

# STABILITÀ DEI PARAMETRI NELLO *SPEAKER RECOGNITION*. LA VARIABILITÀ INTRA E INTER PARLATORE: F0, DURATA E ARTICULATION RATE

Luciano Romito <sup>a</sup>, Rosita Lio <sup>a</sup>, Pier Francesco Perri <sup>b</sup>, Sabrina Giordano <sup>b</sup>

<sup>a</sup> Laboratorio di Fonetica, <sup>b</sup> Dipartimento di Economia e Statistica

Università della Calabria

*luciano.romito@unical.it, lio.rosita@libero.it,*

*pierfrancesco.perri@unical.it, sabrina.giordano@unical.it*

## 1. SOMMARIO

La tendenza della ricerca attuale in ambito di *Speaker Recognition* (SR) è volta a individuare informazioni quanto più oggettive possibili presenti nella voce umana analizzando la produzione di un parlatore senza occuparsi della sfera semantica, della produzione linguistica o della struttura sintattica e morfologica. In aggiunta i metodi noti come semiautomatici e parametrici si occupano di dati *considerati* statici. Tale scelta in primo luogo è giustificata dalla relativa facilità della misura e dal trattamento di un ristretto numero di parametri (cfr. Barlow & Wagner, 1998) e in secondo luogo perché la misura di dati statici è la naturale evoluzione di una tradizionale analisi linguistica (cfr. McDougall, 2006).

Sono i segmenti statici quelli utilizzati per lo studio delle lingue, si pensi agli inventari fonologici, alle aree di esistenza delle vocali costruite su porzioni stazionarie (*mid point* o *steady state*), alle rotazioni consonantiche o alle regole fonologiche. Tale analisi prende lo spunto dalla necessità di differenziare due lingue, due dialetti o una lingua da un dialetto. Così, grande spazio nelle riviste, occupano concetti quali isoglosse o isofone utilizzati per identificare confini ideali tra due lingue o tra due dialetti.

Quanto detto risulta funzionale per differenziare ma non per riconoscere, o addirittura identificare. Di fatto anche il concetto di isoglossa oggi viene sostituito dall'idea più 'analogica' di corridoio di transizione, una larga fascia dove coesistono variabili differenti che caratterizzano entrambe le lingue o i dialetti contigui.<sup>1</sup>

Un parlante nel produrre un messaggio o un atto comunicativo attraverso un meccanismo astratto (linguistico), organizza *target* e *goal* che, in seguito, verranno tradotti in azioni che si realizzeranno in un 'progetto fonetico'. Il meccanismo linguistico è essenzialmente l'insieme delle regole e della grammatica del parlante; è la lingua costituita dal lessico, dalla morfologia, dalle opposizioni fonologiche, dalla sintassi, ecc. Tale meccanismo è fortemente influenzato dall'età, dal sesso, dal controllo fonologico, da fattori sociali quali l'origine geografica, lo stato economico, il contesto, la scolarizzazione, ecc. Nolan, a tal proposito, nel 1997 (p.749) scrive: "In implementing the resources of their linguistic mechanism, speakers have to map them onto their individual anatomy. Whilst the requirements of communication may determine many of the details of speech articulation, we may hypothesize that there may be aspects of speech production where each individual

---

<sup>1</sup> In un paese della preSila catanzarese (Soveria Mannelli) coesistono due variabili per il passato e l'imperfetto: la variabile catanzarese [ˈji:vi] e quella cosentina [ˌsiɲuˈju:tu] 'sono andato'.

is free to find his or her own articulatory solution. The speaker's behavior here is not 'learned' as part of the shared knowledge of the linguistic community; rather it is acquired, probably by trial and error".

Due differenti parlanti possono eseguire progetti fonetici differenti per lo stesso scopo linguistico e le conseguenze acustiche di tali progetti possono aiutare molto nel differenziare, anche se, a nostro avviso le modifiche non riguarderanno la parte statica del segnale.

Lo scopo di questo progetto di ricerca, i cui primi risultati sono stati presentati ai Convegni AISV 2006 e 2007, è quello di studiare la variabilità interna di alcuni parametri acustici, di verificare la correttezza di un confronto o di una comparazione basata sul progetto fonetico e quindi su parametri dinamici e di comparare i risultati ottenuti con quelli basati su parametri considerati statici. Verrà analizzato soprattutto l'effetto prodotto da differenti canali di registrazione, da differenti stili di parlato e da differenti software di analisi. In questa ricerca l'attenzione non è focalizzata sul numero degli intervistati bensì sulla varietà dei canali di registrazione investigati e degli stili di parlato considerati.

## 2. PREMESSA

La voce è molto più di una semplice sequenza di suoni. Essa è intrinsecamente articolata e gran parte della sua complessità è legata ai rapporti tra le singole variabili che operano al suo interno come ad esempio il senso, il significato, le intenzioni, le emozioni, lo stato di salute, lo stato sociale, il livello di autostima, il livello di scolarizzazione, ecc. Tutto ciò, ovviamente, è molto importante dal punto di vista forense, almeno potenzialmente, visto che è difficile isolare acusticamente le variabili legate ad ogni singolo livello e considerato che tutte vengono veicolate in un unico canale: quello acustico. J. Laver a riguardo scriveva "The voice is the very emblem of the speaker, indelibly woven into the fabric of speech. In this sense, each of our utterances of spoken language carries not only its own message, but through accent, tone of voice and habitual voice quality it is at the same time an audible declaration of our membership of particular social and regional groups, of our individual physical and psychological identity, and of our momentary mood" (Laver, 1994: 2). Per effettuare una corretta misurazione quindi, è necessaria, una profonda conoscenza della correlazione esistente tra singola variabile ed effetto acustico prodotto. Solo grazie a questa competenza è possibile identificare i parametri da estrapolare, interpretare i segnali sonori e decidere se i campioni di voce sono comparabili.

La tendenza della ricerca italiana e del mondo occidentale in genere, in ambito di SR, è volta ad individuare informazioni quanto più possibili oggettive presenti nella voce umana. L'attenzione è quindi concentrata su tutte quelle informazioni acustiche presenti nella voce, trascurando, da una parte la sfera semantica, morfologica, fonologica e sintattica e dall'altra variabili stilistiche come quelle diafasiche, diastratiche o diatopiche presenti nel segnale sonoro.

Le indagini peritali fino ai primi anni '80 si basavano essenzialmente su aspetti strutturali (paradigmatici) più che su caratteristiche acustiche. Tali indagini linguistiche, glottologiche o sociolinguistiche, che diedero origine anche ad un filone di studio e di ricerca noto come 'sociolinguistica giudiziaria' (v. Trumper, 1979), furono utilizzate per identificare una comunità linguistica più che l'identità di un singolo parlante. Venivano utilizzate soprattutto in fase di indagini preliminari al fine di ricercare la provenienza di voci anonime presenti in rivendicazioni a sfondo terroristico o in casi di sequestro di persona. Tutto ciò era possibile soprattutto perché il materiale sonoro da analizzare e da studiare era abbon-

dante (si veda cosa è successo dopo l'applicazione del 'blocco telefonico'),<sup>2</sup> al contrario di quanto accade oggi (cfr. Romito *et al.*, 1996a, e 1996b). Il tempo dedicato ad ogni singola indagine (consulenza) dagli esperti, era molto maggiore e spesso frutto di stretta collaborazione tra competenze diverse quali la musica, la linguistica, la fisica acustica e la statistica.

Oggi sarebbe impossibile, e forse anche inutile, effettuare una consulenza (o perizia) linguistica; i motivi sarebbero da ricercare nella competenza dialettologica dell'esperto, nei tempi molto lunghi richiesti per l'analisi che poco si conciliano con la velocità imposta dalle indagini ed infine nella richiesta di consulenze che riguardano più il singolo parlante che la comunità di appartenenza; inoltre con materiale sempre più scarso sia in quantità che in qualità, sarebbe come attribuire un breve articolo anonimo pubblicato su un giornale ad uno scrittore noto.<sup>3</sup> Questo tipo di inchieste non si basano su ricorrenze lessicali o su intercalari, né tanto meno su minuzie o articolazioni particolari ma sull'individuazione di regole fonologiche, morfologiche e sintattiche presenti in una registrazione. Ad esempio, analizzando (sintatticamente) una registrazione anonima di un parlante meridionale calabrese potremmo concentrare la nostra attenzione sulla posizione che assume il pronome possessivo in alcuni particolari contesti come i nomi parentali. Potremmo constatare che i dialetti della costa calabrese Tirrenica (per esempio: Palmi, Delianuova) antepongono il pronome possessivo (*me patre, me frate, me soru* "mio padre, mio fratello, mia sorella") mentre al contrario i dialetti della costa calabrese Ionica (per esempio: Catanzaro lido, Soverato, Siderno, Locri) post pongono il pronome con diversi esiti di suffissazione (*patrimma, fratimma, soemma o sorma* "mio padre, mio fratello, mia sorella"). Ipotizzando di aver identificato nella registrazione anonima una provenienza del parlante dalla costa ionica calabrese, una seconda analisi (fonologica-fonetica) sempre sulla stessa registrazione potrebbe riguardare l'esito della doppia -LL- latina nei dialetti in questione. I dialetti del catanzarese prevedono un esito occlusivo retroflesso sonoro [d̥d̥] quindi ILLUM > *id̥du* "lui"; la zona più a sud sempre sulla costa Jonica come Roccella Jonica ecc. prevede un esito approssimante palatale sonoro [j] *iju* "lui" mentre ancora più a Sud (Badolato)

---

<sup>2</sup> Possibilità di rintracciare la provenienza di una telefonata, il numero e l'apparecchio telefonico. Tale operazione è possibile solo se la conversazione è sufficientemente lunga. Da allora i malviventi iniziarono a ridurre drasticamente la durata delle telefonate.

