Studi AISV 1

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

a cura di
Mario Vayra, Cinzia Avesani e Fabio Tamburini

AISV
Associazione Italiana Scienze della Voce

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

a cura di
Mario Vayra, Cinzia Avesani e Fabio Tamburini

Milano 2015

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 Schmidt (Università di Zurigo), Carlo Semenza (Università di Padova), Claudio Zmarich (CNR-ISTC).

© 2015 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: 978-88-97657-11-8

Indice

Prefazione di *Mario Vayra e Cinzia Avesani* 7

PARTE I
Variazione fonetica, restrizioni sulla variazione e mutamento

DANIEL RECASENS
Phonetic factors contributing to the inception and evolution of sound change 23

MARIANNE POUPLIER
Between 'whims of fashion' and 'phonetic law': Performance constraints in speech production in the face of linguistic diversity 41

LORENZO FILIPPONIO, SONIA CAZZORLA
The vowels of Bari. A comparison between local dialect and regional Italian 59

CINZIA AVESANI, VINCENZO GALATÀ, MARIO VAYRA, CATHERINE T. BEST, BRUNO DI BIASE, OTTAVIA TORDINI, GRAZIANO TISATO
Italian roots in Australian soil: coronal obstruents in native dialect speech of Italian-Australians from two areas of Veneto 73

FRANCESCA PINTO
High vowels elision in Japanese. A perceptive approach 99

CAMILLA BERNARDASCI
Aspetti quantitativi del vocalismo tonico del dialetto di Piandelagotti 113

ALESSANDRO VIETTI, LORENZO SPREAFICO, VINCENZO GALATÀ
Sull'allofonia di /ʀ/ in tirolese: metodo e dati da un'analisi articolatoria ultrasonografica preliminare 129

ROSALBA NODARI
Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi 139

VALENTINA PAGGINI, SILVIA CALAMAI
L'anafonesi in Toscana. Il contributo degli archivi sonori del passato 155

LUCIANO ROMITO, MARIA ASSUNTA CIARDULLO, ANDREA TARASI
Analisi acustica delle occlusive sorde aspirate del dialetto di San Giovanni in Fiore (CS) 169

PARTE II
Acquisizione, destrutturazione e codifica neurale della struttura sonora del linguaggio

Parte II – Sezione A
Acquisizione di L2

ANNA DE MEO, MARILISA VITALE, YING XU
Lo sviluppo del sistema vocalico dell'italiano nell'interlingua di adulti sinofoni 191

STEPHAN SCHMID, SARAH WACHTER
Le ostruenti sonore nella pronuncia dell'italiano di apprendenti svizzero-tedeschi 203

YILUN TANG, BARBARA GILI FIVELA
Il cinese come L2: sulla percezione di alcune vocali da parte di studenti italiani poliglotti 219

Parte II – Sezione B
Destrutturazione della forma sonora

TYPHANIE PRINCE
On the Jakobson's Postulate: the notion of markedness. A Case study of French Phonology of Aphasia and Acquisition 235

MANUELA BOLLANI
La percezione dello stimolo ritmico tra acquisizione, sintonizzazione e perdita: stato dell'arte e alcune riflessioni 255

ELISA PELLEGRINO, VALERIA CARUSO, ANNA DE MEO
Discriminating between pathological and foreign-accented speech: an acoustic-perceptual study on Italian 271

BARBARA GILI FIVELA, MASSIMILIANO M. IRACI, MIRKO GRIMALDI, CLAUDIO ZMARICH
Consonanti scempie e geminate nel morbo di Parkinson: la produzione di bilabiali 289

GIOVANNA LENOCI, CATERINA PISCIOTTA, CLAUDIO ZMARICH
Profili evolutivi della coarticolazione per una diagnosi precoce di balbuzie cronica in età prescolare 313

DANIELA BELTRAMI, LAURA CALZÀ, GLORIA GAGLIARDI, ENRICO GHIDONI, NORINA MARCELLO, REMA ROSSINI FAVRETTI, FABIO TAMBURINI
Strumenti di *screening* linguistico per l'identificazione precoce della fragilità cognitiva 329

CLAUDIO ZMARICH, CINZIA QUARTARONE, SIMONA BERNARDINI, SABRINA BONICHINI, DANIEL COLANTONIO, GIULIA NATARELLI, DEBORA STOCCO, CATERINA PISCIOTTA
Fluenza verbale e produzione lessicale in bambini di 24 mesi d'età con familiarità alla balbuzie: confronto a scopo di prognosi 345

Parte II – Sezione C
Codifica neurale

ANNA DORA MANCA, FRANCESCO DI RUSSO, MIRKO GRIMALDI
Orderly organization of vowels in the auditory brain: the neuronal correlates of the Italian vowels 357

Parte II – Sezione D
Lingua dei segni e linguaggio musicale

SABINA FONTANA, MARIA ROCCAFORTE
Lo strutturarsi e il destrutturarsi dei suoni nell'interazione con la lingua dei segni italiana (LIS) 371

MARIA ROCCAFORTE
Componenti orali e riconoscibilità nella Lingua dei Segni Italiana (LIS) 383

PAOLO BRAVI
The Construction of the Individual Style of Singing in the Sung Poetry of Southern Sardinia: Reflections upon a Critical Case 391

PARTE III
Trattamento automatico delle lingue naturali

RODOLFO DELMONTE
Expressivity in TTS from Semantics and Pragmatics 407

PIERO COSI, GIULIO PACI, GIACOMO SOMMAVILLA, FABIO TESSER
KALDI: Yet Another ASR Toolkit? Experiments on Italian children speech 429

MASSIMILIANO TODISCO, FABIO POROLI, MAURO FALCONE
Uno strumento per la prototipizzazione rapida di "dialoghi pratici" 439

PIERO COSI, GIULIO PACI, GIACOMO SOMMAVILLA, FABIO TESSER
Building Resources for Verbal Interaction Production and Comprehension within the ALIZ-E Project 455

ANTONIO ORIGLIA, FRANCESCO CANGEMI, FRANCO CUTUGNO
WORDEN: a PYTHON interface to automatic (non)word generation 459

ANDREA PAOLONI, MASSIMILIANO TODISCO
Calcolo del rapporto di verisimiglianza tra parlanti in prove soggettive di ascolto 471

ALESSANDRO VIETTI, VITTORIO ANSELMI, LORENZO SPREAFICO
Verso un sistema di riconoscimento automatico del parlato tramite immagini ultrasoniche 477

Autori 491

DOI: 10.17469/O2101AISV000001

Prefazione

Il volume presenta contributi che possono essere raccolti intorno ad alcuni nuclei tematici considerati centrali nell'ambito della *Laboratory Phonology*: le restrizioni fonetiche sul mutamento in diacronia e sulla variazione sincronica linguistica e sociofonetica (con particolare riferimento alle varietà linguistiche d'area italiana e romanza, ma non solo); l'acquisizione fonetico-fonologica di una lingua seconda e i processi di destrutturazione sonora della lingua prima. A questi nuclei principali si affiancano contributi sulla codifica neurale dei suoni del parlato e sul rapporto speciale che lega, entro una comunicazione di tipo multimodale, la componente orale alla componente gestuale nella Lingua dei Segni Italiana. In chiusura, una sezione dedicata alla "generazione" automatica della lingua parlata, nelle sue strutture e nelle sue variazioni, per mezzo dei sistemi di trattamento automatico delle lingue naturali.

Il libro trae origine dalle tre intense giornate dell'XI Convegno dell'AISV (Associazione Italiana di Scienze della Voce), svoltosi a Bologna nel gennaio 2015 su una sollecitazione di ispirazione jakobsoniana. Intitolato *Il farsi e il disfarsi del linguaggio. L'emergere, il mutamento e la patologia della struttura sonora del linguaggio*, il convegno intendeva infatti proporre all'attenzione degli studiosi il tema del classico e insieme rivoluzionario saggio di Roman Jakobson, *Kindersprache, Aphasie und allgemeine Lautgesetze*, che nel 1941 lanciava l'idea di un nesso fra acquisizione, afasia e leggi del mutamento fonico, generali e diacroniche. La proposta di discutere questi temi alla luce delle teorie linguistiche attuali è stata raccolta da studiosi affermati insieme a giovani e valenti ricercatori, italiani e stranieri. Il volume ospita una selezione dei lavori presentati in quella sede, sottoposti a *blind peer review* dal nutrito comitato scientifico del convegno: nella veste che qui si propone essi costituiscono il frutto del confronto a tratti serrato e sempre costruttivo che ha caratterizzato in un clima informale le giornate di studio bolognesi.

I saggi sono raccolti in tre parti, dedicate a: "Variazione fonetica, restrizioni sulla variazione e mutamento"; "Acquisizione, destrutturazione e codifica neurale della struttura sonora del linguaggio"; "Trattamento automatico delle lingue naturali".

Aprono la prima parte del volume i contributi ad invito di due studiosi, Daniel Recasens e Marianne Pouplier, che presentano lo stato dell'arte rispettivamente sul tema dei fattori fonetici che innescano il mutamento storico dei suoni di una lingua e sulle "restrizioni a livello di esecuzione" associate alla dinamica del meccanismo di produzione.

In *Phonetic factors contributing to the inception and evolution of sound change* Daniel Recasens tratta, sulla base di robuste evidenze sperimentali, fondate su una conoscenza profonda della linguistica romanza, la questione delle cause fonetiche del mutamento fonico (*sound change*: SC). Il saggio si articola in quattro sezioni. La prima mostra come, mentre in casi specifici il SC possa essere innescato primariamente o da variazioni di tipo articolatorio (come assunto dalla maggior parte di linguisti e filologi della prima metà del XX secolo) o da fattori strettamente acustici, secondo un principio di *acoustic equivalence* (Ohala, 1981), spesso le due opzioni risultino di fatto mutuamente compatibili.

La seconda suggerisce che in alcuni casi specifici, il SC può seguire un percorso evolutivo molteplice (è il caso della elisione di /l/ "scura" o della palatalizzazione dei nessi latini /kt, ks/ nelle lingue romanze). La terza e quarta sezione sono dedicate rispettivamente al ruolo della prominenza accentuale (acustica/articolatoria) e della posizione entro la parola nelle assimilazioni e dissimilazioni vocaliche, e al modo in cui il SC ha origine nell'individuo.

Mi soffermerò esclusivamente sulla prima e seconda parte, e in particolare sulla trattazione della vocalizzazione di /l/ in Romanzo; al di là del ruolo cruciale che la vocalizzazione di /l/ svolge nella argomentazione di Recasens, questo caso mi consentirà di illustrare la solidità e la finezza del metodo con cui lo studioso muove nel vastissimo *data-base* offerto dal dominio romanzo.

Tipici casi di SC di origine articolatoria, ci ricorda Recasens, sono quei mutamenti in cui l'ascoltatore fallisce nel " compensare" gli effetti della coarticolazione. In questi casi,

> ...[c]ontext-dependent coarticulatory effects may give rise to articulatory changes in the phonetic implementation of a given phoneme which may lead listeners to believe that a new phoneme has been produced by the speaker.

Un caso illuminante è costituito dal

> ... raising of stressed /a/ to a mid front vowel before an (alveolo)palatal consonant in Early Romance, as in the developments Latin ['akse] AXE "axis" > *['ajse] > Catalan [e(j)'] and Latin ['fakto] FACTU "done" > *['fajto] > Catalan [fet], where [j] in the intermediate phonetic variants *['ajse] and *['fajto] corresponds to a lenited realization of the original syllable-final velar stop.

Ma lo scenario, sotto la lente originale e potente dello studioso catalano, si rivela assai più complesso.

> In partial disagreement with linguists and philologists from the XIX and XX centuries, data reported in sections 1.1.1 through 1.1.3 show that there are sound changes which appear to be acoustically driven and thus are not triggered by articulatory variations.

Come noto, la vocalizzazione di /l/ in Romanzo può produrre gli esiti [w] o [j] a seconda che la laterale sia rispettivamente chiara o scura. "Thus, the replacement of dark /l/ by [w] accounts for the dialectal Catalan variants a[w]ba ALBA "dawn" and esca[w]far EXCALEFACERE "to heat", and the substitution of /l/ by [j] for

Italian amp[j]o derived from Latin AMPLU". In accordo alla *acoustic equivalence hypothesis* (Ohala, 1974), questi mutamenti, osserva Recasens, sono associati al fatto che lo stadio stabile di /l/ abbia una frequenza di F2 bassa, intorno a circa 800-1000 Hz ([ɬ] > [w]) oppure una F2 alta, fra i 1600-1800 Hz ([l] > [j]).

La lezione che si trae dall'insieme dei dati articolatori e percettivi riportati in questo lavoro è tuttavia che conviene indagare con attenzione l'esistenza di varianti fonetiche intermedie, valendosi in pieno anche del dettaglio fonetico più fine, prima di accantonare una spiegazione primariamente articolatoria del SC. Tale strategia infatti, consente di in luce gli elementi chiaramente a favore di una interpretazione articolatoria dei due processi di vocalizzazione di /l/ nel Romanzo.

> ...[t]hus, it has been hypothesized that the replacement of clear /l/ by [j] in the tautosyllabic clusters /pl/ and /kl/ has not taken place directly but through the intermediate realizations [pʎ] and [kʎ] which are still available in several Romance dialects, e.g., Ribagor.an Catalan [kʎaw] CLAVE "key", [ˈpʎɔwɾe] PLOVERE "to rain", Francoproven.al areas [ˈpʎema]PLUMA "feather", [kʎa] CLAVE (Repetti and Tuttle, 1987). In particular, the development /kl/> [kʎ]> [kj] involves two consecutive production mechanisms: blending between the tongue front gesture for the alveolar lateral and the dorsal gesture for the velar stop into the alveolopalatal lateral realization [ʎ], followed by delateralization of [ʎ] into [j] through central tongue contact loss. Regarding the change [ɬ] > [w], there is articulatory evidence indicating that dark /l/ vocalization has been induced by apical contact loss, which contributes to lower the F2 frequency for the lateral down to 600-800 Hz thus rendering the alveolar lateral spectrally similar to /w/ (Lin et al., 2014).

