Paci – Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Unità Organizzativa di Supporto di Padova, Padova e MIVOQ S.R.L, Padova.
giulio.paci@pd.istc.cnr.it; giulio.paci@mivoq.it

Filippo Pasqualetto – Dipartimento di Neuroscienze, Università di Padova, Padova.
filippo.pasqualetto.1@studenti.unipd.it

Giulia Pedrazzini – Liceo cantonale di Bellinzona.
giulia.pedrazzini@edu.ti.ch

Elisa Pellegrino – Dept. of Computational Linguistics, University of Zurich, Zurich, Switzerland.
epellegrino@unior.it

Marco Pellin – Fondazione Bruno Kessler (FBK-irst)

Massimo Pettorino – Dipartimento di Studi Letterari, Linguistici e Comparati, Università degli Studi di Napoli L'Orientale, Napoli.
mpettorino@unior.it

Caterina Pisciotta – Centro Medico di Foniatria, Padova e A.I.BA.COM. ONLUS, Pisa.
cpisciotta@centrofoniatria.it

Mirco Ravanelli – Fondazione Bruno Kessler (FBK-irst).
mravanelli@fbk.eu

Luciano Romito – Laboratorio di Fonetica, Dipartimento di Lingue e Scienze dell'Educazione, Università della Calabria, Cosenza.
luciano.romito@unical.it

Renata Savy – Università di Salerno.
rsavy@unisa.it

Stephan Schmid – Phonetisches Laboratorium, Institut für Vergleichende Sprachwissenschaft.
stephan.schmid@uzh.ch

Bianca Sisinni – Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) – Università del Salento & Laboratorio Diffuso di Ricerca Applicata alla Medicina – (DReAM), Lecce.
b.sisinni@gmail.com

Giacomo Sommavilla – Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Unità Organizzativa di Supporto di Padova, Padova e MIVOQ S.R.L, Padova.
giacomo.sommavilla@pd.istc.cnr.it; giacomo.sommavilla@mivoq.it

Patrizia Sorianello – Dipartimento di Lettere Lingue e Arti, Italianistica e culture comparate, Università degli Studi di Bari, Bari.
patrizia.sorianello@uniba.it

Alessandro Sosi – Fondazione Bruno Kessler (FBK-irst).
alesosi@fbk.eu

Fabio Tesser – Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Unità Organizzativa di Supporto di Padova, Padova.
fabio.tesser@pd.istc.cnr.it

Graziano Tisato – Istituto di Scienze e Tecnologie della Cognizione (ISTC), C.N.R. di Padova.
graziano.tisato@pd.istc.cnr.it

Pascal van Lieshout – University of Toronto, Oral Dynamics Lab, Toronto, Canada.
p.vanlieshout@utoronto.ca

Mario Vayra – Dipartimento di Filologia Classica e Italianistica, Università di Bologna, Bologna e Istituto di Scienze e Tecnologie della Cognizione del CNR (ISTC-CNR), Padova.
mario.vayra@unibo.it

Debora Vigliano – Graduate School of International Cultural Studies, Tohoku University, Kawauchi, Sendai, Japan.
debora.vigliano.q7@dc.tohoku.ac.jp

Laura Zampini – Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Milano.
laura.zampini1@unimib.it

Paola Zanchi – Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Milano.
p.zanchi@campus.unimib.it

Claudio Zmarich – Istituto di Scienze e Tecnologie della Cognizione (ISTC), C.N.R. di Padova e Dipartimento di Neuroscienze, Università di Padova, Padova.
claudio.zmarich@cnr.it; claudio.zmarich@pd.istc.cnr.it

Studi AISV è una collana di volumi collettanei e monografie dedicati alla dimensione sonora del linguaggio e alle diverse interfacce con le altre componenti della grammatica e col discorso. La collana, programmaticamente interdisciplinare, è aperta a molteplici punti di vista e argomenti sul linguaggio: dall'attenzione per la struttura sonora alla variazione sociofonetica e al mutamento storico, dai disturbi della parola alle basi cognitive e neurobiologiche delle rappresentazione fonologiche alle applicazioni tecnologiche. I testi sono selezionati attraverso un processo di revisione anonima fra pari e vengono pubblicati nel sito dell'Associazione Italiana di Scienze della Voce con accesso libero a tutti gli interessati.

Renata Savy è Professore Associato di Linguistica generale e Linguistica Applicata presso l'Università di Salerno. È membro dell'AISV dalla sua fondazione. I suoi interessi di ricerca vertono su temi di fonetica segmentale e prosodica dell'italiano e delle sue varietà e su aspetti contrastivi di fonetica e fonologia di italiano e altre lingue europee, anche in prospettiva di apprendimento linguistico. Di recente ha pubblicato lavori di analisi delle strategie di interazione dialogica e di interfaccia pragmatica-prosodia in chiave comparativa interlinguistica. Ha partecipato, inoltre, a vari progetti di interesse nazionale per la costituzione e l'analisi di corpora di parlato, occupandosi degli aspetti di codifica ed etichettatura.

Iolanda Alfano è Assegnista di ricerca presso il Dipartimento di Studi Umanistici dell'Università di Salerno, si è addottorata presso l'Università Autonoma di Barcellona. Ha lavorato alla raccolta e all'analisi di corpora di parlato, in particolare in merito ad aspetti fonetici, pragmatici e prosodici. È autrice di studi contrastivi tra l'italiano e lo spagnolo, considerate come lingue prime e come lingue straniere.

AISV - Associazione Italiana Scienze della Voce
sito: www.aisv.it
email: aisv@aisv.it| redazione@aisv.it
ISBN: 978-88-97657-16-3

Edizione realizzata da
Officinaventuno
info@officinaventuno.com | sito: www.officinaventuno.com
via Doberdò, 21 - 20126 Milano - Italy