<sup>3</sup> Si pensi ad esempio alla grande difficoltà nel'attribuire alcuni articoli di giornale a Antonio Gramsci (cfr. Basile & Lana, 2009). C'è da aggiungere però che nonostante quanto affermato in Italia vengono effettuate senza alcun fondamento scientifico, delle perizie linguistiche finalizzate al riconoscimento del parlante basate su poche frasi e su analisi basate su idioletto. È inutile affermare che anche se in Italia non esiste nessun controllo nelle Aule di Tribunale sul fondamento scientifico di alcune consulenze questo non significa che le stesse abbiano qualche valore. Negli Stati Uniti fino al 1997, l'ammissibilità del giudizio era essenzialmente basata sugli standard di Frye o di McCormik (Frye, 1923), dopo tale data vengono riportati i criteri essenziali per l'accettabilità di un metodo in ambito forense. I criteri per la scientificità di un metodo sono: qualunque teoria o tecnica che vuole essere tale deve essere testata; una tecnica deve essere stata pubblicata o sottomessa ad un *peer review*; deve essere stato considerato il potenziale errore; deve essere dichiarato se esiste uno standard e se questo è sotto il controllo dell'operatore della tecnica; il livello di accettazione della tecnica all'interno della comunità scientifica (cfr. Rose, 2002: 121).

l'esito è monovibrante alveolare sonoro [r] *iru* "lui". Individuata la zona si potrebbe andare ancora più nello specifico concentrandosi su processi metafonetici o su strutture sintattiche come l'uso dell'infinito opposto al /*ma, ca, u*/ + verbo al presente [po'ter(r)ə cjo'viri] (base latina) versus [po'tera ma 'cjoval] (base greco-bizantina) "potrebbe piovere". L'incrocio e la coesistenza di una serie di variabili (correttamente identificate) conduce alla identificazione di una precisa comunità linguistica. Tanto precisa sarà l'identificazione della comunità linguistica quanto unica e particolare risulterà essere la variabile considerata come nel caso dell'esito di -LL- in [ʎ] presente solo in un piccolo paese della Calabria aspro montana o ancora il processo di tritongazione riscontrato solo in un quartiere della città di Reggio Calabria.<sup>4</sup>

Ovviamente è cosa molto differente identificare un singolo parlatore attraverso la sola analisi linguistica.

### 3. EXCURSUS SUI METODI DI SR

I metodi di SR utilizzati oggi in ambito forense, vengono suddivisi in automatici, semiautomatici e soggettivi. In questa tassonomia, l'attenzione è rivolta all'intervento dell'operatore sull'analisi e sulla estrapolazione dei parametri utili alla comparazione. In questa sede si è scelto, invece, di basare la categorizzazione sui parametri utilizzati definiti 'statici', 'dinamici' e 'dinamico-selettivi'; un tentativo di avvicinare i metodi automatici con il controllo dell'operatore.

#### 3.1 Metodi che utilizzano dati statici

In questa sezione possono essere sicuramente annoverati tutti i metodi semiautomatici e manuali definiti parametrici ed alcuni metodi automatici come quello basato sulla funzione dissipativa (*functional dissipation*).<sup>5</sup>

Per la maggior parte dei metodi semiautomatici e parametrici oggi utilizzati in ambito forense, i dati definiti statici sono identificati nelle porzioni stazionarie delle vocali (in genere quelle toniche). La scelta è giustificata dalla *relativa* facilità della misura e del trattamento di un ristretto numero di parametri (cfr Barlow & Wagner, 1998). Inoltre la misura delle parti stazionarie delle vocali toniche (dati statici) è la naturale evoluzione di una tradizionale analisi linguistica/dialettologica (cfr. McDougall, 2006), basti pensare al concetto già presentato di isoglosse e isofone o alle mappe tematiche basate su singole variabili o su inventari fonologici, e sistemi vocalici di derivazione latina (v., ad esempio, Tagliavini, 1982). Altro motivo invece riguarda la correlazione, anche questa di tradizione

---

<sup>4</sup> Si confrontino gli studi dialettologici a partire dagli anni '80.

<sup>5</sup> Il metodo attraverso la funzione dissipativa è ancora sperimentale e ad oggi non è mai stato utilizzato in ambito forense; cfr. Napolitano *et al.* (in stampa) per l'ambito forense e Napolitano *et al.* (2002) per l'ambito medico: "Functional dissipation is based on signal transforms, but uses the transforms recursively to uncover new features. We generate a variety of masking functions and 'extract' features with several generalized matching pursuit iterations. In each iteration the recursive process modifies several coefficients on the transformed signal with the largest absolute values according to the specific masking function; in this way the greedy pursuit is turned into a slow, controlled, dissipation of the structure of the signal that for some masking functions, enhances separation among classes".

linguistico-fonetica (cfr. Fant, 1960), tra l'impostazione articolatoria e il relativo effetto acustico (nel nostro caso il parametro da estrapolare). Così, ad esempio, il valore acustico della prima formante vocalica corrisponderà (con una relazione inversamente proporzionale) all'altezza della lingua lungo un asse basso-alto all'interno dell'apparato boccale, ecc. Nel nostro esperimento le misure vengono effettuate solo sulla porzione stazionarie delle vocali caratterizzate da accento frasale e meglio rispondenti al concetto di target articolatorio.

### 3.2 Metodi che utilizzano dati dinamici

I metodi che utilizzano esclusivamente dati (acustici) dinamici sono quelli automatici o semiautomatici.<sup>6</sup> Questi, molto diversi tra loro, si basano essenzialmente su spettri a lungo termine (LTS) o su coefficienti Melcepstrali e al momento non vengono utilizzati in ambito forense.<sup>7</sup> Esiste però la possibilità di associare a metodi che utilizzando dati statici un dato dinamico come la stima della velocità di eloquio o meglio di articolazione (*Articulation Rate*; cfr. Künzel, 1997; Zavattaro, 2005).

### 3.3 Altri metodi

Un discorso differente deve essere effettuato per i metodi soggettivi come i *voiceprints* (o confronto dei sonogrammi) e i metodi percettivi uditivi. Vengono entrambi utilizzati in ambito forense in Italia, nonostante la comunità scientifica internazionale ne sconsigli l'uso (soprattutto per il confronto dei sonogrammi) visto l'alta probabilità di errore. Per quanto riguarda il confronto dei sonogrammi in ambito forense (v. Tosi, 1979), la comunità scientifica si è più volte pronunciata sulla sua inaffidabilità (cfr. Gruber & Poza, 1995: 54-71). Tale metodo si basa essenzialmente su due protocolli: il primo protocollo è stato sviluppato da VIAAS (*Voice Identification and Acoustic Analysis SubCommittee*, della *International Association for Identification*) ed è stato pubblicato negli atti dell'associazione VCS 1991;<sup>8</sup> il secondo protocollo schematizzato dell'FBI è stato pubblicato in Koenig (1986: 2089-90).<sup>9</sup> I protocolli sono molto simili, entrambi sono soggettivi e basati

---

<sup>6</sup> Per un giudice, in maniera naturale e istintiva, è più facile comprendere il metodo uditivo o quello parametrico. Molto difficile è, invece, la comprensione del metodo automatico se non si ha una profonda conoscenza sia del metodo che dello speech processing.

<sup>7</sup> Un tentativo di correlare l'impostazione articolatoria e il coefficiente cepstrale è da imputare al lavoro di Clermont & Itahashi (1999), secondo cui la qualità vocalica (quindi i valori formantici) sono in stretta correlazione con la variazione del II e del III coefficiente cepstrale.

<sup>8</sup> VCS (1991:373-9): "Ideally, the exemplar should be spoken [by the suspect] in a manner that replicates the unknown talker, to include speech rate, accent, (whether real or feigned), hoarseness, or any abnormal vocal effect. In general, the suspect is instructed to talk at his or her natural speaking rate: if this is markedly different from the unknown sample, efforts should be made through recitation to appropriately adjust the speech rate of the exemplar. Spoken accents or dialects, both real and feigned should be emulated by the known speaker. If any other unique aural or spectrally displayable speech characteristics are present in the questioned voice, then attempts should be made to include them in the exemplars".

<sup>9</sup> "AFTI: Visual comparison of spectrograms involves, in general, the examination of spectrograph features of like sounds as portrayed in spectrograms in terms of time, frequency and amplitude... Aural cues... include resonance quality, pitch, temporal factors,

sull'esperienza dell'esperto. Le critiche mosse a tale metodo riguardano l'identificazione degli elementi minimi utilizzati per la comparazione (Hollien, 1990: 215), l'impossibilità di presentare le evidenze dell'esaminatore o le caratteristiche numerabili e, infine, l'utilizzo di parametri qualitativi (Aitken, 1995: 14-15). Al momento il metodo sembra essere più intuitivo che analitico.<sup>10</sup>

Il metodo Percettivo-Uditivo sfrutta la capacità del singolo individuo a riconoscere la similitudine o la differenza tra due voci. Alcuni tra i metodi utilizzati sono; il *Panel Approach* (comparazione di coppie di frasi anche di diversa durata e tipo; le risposte sono in percentuale e si basano su caratteristiche stilistiche, linguistiche e acustiche); il *Direct Processing* (un ascoltatore esperto ascolta un intero brano e ne identifica la voce) e l'*Aural-Perceptual Approach* o *Aural-Spectrographic Method*<sup>11</sup> (che prevede una combinazione del Metodo Percettivo-Uditivo e del confronto dei *Voiceprints* o sonogrammi; cfr. Hollien, 1990: 215; McDermott *et al.*, 1996).<sup>12</sup> Il metodo è quello di più facile comprensione per un giudice e una Corte.

#### 4. PARAMETRI STATICI E PARAMETRI DINAMICI

Un parlante nel produrre un atto linguistico mette in campo tutta una serie di aggiustamenti e di processi coarticolatori sia segmentali che sovrasegmentali. Ogni produzione è influenzata oltre che da semplici processi fonologici e fonetici anche da variabili diafasiche, diastatiche, diatopiche, come già detto. Ad esempio, il tempo e la velocità di eloquio sono variabili che influenzano ed incidono molto sulla produzione del parlato spontaneo. Senza entrare molto nello specifico le lingue del mondo vengono classificate e divise proprio in base alla gestione del tempo. In tutte le lingue aumentare la velocità vuol dire ridurre la precisione nell'articolazione di alcuni segmenti o sillabe o addirittura cancellarne alcune ritenute ridondanti per la comprensione del messaggio linguistico. Quindi, mentre da una parte in molti dialetti meridionali si registra la sola

---

inflection, dialect, articulation, syllable grouping, breath pattern disguise, pathologies and other peculiar speech characteristics.”

<sup>10</sup> V. Kersta (1962: 1253): “Voiceprint identification is a method by which people can be identified from a spectrographic examination of their voice. Closely analogous to fingerprint identification, which uses the unique features found in people’s fingerprints, voiceprint identification uses the unique features found in their utterances”. V. anche Nash, citato in Hollien (1990: 224): “As each one of the ridges of your fingers or on the palm of your hand differ from each other, so do all of the other parts of your body. They are unique to you ... including your voice mechanism”. Tale metodo è stato sviluppato e commercializzato da Kersta (1962). Infine, cfr. Tosi (1979): “... the legal application of speaker identification, which at present still consists mainly in the practice of visual examination of spectrograms...”.