Sono compatibili con questa interpretazione i risultati di uno studio percettivo di Recasens e Espinosa (2010a), in cui numerosi soggetti catalani avevano, tra i compiti, quello di identificare come /l/ o come /w/ diversi esemplari di [ɬ] inseriti in stimoli di tipo [VɬCV], estratti da sequenze di parlato reale prodotte in catalano majorchino, una varietà romanza in cui la laterale alveolare presenta una realizzazione fortemente scura. I risultati del test di identificazione indicano che la vocalizzazione di /l/ può essere innescata sia da una perdita di contatto alveolare unita a una bassa F2 (equivalenza acustica), sia dalla sola mancanza di contatto alveolare (senza la presenza di una bassa F2), soprattutto laddove /l/ sia seguita da /b/ (57%) piuttosto che da /d, s, k/ (15%-35%).

> While these identification results do not seem to provide clear support for either of the two explanatory hypotheses, there are reasons to believe that the articulatory account has higher predictive power than the acoustico-perceptual one: alveolar contact loss appears to cause F2 to lower to a greater extent than other articulatory maneuvers such as tongue body retraction thus giving rise to a /w/-like spectral configuration more effectively and this frequency lowering effect is especially prone to occur before labial consonants which is where dark /l/ vocalization occurs most often.

Con originalità il saggio di Recasens contribuisce a sfatare un'altra assunzione tradizionale: quella che i SC debbano seguire un unico percorso evolutivo. Molti casi,

osserva lo studioso, possono invero essere spiegati in termini esclusivamente di rafforzamento articolatorio:

> [t]hus, articulatory enhancement may account for why word-initial /n, l/ or their geminate cognates /nn, ll/ may be replaced by the alveolopalatals /ʎ,ɲ/ (Argentinian Spanish [ˈɲu.o] for [ˈnu.o] NODU "knot", Catalan [ʎas] LAQUEU "bow", [aɲ] ANNU "year"); oppure di riduzione articolatoria: "articulatory reduction, on the other hand, can explain why /ɲ/ may yield [n] in syllable- or word-final position, as shown by the alternation between syllable-initial intervocalic [ɲ] and word-final [n] in the Spanish lexical pair desdeñar "treat with disdain" – desdén "disdain.

Tuttavia, sostiene Recasens richiamando il lavoro di Leigh Lisker sugli indici percettivi di sonorità (Lisker, 1986, ad es.), l'attenzione al dettaglio fonetico suggerisce che il SC possa procedere attraverso più di una trafila evolutiva, in funzione della prominenza relativa delle proprietà acustiche che il parlante-ascoltatore utilizza nella identificazione fonemica. Seguendo l'argomentazione di Recasens, riprendiamo il caso della identificazione di [ɫ] come /w/: mutamento che appare con la maggior evidenza laddove la laterale è preceduta da una vocale bassa, e che può seguire due percorsi:

> a) The misperception of dark /l/ as /w/ whenever the steady-state spectrum for the alveolar lateral exhibits a sufficiently low F2 frequency, i.e., /al/ > [aw]. This is by far the most preferred option in the world's languages (see section 2.2). (b) The integration of the VC transitions as /w/ (/al/ > [awɫ]) followed by [ɫ] elision whenever the apical closure is severely reduced or occurs after voicing offset utterance-finally ([awɫ] > [aw]). This phonetic development occurs in some Romance dialects which still keep the intermediate phonetic variant aul, as for [awɫt] ALTU "high" (Romansh) and *aurdeia* 'aldea' "village" (Northern Portuguese dialects).

I risultati qui riportati di un studio spettrografico e percettivo (Recasens e Espinosa, 2010b) mostrano che /l/ viene identificato come /w/ con maggiore probabilità quando F2 associata al periodo di stabilità di /l/ si abbassa da 1050 Hz a 650 Hz (5-45%), rispetto a quando la fine delle transizioni VC viene temporalmente anticipata del 15/35%. Inoltre quest'ultima caratteristica acustica contribuisce a un incremento nelle percentuali di identificazione di /l/ come /w/ solo quando la frequenza di F2 durante lo stadio-stabile della laterale alveolare è minima (650-750 Hz).

> These perception results are consistent with the outcome of the vocalization process being [w] rather than [wl] in the world's languages, and suggest that the vowel transitions may act as a secondary vocalization cue which is used by speakers to enhance the perceptual effectiveness of the steady-state F2 frequency whenever /l/ is strongly dark.

Per limitarsi al tema cruciale delle origini fonetiche, articolatorie o acustico-percettive del SC, il saggio di Recasens ci consegna uno scrigno di dati e di idee originali su cui riflettere. Ad esempio, che occorre far tesoro di varianti articolatorie intermedie, prima di abbacciare la tesi di un'origine esclusivamente acustico-percettiva del SC (*acoustic equivalence*). Che i SC possono avere più di una origine acustica e quindi

possono procedere attraverso molteplici percorsi evolutivi. Che gli indici acustici di un dato fonema possono svolgere un ruolo primario o secondario nell'attivare un SC, in funzione della loro salienza relativa, la quale può variare da dialetto a dialetto.

Marianne Pouplier, in *Between 'whims of fashion' and 'phonetic law': Performance constraints in speech production in the face of linguistic diversity* delinea con acuto spirito critico progressi e questioni a tuttoggi apertissime nella ricerca sulle forze biologiche che delimitano i "confini" della diversità linguistica. Di fatto, la domanda che la Pouplier solleva concerne, ancor più crucialmente, la questione se le restrizioni indotte dai processi di percezione e produzione possano non soltanto delimitare, ciò che oggi la teoria fonetica dà di fatto per acclarato, ma ancor più *plasmare* la forma sonora delle lingue, all'interno dei confini che circoscrivono Babele.

Il lavoro gravita intorno alla sillaba, rappresentata in numerosi modelli teorici biologicamente orientati come il dominio entro cui operano restrizioni motorie universali, e che diviene in questo saggio il banco di prova privilegiato su cui esaminare l'interazione fra aspetti linguo-specifici e universali della grammatica e del controllo motorio. Il fatto è che, come avverte Pouplier, nonostante molti indizi

> for the existence and the formative role of production constraints, languages have been known to differ systematically in their articulatory organization of the vowel cycle.

Con l'applicazione, fin dagli anni '80, della nuova prospettiva dei sistemi dinamici (*dynamical system approach)* alla produzione dei suoni del linguaggio, l'attenzione si è rivolta al ruolo della coordinazione di tipo *in-fase* (0°) o *anti-fase* (180°), considerata come modalità biologicamente preferita di coordinazione motoria in compiti non linguistici (Turvey, 1990). Nell'ambito della A.P.,

> ... [a] rather direct link has been drawn between the basic stability of in-phase coordination and linguistic preference, i.e. statistical asymmetries in cross-linguistic patterns. Typological CV preference, the order of acquisition of onsets before codas, onsets having a greater resistance to sound change compared to codas and moraic weightlessness of onsets have all been linked to the hallmark stability of in-phase coordination (Nam, 2007; Nam, Goldstein & Saltzman, 2009). In this model, onset clusters require more complex coordination patterns than singleton onsets: they are characterized by two competing phase relationships which prevent the multiple consonants from being produced on top of each other. All onset consonants are coupled in-phase to the vowel, but to ensure their recoverability, they are coupled anti-phase to each other. This underlying coupling topology gives rise to the so-called 'c-center effect' (Browman, Goldstein, 1995; 2000).

Il cuore del problema risiede nel fatto che

> the basic gestural syllable model is not able to fully capture onset-vowel timing; a number of exceptions have been reported which are not straightforwardly accommodated by the model.

Il lavoro prende le mosse da un precedente studio articolatorio (EMA) sul rumeno (Marin, 2013), da cui emergeva il comportamento "irregolare" rispetto all'ipotesi

del *c-center effect* di nessi in attacco sillabico quali /ps, ks-, kt-, kn-/ rispetto a nessi quali /sp-, sk-, sm-/. Nel primo caso non si assisteva a un aumento della sovrapposizione consonante-vocale in condizioni di tipo CCV vs CV (ad es. in *salt* vs. *psalm*) mentre nel secondo caso, in accordo col modello gestuale, la sovrapposizione della consonate alla vocale aumentava in contesto CCV rispetto a CV (/palə/vs. /spalə/).

Per spiegare tali "irregolarità, Pouplier avanza una proposta originale, che chiama in causa una re-interpretazione, nell'ambito del modello gestuale, della nozione di "complessità fonetico-fonologica" e l'ipotesi di un'interazione (linguo-specifica) fra forze *articolatorie* diverse.

> Complexity in the gestural model is defined as any deviation from the fundamental in-phase/anti-phase pattern and these deviations have often been motivated on perceptual grounds (Browman, Goldstein, 2000; Chitoran, Goldstein & Byrd, 2002).

La proposta avanzata da Pouplier è invece di natura articolatoria. Pouplier riporta i risultati di un suo recente studio articolatorio (EMA) sul polacco, una lingua che esibisce un numero inconsueto di nessi consonantici (Pastätter e Pouplier, 2015). In polacco, osserva la studiosa, in una serie di nessi C1C2 le consonanti possono comparire in entrambi gli ordini sia in attacco che in coda, consentendo in tal modo di studiare in modo sistematico gli effetti dell'ordine C1-C2 entro la medesima posizione sillabica.

Scopo dello studio è verificare l'ipotesi che l'incremento nella sovrapposizione tra consonante e vocale condizionato, secondo le predizioni del modello gestuale della sillaba, dalla crescente complessità dell'attacco sillabico, sia in relazione inversa alla resistenza coarticolatoria opposta dalla consonante adiacente alla vocale (C2). La predizione empirica è che una consonante fortemente resistente alla coarticolazione possa bloccare la sovrapposizione crescente tra vocale e consonante e che in tali condizioni non sia osservabile un effetto c-center. Dal momento che le sibilanti sono tra le consonanti più resistenti alla coarticolazione, in termini di posizione sia linguale che mandibolare (Recasens & Espinosa, 2009), questo studio si focalizza pertanto sui nessi SCV e CSV (dove S = /s, ʃ/) in posizione di attacco sillabico dove il polacco ammette una vasta gamma di nessi simmetrici SCV and CSV (/ks-, sk-; mʃ-, ʃm-; ps-, sp-; pʃ-, ʃp-/), tutti inclusi nello studio, insieme ai corrispondenti attacchi sillabici semplici (ad esempio, mʃalik - ʃalik; ʃmata - mata).

> Results confirmed the predicted significant main effect of sibilant position (vowel adjacent vs. vowel remote) on onset-vowel timing (Figure 2). For the CSV condition, there was no change in onset-vowel overlap as onset complexity increased from SV to CSV. For SCV conditions, on the other hand, such an increase in overlap was observed between CV => SCV, in accordance with the gestural model (Figure 2). That is, SCV clusters conformed to the preferred onset-vowel coordination pattern, while CSV clusters did not.

Si tratta, per Pouplier, di un argomento forte a favore dell'esistenza di "relazioni di compromesso" (*trade-off relations*) fra tipi diversi di restrizioni sulla *produzione* dei suoni linguistici: relazioni che possono indurre le deviazioni osservate dai pattern

articolatori previsti come preferiti, se si assume che i "compromessi" siano linguo-specifici e non universali.

> If coarticulation resistance interacts with syllable-level timing, this provides evidence that different gestural coordination patterns may be stabilized by languages on grounds of speech motor control, i.e. there may be trade-offs in speech motor control itself which lead to a contextual conditioned redefinition of preferred articulatory coordination patterns without necessarily implying a lesser stability thereof (Pouplier, 2012a). This in turn would invite us to take a broader view of what are deemed to be preferred articulatory patterns across languages.

Il volume prosegue con una ricca serie di interessanti contributi sulla variazione di tipo diacronico, fonetico, allofonico nei repertori linguistici, dall'italiano e suoi dialetti al giapponese.

Il primo articolo (Filipponio, Cazzorla) prende in esame il sistema vocalico del dialetto e dell'italiano regionale barese e, attraverso l'analisi acustica dei sistemi vocalici delle due varietà, getta luce su due distinti cambiamenti attualmente in atto. Il primo riguarda l'opposizione tra vocali medio-alte e medio-basse che, assente nel sistema dell'italiano regionale e già instabile nel dialetto barese, si è in quest'ultimo ormai completamente neutralizzata. Il secondo cambiamento rilevato dagli autori, di tipo squisitamente sociofonetico, riguarda l'abbassamento generalizzato delle vocali (osservato attraverso la variazione dei valori di F1) rilevato nelle nuove generazioni, che in tutti i giovani colpisce l'italiano regionale, e che specificamente le giovani donne estendono anche al dialetto.

Il contributo di Paggini e Calamai contribuisce allo studio della variazione diacronica nei dialetti con uno studio geolinguistico sui dialetti toscani. L'analisi si concentra sull'anafonesi, studiata utilizzando l'archivio sonoro della Carta dei Dialetti Italiani. Il quadro geolinguistico che ne emerge collide con le descrizioni tradizionali della distribuzione del fenomeno in area toscana: nel decennio coperto dai dati dell'archivio (1960-1970), la Toscana risulta infatti già quasi completamente anafonetica, con l'unica eccezione del territorio non urbano meridionale, limitato alle province di Siena e Arezzo.

Lo studio di Pinto si concentra invece sul giapponese, nel quale l'autrice osserva un probabile cambiamento in atto indotto, nel parlato, dalla elisione delle vocali alte, con il conseguente aumento della quantità e della complessità dei nessi consonantici secondari. Attraverso i risultati di uno studio di tipo percettivo, il lavoro conferma la capacità dei locutori di discriminare correttamente le parole che presentano elisione delle vocali alte.

Il lavoro di Nodari, che si colloca pienamente nell'ambito degli studi sulla variazione sociofonetica, tratta il ruolo che fattori interni (linguistici) ed esterni (sociolinguistici) giocano nell'aspirazione delle consonanti occlusive sorde in un gruppo di adolescenti della Calabria centrale. I risultati dell'analisi del Voice-Onset-Time (VOT) mostrano con chiarezza che la variazione nella sua durata è condizionata sia dal tipo di consonante e di contesto fonetico, sia dalle caratteristiche socio-culturali dei parlanti.