<sup>11</sup> Questo metodo è ancora usato almeno fino al 2001 dall’FBI (cfr. Nakasone & Beck, 2001), dalla Polizia Giapponese (cfr. Osanai, 1995), in Israele, Italia, Spagna, Columbia (cfr. Rose, 2002), non viene più usato in Olanda e Germania (cfr. Künzel, 1994: 138).

<sup>12</sup> Sentenza dello Stato della California: “That the aural spectrographic analysis of the human voice for the purposes of forensic identification has failed to find acceptability and reliability in the relevant scientific community, and that therefore, there exists no foundation for its admissibility into evidence in this hearing pursuant to the law of California”.

centralizzazione delle vocali finali (ipoarticolazione) e quindi parole diverse come *i chiodi*, *il chiodo* e *piove* possono essere prodotte indistintamente [ˈcʲjovə] perché comunque ci penserà il contesto ad esplicitare e differenziare le produzioni, in altre lingue la velocità si ottiene sia centralizzando che cancellando sillabe ritenute ridondanti. Così una frase francese come *je ne sais pas* può, se prodotta velocemente diventare [ʒəneseˈpa] > [j̥seˈpa] > [ʃeˈpa] fino a raggiungere la sua massima influenza e minima produzione in [ˈʃpa]. Le produzioni oscillano quindi da una iperarticolazione dove tutta l'informazione è veicolata dal segnale, ad una ipoarticolazione dove invece l'informazione è data dal contesto o dalle conoscenze pregresse. Come si può notare, gli elementi che non vengono mai intaccati o indeboliti sono i segmenti Tonici (sia vocali che sillabe), quelli definiti Target, Goal e quelli all'interno dei quali troveremo la parte stazionaria da misurare.

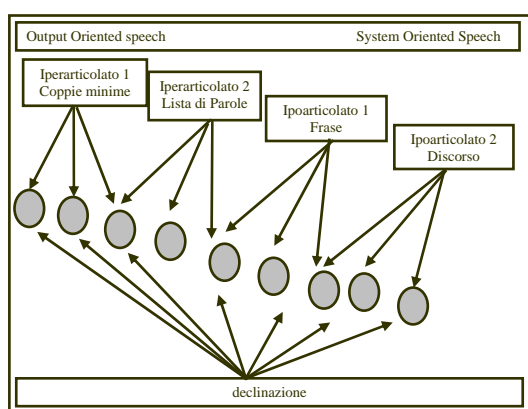


Figura 1: Grafico adattato da Romito *et al.*, 1997

La vocale (o sillaba) tonica è, quindi, (almeno teoricamente) l'unico elemento che raggiunge il target articolatorio, meno intaccato e influenzato dalla coarticolazione, sempre presente e mai cancellata e più vicino a quella 'idea' acustica che abbiamo nella testa e nella *langue* degli strutturalisti. Durante la produzione di questi elementi, ritenuti massimi portatori di informazioni linguistiche, fisiologicamente assistiamo al 'congelamento' delle posizioni geometriche e dei volumi creati all'interno dell'apparato boccale. Tale porzione temporale viene definita *steady state*, *mid point* o semplicemente 'parte stazionaria'. In queste porzioni le formanti hanno un andamento lineare stabile e soprattutto (quasi) parallelo. Questi dati acustici definiti 'stabili' (da noi definiti 'statici'), vengono considerati, nei metodi di comparazione, rappresentativi di ogni singolo parlante. Tale concezione teorica non tiene conto delle aree di esistenza vocaliche, della qualità fonetica nonché della qualità personale introdotta da Ladefoged.<sup>13</sup> Le differenti vocali vengono identificate e riconosciute dall'ascoltatore perché i valori delle formanti ricadono nell'area di esistenza 'propria' di quella vocale. Tali valori, quindi, così come le impostazioni articolatorie correlate, sarebbero, sempre (teoricamente), rappresentative più della lingua parlata che del

<sup>13</sup> Qualità personale e qualità fonetica: tutte le vocali che ricadono nella stessa area di esistenza hanno uguale qualità fonetica e diversa qualità personale se prodotte da differenti persone.

singolo parlante (cfr. Romito *et al.*, 1996, 1997; ma v. soprattutto Lindbloom, 1990). La mutua comprensione tra appartenenti alla stessa comunità linguistica è possibile proprio grazie, alla condivisione delle stesse aree di esistenza e quindi degli stessi target articolatori.<sup>14</sup>

## 5. IL PROGETTO FONETICO

Un parlante nel produrre un messaggio o un atto comunicativo, attraverso un meccanismo astratto (sul piano linguistico), organizza, target e goal che in seguito verranno tradotti in azioni che si realizzeranno in un 'progetto fonetico'. Come già detto, il meccanismo linguistico è essenzialmente l'insieme delle regole e della grammatica del parlante; è la lingua – costituita da lessico, morfologia, opposizioni fonologiche, sintassi, ecc. – ed è fortemente influenzato dall'età, dal sesso, dal controllo fonologico, da fattori sociali, dalla provenienza geografica, dallo stato economico, dal contesto e dalla scolarizzazione.

Due differenti parlanti possono attuare progetti fonetici differenti per lo stesso scopo o *target* linguistico e le conseguenze acustiche di tale progetto possono aiutare molto nel differenziare i parlanti ma, fatto estremamente importante non interesseranno la parte statica del segnale (cioè la qualità linguistica).

Lo scopo o il *target* può essere comune e avere caratteristiche linguistiche comuni, al contrario il progetto e il percorso che conduce (o guida) allo scopo o al *target*, è individuale e personale. In questa sede il percorso viene chiamato 'progetto fonetico'. Ipotizziamo inoltre, che il progetto fonetico si differenzi maggiormente nei dati dinamici di quanto non faccia nei dati statici e che le transizioni siano più individuali e personali di quanto non lo siano le parti stazionarie (con valore linguistico). Questo lavoro vuole verificare se è più corretto effettuare un confronto o una comparazione (cfr. Rose, 2002; ma anche Ezzaidi, Rouat & O'Shaughnessy, 2001), utilizzando il 'progetto' e i dati 'dinamici' o il metodo parametrico e i dati statici. Lo scopo di questo lavoro, è quindi quello di comparare i dati statici con i dati dinamici, in seguito verranno anche effettuate sperimentazioni sul confronto dei dati acustici articolatori (formanti, F0 ecc.) con dati più prettamente acustici (MFCC, LTS, ecc.). Nel prossimo futuro pensiamo di comparare solo ed esclusivamente le transizioni.

## 6. MATERIALI E METODI

Lo scopo di questo progetto di ricerca è quello di studiare la stabilità di alcune variabili definite statiche, dinamiche e dinamico-selettive.

In questa sede, ci occuperemo della frequenza fondamentale (F0) e della dimensione temporale nel parlato attraverso lo studio di registrazioni avvenute attraverso differenti canali e differenti modalità di produzione. Gli esperimenti si basano (come anche quelli relativi a Romito *et al.* (2007, 2008) su una selezione del *corpus* di voci intercettate PRIMULA.<sup>15</sup>

---

<sup>14</sup> Rammentiamo che la qualità vocalica cioè la posizione sullo spettro delle diverse formanti viene definita qualità linguistica.

<sup>15</sup> Cfr. [http://www.linguistica.unical.it/labfon/home\\_corpus\\_primula.html](http://www.linguistica.unical.it/labfon/home_corpus_primula.html) e il paragrafo successivo per una maggiore definizione.



Le ipotesi di partenza presuppongono (cfr. Romito & Lio, 2008) che una variabile sia portatrice di informazione in ambito di SR quando:

- a) mostra una alta variabilità interparlatore e una bassa variabilità intraparlatores;
- b) è resistente al camuffamento;
- c) ha una alta frequenza di occorrenza;
- d) è robusta durante la trasmissione;
- e) è relativamente facile da identificare e misurare.

### 6.1 Scelta del Campione

PRIMULA è un corpus ristretto di voci calabresi ideato e creato presso il Laboratorio di Fonetica dell'Università della Calabria per la valutazione delle metodologie e dei sistemi di riconoscimento del parlato con particolare attenzione all'ambito forense. Il corpus costituisce una base comune sulla quale misurare le tecniche e le metodologie utilizzate in ambito di SR. Ciò che caratterizza PRIMULA è la modalità con cui lo stesso è stato creato. Durante la fase di ideazione del *corpus* e successivamente durante la fase di acquisizione dello stesso si è ritenuto di dover simulare una situazione reale al fine di avere, a prodotto finito, situazioni simili o quantomeno assai vicine a quelle che si presentano di norma nella maggior parte dei casi forensi. Proprio in virtù di ciò una parte delle registrazioni è stata effettuata con attrezzature normalmente utilizzate per le intercettazioni (grazie all'ausilio di alcuni Commissariati di Polizia e ditte private normalmente utilizzate nelle fasi di intercettazione dagli organi inquirenti). È stato così possibile registrare contemporaneamente ed in parallelo lo stesso materiale prodotto attraverso una microspia installata su un'autovettura e attraverso un cellulare collegato con un telefono fisso. Tale materiale registrato costituisce un'intercettazione 'ambientale' (in automobile) e una registrazione telefonica (tra utenza cellulare e utenza di rete fissa). Il corpus contiene poi, una serie di registrazioni di 'controllo' in condizioni differenti.

Il *corpus* PRIMULA consta di oltre 900 files raggruppati in differenti tipologie di registrazione: camera silente (segnale di alta qualità, utilizzato come training test o come saggio fonico); telefonata in ambiente rumoroso, in strada e in auto; infine intercettazione ambientale in auto (in movimento, durante una sosta e fuori dall'auto). Il tutto per 4 parlatori di sesso maschile con tipologie di produzione differenti sia di parlato letto (singole frasi lette per dieci volte e lettura di singole frasi), parlato spontaneo (sia in dialetto calabrese che italiano regionale) ed inoltre con tre diverse tipologie di voce: voce alta,<sup>16</sup> voce normale e voce bassa.<sup>17</sup>

### 6.2 Scelta dei Parametri

#### 6.2.1 Parametri acustici: frequenza fondamentale (F0)

I parametri scelti sono stati differenziati in 'statici', 'dinamici' e 'dinamico-selettivi'. La differenza risiede, non certo nella variabile, ma nella sua misurazione. I parametri statici comprendono le misurazioni medie di F0 effettuate nella porzione stazionaria della vocale tonica con accento primario con una finestra di analisi non inferiore a 20 ms. Ciò rende

---

<sup>16</sup> Per ottenere una voce alta in maniera naturale, il parlante ha letto le frasi oggetto del test con una cuffia che diffondeva musica con un preciso valore di dB.

<sup>17</sup> Per voce bassa si è preferito una voce mormorata.

relativamente semplice la misurazione ma per ottenere un numero sufficiente di dati estrapolati è necessario avere un segnale molto lungo.

Per i parametri dinamici invece è stato considerato l'andamento globale della F0 con valori misurati ogni 0,01 sec. (i valori vengono estrapolati automaticamente su tutto il segnale: vocali, approssimanti, glides e consonanti sonore); l'analisi è molto veloce e si riesce ad ottenere un gran numero di dati anche su un segnale molto breve.

Per i parametri dinamico-selettivi è stata misurata la F0 in maniera dinamica ma solo ed esclusivamente su porzioni vocaliche (siano esse toniche che atone). In questo ultimo caso le vocali sono state differenziate sotto il profilo sia percettivo che spettrale in vocali toniche (VT), divise in *bad* (BVT), cioè vocali toniche che hanno subito un processo di deaccentuazione o caratterizzate da accento frasale secondario e *good* (GVT), cioè vocali sicuramente toniche con accento primario, e vocali atone (VA), suddivise anch'esse in *bad* (BVA) cioè vocali atone finali di parola o di frase e *good* (GVA) cioè vocali che nonostante lo status fonologico di atone vengono percepite come vocali qualitativamente con buona dispersione sul quadrilatero vocalico e buona rappresentazione spettrale. Anche su un segnale relativamente breve è possibile rilevare un sufficiente numero di dati. In un secondo è possibile trovare da 5 a 7 elementi vocalici.

#### 6.2.2 Parametri linguistici: la velocità di eloquio e l'Articulation Rate

Il valore dell'*Articulation Rate* (AR) è stato misurato, sempre sullo stesso materiale sonoro, secondo la seguente formula:<sup>18</sup>

$$AR = \frac{\text{numero di sillabe fonetiche}}{\text{durata della catena fonica}}$$

dove per 'catena fonica', gruppo di respiro (o *run*) si intende tutto ciò che è presente tra due pause. Inoltre, in accordo con Künzel (1997) e Zavattaro (2005: 30), ai fini dello SR non vengono utilizzate le catene foniche con un numero di sillabe inferiore a 6. È stato infatti dimostrato che in questi casi la variabilità interna aumenta enormemente (Zavattaro, 2005).

#### 6.3 Scelta degli algoritmi e dei metodi

Per quanto riguarda gli algoritmi scelti (soprattutto per la misura dei dati statici) la nostra attenzione si è soffermata su alcuni software normalmente utilizzati in ambito forense per l'analisi del segnale e l'estrapolazione dei parametri formantici: IDEM<sup>19</sup> (soprattutto nella sua sezione chiamata *Ares*), *Praat*<sup>20</sup> e *Multi-Speech*<sup>21</sup>.

---

<sup>18</sup> Per una discussione completa sull'*Articulation Rate*, nonché sulle diverse definizioni presenti in letteratura si veda Romito et al., (2006).

<sup>19</sup> IDEM è un software per l'analisi del segnale e per l'identificazione del singolo parlante. È stato sviluppato dalla Fondazione Ugo Bordoni in collaborazione con l'Arma dei Carabinieri. Il software è composto da tre differenti moduli: *Ares* per la misura e l'estrapolazione dei dati parametrici, *Edit* per le funzioni di editing del segnale e *Spread* per la comparazione statistica dei dati. Tale software utilizza contemporaneamente l'analisi LPC e l'analisi cepstrale rappresentati su una finestra di analisi FFT. Il numero dei coefficienti e la finestra di analisi è stata mantenuta fissa, la ricerca della porzione da analizzare è manuale.

<sup>20</sup> *Praat* – che in olandese significa “parola” o “parlare” – è un *software open source* per l'analisi del segnale. È stato sviluppato da Paul Boersma e David Weenink dell'Università

#### 6.4 Le analisi statistiche

Ci preme innanzitutto sottolineare che le analisi statistiche effettuate in questo lavoro non mirano all'identificazione del parlatore in senso forense, ma esclusivamente alla comparazione di campioni di dati al fine di misurare la variabilità intra e inter parlatore.<sup>22</sup> Per l'analisi descrittiva dei dati ci avvarremo di misure di sintesi, dell'ausilio grafico e dell'indice di affidabilità  $\alpha$  di Cronbach, mentre per il confronto tra medie si farà uso dell'ANOVA ad una e a due vie e dell'ANOVA multivariata o MANOVA.<sup>23</sup>

### 7. LEGENDA E TABELLE

Le tabelle, che verranno presentate, conterranno i risultati dei confronti effettuati tra le medie di F0 o di *Articulation Rate* al variare dei parlanti, dei canali degli stili e dei segmenti. Il livello di significatività è sempre fissato a 0,05.

La legenda seguente favorisce la lettura delle tabelle statistiche che verranno presentate nei §§ successivi e l'individuazione delle variabili di volta in volta utilizzate.

I parlanti maschili sono 4, tutti meridionali e provenienti da 4 province differenti. Saranno segnalati attraverso le seguenti etichette LR, SC, AM e VG.

I canali e gli stili utilizzati sono i seguenti:

Micro LS = Canale: *Ambientale in auto - Microspia*; Stile: *lettura*;

Micro PS = Canale: *Ambientale in auto - Microspia*; Stile: *Parlato Spontaneo*;

Micro PS out = Canale: *Ambientale fuori dall'auto - Microspia*; Stile: *Parlato spontaneo*;

Tel Auto LS = Canale: *in Auto- Telefonico*; Stile: *lettura*;

Tel Strada LS = Canale: *in Strada- Telefonico*; Stile: *lettura*;

Tel Aula LS = Canale: *in Aula- Telefonico*; Stile: *lettura*;

Lezione = Canale: *in Aula-microfono*; Stile: *spontaneo formale*.

---

di Amsterdam e viene presentato nell'intro del *software* come "a computer program with which you can analyze, synthesize and manipulate speech".

<sup>21</sup> *Multi-Speech, Model 3700*, è un *software* per Windows che permette di campionare, analizzare e ascoltare. È completo di simboli IPA, e permette di editare, estrarre il pitch ed effettuare analisi formantiche attraverso LPC o FFT. È stato prodotto e sviluppato dalla *Kay Elemetrics Corp.* e dalla *Speech Technology Research Ltd.*

<sup>22</sup> In ambito di *speaker recognition* (anche se al momento in Italia non esiste un protocollo standard), le variabili utilizzate per l'identificazione del parlatore sono esclusivamente acustiche e prevedono la misura media della frequenza fondamentale (F0) e delle frequenze formantiche su segmenti vocalici. I risultati vengono forniti attraverso il *likelihood ratio*. In alcuni casi viene anche fornito la stima dell'errore di falsa identificazione e di mancato riconoscimento utilizzando una popolazione di riferimento.

<sup>23</sup> Per applicare la metodologia dell'ANOVA bisogna verificare che vengano soddisfatte le ipotesi su cui poggia: normalità (verificata con il test di Smirnov o Shapiro) e omogeneità delle varianze (attraverso il test di Lavene. In quest'ultimo caso sono stati scelti test robusti per varianze non omogenee come Welch e Brown – Forsythe e Tamhane nel caso in cui l'omogeneità non venga soddisfatta. Le tabelle, sia relative alla normalità che alla omogeneità della varianza, per motivi di spazio non verranno presentate ma verrà segnalata di volta in volta nella tabella il test post hoc utilizzato.

Altra variabile è la porzione di segnale utilizzata per la misurazione e la stima della frequenza fondamentale. Essa è infatti stata misurata sui seguenti segmenti fonici:

VTG = Vocali Toniche ritenute *Good* sia percettivamente che spettrograficamente;

VTB = Vocali Toniche ritenute *Bad* sia percettivamente che spettrograficamente;

VAG = Vocali Atone ritenute *Good* sia percettivamente che spettrograficamente;

VAB = Vocali Atone ritenute *Bad* sia percettivamente che spettrograficamente.

## 8. CONTROLLO DEI DATI

### 8.1 Variabile Metodo-Algorithmo

La prima comparazione effettuata riguarda i dati estrapolati sullo stesso materiale attraverso diversi *software* con diversi algoritmi (per le specifiche relative alle misure con i singoli *software* e alla segmentazione dei segnali si veda Romito & Galatà, 2008).

Parlanti, Canali e Stili	IDEM	Multispeech	Praat
LR Micro Letto	157.78	156.36	158.80
LR Micro PS in	142.30	144.20	145.93
LR Micro PS out	188.42	189.16	190.14
LR Tel Auto Letto	158.17	159.69	164.85
LR Tel Strada Letto	132.44	133.72	134.25
VG Micro PS	162.22	163.87	174.11
VG Tel Auto Letto	144.30	143.94	150.64
AM Tel Auto Letto	152.9	150.07	152.30
AM Micro PS	151.11	151.30	146.16
SC Micro Letto	124.31	128.00	125.32
SC Tel Auto Letto	152.15	150.88	154.95

Tabella 1: Valori medi di F0 statico misurato solo sulle vocali toniche per parlanti (canale e stile) e *software*

L'analisi scelta è quella multivariata (MANOVA); la variabile utilizzata è la frequenza fondamentale, i parlanti sono LR, SC, AM e VG (divisi anche per canale e stile). Le misure sono state effettuate con i seguenti software: *Idem*, *Praat* e *Multi-Speech*.

Il risultato della comparazione tra le medie della frequenza fondamentale ottenute per i diversi metodi-algoritmi dimostra che non vi è alcuna differenza (Sig=0,7873).<sup>24</sup> Se ne deduce che il metodo-algoritmo non influenza la misura di F0; ciò rende possibile comparare campioni misurati con metodi-algoritmi differenti.<sup>25</sup>

Anche il risultato ottenuto con la stima dell'indice di affidabilità ( $\alpha$  di Cronbach) conferma quanto precedentemente affermato; il risultato di 0,980 confrontato con la tabella di riferimento presente in George & Mallery (2003: 231) rileva che il giudizio di affidabilità può essere definito *Excellent*.

<sup>24</sup>La differenza risulta essere altamente significativa per un valore di Sig <0.01, moderatamente significativa per un valore di Sig compreso tra 0.05 e 0.01 e non significativa per valori di Sig > 0.05.