Anche il lavoro di Romito e colleghi si occupa di VOT, questa volta analizzando specificamente le occlusive del dialetto calabrese di San Giovanni in Fiore. I risultati inducono gli autori a riconsiderare le isoglosse tradizionali della Calabria e a proporre, la collocazione di San Giovanni in Fiore nell'area di Catanzaro e Reggio Calabria, per quanto concerne il tratto dell'aspirazione.

Il contributo di Bernardasci si concentra invece su un dialetto settentrionale, quello di Piandelagotti, un dialetto apenninico emiliano-occidentale di tipo conservativo a ridosso del fascio di isoglosse La Spezia-Rimini. Il lavoro affronta sperimentalmente il tema della quantità vocalica nel sistema tonico, e considera il rapporto di tipo compensativo che lega la durata della vocale tonica e quella della consonante seguente. I risultati delle analisi acustiche portano l'autrice ad attestare come la quantità vocalica sia ancora saldamente distintiva e pertanto a collocare tale varietà dialettale entro il gruppo dei dialetti Gallo-Italici.

Anche il saggio di Avesani, Galatà, Vayra, Best, Di Biase, Tordini e Tisato rientra nell'ambito della sociofonetica. L'argomento è la permanenza del dialetto materno in una popolazione di italo-australiani, emigrati in in Australia dal Veneto nel primo dopoguerra. In particolare ci si chiede quale sia l'attrito esercitato dall'inglese-L3 sul dialetto-L1 e in che misura la variazione indotta su L1 dal contatto linguistico con l'inglese sia riportabile a fatti sistemici (linguistici) e a fatti sociolinguistici. I risultati di uno studio acustico sul sistema delle fricative coronali venete indicano che dopo più di 50 anni di residenza nel paese straniero, l'attrito esercitato dall'inglese-L3 sul dialetto nativo risulta essere solo in uno stadio iniziale e limitato al solo sistema del veneto centrale, con esclusione di quello settentrionale.

Dell'allofonia di /R/, si occupa il contributo di Vietti, Spreafico e Galatà, che collocano questo studio nell'ambito del dibattito sull'interfaccia tra fonetica e fonologia e analizzano la variazione delle rotiche in tirolese La distribuzione di /R/ è affrontata raccogliendo nuovi dati di tipo articolatorio, mediante tecniche di analisi ecografica (ultrasuoni). I risultati indicano l'esistenza di almeno cinque varianti di rotiche uvulari, caratterizzate a livello fonologico da una distribuzione allofonica non pienamente regolare, e a livello fonetico da differenze sistematiche nella loro configurazione articolatoria (linguale).

La seconda parte del volume, "Acquisizione, destrutturazione e codifica neurale della struttura sonora del linguaggio", è dedicata all'acquisizione fonetico-fonologica di una lingua seconda e ai processi di destrutturazione sonora della lingua prima in varie patologie del linguaggio. A queste due sezioni seguono lavori sulla codifica neurale dei suoni del parlato e sul rapporto speciale che lega, entro il tipo di comunicazione multilineare e multimodale che caratterizza le lingue dei segni, la componente orale alla componente gestuale nella LIS, la Lingua dei Segni Italiana.

La prima sezione, "Acquisizione di L2", si apre con tre lavori sullo sviluppo della competenza fonetica e fonologica in L2. Due di questi sono dedicati, specularmente, all'acquisizione del sistema vocalico dell'italiano-L2 da parte di apprendenti cinesi (De Meo, Vitale, Xu), e alla percezione delle vocali del cinese-L2 da parte di appren-

denti italiani plurilingui (Yilun, Gili Fivela). Il terzo lavoro (Schmid, Wachter) è dedicato invece all'aquisizione del tratto di sonorità nelle ostruenti italiane da parte di parlanti zurighesi, la cui L1, un dialetto svizzero-tedesco, realizza l'opposizione di sonorità attraverso un contrasto di "forza" tra foni "forti" e "leni".

In tutti i lavori di questa sezione, i dati sperimentali oggetto dell'indagine acustica (De Meo et al.; Schmid, Wachter) o dell'indagine percettiva (Yilun, Gili Fivela) sono interpretati nei temini dei principali modelli acquisizionali della fonologia di L2, esplorando il ruolo di filtro operato dalla L1 nella capacità di produzione e percezione dei suoni di una L2; l'interferenza negativa o la facilitazione indotta dalla conoscenza di una L3; il ruolo giocato della marcatezza nel processo di apprendimento.

La seconda sezione, "Destrutturazione della forma sonora", più corposa, è dedicata al mutamento della struttura sonora della lingua nelle patologie del linguaggio.

Il lavoro di Prince affronta direttamente la questione della validità e attualità del postulato jakobsoniano, concernente il rapporto speculare fra acquisizione e perdita del linguaggio. Lo studio, che prende in considerazione 20 bambini e 20 pazienti afasici francesi, pone al centro dell'analisi fonologica la nozione di marcatezza, fondamentale nel lavoro di Jakobson. L'ipotesi è che la marcatezza svolga un ruolo fondamentale sia nell'acquisizione che nell'afasia, e che bambini e afasici rispettino una gerarchia di complessità, definita da proprietà della marcatezza. Sotto questa ipotesi i diversi processi fonologici sarebbero diretti a ridurre la complessità, adottando strutture meno marcate. Da questa prospettiva, la complessità deriva dall'interfaccia fra restrizioni relative al contenuto dei segmenti e rappresentazione sillabica. I risultati mostrano con chiarezza che, se molti dei dettagli del postulato di Jakobson necessitano di una revisione, i dati sull'afasia esaminati offrono un'immagine speculare dei fenomeni di acquisizione riportati dalla letteratura recente. Il quadro che ne emerge è avvincente e complesso insieme. I processi di cancellazione, epentesi, sostituzioni e riduzione operano *de facto* sia durante l'acquisizione, sia nei disordini fonologici. Tuttavia, mentre le cancellazioni risultano essere i processi di gran lunga più importanti rispetto ad epentesi e metatesi, sia nell'acquisizione che nell'afasia lo stadio delle cancellazioni sembra precedere quello delle sostituzioni nell'acquisizione laddove nell'afasico appare l'ordine contrario.

Anche il saggio di Bollani si ispira all'intuizione jakobsoniana, rivelando attraverso un'attenta disamina della letteratura, le significative analogie che emergono tra acquisizione del ritmo e problemi della sua percezione nei pazienti con disturbi del linguaggio.

I quattro saggi successivi si concentrano sul parlato patologico. Indubbia la loro grande rilevanza sul piano applicativo e non solo teorico, per le potenziali implicazioni sociali, sanitarie, economiche.

Le anomalie nella produzione dei suoni linguistici in pazienti adulti affetti da Parkinson sono studiate nel lavoro di Gili Fivela, Iraci, Grimaldi, Zmarich, mentre le anomalie di produzione in bambini affetti da balbuzie sono studiate in due lavori

collegati, quello di Lenoci, Pisciotta, Zmarich e quello di Zmarich, Quartarone, Bernardini, Bonichini, Colantonio, Natarelli, Stocco, Pisciotta. Il lavoro di Beltrami, Calzà, Gagliardi, Ghidoni, Marcello, Rossini Favretti, Tamburini si occupa del parlato di soggetti anziani affetti da fragilità cognitiva.

Obiettivo dello studio articolatorio che Gili Fivela e colleghi si propongono è di verificare se i gesti articolatori di soggetti affetti dal morbo di Pakinson mostrino alterazioni spaziali e/o temporali rispetto ai gesti di soggetti non patologici. Dai dati sperimentali (raccolti con EMA) emerge prima di tutto che i soggetti affetti da Parkinson producono le consonanti con gesti articolatori di ampiezza ridotta rispetto ai locutori normali; e in secondo luogo che anche la sonorità e il modo di articolazione risultano alterati, in grado maggiore in rapporto all'età del paziente.

L'insorgenza della balbuzie nel bambino, la sua manifestazione fonetica e l'individuazione di indici predittivi che ne permettano un trattamento precoce sono i temi dei due articoli di Lenoci et al. e Zmarich et al., disegnati nell'ambito di un ampio progetto di ricerca longitudinale volto ad individuare precocemente i bambini a rischio di balbuzie cronica. I due lavori trattano i profili evolutivi della coarticolazione intrasillabica, il numero di disfluenze da balbuzie e l'ampiezza del lessico, quali manifestazioni linguistiche che possono in prospettiva contribuire a distinguere quei bambini che in futuro cronicizzerano il disturbo da quelli che presenteranno una remissione spontanea. L'immediata valenza applicativa di questi studi risiede nella possibilità di identificare precocemente i casi su cui praticare un trattamento mirato, prevenendo con le maggiori possibilità di remissione l'insorgenza di balbuzie cronica

Altrettanto rilevante l'obiettivo del lavoro di Beltrami et al., che è quello di identificare in modo automatico i vari stadi e tipi di demenze a partire dall'analisi statistica della produzione linguistica spontanea dei pazienti. Scopo ultimo del progetto entro il quale questo contributo si colloca è lo sviluppo di tecniche facilmente somministrabili di raccolta e analisi di campioni di linguaggio spontaneo per lo *screening* sistematico della popolazione potenzialmente a rischio.

Le due sezioni finali di questa parte del volume, "Codifica neurale e Lingua dei segni e linguaggio musicale", presentano un lavoro di neurolinguistica sulla rappresentazione del sistema vocalico nella corteccia cerebrale (Manca, Russo, Grimaldi) e due lavori sulla doppia natura, multineare e multimodale, della lingua dei segni (Fontana, Roccaforte); Roccaforte). Chiude la sezione il lavoro di Bravi sul rapporto tra l'aspetto sonoro-musicale e la struttura metrica del verso nelle *cantadas campidanesas*, la poesia cantata d'improvvisazione della Sardegna meridionale.

I due lavori sulla Lingua Italiana dei Segni (LIS) affrontano il problema di definire natura e ruolo che le componenti orali hanno nelle lingue dei segni rispetto alla lingua parlata. Nella LIS così come in altre lingue segnate, è stata rilevata la massiccia presenza di componenti orali prodotte in concomitanza con le unità segniche. Tali componenti sono state definite come "gesti labiali" e "labializzazioni" sulla base della presenza/assenza di un nesso diretto con il parlato (i gesti labiali

sono forme di oralità prive di un rapporto con la lingua vocale; le labializzazioni sono frammenti di parole della lingua parlata articolate con o senza l'apporto della componente glottica).

Il saggio di Fontana e Roccaforte si pone lo scopo di analizzare queste componenti orali e di modificarne l'attuale classificazione. Considerando la totalità del segno come un complesso articolatorio che comprende sia il piano gestuale che quello non-manuale le autrici definiscono quali e quante siano le configurazioni articolatorie più frequenti nella componente orale, come si strutturino in relazione al flusso segnico e come si destrutturino rispetto al parlato.

Il lavoro di Roccaforte si concentra in modo specifico sulle labializzazioni e, con un esperimento pilota su un gruppo di informanti sordi segnanti, studia in quale misura la presenza della componente labiale faciliti la riconoscibilità del segno, al fine di dimostrare se sia possibile attribuire alle labializzazioni una funzione di "disambiguazione".

Nell'articolo di Manca, Russo e Grimaldi, la elettoencefalografia (EEG) viene impiegata per esplorare la mappatura delle vocali nella corteccia uditiva. Il lavoro presenta i risultati di un esperimento neurolinguistico nel quale sono state rilevate le risposte dei segnali evento-correlati (ERP) alle vocali italiane [a] e [i]. Le proprietà della componente N1 degli ERP sono state analizzate in risposta a vocali prodotte ad alta voce, articolate senza emettere suono e solamente pensate. I risultati confermano che l'attività corticale riflette le diverse proprietà acustiche spettrali delle vocali (F1 e F2) e supporta la tonotopia quale princiale principio organizzatore primario della corteccia uditiva.

La terza parte del volume, "Trattamento automatico delle lingue naturali", "inverte" il punto di osservazione sulla lingua, e dall'analisi della variazione, del mutamento e della destrutturazione in un contesto naturale si concentra invece sulla "generazione" della lingua parlata ottenuta mediante sintesi automatica da testo, sul suo riconoscimento automatico e sui sistemi di interfaccia uomo-macchina che trovano applicazione in vari settori della nostra società, *in primis*, quali ausilii comunicativi per i portatori di handicap.

Apre questa sezione un contributo ad invito, *Expressivity in TTS from Semantics and Pragmatics,* di Rodolfo Delmonte. In questo saggio viene presentato un sistema completo di sintesi della voce da testo scritto (TTS) basato su una complessa rappresentazione linguistica del testo che ne garantisce una lettura espressiva e naturale. La novità e direi l'unicità di questo tra i sistemi TTS esistenti è che la ricchezza delle rappresentazione sintattica semantica e prosodica sulla quale è basato garantisce una lettura espressiva, fortemente aderente alla resa naturale di un locutore umano, sia di testi in prosa che di testi in poesia. Il lavoro, oltre a rappresentare lo stato dell'arte in questo campo, costituisce la summa dell'esperienza pluridecennale di Delmonte nella creazione di sistemi di sintesi da testo, ed esprime la grande competenza sintattica, semantica e prosodica dell'Autore.

In armonia con l'originale ispirazione jakobsoniana del convegno, Delmonte mette in evidenza la centralità della funzione espressiva del linguaggio. Egli osserva che, in generale, salvo rarissime eccezioni, la voce sintetica dei sistemi TTS è monotona e priva di un'espressività vera, in grado di rendere compiutamente gli effetti di senso del testo generato. Riprendendo Huckvale (2002), Delmonte ci ricorda che:

> the lack of expression or interest or emotion in the speech of current systems is due to the fact that the systems don't actually understand what they are saying, see no purpose to the communication, nor actually have any desires of their own. It makes no sense to add "expressiveness" or "emotion" to voice output from a mindless system (*ibi*, p.1262).