<sup>25</sup> Ricordiamo che i risultati riguardano esclusivamente la misura del valore della frequenza fondamentale. Molto differenti sono, invece, i risultati relativi alle frequenze formantiche che presenteremo in un prossimo lavoro.

Valori	Giudizi di affidabilità
0,9	Excellent
> 0,8	Good
> <b>0,7</b>	<b>Acceptable</b>
> <b>0,6</b>	<b>Questionable</b>
> 0,5	Poor, and
< 0,5	Unacceptable

Tabella 2: Giudizi di affidabilità secondo George & Mallery (2003:231)

### 8.2 Variabile 'canale'

Abbiamo testato l'affidabilità della variabile canale (telefono, ambiente rumoroso, auto ecc.) attraverso la stima dell' $\alpha$  di Cronbach. L'ipotesi formulata è la seguente: quanto la differenza del canale utilizzato influenza la misura della frequenza fondamentale?

I risultati mostrati nella tabella 2 rivelano un giudizio *questionable* associato all'indice di affidabilità per i parlanti LR ( $\alpha = 0.592$ ) e SC ( $\alpha = 0.576$ ), mentre il livello diventa *acceptable* per il parlante VG ( $\alpha = 0.722$ ). Le misure effettuate sullo stesso materiale sonoro, con gli stessi algoritmi sullo stesso parlante, vengono parzialmente influenzate dal canale utilizzato per la registrazione. Quanto affermato è solo una controprova del fatto che i dati ottenuti da misurazioni di materiale sonoro registrato su canali differenti non devono essere comparati tra di loro vista la difficoltà nello stabilire se la eventuale differenza riscontrata sia da attribuire al parlante o al canale di registrazione.

### 8.3 Distribuzione dei dati

Un ultimo controllo riguarda gli stili di voce utilizzati. Bisogna premettere che la frequenza fondamentale è il correlato acustico della vibrazione delle corde vocali e che il mormorio, il bisbiglio, il sussurro, e in genere la voce prodotta con bassa intensità, prevedono un'assenza di tale vibrazione (o almeno non completa; in alcuni casi infatti, le corde non vibrano per tutta la loro lunghezza vista anche la presenza di un'apertura attraverso le cartilagini aritenoidee che rende difficile la realizzazione del processo di Bernoulli).

La distribuzione dei dati, presentata di seguito, è relativa alla stessa frase, letta dallo stesso parlante, misurata con lo stesso algoritmo, nello stesso ambiente (camera silente), ma con livelli di voce differente: alta, normale e bassa. Come si evince dai grafici di seguito presentati, lo stile di voce influenza molto le misure della frequenza fondamentale. Innanzitutto nel caso di voce bassa il numero delle occorrenze è ridotto (come era prevedibile), i dati non hanno una distribuzione normale ed inoltre il valore medio e la mediana sono molto più bassi rispetto a quelli relativi alla voce normale e della voce alta (che invece innalza i valori di F0).

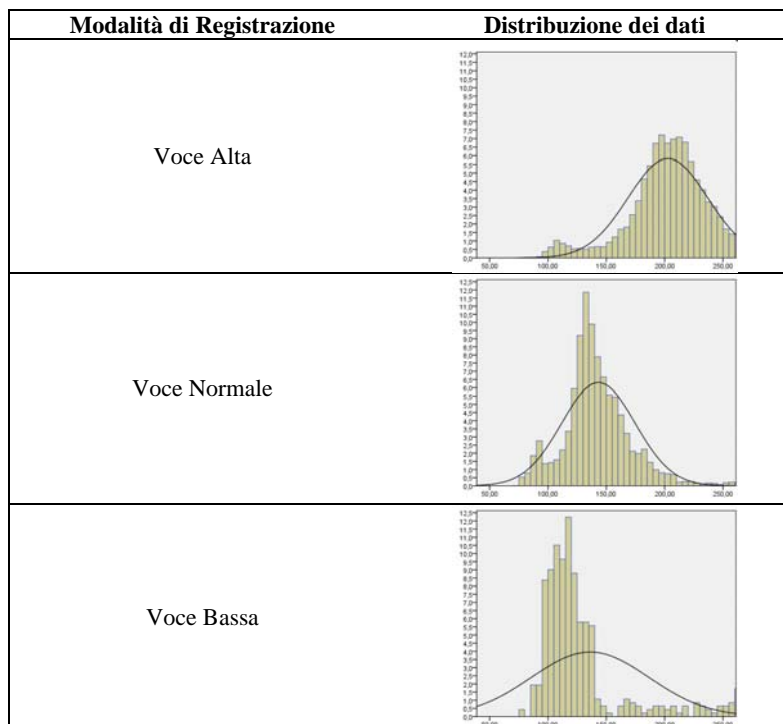


Figura 2: Curve di normalità con identica scala delle misure in Hz della F0 relative allo stesso parlatore sulle stesse frasi in camera silente con Voce Alta, Normale e Bassa<sup>26</sup>

Nel caso della comparazione delle medie di F0 estrapolate dalle registrazioni effettuate su diversi canali e da diversi parlanti si è deciso di omettere lo stile di voce basso. Lo stesso stile verrà, invece, considerato nella comparazione della *Articulation Rate*, in quanto la mancanza di vibrazione delle corde vocali non influenza fisiologicamente la velocità di eloquio.

## 9. ANALISI DEI DATI STATICI

Effettuati i test di controllo sui dati e sui metodi di estrapolazione degli stessi, concentriamo la nostra attenzione sulla variabilità interna di ogni singolo parlante in funzione dei canali e degli stili di parlato.

Verranno confrontati i valori della frequenza fondamentale estrapolati da 4 parlanti (LR, SC, VG e AM) in funzione dei diversi canali (registrazione ambientale, in auto e fuori dall'auto, tramite telefono in strada e in auto). Contemporaneamente verranno considerate le registrazioni in funzione dello stile di parlato (lettura di frasi e parlato spontaneo). Il materiale in nostro possesso non è uniforme quindi anche i confronti effettuati all'interno di

<sup>26</sup> La scala è stata volutamente uniformata anche se ciò ha causato il taglio della coda destra nel caso di voce alta.

ogni singolo parlante saranno differenti. Nelle tabelle seguenti verranno riportate alcune misure statistiche calcolate sui *dati statici* di ogni parlante (le tipologie di registrazione non sono uguali per tutti i parlanti analizzati).

	N	Media	d. std.	Min	Max
LR micro LS	59	157,92	29,223	103	216
LR Micro PS in	17	158,18	28,432	114	211
LR Micro PS in(2)	20	144,30	20,055	113	211
LR Tel Strada LS	20	152,90	36,336	103	222
Lr Micro PS out	17	151,12	31,774	99	205
LR Tel Auto LS	20	152,15	32,300	104	211
Totale	153	154,00	29,745	99	222

Tabella 3: Misure di F0 per LR

	N	Media	D. std.	Min	Max
SC Tel Auto LS	63	142,30	30,238	98	216
SC Micro LS	59	142,32	29,966	99	216
Totale	122	142,31	29,982	98	216

Tabella 4: Misure di F0 per SC

	N	Media	D. std.	Min	Max
VG Tel Auto LS	21	188,43	29,072	129	242
VG Micro PS	18	162,22	26,559	136	216
Totale	39	176,33	30,587	129	242

Tabella 5: Misure di F0 per VG

	N	Media	D. std.	Min	Max
AM Micro PS	25	132,44	19,744	90	174
AM Tel Auto LS	19	124,32	25,684	91	170
Totale	44	128,93	22,590	90	174

Tabella 6: Misure di F0 per AM

Al fine di verificare la presenza di un effetto ‘canale-stile’ sulla frequenza fondamentale si utilizza un’analisi ANOVA ad una via (F0 by ‘canale-stile’) per ciascun parlante.

L’analisi ANOVA fornisce per il parlante LR un valore della statistica F pari a 0.743 con Sig = 0.592. Questo indica che la variabilità interna al parlante non è molto elevata e

quindi diversi canali e stili non provocano differenze significative nei valori medi di F0. Analoghe conclusioni valgono per i parlanti SC (F = 0.00 e Sig = 0.997) e AM (F = 1.410 e Sig = 0.242).

Unico dato contrastante riguarda, invece, il parlante VG per il quale si è riscontrato un effetto rilevante del canale-stile sulla frequenza fondamentale (F = 8.524) evidenziando una significativa variabilità intraparlatore.

Il confronto in questo caso ha riguardato sia i canali che gli stili differenti. È stata comparata la media dei valori estrapolati da una registrazione avvenuta in auto per mezzo telefono con uno stile 'lettura frase' con una registrazione avvenuta in auto attraverso microspia con uno stile 'parlato spontaneo'.

Effettuata questa prima comparazione e questo controllo sulla variabilità interna ad ogni singolo parlatore in funzione del canale e dello stile, abbiamo successivamente comparato i valori medi delle F0, definite statiche ed estrapolate attraverso i diversi canali e i differenti stili (per i dati descrittivi si faccia riferimento alla tabella 1), per i diversi parlatori. A tal fine l'analisi ANOVA fra i gruppi (LR, SC, VG e AM) evidenzia una differenza significativa tra i valori medi di F0 registrati per i diversi parlanti come mostra la tabella seguente:

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	66777,298	11	6070,663	7,241	,000
Within Groups	290066,032	346	838,341		
Total	356843,330	357			

Tabella 7: ANOVA per F0 by parlante

Il passo successivo ha riguardato l'analisi puntuale di ogni singola comparazione, incrociando gli stili dei parlatori come ad esempio **LR Micro LS** vs **SC Tel Auto LS**, o **LR Micro LS** vs **SC Micro LS**.<sup>27</sup>

Dalle analisi effettuate risulta che in molti casi la differenza non è significativa. Questo implica che il valore di F0 statico isolatamente non può essere considerato utile per differenziare i singoli parlanti del nostro esperimento, soprattutto se i dati appartengono a stili e tipologie differenti.

Dal punto di vista della identificazione del parlatore, il valore di F0 come variabile statica risponde a quanto da noi affermato in precedenza rispetto al valore linguistico delle porzioni stazionarie. Il valore statico della F0 delle vocali toniche ha principalmente un valore linguistico, inoltre i 4 parlanti di caratteristiche fisiche e di provenienza geografica simile non producono tutti valori di F0 significativamente diversi. I risultati dell'ANOVA risultano essere significativi solo per il confronto della voce VG vs AM. In tutti gli altri casi il risultato del test è non significativo.

Il grafico seguente costruito con i valori medi della frequenza fondamentale in ogni singola voce mostra, forse meglio della tabella post hoc dell'ANOVA, come esistano tipi diversi di voce (VG e AM) e voci invece molto più simili tra loro (LR e SC) almeno sotto il profilo dei valori di F0.

<sup>27</sup> Le voci messe a confronto sono quelle presenti in tabella descrittive statici.



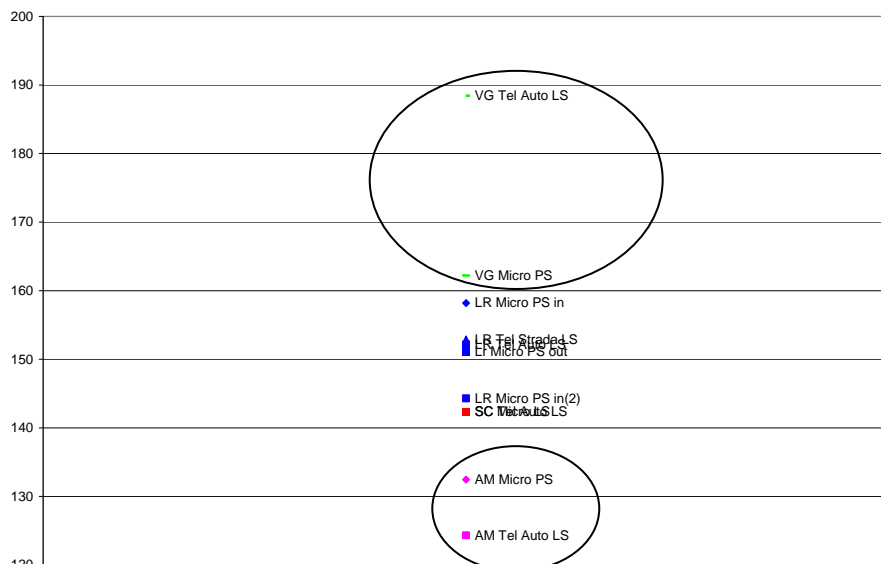


Figura 3: Il grafico presenta i valori medi per ogni parlante in riferimento dello stile e del canale di registrazione

I valori relativi al parlato spontaneo registrati in auto dal parlante LR sono molto più simili a quelli del parlante SC rispetto a quelli registrati dallo stesso parlante LR per telefono o in strada.

### 10. ANALISI DEI DATI DINAMICI

Comparazioni analoghe a quelle effettuate per i dati statici sono state condotte anche sui dati dinamici (cfr § 4). Di seguito vengono fornite le tabelle riportanti i dati descrittivi per singolo parlante (N=593) e successivamente per ogni singolo canale e stile di voce (N=98). I parlanti scelti per questa comparazione sono: LR, SC e VG.

	N	Media	Dev. std.	Min.	Max.
Parlante LR	593	157,3571	31,56250	77,49	228,68
Parlante SC	593	160,0232	35,62091	79,72	294,30
Parlante VG	593	183,1661	41,77483	74,67	299,35

Tabella 8: Valori medi di F0 dinamico misurato su tutto il segnale, relativi ai singoli parlanti (N=593)

Parlante	LR	SC	VG
Tel. Aula LS	155,6433	148,0813	173,1645
Camera Silente; Voce Alta; LS	188,7603	205,0760	214,3250
Camera Silente; Voce Normale; LS	134,8788	137,8921	174,8402
Amb. Micro LS	160,1450	153,1690	159,0089
Tel. Strada LS	144,0436	157,2374	190,9251
Tel. Auto; LS	161,5613	160,3320	187,8102

Tabella 9: Valori medi di F0 dinamico misurato su tutto il segnale, relativi ai singoli parlanti (N=98) differenziati per canali e stili

Con l'obiettivo di valutare l'effetto su F0 delle variabili parlante e canale-stile si è condotta un'analisi a due vie F0 *by* parlante *by* 'canale-stile'. L'analisi evidenzia un impatto significativo di entrambe le variabili. Infatti, si rifiuta l'ipotesi che mediamente la frequenza fondamentale non differisca tra parlanti. Analogamente si evince che anche il canale-stile contribuisce a differenziare i valori medi di F0. In particolare, invece di verificare il confronto tra i singoli canali e stili, abbiamo testato se l'effetto apportato dal canale avesse un peso maggiore o minore di quello apportato dal parlante. L'*Eta quadrato medio* riporta un valore di 0,230 per il canale e 0,119 per il parlante. Questo rivela che la variabilità intra-parlatore con dati estrapolati in maniera dinamica è maggiore di quella interparlatore.

Sorgente	F	Sig.	Eta quadrato parziale
Modello corretto	50,310	,000	,327
Intercetta	49564,524	,000	,966
Parlante	119,018	,000	<b>,119</b>
Canali	105,297	,000	<b>,230</b>
Parlante * Canali	8,972	,000	,049

Tabella 10: Test degli effetti fra soggetti (variabile dipendente: frequenza), Indice Eta Quadrato Medio

Un'ulteriore conferma si ricava dall'*effect size* di Cohen (0,29 per il canale e 0,13 per il parlante). In entrambi i casi il valore algebrico maggiore indica il peso maggiore della variabile considerata. La variabile canale apporta maggiore differenza di quanto non faccia la variabile parlante.

Abbiamo quindi mantenuto costante la variabile canale e effettuato la comparazione tra i diversi parlanti. Nella tabella seguente viene presentata la comparazione tra i parlanti LR, SC e VG all'interno della variabile 'Telefono Aula'. Il risultato per questa variabile è confermato: la differenza tra i singoli parlanti è significativa.

### Confronti multipli

Variabile dipendente: TelefonoAula

	(I) Parlante	(J) Parlante	Differenza fra medie (I-J)	Errore std.	Sig.	Intervallo di confidenza 95%	
						Limite inferiore	Limite superiore
Tamhane	LR	SC	7,56202*	3,03468	,040	,2528	14,8713
		VG	-17,52121*	3,58314	,000	-26,1560	-8,8864
	SC	LR	-7,56202*	3,03468	,040	-14,8713	-,2528
		VG	-25,08324*	3,52395	,000	-33,5771	-16,5894
	VG	LR	17,52121*	3,58314	,000	8,8864	26,1560
		SC	25,08324*	3,52395	,000	16,5894	33,5771

Tabella 11: Confronto dei diversi parlanti all'interno dello stesso canale (telefono aula)

Non viene però confermato anche all'interno degli altri canali come 'microspia' (stile 'lettura') come mostra la tabella seguente. I valori di significatività mostrano che non vi è alcuna differenza significativa tra i singoli parlanti.

### Confronti multipli

Variabile dipendente: Microspia Letto

	(I) Parlante	(J) Parlante	Differenza fra medie (I-J)	Errore std.	Sig.	Intervallo di confidenza 95%	
						Limite inferiore	Limite superiore
Tamhane	LR	SC	6,97607	3,95194	,219	-2,5428	16,4950
		VG	1,13613	5,52730	,996	-12,1997	14,4720
	SC	LR	-6,97607	3,95194	,219	-16,4950	2,5428
		VG	-5,83994	5,41031	,630	-18,9004	7,2205
	VG	LR	-1,13613	5,52730	,996	-14,4720	12,1997
		SC	5,83994	5,41031	,630	-7,2205	18,9004

Tabella 12: Confronto dei diversi parlanti all'interno dello stesso canale (microspia lettura)

## 11. ANALISI DEI DATI DINAMICO-SELETTIVI

I dati presentati in questo § rappresentano, in parte, il nostro concetto di 'progetto fonetico'. Essi evidenziano l'evoluzione dinamica di tutta la produzione, sebbene la selezione porti a considerare solo il valore medio estrapolato in una parte statica del segnale e solo in quelle porzioni vocaliche dove la sonorità risulta essere un dato rilevante. Inoltre è importante ricordare che con questa analisi vi è la possibilità di avere numerosi dati anche in presenza di poco materiale sonoro. Di seguito come negli altri casi esaminati vengono presentate le tabelle riassuntive dei dati descrittivi per i singoli parlanti e le singole voci (canali e stili), nonché per i singoli segmenti fonici (le vocali).<sup>28</sup>

<sup>28</sup> In questo caso abbiamo aggiunto una variabile, la differenza tra le vocali, per sottolineare l'aspetto selettivo della analisi.

	Media	N	Dev. std.	Mediana	Minimo	Massimo
LR micro Letto	162,0996	231	42,54	154	76,00	342,00
SC Telefono Auto Letto	142,9912	226	24,19	140	97,00	229,00
VG Telefono Auto Letto	180,9249	253	31,47	177	120,00	302,00
VG Micro Parlato Spontaneo	178,8148	135	57,03	171	88,00	348,00
LR Telefono Strada Letto	151,4731	260	32,25	149	81,00	323,00
LR Telefono Auto Letto	154,7761	259	26,99	153	94,00	249,00

Tabella 13: Tabella riassuntiva dei dati descrittivi per i singoli parlanti e le singole voci

	Vocale /a/	Vocale /e/	Vocale /i/	Vocale /o/
LR Micro Letto	150,11	158,27	174,55	153,38
SC Tel Auto Letto	131,40	158,50	140,30	140,84
SC Micro Letto	132,94	156,66	147,62	140,11
VG Tel Auto Letto	173,33	200,60	200,80	182,00
LR Micro PS	155,20	161,33	150,33	164,66
AM Micro PS	136,00	118,00	147,00	133,12
VG Micro PS	166,50	147,00	171,25	163,33
LR Micro PS	152,40	134,20	156,60	134,00
LR Tel Strada Letto	138,00	184,40	155,80	133,40
LR Micro PS out	153,0000	154,40	135,66	156,25
AM Tel Auto Letto	124,0000	133,40	121,75	117,60
LR Tel Auto Letto	141,4000	178,00	165,00	124,20

Tabella 14: Valori medi di F0 statico misurato solo sulle vocali toniche diviso per vocale, canale e parlante

Concentrando l'attenzione sulla differenza tra le vocali, abbiamo confrontato i dati estrapolati dalle vocali toniche con quelli estrapolati dalle vocali atone ed inoltre i dati estrapolati dalle vocali definite *good* (cfr. Legenda) e quelle definite *bad*. Ovviamente la qualità vocalica non entra in gioco nelle misure della frequenza fondamentale (o almeno non è così rilevante come la struttura formantica).