Il problema diviene allora per lo studioso quello di riformulare lo scopo di un sistema di sintesi da testo, spostando il centro dell'attenzione dal *replicare* come gli esseri umani parlano al cercare di *comprendere* come lo fanno.

Il valore del sistema creato da Delmonte risiede proprio nell'aver compiuto questo passo, interfacciando sapientemente le proprietà sintattiche, semantiche e discorsive del testo con una rappresentazione prosodica fine, la cui implementazione permette una resa estremamente naturale perfino dei testi poetici. La prosodia gioca un ruolo centrale in questo prototipo, che si avvicina ad un "ideale" di TTS nel quale il *phrasing* è coerente con la semantica del testo; l'intonazione varia in funzione delle proprietà strutturali della frase nel discorso e con le intenzioni del parlante; le relazioni di prominenza sono assegnate in base alla struttura informativa del discorso e allo status informativo degli elementi discorsivi; l'espressività veicola variazioni di attitudine e modalità derivate da un'analisi soggettiva e affettiva profonda.

Il lavoro di Delmonte ci schiude il complesso intreccio di conoscenze sintattiche semantiche pragmatiche e prosodiche che governano il sistema di sintesi espressiva proposto. Di particolare rilevanza è il fatto che il tipo di rappresentazione semantica fornita dal sistema non si limiti a identificare automaticamente nel testo la struttura predicato-argomento, il dominio della negazione, le quantificazioni e a risolvere le anafore, ma che la rappresentazione semantica venga estesa a includere il livello del discorso attraverso un'analisi di tipo proposizionale. Delmonte, inoltre propone che:

> [i]t also paves the way for a deep sentiment or affective analysis of every utterance, which alone can take into account the various contributions that may come from syntactic structures like NPs and APs where affectively marked words may be contained. Their contribution needs to be computed in a strictly compositional manner with respect to the meaning associated to the main verb, where negation may be lexically expressed or simply lexically incorporated in the verb meaning itself.

È proprio grazie a rappresentazioni linguistiche di tale ampiezza, fornite direttamente dal sistema e interfacciate con un'appropriata rappresentazione prosodica, che il prototipo di TTS proposto può restituire in modo naturale la maggior parte dei movimenti intonativi, inclusi quelli derivanti dagli spostamenti di *focus*.

Molti sono i contributi empirici che offre il saggio di Delmonte; tra questi, uno riguarda la corrispondenza tra i correlati acustici che il modulo prosodico del TTS assegna al testo in base alla rappresentazione linguistica (F0, intensità, pause, velocità di eloquio) e gli elementi fonologici specificati dal sistema di analisi e trascrizione intonativa noto come *Tone and Break Indeces* (ToBI). Delmonte rileva la necessità di arricchire l'inventario tonale di ToBI con nuovi elementi fonologici, e illustra come implementarlo in un TTS fornendo la corrispondenza tra i parametri acustici e la loro rappresentazione simbolica in ToBI.

Così come il sistema precedente può costituire un ausilio importante per i non vedenti, il lavoro di Vietti, Anselmi e Spreafico presenta un modello di interfaccia silente per la lingua italiana che permette la comunicazione orale, anche se limitata, a coloro che hanno problemi nel produrre un segnale acustico udibile. In particolare, il sistema interpreta le immagini ultrasoniche dei movimenti della lingua di un soggetto e le converte in rappresentazioni che possono essere poi usate come input di un sistema di sintesi automatica da testo.

I due articoli di Cosi e collaboratori illustrano un sistema di riconoscimento automatico della voce basato sull'uso di reti neurali che è stato sperimentato sul parlato di bambini italiani, e il relativo corpus di parlato infantile servito per addestrare il sistema. Scopo dei lavori è studiare l'interazione bambino-robot con il fine ultimo di implementare il sistema in un robot-bambino da usare come ausilio negli ospedali pediatrici.

Il lavoro di Todisco, Poroli e Falcone affronta un problema collegato ai precedenti e centrale per il successo dei cosiddetti "assistenti virtuali": la tipizzazione del dialogo nei sistemi intelligenti che uniscono il riconoscimento automatico della voce, l'elaborazione del testo riconosciuto e la generazione automatica delle risposte. Il contributo di Paoloni e Todisco si centra invece sull'identificatore del parlatore, un tema fondante per le applicazioni delle *speech sciences* in ambito giudiziario. Infine, Origlia, Cangemi e Cutugno si muovono nel campo della fonologia computazionale e presentano WORDEN, un "sistema esperto" che assiste i linguisti nel trovare stimoli adeguati da usare negli esperimenti di produzione e percezione.

Desideriamo congedarci da questo volume con le parole di Roman Jakobson, che nel saggio del 1941, esplicitamente dichiarava:

> (...) La linguistica ha così dimostrato che le lingue del mondo nel loro sviluppo possono essere in relazione con certi mutamenti del linguaggio infantile, e la ricerca ha d'altro lato stabilito che quasi tutte le mutilazioni del linguaggio ordinario osservate nei bambini trovano stretti paralleli nei mutamenti fonetici delle diverse lingue del mondo. (...) Ma oltre a tutti questi possibili e occasionali punti di contatto, ci sono delle necessarie e costanti concordanze fra il linguaggio infantile e lo sviluppo storico delle lingue del mondo, ed è appunto questo problema fondamentale che fino ad oggi è stato scarsamente sollevato (...) (p. 17).

> (...) Le stesse considerazioni valgono, *mutatis mutandis*, per i disturbi del linguaggio nell'afasia (...)" (p. 30).
>
> (Entrambe le citazioni sono tratte dall'edizione Einaudi del 1971: Jakobson, R. (1971). Linguaggio infantile, afasia e leggi fonetiche generali. In Jakobson, R., *Il farsi e il disfarsi del linguaggio. Linguaggio infantile e afasia*. Torino: Einaudi.)

I curatori, infine, ringraziano sentitamente gli altri membri del comitato scientifico del Convegno AISV 2015, per la loro generosa disponibilità e per il prezioso contributo critico che ha arricchito questo volume:

Federico Albano Leoni (Università La Sapienza); Valentina Bambini (IUSS Istituto Universitario di Studi Superiori, Pavia) Carla Bazzanella (Università di Torino); Pier Marco Bertinetto (Scuola Normale Superiore di Pisa); Catherine Best (University of Western Sydney); Giuliano Bocci (Université de Gèneve); Luciana Brandi (Università di Firenze); Maria Grazia Busà (Università di Padova); Silvia Calamai (Università di Siena, sede di Arezzo); Laura Calzà (Università di Bologna); Chiara Celata (Scuola Normale Superiore di Pisa); Anna Chilosi (IRCCS Fondazione Stella Maris, Pisa); Piero Cosi (ISTC-CNR, Padova); Lidia Costamagna (Università per Stranieri di Perugia); Franco Cutugno (Università di Napoli "Federico II"); Amedeo De Dominicis (Università La Tuscia); Anna De Meo (Università di Napoli L'Orientale); Rodolfo Delmonte (Università di Venezia); Mariapaola D'Imperio (Université de Provence); Vincenzo Galatà (Libera Università di Bolzano e ISTC-CNR); Barbara Gili Fivela (Università di Lecce); Mirko Grimaldi (Università di Lecce); Michele Loporcaro (Universität Zürich); Marinella Majorano (Università di Verona); Giovanna Marotta (Università di Pisa); Pietro Maturi (Università di Napoli); Alberto Mioni (Università di Padova); Antonio Origlia (Università di Napoli "Federico II"); Andrea Paoloni (Fondazione Ugo Bordoni, Roma); Massimo Pettorino (Università di Napoli L'Orientale); Marianne Pouplier (Ludwig Maximilians Universität); Daniel Recasens (Universitat Autònoma de Barcelona); Antonio Romano (Università di Torino); Luciano Romito (Università della Calabria); Pier Luigi Salza (Socio onorario AISV - ex dirigente Loquendo); Alessandra Sansavini (Università di Bologna); Leonardo Savoia (Università di Firenze); Renata Savy (Università di Salerno); Carlo Schirru (Università di Sassari); Stefan Schmidt (Universität Zürich); Carlo Semenza (Università di Padova); Patrizia Sorianello (Università di Bari); Rosanna Sornicola (Università di Napoli); Lorenzo Spreafico (Libera Università di Bolzano); Fabio Tesser (ISTC-CNR, Padova); Alessandro Vietti (Libera Università di Bolzano); Claudio Zmarich (ISTC-CNR).

Bologna, Padova
Mario Vayra e Cinzia Avesani

PARTE I

VARIAZIONE FONETICA, RESTRIZIONI SULLA VARIAZIONE E MUTAMENTO

DOI: 10.17469/O2101AISV000002

DANIEL RECASENS

Phonetic factors contributing to the inception and evolution of sound change

This paper uses experimental evidence for showing that, depending on the case, sound changes may be triggered primarily by either articulatory variation (as for changes occurring through segmental weakening or strengthening) or by acoustic equivalence (as for vowel nasalization or for segmental substitutions involving syllable-final stops of different places of articulation). It also argues for a multiple evolutionary pathway in the case of specific sound changes such as dark /l/ elision or the palatalization of Latin /kt, ks/ in Romance.The role of word prominence in vowel assimilations and dissimilations and how sound changes originate in the individual are also looked into.

In recent times more attention has been paid to how sound changes spread through the lexicon and the linguistic community (Labov, 1994; Phillips, 2006; Wang, 1969) than to the phonetic factors which contribute to the origin of sound change. Fortunately the situation is improving thanks to the pioneering work of Ohala (see section 1.2) and an increasing number of books and articles on this research topic (Solé, Recasens, 2012; Yu, 2013). This scenario is in clear contrast with the one existing during the last decades of the XIX century and the first decades of the XX century, as revealed by outstanding publications on the articulatory causes of sound change which appeared at that time (Grammont, 1933; Millardet, 1910; Rousselot, 1897-1901). One reason for the present lack of interest in the phonetic causes of sound change is to be sought in the special emphasis that structuralist and generative linguists have put into the study of synchronic phonology much to the exclusion of diachronic aspects and phonetic variation. Another reason may lie in the difficulty involved in identifying those articulatory and acoustic characteristics which render the replacement of a linguistic sound by another feasible and in determining the degree to which speakers become more or less sensitive to spectral changes which may occur during the phonetic realization of phonemes. Keeping these aspects in mind, the present study deals with the sound change inception problem by looking into the following issues: whether sound changes have an articulatory and/or an acoustic origin (section 1) and their implementation follows a single or a multiple pathway (section 2); the contribution of articulatory/acoustic prominence to sound change (section 3); how sound change originates in the individual (section 4).

1. *Articulation- vs acoustic-based sound changes*

Linguists and philologists working on sound change during the first half of the XX century thought that sound shifts were essentially articulation-based. The general principle underlying this belief was that listeners may fail to identify a given phoneme intended by the speaker whenever subject to prominent articulatory and spectral changes in specific contextual, positional and/or prosodic conditions. According to an alternative acoustico-perceptual based approach, on the other hand, a number of sound changes are triggered by acoustic equivalence rather by articulatory variation (Ohala, 1981). The rationale behind this assumption is that two phonemic units are prone to be confused perceptually whenever exhibiting similar spectral cues in favorable segmental contexts and positions within the word and the utterance. A critical evaluation of these two explanatory hypotheses of sound change is presented next.

1.1 Articulatory origin

Most sound changes originate whenever listeners fail to compensate for coarticulation. Context-dependent coarticulatory effects may give rise to articulatory changes in the phonetic implementation of a given phoneme which may lead listeners to believe that a new phoneme has been produced by the speaker. An illustrative example is the raising of stressed /a/ to a mid front vowel before an (alveolo)palatal consonant in Early Romance, as in the developments Latin [ˈakse] AXE "axis" > *[ˈajse] > Catalan [e(j)ʃ] and Latin [ˈfakto] FACTU "done" > *[ˈfajto] > Catalan [fet], where [j] in the intermediate phonetic variants *[ˈajse] and *[ˈfajto] corresponds to a lenited realization of the original syllable-final velar stop. In this particular case productions of /a/ involving some tongue body fronting and raising, and consequently an increase in F2 frequency, induced by following [j] must have been heard as realizations of the mid front vowel phoneme by the listeners. Another clear example of an articulation-based sound change is the palatalization of a dentoalveolar consonant by an adjacent front vowel or glide, as in the above development Latin [ˈakse] AXE "axis" > *[ˈajse] > Catalan [e(j)ʃ] where the glide has caused following /s/ to turn to a palatoalveolar fricative after which it has stayed or dropped (both variants [ejʃ] and [eʃ] occur in Catalan depending on dialect).

In partial disagreement with linguists and philologists from the XIX and XX centuries, data reported in sections 1.1.1 through 1.1.3 show that there are sound changes which appear to be acoustically driven and thus are not triggered by articulatory variation.

1.1.1 Syllable-final stops

The exchange between oral stops of different places of articulation in syllable-final position whether word-internally or word-finally is prone to be caused by acoustic equivalence (see also Gauchat, 1925). Thus, for example, the spectral similarity between the stop bursts and the vowel formant transitions for /k/ and /t/ next to a front vowel may account for the changes /t/ > [k] and /k/ > [t] in this vowel con-

text whenever the originary stop occurs at the end of the syllable and is thus not too perceptually salient. The former change has taken place in Standard Catalan *ànec* ANATE "duck" and *préstec* PRAESTITU "loan" where /t/ in the Latin forms was replaced by [k], and the latter in the dialectal Catalan variants *coït* CULICINU "mosquito" and *présset* PERSICU "peach" where /k/ in the Latin lexical items shifted to [t] (the lexical cognates for *coït* and *présset* in Standard Catalan are *coïc* and *préssec* and therefore have a velar stop, as in Latin). The fact that /t/ and /k/ may be exchanged in the two directions, i.e., from dental to velar and from velar to dental, suggests that the replacements in question are triggered by spectral equivalence rather than by variations in closure fronting. Indeed, an articulatory account of this reversable change is feasible for /k/ > [t] (the velar may be confused with the dental when fronted to a [c]-like articulation, e.g., before a front vocalic segment) but not for /t/ > [k] (no tongue closure retraction is expected to occur in this case, e.g., before a back rounded vocalic segment).

1.1.2 Vowel nasalization

Changes in vowel quality induced by nasalization such as Old French /ỹ/ > [œ̃] (*un* UNU "one") and /ĩ/ > [ɛ̃] (*vin* VINU "wine") have also been triggered by spectral modifications. While Grammont (1933) suggested that those changes were articulation-based, Beddor (1983) has argued convincingly that they arise when nasal formants are added to the first vowel formant region thus causing higher oral vowels to be heard as if they were lower (and low vowels to be perceived as if they were higher) when nasalized. Recent evidence indicates that the spectral changes associated with nasalization may be enhanced or attenuated by changes in articulatory configuration which may occur during the production of nasalized vowels (Carignan, Shosted, Shih & Rong, 2013).

1.1.3 Sonorants

A third example of an acoustically driven sound change is the shift from clear /l/ to an alveolar tap ([ɾ]), as in the case of dialectal Spanish *alto* becoming *arto* ALTU "high". It may be argued that this sound change is conditioned by segmental shortening (the alveolar tap is shorter than the clear alveolar lateral) and gestural reduction (whenever too short, the tongue front may fail to make a complete closure during the production of clear /l/). Whether these articulatory maneuvers apply or not, the replacement of /l/ by an alveolar tap appears to be finally determined by the fact that the two consonants share a relatively high F2 about 1600 Hz or higher and are thus spectrally similar.

1.2 Acoustic origin

While, as shown in section 1.1, several sound changes which were traditionally thought to be articulation-based appear to have an acoustic origin, there also are sound changes which have been attributed to acoustic equivalence and are more prone to have an articulatory origin instead. As exemplified next, evidence for the articulatory origin of specific sound changes is to be sought in the existence of inter-

mediate variants between the input and output forms which may have been available at different historical periods.

1.2.1 Syllable-initial stops

Segmental substitutions involving syllable-initial stops of different places of articulation may occur in very specific vocalic contexts, as for example /k/ > [p] before a back rounded glide (Indo-European **ekwos* > Classical Greek *hippos* "horse"; Latin [ˈkwattuor] QUATTUOR > Romanian *patru* "four", Sardinian *báttoro*) and palatalized /p/ > [(t)ʃ] (Latin *sap*[j]*a* SAPIAM > Fr. *sache* "know, 1st person, subjunctive tense"; Latin *p*[j]*eno* PLENU > Portuguese *cheo* "full"). These sound changes have been attributed to the existence of similar burst spectra and formant transitions for the originary and outcoming consonants (Ohala, 1978; 1989). Indeed, both the burst and the F2 vowel transition locus are low frequency for labial and back velar stop consonants, and high frequency for the palatalized labial stop and the palatoalveolar affricate (also for /ʧ/).

A closer analysis of these phonetic developments reveals, however, that, at least in a subset of languages (Hickley, 1984), they may have involved an increase in glide constriction narrowing and thus an articulatory mechanism. Under this assumption the phonetic derivation for palatalized /p/ would then be /pj/ > [pç] > [pc] > [pʧ] > [(t)ʃ], which is consistent with lexical variants exhibiting the intermediate stages [pç], [pc] and [pʧ] (Engadin Rhaetoromance [ˈsapca] SAPIAM, Occitan [ˈsapʧa] SAPIAM, Romanian [ˈpçatɾə], [ˈ(p)catɾə] PETRA "stone"; Lausberg, 1970: 398; Rankin, 1974). Given this scenario, it may be suggested that the replacement of /kw/ by [p] has been triggered by an increase in labiovelar constriction degree (/w/ > [ɸ]) and therefore has proceeded through the pathway /kw/ > [kɸ] > [kp] > [p].

1.2.2 /l/ vocalization

The vocalization of /l/ may yield the outcomes [w] or [j] depending on whether the lateral is dark or clear, respectively. Thus, the replacement of dark /l/ by [w] accounts for the dialectal Catalan variants *a*[w]*ba* ALBA "dawn" and *esca*[w]*far* EXCALEFACERE "to heat", and the substitution of /l/ by [j] for Italian *amp*[j]*o* derived from Latin AMPLU. According to the acoustic equivalence hypothesis (Ohala, 1974), these changes are associated with whether the /l/ steady-state period has a low F2 frequency about 800-1000 Hz ([ɬ] > [w]) or else a high F2 about 1600-1800 Hz ([l] > [j]).

There is however evidence for an articulation-based interpretation of these two consonant vocalization processes in Romance. Thus, it has been hypothesized that the replacement of clear /l/ by [j] in the tautosyllabic clusters /pl/ and /kl/ has not taken place directly but through the intermediate realizations [pʎ] and [kʎ] which are still available in several Romance dialects, e.g., Ribagorçan Catalan [kʎaw] CLAVE "key", [ˈpʎɔwɾe] PLOVERE "to rain", Francoprovençal areas [ˈpʎema] PLUMA "feather", [kʎa] CLAVE (Repetti, Tuttle, 1987). In particular, the development /kl/> [kʎ]> [kj] involves two consecutive production mechanisms: blending between the tongue front gesture for the alveolar lateral and the dorsal gesture

for the velar stop into the alveolopalatal lateral realization [ʎ], followed by delateralization of [ʎ] into [j] through central tongue contact loss. Regarding the change [ɫ] > [w], there is articulatory evidence indicating that dark /l/ vocalization has been induced by apical contact loss, which contributes to lower the F2 frequency down to 600-800 Hz thus rendering the alveolar lateral spectrally similar to /w/ (Lin, Beddor & Coetzee, 2014). This articulatory account is also consistent with the fact that [ɫ] vocalization is prone to occur before labial and velar consonants (Gevaudanès Occitan ['awbo] ALBA, [faw'ku] FALCONE), where the lateral may exhibit less apicoalveolar contact and a lower F2 than before consonants of other places of articulation (Hardcastle, Barry, 1985; Scobbie, Wrench, 2003).

In order to seek for a more thorough understanding about whether [ɫ] vocalization is acoustic- or articulation-based we conducted a perception experiment where several Catalan subjects were asked to identify as /l/ or as /w/ several tokens of [ɫ] embedded in [VɫCV] stimuli excised from real speech sequences produced by several speakers of Majorcan Catalan where the alveolar lateral has a strongly dark realization (Recasens, Espinosa, 2010a). A relevant aspect about the perception stimuli was that, since they had been produced with an artificial palate in place, we could test the effect of changes in F2 frequency and in alveolar contact size in the phonemic categorization of [ɫ]. As exemplified in Table 1, the perception stimuli subject to identification had the following spectral and articulatory properties: their F2 could range between 650-750 Hz (more /w/-like) and 800-950 Hz (less /w/-like); they could be articulated with minimal or maximal apicoalveolar contact if showing, respectively, 0 to 3 or 6 to 8 'on' electrodes at any of the alveolar rows of electrodes on the artificial palate (the maximal number of electrodes on a given alveolar row is 8). Given these F2 frequency and alveolar contact ranges, the [VɫCV] stimuli put to test were characterized by either a low or a high F2 and minimal or maximal alveolar contact. The consonant following [ɫ] could be the stops /b, d, k/ and the alveolar fricative /s/ (only /ld/ stimuli specified for the maximal alveolar contact condition were included since /l/ is always articulated with a full closure before /d/ in Catalan). The initial testing hypothesis was that, if spectrally dependent, [ɫ] would be heard as /w/ not only in the case of stimuli showing minimal alveolar contact and a low F2 but also a low F2 and maximal alveolar contact; on the other hand, if vocalization was mainly dependent on alveolar contact loss and thus articulation based, [ɫ] would be heard as /w/ when produced with minimal alveolar contact independently of whether F2 was low or high.

The identification test yielded highest /w/ identification percentages for the low F2 – minimal alveolar contact stimuli, more so when /l/ was followed by /b/ (57%) than by /d, s, k/ (15%-35%). The /w/ identification percentages for all clusters in the three other acoustico-articulatory conditions were generally below 20%. While these identification results do not seem to provide clear support for either of the two explanatory hypotheses, there are reasons to believe that the articulatory account has higher predictive power than the acoustico-perceptual one: alveolar contact loss appears to cause F2 to lower to a greater extent than other articulato-

ry maneuvers such as tongue body retraction thus giving rise to a /w/-like spectral configuration more effectively and this frequency lowering effect is especially prone to occur before labial consonants which is where dark /l/ vocalization occurs most often.

Table 1 - *Articulatory and spectral conditions for the identification test of dark /l/ as /w/ in the cluster/lb/*

speaker	linguoalveolar contact		F2 frequency	
1	low	0 'on' electrodes	low	657 Hz
2	low	0 'on' electrodes	low	666 Hz
3	low	1 'on' electrodes	high	882 Hz
1	low	0 'on' electrodes	high	824 Hz
4	high	8 'on' electrodes	low	755 Hz
4	high	8 'on' electrodes	low	762 Hz
3	high	8 'on' electrodes	high	898 Hz
4	high	7 'on' electrodes	high	827 Hz

1.2.3 Velar softening

Another sound change process which may be accounted for on an articulatory basis despite having been attributed to acoustic equivalence is velar softening, i.e., the shift from a front allophone of /k/ before a front vowel or glide to the palatoalveolar affricate [tʃ]. By virtue of this sound change Old English [kin] became [tʃɪn] in Modern English and Latin ['kɛnto] CENTU "one hundred" turned to ['tʃɛnto] in Italian. According to the acoustico-perceptual hypothesis, this change is rendered possible due to the similarity between the 2500-3500 Hz spectral peak for the stop burst of front /k/, which is typically articulated at the palatovelar zone, and the frication noise of the palatoalveolar affricate /tʃ/. Spectral similarity ensures that front /k/ will be heard as /tʃ/ whenever the stop burst is sufficiently long and/or intense as the central lingual constriction at stop release undergoes considerable narrowing in anticipation of the front vowel or glide (Guion, 1998). According to the articulation-based account, on the other hand, velar softening may take place if closure location for the front velar stop shifts from the postpalato-velar zone to the palatal and even alveolopalatal zone thus giving rise to [c] (Rousselot, 1897-1901). This (alveolo)palatal stop realization, which may be articulated at a very close location to /tʃ/, may then be identified as the palatoalveolar affricate whenever its burst is sufficiently prominent. There are several events in support of an articulation-based account of velar softening through the intermediate realization [c] in Romance: a considerable number of Romance languages (Romanian and Majorcan Catalan, as well as Francoprovençal, Northwestern French, Northern Italian and Rhaetoromance dialects) have the allophone [c] of /k/ (also the allophone [ɟ] of /g/) in their consonant system; velar softening may occur not only before a front

vocalic segment but also before /a/ and word finally where it can only be possibly triggered by an (alveolo)palatal realization of the velar stop which may indeed be found in those two vowel and word conditions in some of the Romance dialects listed above (Recasens, 2011).

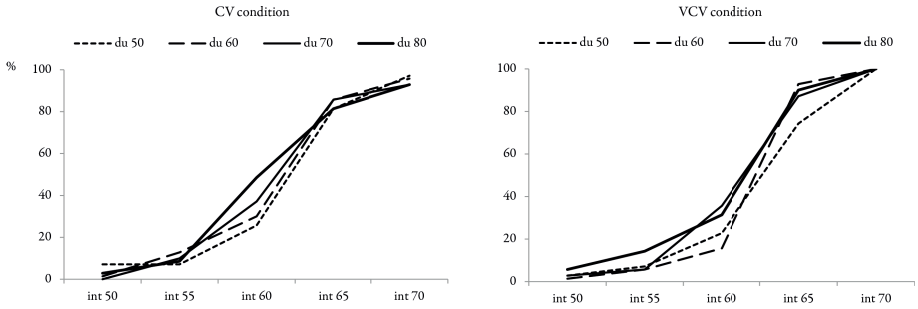
Two perception experiments (see (a) and (b) below) were carried out in order to investigate whether [c] may be heard as /tʃ/ and, if so, which acoustic cues are primarily responsible for velar softening.

(a) Real speech [cV] excerpts excised from a subset of [V#cV], [Vc#V], [VcV] and [#cV] sequences with the vowels /i/ and /a/ produced by several Majorcan Catalan speakers were presented for identification as /k/ or as /tʃ/ (in Majorcan Catalan, [c] is an allophone of /k/ occurring before /i, e, ɛ, a, j/ and word finally). According to the perceptual identification results, [c] is prone to be heard as /tʃ/ before /i/ rather than before /a/ (affricate responses amounted to 40-70% in the former vowel condition and to less than 40% in the latter), which is in line with velar softening occurring mostly before front vocalic segments in the world's languages (Recasens, 2014).

(b) Since the frequency characteristics of the stop burst and the vowel formant transitions for [c] approach those for /tʃ/ independently of segmental context and word position, it was hypothesized that the identification of [c] as /tʃ/ before front vowels and [j] ought to depend on an increase in burst duration and/or intensity associated with the articulatory and aerodynamic factors referred to above. In order to elicit which one of these two acoustic characteristics would yield higher affricate identification percentages, Catalan listeners were asked to identify as /k/ or /tʃ/ a subset of the [#ci] and [i#ci] stimuli used in the previous perception test whose burst duration and intensity values were more appropriate for the affricate and thus ranged between 50 ms and 80 ms and 50 dB and 70 dB, respectively. The identification results show that the intensity level is the most prominent acoustic cue for the identification of [c] as /tʃ/: the affricate is heard about 80%-100% of the time when burst intensity is about 65-70 dB, which is comparable to the frication noise intensity for the affricate. On the other hand, burst duration may also play a role though only when the intensity level is about 60 dB and therefore lies somewhere between the intensity levels for unaspirated [k] and for [tʃ].

The replacement of [c] by /tʃ/ before /a/ and word finally may be achieved through an increase in linguopalatal contact towards the alveolar zone and thus articulatory enhancement (see also Garrett, Johnson, 2010). This tongue contact increase before /a/ (and thus the chances that velar softening applies) is expected to take place mostly when the stop occurs in the phonetically prominent word initial and stressed syllable conditions.