Le differenze risultano essere significative nel caso del confronto dei singoli parlanti, mentre il confronto delle singole vocali o dei tipi *bad* e *good* non risulta essere statisticamente significativo.

Vocale	Media	N	Dev. std.	Mediana	Minimo	Massimo	Varianza
VTG	160,12	74	36,217	153,50	101	335	1311,697
VTB	168,35	23	58,255	147,00	96	333	3393,601
VAG	160,79	80	36,237	157,50	80	342	1313,131
VAB	162,05	44	54,759	146,00	76	339	2998,603
Totale	161,60	221	42,872	153,00	76	342	1838,004

Tabella 15: Parlante LR valori di F0 misurato su tutte le vocali presenti differenziate per tipologia<sup>29</sup>

I dati dinamico-selettivi al momento hanno mostrato una buona rappresentazione dei singoli parlanti anche con dati estrapolati da registrazioni su canali differenti e con stili diversi. Gli stessi dati, come già detto, rispecchiano più degli altri la nostra idea di 'progetto fonetico', presentando il percorso che porta al raggiungimento del *target* più che il target stesso. La variabilità inter-parlante risulta essere maggiore di quella intra-parlante.

## 12. ANALISI DELLA ARTICULATION RATE

La velocità è sicuramente un parametro importante nell'eloquio e nel riconoscimento ed identificazione di un parlante. Al momento viene utilizzato solo congiuntamente con altri parametri come la frequenza fondamentale e le frequenze formantiche. L'importanza dell'utilizzo di tale parametro risiede nel fatto che né il canale, né il rumore di sottofondo, dovrebbero influire sulla velocità e sulla estrapolazione dei dati numerici. Ciò che potrebbe influire è invece lo stile (lettura rispetto a parlato spontaneo) e la variabile diafasica (il contesto situazionale). Di seguito sono riportati i valori descrittivi dei singoli parlanti, dei singoli canali divisi per parlante e infine degli stili divisi per parlante.

Per SC e VG non disponiamo di tutti i dati come si può facilmente notare nella tabella 13.

Parlante	Media	N	Dev std.	Mediana	Minimo	Massimo	Varianza
LR	6,52	192	,81113	6,58	3,50	8,04	,658
VG	7,33	192	,80045	7,38	3,47	12,00	,641
SC	6,27	192	1,02779	6,13	4,29	9,53	1,056

Tabella 16: Misure descrittive sull'*Articulation Rate* differenziati per parlante

<sup>29</sup> Le misure sono state effettuate con *Praat* su tutto il materiale registrato.

	LR	SC	VG
Canale	Media	Media	Media
Telefono Aula	7,2758	7,1100	7,3254
Camera silente voce A	6,0371		
Camera silente voce N	7,2792	6,4796	7,4992
Camera silente voce B	6,1608		
Microspia Letto	6,3558	5,5862	7,5996
Microspia Spontaneo in	6,0537	7,5533	7,3175
Microspia Spontaneo out	6,6825		
Lezione	6,4054		
Telefono Strada	6,7517	6,4096	7,5146
Telefono Auto	6,2425	5,6792	7,5713

Tabella 17: Valori medi di AR differenziati per parlante, per modalità e per canale

Stile	LR		SC		VG	
	Media	N	Media	N	Media	N
frasi lette	6,7810	120	6,2587	121	7,5020	120
Parlato Spontaneo	6,3681	48	7,5796	23	7,3175	24

Tabella 18: Valori medi di AR differenziati per parlante e per stile

	Somma dei quadrati	df	Media dei quadrati	F	Sig.
Fra gruppi	119,192	2	59,596	75,919	<b>,000</b>
Entro gruppi	449,805	573	,785		
Totale	568,997	575			

Tabella 19: Analisi ANOVA tra i diversi parlanti (variabile AR)

Le differenze tra i parlanti senza considerare i diversi canali e stili sono statisticamente significative.

Nei confronti multipli però, la significatività persiste esclusivamente tra LR e VG e tra VG e SC, negli altri casi invece la differenza non è significativa. Sicuramente LR e VG hanno due velocità molto differenti mentre non è lo stesso per LR e SC.

	(I) Parlante	(J) Parlante	Differenza fra medie (I-J)	Errore std.	Sig.
Tamhane	LR	VG	-,81505*	,08224	,000
		SC	,25047*	,09449	,025
	VG	LR	,81505*	,08224	,000
		SC	1,06552*	,09402	,000
	SC	LR	-,25047*	,09449	,025
		VG	-1,06552*	,09402	,000

Tabella 20: ANOVA tra i parlanti senza considerare gli stili e i canali (variabile dipendente: AR – la differenza media è significativa al livello 0.05)

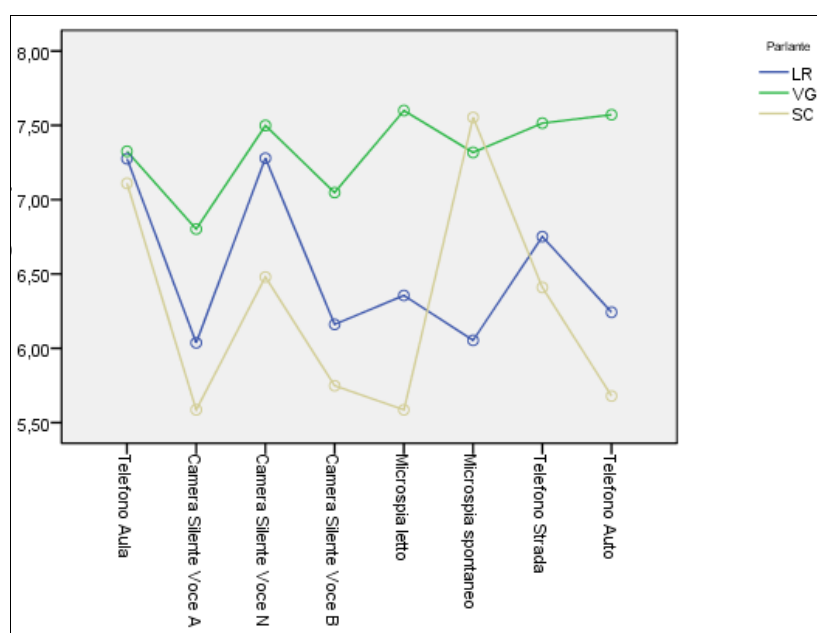


Figura 4: La figura mostra i valori medi per parlante, per canale e per stile

Osservando il grafico delle medie per singola voce e singolo parlante, si nota come sia presente una velocità differente tra i parlanti VG e LR. SC invece in parte segue l'andamento di LR almeno per gli stili 'Telefono Aula', e 'Camera Silente' (Voce A, N e B), mentre registra al suo interno, il valore più basso nella lettura ('Microspia Lettura') e il più alto valore nel parlato spontaneo ('Microspia Parlato Spontaneo'). Tale dato risalta e negli altri due casi LR e VG ha valori contrari. Certo molte sperequazioni potrebbero essere fatte

sulle caratteristiche dei parlanti, sulla loro competenza e abitudine alla lettura o anche sulla gestione del parlato spontaneo e della loro fluenza.

Un ulteriore dato riguarda lo stile voce Alta, Normale e Bassa. In questo caso visto che il parametro non interessa le frequenze, anche la voce con modalità bassa è stata considerata. Come si nota in tutti e tre i parlanti la modalità normale raggiunge velocità maggiori, molto differenti rispetto alla velocità Alta o Bassa. Rammentiamo che il materiale prodotto è identico così come il canale e la modalità di registrazione.

Anche in questo caso abbiamo misurato il peso della singole variabili: parlante e canale. L'Eta quadrato medio rivela, con il suo valore di 0,287 rispetto a 0,208, la maggiore (anche se relativa) importanza del parlante rispetto al canale.

Anche nel confronto tra lo stile 'Lettura' e lo stile 'Parlato Spontaneo' (per il parlante LR è presente anche una registrazione in contesto differente: 'Lezione in Aula') le differenze non sono statisticamente significative.

### 13. CONCLUSIONE

L'identificazione è il risultato secondario di un processo di discriminazione di una voce. Se due entità devono essere discriminate attraverso i loro attributi allora queste, se differenti, devono differire nei loro attributi. Così, se due persone vengono discriminate e riconosciute attraverso la loro voce allora devono differenziarsi ed essere riconosciute attraverso la loro voce.

La voce, così intesa, è un oggetto multidimensionale e come tale deve essere trattato. Riteniamo che solo la competenza di un esperto possa aiutare a scegliere la dimensione più adeguata e la composizione delle differenti dimensioni. Non tutte le caratteristiche, infatti, aggiungono informazione al processo di comparazione, e non tutte le caratteristiche hanno lo stesso peso (statistico) e lo stesso carico informativo.

Questo lavoro non ha le pretese di modificare le condizioni generali delle comparazioni foniche (SR) ma solo di verificare sperimentalmente il peso di ogni singola variabile e soprattutto di valutare la variabilità inter e intraparlante in funzione degli stili di parlato e dei canali di registrazione.

I risultati ottenuti in questo lavoro si differenziano in base ai parametri e alle variabili considerate.

Considerando le variabili definite statiche, i risultati ottenuti dimostrano che la modalità della voce influenza consistentemente i valori della frequenza fondamentale (parametro considerato molto importante nelle comparazioni foniche). Voce Alta, Normale o Bassa producono dati acustici molto differenti (statisticamente rilevanti e significativi) tra loro anche all'interno dello stesso parlante.

Inoltre anche il canale e lo stile influenzano molto la produzione tanto da far sì che lo stile 'letto' di un parlante venga confuso con il parlato 'spontaneo' di un parlante differente. Nei casi in cui le voci considerate e da comparare, sono molto differenti tra di loro, allora la differenza risulterà statisticamente significativa anche se i dati provengono da stili differenti o da registrazioni attraverso canali differenti. In tutti gli altri casi (cioè con voci mediamente simili) il rischio di falsa identificazione è troppo alto. Spesso in tutti i dati da noi analizzati sembra essere molto più influente il canale di registrazione che il parlante.

La variabilità interna della frequenza fondamentale sembra essere maggiore rispetto a quella misurata tra i differenti parlanti. Qualora si ritenesse utile utilizzare le variabili statiche sarà necessario considerare gli stessi stili di produzione e le stesse modalità.



Per quanto riguarda invece i dati dinamici, i risultati ottenuti non sono affatto confortanti. Le voci vengono spesso confuse e i canali di registrazione risultano avere un carico informativo maggiore dello stesso parlante.

I dati dinamico-selettivi al contrario, al momento hanno mostrato una buona rappresentazione dei singoli parlanti anche con dati estrapolati da registrazioni su canali differenti e con stili diversi. Tali dati come già rappresentato in precedenza, rispecchiano maggiormente la nostra idea di 'progetto fonetico', presentando di fatto il percorso che porta al raggiungimento di un target più che il target stesso. In aggiunta, in questo caso non è stata riscontrata differenza significativa neppure all'interno della diversa qualità vocalica.

Riconsiderando le nostre ipotesi di partenza possiamo concludere che i dati dinamico-selettivi mostrano una minore variabilità intra-parlante ed una maggiore variabilità inter-parlante. Non possiamo affermare lo stesso né per i dati statici né per quelli dinamici.

Riguardando le caratteristiche che nei paragrafi precedenti sono state esposte nei confronti delle variabili, possiamo concludere con la tabella seguente:

<b>Caratteristiche Variabile</b>	<b>Risultati parziali</b>
mostra una alta variabilità <b>inter</b> parlante e una bassa variabilità <b>intra</b> parlante	questo è vero esclusivamente per i valori di F0 del parametro Dinamico-Selettivo
è resistente al camuffamento	è stata evidenziata una grande differenza tra Voce Alta e Voce Bassa sia per F0 che per l'AR. Per quanto riguarda i canali invece questi non sembrano camuffare molto i parametri AR e Dinamico-Selettivi
ha una alta frequenza di occorrenza	questo è vero solo nel caso dei parametri dinamici.
è robusta durante la trasmissione	l'unico parametro che sembra non essere influenzato dal canale è l'AR.
è relativamente facile da identificare e misurare	sia F0 (dinamico) che la durata sono molto facili da identificare e misurare.

Ovviamente questo non è che il primo lavoro basato solo ed esclusivamente sulla frequenza fondamentale e sull'*Articulation rate*. Nel prossimo futuro ci concentreremo anche sulle frequenze formantiche e sulla qualità vocalica.

#### 14. BIBLIOGRAFIA

- Aitken, C.G.G. (1995), *Statistics and evaluation of evidence for forensic scientist*, Chichester: Wiley.
- Barlow, M. & Wagner, M. (1998), Measuring the dynamic encoding of speaker identity and dialect in prosodic parameters, in *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Nov 30 – Dec 4, 1998 (R.H. Mannell & J. Robert-Ribes, editors), Sydney: Australian Speech Science and Technology Association, 81-84.
- Basile, C. & Lana, M. (2009), L'attribuzione di testi con metodi quantitative: riconoscimento di testi gramsciani, in *Atti del Convegno Nazionale Ass.I.Term. Terminologia analisi testuale e documentazione nella città digitale*, Aida Informazioni: Roma.
- Clermont, F. & Itahashi, S. (1999), Monophthongal and diphthongal evidence of isomorphism between formant and cepstral spaces, in *Proceedings of the Spring Meeting of the Acoustical Society of Japan*, Meiji University Press, 2005-6.
- Fant, G. (1960), *Acoustic Theory of speech production*, Hague: Mouton.
- Frye (1923), Frye vs. United States (1923), *293 Federal Reports (1<sup>st</sup> series)*, 1013, 1014 (CA).
- George, D. & Mallery, P. (2003), ED425201, *SPSS for Windows Step by Step: A Simple Guide and Reference 11.0 update*, Allyn & Bacon, A Viacom Company: Needham Heights, MA.
- Gruber, J.S. & Poza, F.T. (1995), Voicegram identification evidence, *American Jurisprudence Trials*, 54, Lawyers Cooperative Publishing.
- Ezzaidi, H., Rouat, J. & O'Shaughnessy, D. (2001), Towards combining pitch and MFCC for speaker identification systems, in *Proceedings of Eurospeech 2001*, Aalborg, Denmark, September 3-7.
- Hollien, H. (1990), *The acoustics of crime*, New York: Plenum.
- Kersta, L.G. (1962), Voiceprint identification, *Nature*, 196, 1253-7.
- Künzel, H.J. (1994), Current approaches to forensic speaker recognition, in *Proceedings ESCA Workshop on Automatic Speaker Recognition Identification Verification*, Martigny, Switzerland, April 1994, 135-41.
- Künzel, H.J. (1997), Some general phonetic and forensic aspects of speaking tempo, *Forensic Linguistics*, 4, 48-83.
- Ladefoged, P. (1993), *A Course in Phonetics*, 3rd edition, Sydney: Harcourt Brace College Publishers.
- Laver, J. (1994), *Principles of Phonetics*, Cambridge: Cambridge University Press.
- Lindbloom, B. (1990a), Explaining phonetic variation: a sketch of the H&H Theory, in *Speech Production and Speech Modelling* (W.J. Hardcastle & A. Marchal, editors), Dordrecht: Kluwer Academic Publishers, 135-152.

- Lindbloom, B. (1990b), On the notion of possible speech sound, *Journal of Phonetics*, 18, 135.
- McCormk, P. & Russell, A. (editors) (1996), *Proceedings of the 6th Australian International Conference on Speech Science and Technology*, Canberra: ASSTA.
- McDermott, M.C., Owen, T. & McDermott, F.M. (1996), *Voice Identification: The Aural Spectrographic Method*, Colonia, NJ: Owl Investigations Inc.  
<http://tapeexpert.com/pdf/voiceidauralspectro.pdf>
- McDougall, K. (2006), Dynamic features of speech and the characterization of speaker: Toward a new approach using formant frequencies, *The International Journal of Speech, Language and the Law*, 13, 89-126.
- Nakasone, H. & Beck, S.D. (2001), Forensic automatic Speaker recognition, in *Proceedings of the 2001 Odyssey Speaker and Language Recognition Workshop*, 1-6, Crete, Greece, June 18-22.
- Napoletani, D., Romito, L., Sauer, T. & Struppa, D. (in stampa), *Functional dissipation for speaker recognition*.
- Napoletani, D., Sauer, T. & Struppa, D. (2002) Patent title: *Functional Dissipation Classification of Retinal Images*.
- Nolan, F. (1997), Speaker recognition and forensic phonetics, in *The Handbook of Phonetic Sciences* (W.J. Hardcastle & J. Laver, editors), Cambridge: Cambridge University Press, 744-67.
- Osanai, T., Tanimoto, M., Kido, H. & Suzuki, T. (1995), Text-Dependent Speaker Verification using Isolated Word Utterances based on Dynamic Programming, *Reports of the National Research Institut of Police Science*, 48, 15-19.
- Romito, L. & Galatà, V. (2008a), Speaker Recognition in Italy: evaluation of methods used in forensic cases, in *Actas del IV Congreso de Fonética Experimental* (A. Pamies & E. Melguizo, editors), Granada, Spagna, 11-14 febbraio 2008 (= *Language Design*, Special Issue, vol. 1), Método Ediciones: Granada, 229-240.
- Romito, L. & Galatà, V. (2008b), *Primula: un corpus ristretto di voci calabresi per la valutazione delle metodologie e dei sistemi di riconoscimento del parlatore*, Università della Calabria.
- Romito, L., Maddalon, M. & Trumper, J. (1996a), Atteggiamento della Magistratura nei confronti delle perizie foniche, in *Caratterizzazione del parlatore* (F. Fedi & A. Paoloni, editors), Atti delle VI Giornate di Studio del Gruppo di Fonetica Sperimentale (Roma, 23-24 novembre 1995), Roma: Fondazione Ugo Bordoni, 34-45.
- Romito, L., Maddalon, M. & Trumper, J. (1996b), La parametrizzazione nei test di riconoscimento, in *Caratterizzazione del parlatore* (F. Fedi & A. Paoloni, editors), Atti delle VI Giornate di Studio del Gruppo di Fonetica Sperimentale (Roma, 23-24 novembre 1995), Roma: Fondazione Ugo Bordoni, 87-93.
- Romito, L. (2009), *'Le intercettazioni'*, contributo a *'Ndrangheta: L'educazione e le istituzioni per un progetto comune*, Università della Calabria, 6 marzo 2009.

- Romito, L., Galatà, V. & Lio, R. (2006), Fluency Articulation and Speech Rate as new parameters in the Speaker Recognition, in *Actas del III Congreso de Fonética Experimental*, Santiago de Compostela, 26-24 ottobre 2005, Santiago de Compostela: Xunta de Galicia, 537-549.
- Romito, L. & Lio, R. (2008), Stabilità dei parametri nello *Speaker Recognition*: la variabilità intra e interparlatore, in *La Fonetica Sperimentale: Metodo e Applicazioni*, Atti del 4° Convegno Nazionale dell'Associazione Italiana di Scienze della Voce, Arcavacata di Rende (CS), 3-5 dicembre 2007 (L. Romito, V. Galatà, R. Lio, editors), Torriana: EDK Editore, 125-128 (abstract).
- Romito, L., Tucci, M., & Cavarretta, G. (2009), Verso un formato standard nelle intercettazioni: archiviazione, conservazione, consultazione e validità giuridica della registrazione sonora, *Atti del convegno Convegno Internazionale Ass.I.Term*, Università della Calabria, 6-8 giugno 2008.
- Romito, L., Turano, T., Loporcaro, M. & Mendicino, A. (1997), Micro- e macrofenomeni di centralizzazione vocalica nella variazione diafasica: rilevanza dei dati fonetico-acustici per il quadro dialettologico del calabrese, in *Fonetica e fonologia degli stili dell'italiano parlato* (F. Cutugno, editor), Atti delle VII Giornate di Studio del Gruppo di Fonetica Sperimentale, Napoli, 14-15 novembre 1996, Roma: Esagrafica, 157-175.
- Rose, P. (2002), *Forensic Speaker identification*, London: Taylor & Francis.
- Tagliavini, C. (1982), *Le origini delle lingue neolatine*, Patron: Bologna.
- Tosi, O. (1979), *Voice Identification: Theory and Legal Applications*, Baltimore: University Park Press.
- Trumper, J.B. (1979), *Sociolinguistica giudiziaria*, Padova: CLESP.
- VIAAS, (1991), *Voice Identification and Acoustic Analysis SubCommittee*, della *International Association for Identification*, pubblicato negli Atti dell'Associazione VCS.
- Zavattaro, D. (2005), *Articulation Rate e l'identificazione del parlatore a scopo forense*, Tesi di di dottorato in Scienze Forensi, XVIII ciclo, A.A. 2004-2005, 2a Università di Roma Tor Vergata.