Figure 1 - *Cross-subject affricate identification percentages for [c] as /tʃ/ in the [#ci] and [i#ci] conditions. Percentages are given for five [c] burst intensity levels (int = 50, 55, 60, 65 and 70 dB) and four [c] burst duration values (du = 50, 60, 70 and 80 ms)*



2. *Single vs multiple evolutionary pathways*

It has been traditionally assumed that sound changes proceed along a single pathway. Several changes depend indeed on a single articulatory mechanism which can be identified relatively easily. Thus, articulatory enhancement may account for why word-initial /n, l/ or their geminate cognates /nn, ll/ may be replaced by the alveolo-palatals /ʎ, ɲ/ (Argentinian Spanish [ˈɲuðo] for [ˈnuðo] NODU "knot", Catalan [ʎas] LAQUEU "bow", [aɲ] ANNU "year"). Articulatory reduction, on the other hand, can explain why /ɲ/ may yield [n] in syllable- or word-final position, as shown by the alternation between syllable-initial intervocalic [ɲ] and word-final [n] in the Spanish lexical pair *desdeñar* "treat with disdain" - *desdén* "disdain". A rise in intraoral pressure level conditioned to a greater or lesser extent by articulatory factors explains why oral stop epenthesis may take place in nasal+ fricative clusters (e.g., /ms/ > [mps]) and why syllable-final obstruents are prone to devoice in heterosyllabic consonant clusters (e.g., /zt/ > [z̥t]).

Attention to phonetic detail suggests that several sound changes may proceed through more than one evolutionary pathway depending on the relative prominence of the acoustic characteristics that speaker-listeners use for phonemic identification. This possibility is in line with vowels and consonants being often cued by several cooccurring acoustic characteristics (see Lisker, 1986 for the acoustic properties that cue the stop voicing distinction in the world's languages). This section deals with several phonemic replacements which may be achieved through more than one evolutionary pathway and seeks to determine which acoustic cues are used by listeners for implementing each sound change.

2.1 Dark /l/ vocalization

Languages and dialects may choose between two different pathways for the identification of [ɬ] as /w/, most clearly so when the lateral is preceded by a low vowel:

(a) The misperception of dark /l/ as /w/ whenever the steady-state spectrum for the alveolar lateral exhibits a sufficiently low F2 frequency, i.e., /al/ > [aw]. This is by far the most preferred option in the world's languages (see section 1.2.2).

(b) The integration of the VC transitions as /w/ (/al/ > [awɬ]) followed by [ɬ] elision whenever the apical closure is severely reduced or occurs after voicing offset utterance-finally ([awɬ] > [aw]). This phonetic development occurs in some Romance dialects which still keep the intermediate phonetic variant *aul*, as for [awɬt] ALTU "high" (Romansh) and *aurdeia* 'aldea' "village" (Northern Portuguese dialects).

Spectrographic and perception data reported elsewhere (Recasens, Espinosa, 2010) show that in order for /al/ to be heard as /awl/ the VC transitions must start before the steady-state portion of the alveolar lateral and therefore the tongue dorsum lowering and backing motion for dark /l/ must occur in advance of the tongue tip raising gesture for the consonant. A perception experiment was carried out in order to ascertain whether the identification of dark /l/ as /w/ is likely to proceed either through the pathway /Vl/ > [Vw] when F2 lowering at the consonant steady-state period causes dark /l/ to be perceived as /w/, or else the pathway /Vl/ > [Vwɬ] > [Vw] whenever variations in the relative timing of the VC transitions cause an [w]-like segment to be perceived at /l/ onset. A series of synthetic speech stimuli were built up which differed in the two acoustic parameters as follows: the steady-state F2 frequency values ranged between 650 Hz for a strongly dark realization of [ɬ] and 1050 Hz for a less dark realization of the consonant; the VC transitions ended at four different temporal points: at the onset of the /l/ steady-state period, or at 15 ms, 30 ms and 45 ms before it. Subjects were asked to identify the stimuli in question as /l/ or /w/.

Perceptual identification results show that /l/ is more likely to be identified as /w/ when F2 at the steady-state period of /l/ is lowered from 1050 Hz down to 650 Hz (by 5-45%) than when the offset of the VC transitions is anticipated in time (by 15-35%). Moreover, the latter acoustic characteristic contributes to an increase in the /w/ identification percentages only when the F2 frequency at the steady-state portion of the alveolar lateral is minimal (650-750 Hz). These perception results are consistent with the outcome of the vocalization process being [w] rather than [wl] in the world's languages, and suggest that the vowel transitions may act as a secondary vocalization cue which is used by speakers to enhance the perceptual effectiveness of the steady-state F2 frequency whenever /l/ is strongly dark.

2.2 Dark /l/ elision

Another example illustrating a double evolutionary pathway is the elision of dark /l/ in preconsonantal position mostly after /a/ or a back rounded vowel. It has been assumed traditionally that this change must take place after /l/ vocalization and thus that the final outcome *o*C of /alC/ (Spanish *otro* ALTERU "other", French *haut* ALTU) has been issued through the stages /alC/ > [awC] > [ɔwC] > *o*C

where the lateral vocalizes in the first place, /a/ assimilates to following [w] at a later stage, and [w] ceases to be heard due to perceptual similarity with preceding *o* at the end of the derivation. This development is supported by the existence of the intermediate sequence *ou* in lexical forms such as Portuguese *outro* ALTERU.

Another phonetic evolution is also possible, i.e., /alC/ > [ɔɬC] > *o*C, where consonant-dependent anticipatory tongue body lowering and retraction causes lower pharyngeal /a/ to raise to upper pharyngeal [ɔ] after which [ɬ] ceases to be heard due to the spectral similarity with the mid back rounded vowel. Supporting evidence for this evolutionary pathway comes from dialectal variants such as Lombard *molta* MALTHA "malt" (Rohlfs, 1966: 37), as well as the absence of specific intermediate vocalized variants. Thus, for example, Catalan *cop* COLAPHU "blow" must have been issued directly from *colp* since *colp* and *cop*, but not *coup*, appear in old written texts (Coromines, 1980-1988, 2: 888).

2.3 /Vɲ/ > [VjN]

It has been assumed that the decomposition of /ɲ/ into [jN], which occurs mostly syllable-finally, may operate whenever the alveolopalatal nasal exhibits a reduced realization involving a decrease in dorsopalatal contact. At least among a subset of speakers, this change has become phonologized in preconsonantal word final position in Majorcan Catalan. Thus, in this Catalan dialect, /aɲ/ is realized as [ajm] in the sequence *any bo* "good year" and /paɲ/ as [pajn] in the sequence *pany dur* "tough lock", where regressive place assimilation applies to the nasal consonant outcome of the /ɲ/ decomposition process. However, while weakening could certainly account for the reduction of /ɲ/ to [n] (see section 2), it leaves unexplained the presence of the on-glide in phonetic outcomes such as [jm] and [jn] above. An alternative explanatory account for this on-gliding process is that the replacement of /ɲ/ by [jN] has been triggered not by articulatory weakening but by articulatory reinforcement involving an increase in dorsopalatal contact and consequently in the frequency endpoint and the extent of the F2 vowel transitions (/ɲ/ > [jɲ] > [jN]). Several data speak in support of this phonetic development: electropalatographic data for Majorcan Catalan show that syllable-initial /ɲ/ may be realized not only as an alveolopalatal consonant exhibiting a closure location encompassing the alveolar and prepalatal zones but also as a purely palatal consonant which is articulated at the palatal zone exclusively; in other Romance languages, [j] insertion may also take place occasionally before /ɲ/ and other syllable initial (alveolo)palatal consonants without the consonant becoming depalatalized after on-gliding has taken place (see Recasens, 2014 for examples).

In order to elicit whether the change /ɲ/ > [jN] is triggered by articulatory weakening or by articulatory strengthening, the perceptual role of the two following acoustic cues acting in synchrony was submitted to perceptual evaluation in Recasens, Espinosa (2010):

(a) The duration of the VC transitions, which ought to be longer for [jn] and [jɲ] than for [ɲ] and, therefore, could be associated both with a reduced version of

[ɲ] (/ɲ/ > [jN]) and with an increase in dorsopalatal contact during the alveolopalatal nasal (/ɲ/ > [jɲ]).

(b) The VC transitions endpoint frequency, which should be higher for [jɲ] (though not necessarily for [jN]) than for [ɲ] and therefore ought to be positively related to an increase in dorsopalatal contact. Indeed, the larger the dorsopalatal contact size, the higher the F2 frequency at the offset of the vowel transitions is expected to be.

In order to find out which one of the two acoustic cues contributes most to /ɲ/ decomposition, a series of LPC-generated synthetic stimuli were built up using a representative production of the sequence /aɲ/ by a native Catalan speaker. The stimuli differed both in the endpoint frequency and duration of the F2 vowel transitions, which ranged between 1800 Hz to 2200 Hz and 70 and 170 ms respectively. Subjects were asked to identify these acoustic stimuli as either /ɲ/ or /jn/. Results show that the /jn/ identification percentages increase with the endpoint frequency (by 10%-50%) rather than with the duration (by 20%- 40%) of the F2 vowel transition; moreover, the duration cue was only perceptually effective when the endpoint frequency amounted to about 2100-2200 Hz and was thus highest. This finding is in support of the hypothesis that an increase in dorsopalatal contact may cause /ɲ/ to be heard as /jn/ since, as pointed out above, this articulatory characteristic and the F2 transitions endpoint frequency are positively correlated. Results also reveal that the duration of the F2 vowel transitions acts as a secondary cue by contributing to an increase in the percentage of /jn/ responses whenever the frequency endpoint exceeds 2000 Hz.

2.4 Palatalization of /kt, ks/

Another relevant sound change which illustrates the multiple vs single pathway is the evolution of /kt/ and /ks/ in the Romance languages. Two derivations (a) and (b) have been proposed traditionally in order to account for the phonetic outcomes /jt/, /tʃ/ and [c] of /kt/ (as in Portuguese *leite* LACTE "milk", Spanish *leche* and Rhaetoromance [lac]), and /js/ and /ʃ/ of /ks/ (as in French *fraisse* FRAXINU "ash" and Catalan *freixe*):

(a) The velar stop undergoes lenition and vocalization into [j] after which progressive palatalization causes the following dentoalveolar to turn into an (alveolo)palatal consonant (Menéndez Pidal, 1968: 143-144). At a later stage, the outcome [jc] may yield [c] though glide absorption and [c] may give rise to an affricate.

/kt/ > [çt] > [jt] > [jc] > [c] > [tʃ]
/ks/ > [çs] > [js] > [jʃ] > [ʃ].

(b) The two consonants of the cluster /kt/ blend into an alveolopalatal or palatalized stop ([c], [t^{j}]) and those of the cluster /ks/ into a palatoalveolar or palatalized fricative ([ʃ], [s^{j}], after which several changes (full palatalization, on-gliding, affrica-

tion) may take place. Two slightly different formulations of this development have been proposed: Hall's in (a') (Hall, 1974) and Thomsen's in (b') (Thomsen, 1876).

(a') /kt/ > [xt] > [çt] > [tʲ] > [tʃ, jt]
/ks/ > [xs] > [çs] > [sʲ] > [ʃ, js]

(b') /kt/ > [tʲ:] > [jt, c, tʃ]
/ks/ > [sʲ:] > [js, ʃ].

At least regarding originary /kt/, we would like to propose that the two phonetic derivations, not just one, may have operated in Romance. Indeed, the change /kt/ > [çt] > [jt] looks more plausible than /kt/ > [c] > [jt] in the case of dialects and languages where the final outcome is [(j)t] such as Portuguese, Old Spanish and Catalan, the main reason being that the alveolopalatal stop [c] is not prone to develop an on-glide due to its fast and relatively short vowel transitions. On the other hand, the derivation involving gestural blending is preferable in the case of dialects and languages where the final outcome is [c] (Rhaetoromance, Northern Italian dialects). It should also be noticed that, as the developments above show, the final outcome [tʃ] may derive from [c] in the two derivations and thus whether the (alveolo)palatal stop is generated from [jt] through progressive palatalization (as in Spanish) or through blending (as in Rhaetoromance).

As for /ks/, manner requirements for /s/ render unlikely the implementation of a blending process between the stop and the following fricative consonant the reason being that the tongue dorsum needs to be lowered somewhat for the passage of airflow through a central slit. Therefore, only the vocalization pathway referred to in (a) above and the following developments appear to be possible in this case: /ks/ > [çs] > [js] > [jsʲ] > [js] (French, Occitan) and /ks/ > [çs] > [js] > [jʃ] > [ʃ] (Catalan, Portuguese, Old Spanish).

3. *The contribution of articulatory/acoustic prominence*

This section explores briefly the relevance of positional factors in sound change implementation. Articulatory and acoustic data reveal that consonants are produced with more tongue contact and vowels become more open word-initially than word-internally, and that phonetic segments occurring in stressed syllables are more prominent than those embedded in unstressed ones (Fougeron, Keating, 1997).

In a recent book (Recasens, 2014) we analyzed the effect of word position and stress on sound changes affecting unstressed vowels in Catalan using assimilatory and dissimilatory data collected by dialectologists and linguists since the beginning of the XX century. According to this database, progressive assimilations and dissimilations triggered by consonants on an immediately following unstressed vowel occur mostly in the word syllable-initial position (70.5% of the time) which is in support of the notion that segmental prominence plays an important role in sound change implementation. Table 2 (1a) exemplifies the assimilatory action of consonants placed in the absolute word-initial position: the raising processes /o/ > [u]

and /e/ > [i] triggered by preceding labial and alveolopalatal consonants, respectively, and the lowering process /e/ > [a] after the alveolar trill [r] whose production involves some tongue dorsum lowering and retraction.

The examples under (1b) in Table 2 show that vowel-dependent regressive assimilations are prone to apply when the triggering vowel is stressed and the target vowel is located word-initially. This finding is in agreement with stressed vowels being especially prominent while running against the assumption that unstressed vowels subject to a specific sound change should be located in a non-prominent word position. The relevant examples in the table include two vowel raising instances induced by a contextual high vowel (/o/ > [u], /e/ > [i]) and an instance of vowel lowering triggered by a contextual low vowel (/e/ > [a]).

The third set of examples reported in section (1c) of Table 2 is about changes in unstressed vowel quality which are exerted simultaneously by contextual vowels and consonants. While sound changes have been traditionally attributed to a single contextual phonetic segment, there are cases where several contextual segments may act as sound change triggers.

Table 2 - *Assimilatory changes in unstressed vowels (underlined) triggered by contextual vowels and consonants (in bold). Data have been taken from a Catalan database*

Process	Triggering segment	Examples
(1a) Progressive assimilations induced by word initial consonants		
/o/ > [u]	labial C	[**b**o̲ˈro] “bud”
/e/ > [i]	alveolopalatal C	[ʎe̲sˈto] “lath”
/e/ > [a]	alveolar trill	[**r**e̲ˈse] “shelter”
(1b) Regressive assimilations induced by vowels		
/o/ > [u]	high V	[to̲ˈs**u**t] “stubborn”
/e/ > [i]	high V	[re̲ˈz**i**na] “resin”
/e/ > [a]	lower V	[te̲ˈr**a**t] “terrace”
(1c) Assimilations induced by contextual vowels and consonants		
/o/ > [u]	labial C, alveolopalatal C, high V	[**p**o̲ˈ**ʎi**] “donkey”
	velar C, labial C, high V	[**k**o̲ˈ**mu**] “common”
/e/ > [i]	alveolopalatal C, high vowel	[ʎe̲ˈ**ɣu**m] “vegetable”
		[de̲ˈ**ʒu**] “on an empty stomach”

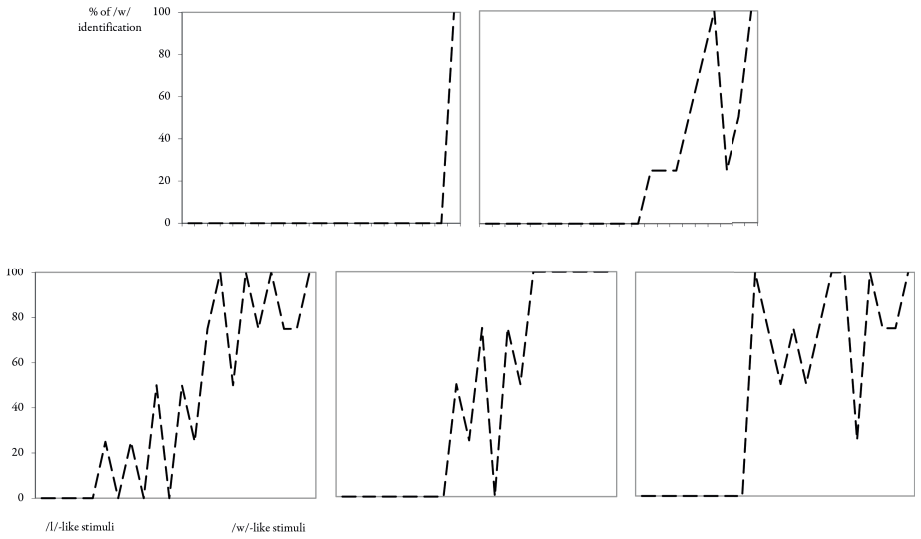
Thus, for example, the raising of /o/ to [u] in the word [poˈʎi] “donkey” in Western Catalan may in principle be exerted at the same time by the preceding labial consonant (which causes an increase in degree of labiality during /o/), and the following alveolopalatal consonant and high vowel (which cause the tongue dorsum to raise). Three contextual segments may also be involved in the raising of /o/ in the word [koˈmu] “common” in the same Catalan dialect: the stressed high back vowel and the following labial consonant, as well as the preceding velar stop which may cause the postdorsal closure for /o/ to shift from the upper pharynx towards the velar zone. Finally, the change /e/ > [i] in the words [ʎeˈɣum] “vegetable” and [deˈʒu]

"on an empty stomach" also in Western Catalan may be induced by both the following high stressed vowel and the preceding or following alveolopalatal or palatoalveolar consonant. All these examples reveal that the regressive assimilatory effect of a distant vowel may be reinforced by the progressive or regressive action of a contextual consonant, and therefore that vowel assimilations may be induced by the coarticulatory influence of several contextual segments acting simultaneously.

4. *How does sound change originate in the individual?*

Ohala's hypocorrection model states that sound change originates from the failure to compensate for coarticulation, which implies that listeners with little knowledge of the language such as children and second language learners are most likely to make perceptual errors and initiate sound changes. The lack of communication among peoples speaking the same language may cause much dialectal diversity to occur and therefore is also the seed for sound change. It has been shown indeed that individuals living in islands or remote regions may exhibit phonetic developments which diverge from the rule (Andersen, 1998). Another factor contributing to the inception of sound change, which is of much concern in the present investigation, is the degree of sensitivity on the part of the listener to small phonetic differences in the realization of a given phoneme which may take place in different contextual, positional and prosodic conditions. In so far as they are highly sensitive to such small variations in the acoustic signal, innovative listeners may lead sound changes (Grosvald, Corina, 2010). The relevance of this latter aspect can be ascertained from results obtained from peceptual tests which require a fine discrimination between acoustic stimuli differing slightly in one or more acoustic cues. Thus, the 17 subjects who took the identification test for preconsonantal [ɬ] referred to in section 1.2.2 could be grouped depending on their degree of sensitivity to small spectral differences which were available in the acoustic stimuli: at one end of the sensitivity scale, five subjects yielded no /w/ identification percentages at all and another two subjects seldom heard /w/ (two top panels in Figure 2); at the other end, ten subjects heard more instances of /w/ as the [ɬ] spectral cues became more /w/-like, with seven of them showing a more gradual increase in /w/ identification percentages than the remaining three (three bottom panels in Figure 2).

Figure 2 - *Subject-dependent patterns of identification of dark /l/ as /w/ according to the perception test described in section 1.2.2 (Top) Poor /w/ identification patterns; (bottom) better identification patterns exhibiting gradual and abrupt rises in the number of /w/ responses. In all graphs, the stimuli placed on the left are /l/-like and those on the right are /w/-like*



5. *Summary and conclusions*

Production and perception data reported in this paper reveal that intermediate articulatory variants need to be looked for before disclaiming an articulation-based explanation for specific sound changes. The relevance of phonetic detail for uncovering the phonetic causes of sound change has been illustrated in the case of two changes which are prone to be articulation-based rather than triggered by acoustic equivalence: /l/ vocalization and velar softening. Along these lines the study has also shown for /l/ vocalization and the segmental decomposition of palatal consonants that sound changes may have more than one acoustic origin and therefore may proceed through two and perhaps more evolutionary pathways. Information about the acoustic cues which contribute to the misidentification of one phoneme by another in these circumstances may be obtained by manipulating fine-grained acoustic characteristics which correlate with articulatory variations that may take place in real speech scenarios.

Results reported in this study also suggest that the acoustic cues for a given phoneme may play a primary or secondary role in sound change inception depending on their relative salience, which may vary from dialect to dialect. Secondary acoustic cues may be used by speakers to enhance the perceptual effectiveness of the primary acoustic cue. Thus, dark /l/ vocalization has been shown to depend on changes in F2 frequency at the steady-state consonant period and thus on tongue body configuration and apical contact loss rather than on variations in the timing of the vowel transitions which result from different degrees of gestural anticipation during the

preceding vowel. On the other hand, velar softening appears to be implemented through a rise in burst intensity level rather than through burst lengthening. In the two cases the secondary acoustic characteristics may assist the primary cue in giving rise to a sound change.

The role of segmental prominence and the individual speaker has also been looked into. As for the former aspect, descriptive data on the frequency of occurrence of contextual segments which may trigger a given change suggest that prominence within the word domain plays a relevant role in the implementation of assimilatory changes affecting unstressed vowels. Moreover, it appears that sound changes may be triggered by more than one segment at a time, a possibility which deserves being explored by means of acoustic analysis and perceptual tests in the future. Regarding the role of the individual speaker, the paper argues that those individuals who are especially sensitive to small variations in the acoustic signal are prone to lead sound changes.

Acknowledgments

This research has been funded by projects FFI2013-40579-P from the Spanish Ministry of Economy and Competitiveness, by ICREA (Catalan Institution for Research and Advanced Studies), and by the research groups 2014SGR61 from the Catalan government.

References

Andersen, H. (1988). Center and periphery: adoption, diffusion, and spread. In Fisiak, J. (Ed.), *Historical dialectology, regional and social*. Berlin: Mouton de Gruyter, 39-83.

Beddor, P.S. (1983). *Phonological and phonetic effects of nasalization on vowel height*. Bloomington: Indiana University Linguistics Club.

Carignan, C., Shosted, R.K., Shih, Ch. & Rong, P. (2013). Compensatory articulation in American English nasalized vowels. In *Journal of Phonetics*, 39, 668-682.

Coromines, J. (1990-1998). *Diccionari etimològici complementari de la llengua catalana,* 8 vols. Barcelona: Curial.

Fougeron, C., Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. In *Journal of the Acoustical Society of America*, 101, 3728-3740.

Garrett, A., Johnson, K. (2010). Phonetic bias in sound change. In Yu, A.C. (Ed.), *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press, 51-79.

Gauchat, L. (1925). Confusions d'occlusives dans les patois de la Suisse romande. In *Homenaje ofrecido a Menéndez Pidal*, 1, Madrid: Hernando.

Grammont, M. (1933). *Traité de phonétique*. Paris: Delagrave.

Grosvald, M., Corina, D. (2010). The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In Solé, M.J., Recasens, D.

(Eds.), *The Initiation of sound shange: Perception, production, and social factors*. Amsterdam: John Benjamins, 77-102.

Guion, S.G. (1998). The role of perception in the sound change of velar palatalization. In *Phonetica*, 55, 18-52.

Hall, R.A. (1974). *External history of the Romance languages*. New York: Elsevier.

Hardcastle, W.J., Barry, W.J. (1985). Articulatory and perceptual factors in /l/ vocalization in English. In *University of Reading Phonetics Laboratory Work in Progress*, 5, 31-44.

Hickley, R. (1984). On the nature of labial velar shift. In *Journal of Phonetics*, 12, 345-354.

Labov, W. (1994). *Principles of linguistic change: Internal factors*. Wiley-Blackwell.

Lausberg, H. (1965). *Lingüística románica*, 1, Madrid: Gredos.

Lin, S., Beddor, P.S. & Coetzee, A.W. (2014). Gestural reduction, lexical frequency, and sound change: a study of /l/ lenition. In *Laboratory Phonology*, 5, 2-36.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. In *Language and Speech*, 29, 3-11.

Menéndez Pidal, R. (1968). *Manual de gramática histórica española*. Madrid: Espasa Calpe.

Millardet, G. (1910). *Études de dialectologie landaise; le développement des phonèmes additionnels*. Toulouse: E. Privat.

Ohala, J.J. (1974). Phonetic explanation in phonology. In Bruck, A., Fox, R. & La Galy, M. (Eds.), *Papers from the parasession on Natural Phonology*. Chicago: Chicago Linguistic Society, 251-274.

Ohala, J.J. (1978). The story of [w]: An exercise in the phonetic explanation for sound patterns. In *Report of the Phonology Laboratory*, University of California at Berkeley, 2, 133-155.

Ohala, J.J. (1981). The listener as a source of sound change. In Masek, C.S., Hendricks, .A. & Miller, M.F. (Eds.), *Papers from the Parasesssion on Language and Behaviour*. Chicago: Chicago Linguistic Society, 178-203.

Ohala, J.J. (1989). Sound change is drawn from a pool of synchronic variation. In Breivik, L.E., Jahr, E.H. (Eds.), *Language change: Contribution to the study of its causes*. Berlin: Mouton de Gruyter, 173-198.

Phillips, B.S. (2006). *Word frequency and lexical diffusion*. New York: Palgrave Macmillan.

Rankin, R.L. (1974). Latin kw, gw Rumanian p,b: an explanation. In Luján, M., Hensey, F. (Eds.), *Current studies in Romance linguistics.* Washington: Georgetown University Press, 14-26.

Recasens, D. (2014). *Coarticulation and sound change in Romance*. Amsterdam: John Benjamins.

Recasens, D. (2011). Velar and dental stop consonant softening in Romance. In *Diachronica*, 28, 186-224.

Recasens, D. (2014). Acoustic and coarticulatory characteristics of (alveolo) palatal stop consonants, and velar softening. In *Journal of Phonetics*, 42, 37-51.

Recasens, D., Espinosa, A. (2010). A perceptual analysis of the articulatory and acoustic factors triggering dark /l/ vocalization. In Recasens, D., Sánchez Miret, F. & Wireback, J. (Eds.), *Experimental phonetics and sound change*. München: Lincom Europa, 71-82.

Recasens, D., Espinosa, A. (2010). The role of spectral and temporal cues in consonantal vocalization and glide insertion. In *Phonetica*, 67, 1-24.

Repetti, L., Tuttle, E.F. (1987). The evolution of Latin PL, BL, and CL and GL in Western Romance. In *Studi mediolatini e volgari*, 33, 53-115.

Rohlfs, G. (1966). *Grammatica storica della lingua italian e dei suoi dialetti*. 1. Fonetica. Turin: Einaudi.

Rousselot, J.-P. (1897-1901). *Principes de phonétiques expérimentale*. Paris: Welter.

Scobbie, J., Wrench, A. (2003). An articulatory investigation of word final /l/ and /l/-sandhi in three dialects of English. In Solé, M.J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the 15th International Congress of the Phonetic Sciences*. Barcelona: Causal, 1871-1874.

Solé, M.J., Recasens, D. (2012). *The initiation of sound change. Production, perception, and social factors*. Amsterdam: John Benjamins.

Thomsen, V.L.P. (1876). Remarques sur la phonétique romane. L'‹i› parasite et les consonnes mouillées en français. In *Mémoires de la Société de Linguistique de Paris*, 3, 106-123.

Wang, W.S.-Y. (1969). Competing changes as a cause of residue. In *Language*, 45, 9-25.

Yu, A. (2013). *Origins of sound change. Approaches to phonologization*. Oxford: Oxford University Press.

DOI: 10.17469/O2101AISV000003

MARIANNE POUPLIER[1]

Between 'whims of fashion' and 'phonetic law': Performance constraints in speech production in the face of linguistic diversity

There is a long tradition in the speech sciences seeking to trace the shaping force of physiological constraints in the world's languages. Particularly the syllable has been in the focus of biologically oriented approaches and has been seen as the primary locus for the manifestation of universal production constraints. We discuss how current research into patterns of articulatory coordination within the syllable has revealed that by hypothesis preferred patterns are across languages more varied than anticipated by these models. Data from Polish are used as a case in point to exemplify how trade-offs between different production constraints may condition a departure from a by hypothesis preferred pattern. We discuss the extent to which these trade-offs may be language specific and the ramifications for modelling speech production.

1. *Physiological preference and the limits of linguistic diversity*

It almost goes without saying that speech perception and production impose absolute limits on linguistic diversity in terms of hearing thresholds, just noticeable differences, anatomical limits on possible tongue shapes, and the like, even though it has not been trivial to identify those hard limits of speech (e.g. the discussion on the (im)possibility of nasalized fricatives, cf. recently Warner et al., 2015). There is less consensus regarding the question whether performance constraints shape sound patterns *within* the available space. Some have taken the view that spoken language is first and foremost to be understood as a cultural product. Ladefoged, for instance, maintained that differences between languages are due to "whims of fashion rather than the rule of phonetic law" (Ladefoged, 1983a: 2); see also Ladefoged (1983b). From such a stance statistical regularities in the languages of the world are primarily "matters of chance and custom" (Ladefoged, 1983a: 1), and so are the exceptions to these statistical tendencies. They arise from just-so stories rooted in each language's individual history and culture (Evans, Levinson, 2009).

Others have taken a different stance: Lindblom, MacNeilage & Studdert-Kennedy (1983), for instance, observed many years ago that the world's languages "fastidiously

[1] Work supported by the ERC under the EU's 7th Framework Programme (FP/2007-2013) / Grant Agreement n. 283349-SCSPL. I would like to thank the organizers and audience of AISV 2015 for valuable discussions and feedback. I am also greatly indebted to my collaborators Manfred Pastätter and Stefania Marin.

underexploit the full range of possibilities" (193). They argued that a random sampling of the possibilities of the universal phonetic space should make all possible sound patterns equally likely. Yet the phonologies the world's languages seem to defy the notion of random sampling. Performance constraints of the speech production and perception systems lead to structuration in a self-organizing system and thus effectively impose limits on linguistic diversity. The authors formulate performance constraints relevant for speech as sufficient perceptual distance at an acceptable articulatory cost (also known by the terms of hyper- and hypo-articulation (Lindblom, 1983b)). The impetus of the research program laid out by these and other authors in the 1980s and early 1990s was to understand spoken language primarily as biological behavior. The phonological and phonetic structure of spoken language is designed to ensure a highly efficient communication system; what efficiency means can only be understood with reference to the auditory and speech motor system. The latter is in turn nothing but a biological motion system – hence explicit parallels are sought between 'talking' and 'walking', posture control, reaching tasks, finger tapping and the laws of evolution (Grimme, Fuchs, Perrier & Schöner, 2011; Haken, Kelso & Bunz, 1985; Kelso, 1986; Lindblom, 1983a; Nelson, 1983; Ostry, Cooke & Munhall, 1987; Pouplier, 2012a). For instance, the wide preference for CV syllable structure or for mirroring consonant order between onset and coda (e.g., onset /bl/, coda /lb/) are argued to be deeply rooted in the physiology of speech production and perception (Browman, Goldstein, 1986; MacNeilage, 1998; MacNeilage, Davis, Kinney & Matyear, 2000; Redford, 1999; Vallée, Rossato & Rousset, 2009).

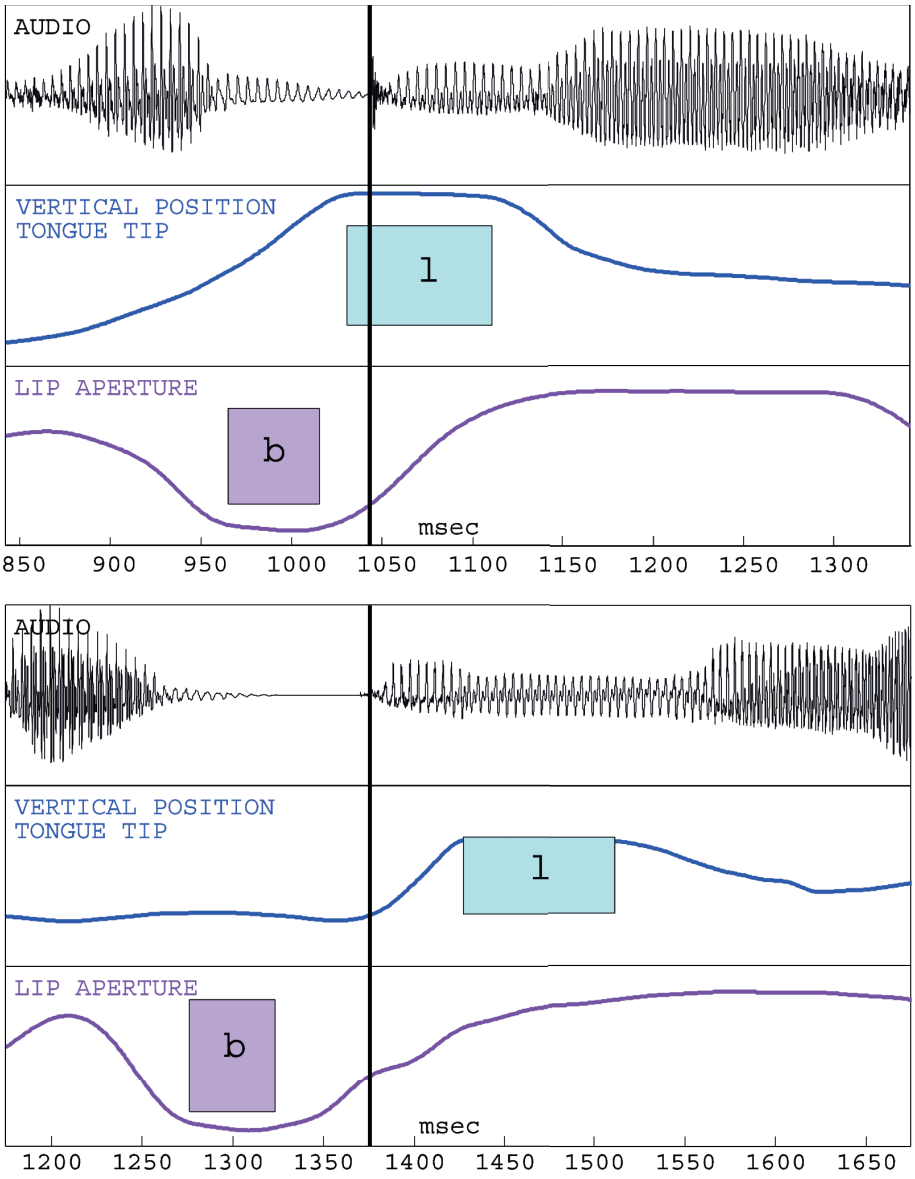
A lot of fruitful research has been informed by the research program looking for the biological foundations of language. Empirical efforts have been devoted to recreate in the laboratory distributional asymmetries in the world's languages in order to understand their conditions of emergence from a physiological viewpoint. A famous example is rate-induced resyllabification of VC to CV, first described by Stetson (1951), which has inspired an approach to spoken language from a dynamical systems perspective (Kelso, Saltzman & Tuller, 1986). Understanding preferred production patterns through rate-scaling is continued as a research paradigm until today (Goldstein, Pouplier, Chen, Saltzman & Byrd, 2007; Rochet-Capellan, Schwartz, 2007). Yet pinning down the relationship between (reflexes of) biological constraint in grammar and linguistic diversity has been very difficult. Languages continue to surprise us as to which sound patterns they incorporate into their grammars: much cited examples are languages like Tashlhiyt Berber or Georgian which in certain parts of their grammar seem to defy any notions of sonority in syllable phonotactics (Butskhrikidze, 2002; Dell, Elmedlaoui, 2002). Also many of the Slavic languages make pervasive use of phonotactic patterns that go against the grain of physiological preference. Languages also continue to surprise us with respect to the breadth of articulatory patterns they employ; the physical implementation of speech is more varied than often assumed, and it has been argued that there is a principled link between the phonotactic rules of a language and which general articulatory coordination pattern the language employs (Pouplier, Beňuš, 2011). In this paper, we will discuss a study on Polish as one instance

of empirical evidence for ways in which languages may transgress hypothesized production preferences.

The syllable is by many seen as a primary domain within which production constraints operate, conditioning systematic phonotactic asymmetries. For instance, the jaw cycle has been proposed as the phylo- and ontogenetic basis for the syllable, and as a conditioning (albeit not a deterministic) factor in consonant phonotactics (MacNeilage, Davis, 2000; 2001; Redford, 1999; 2001). Since the 1980s and the rise of a dynamical systems approach to speech production which is deeply informed by a biological perspective on language, the role of in-phase coordination has received a lot of attention. The results of the above-mentioned experiment by Stetson (1951) in which speakers resyllabify VC under rate pressure to CV have been interpreted as evidence for a biologically preferred mode of speech motor coordination (synchrony) to be part of CV but not VC production (de Jong, 2001; Kelso et al., 1986): In a CV, but not a VC syllable, consonant and vowel are produced in-phase. Also work on speech errors (Goldstein, Pouplier et al., 2007; Pouplier, 2008) and the labial-coronal effect (Rochet-Capellan, Schwartz, 2007) have been linked to in-phase coordination and the jaw cycle. The labial-coronal effect appears in experimentally induced reorganization of CVCV to CCV (in a rate-scaling task). Specifically, both *pata* and *tapa* result in the labial-coronal sequence /pta/ (not /tpa/). The authors claim that this is due to /pta/ allowing for in-phase coordination with the jaw. They draw a direct connection between their experimental results and a statistical preference for labial-coronal phonotactics in the world's languages.

Despite this and similar evidence for the existence and the formative role of production constraints, languages have been known to differ systematically in their articulatory organization of the vowel cycle (vowel-to-vowel coarticulation (Beddor, Harnsberger & Lindemann, 2002; Manuel, 1990; Öhman, 1966)) as well as in consonant-consonant timing (Davidson, 2006; Kochetov, Pouplier & Son, 2007; Pouplier, Beňuš, 2011). This means for one that the same sequence of phonological symbols can be (co-)articulated very differently across languages. Figure 1 gives an illustrative example for such timing differences between German and Georgian for the syllable /ble/. In German (top panel of Figure 1), the /l/ has achieved its target plateau at the time of the release burst of the labial. In contrast, note how for Georgian (bottom panel) the constriction formation for /l/ only begins after the release of the labial, giving rise to an open transition. Secondly, also within a language place and manner differences are associated with significant differences in timing. For instance, Hoole and Bombien provided evidence for tautosyllabic stop-lateral sequences in German and French overlapping more in time than stop-nasal sequences (Bombien, Hoole, 2013; Hoole, Pouplier, Benus & Bombien, 2013). They attribute this to conflicting aerodynamic conditions for stops and nasals; a tight overlap between a stop and a nasal would obscure the burst and hence would make /kn/ perceptually unstable.

Figure 1 - *Illustration of timing differences between languages on the basis of kinematic data. The same segmental sequence, /bl/, is timed very differently in German (top) and Georgian (bottom). Top panel within each figure shows vertical tongue tip position over time, the bottom panel shows lip aperture. The rectangles indicate articulatory constriction plateaus for /b, l/. The black vertical line across panels marks the timepoint of the acoustic burst for /b/*



Many of the systematic articulatory timing differences reported in the literature are related to prosodic organization, i.e. syllable position and constituency. In an ideal case, we can distinguish by articulatory patterning a VC$CV from a V$CCV from a VCC$V sequence in terms of both consonant-vowel and consonant-con-

sonant timing (Krakow, 1999). Since the syllable is also seen as a primary locus of physiological constraint manifestation, it constitutes an ideal testing ground for the interaction between language-specific and universal aspects of grammar and speech motor control. In the next section I will briefly introduce the coupled oscillator model of syllable structure as one particular suggestion of how 'naturally given' speech motor patterns are exploited by linguistic systems. I will then report on recent work on Polish which seeks to understand the conditions under which languages may diverge from these patterns, and explore the ramifications for modelling speech production.

2. *Coupled oscillator view of syllable structure*

Within their articulatory phonology framework, Browman and Goldstein (1985) and subsequent) have proposed to model syllable structure based on systems of coupled oscillators. A comprehensive overview of articulatory phonology is outside the scope of this paper; the interested reader is referred to recent publications (Gafos, Goldstein, 2012; Goldstein, Pouplier, 2014; Pouplier, 2011; 2012b). Articulatory phonology is a 'representational phonology' in the sense of Kisseberth (2011), characterized by laying emphasis on the predictive power of representations rather than on rules and constraints (for a combined articulatory phonology and OT approach see Gafos, 2002). Gestures, the basic representational units of the model, are conceptualized as linear second order mass-spring systems which enter into lexically specified coupling relations to form larger constituents such as segments and syllables. Syllables are seen as arising from characteristic patterns of temporal coordination among underlying gestures.

The gestural framework has relied to a large degree on the explanatory power of the relative phase values of 0° and 180° which have been identified in non-speech tasks as intrinsic modes of a system of coupled oscillators (Turvey, 1990). This means that these coordination patterns can be performed without learning and cannot be destabilized. Humans routinely learn and perform very complex coordination patterns, but when the learnt patterns can no longer be upheld due to e.g., rate pressure or disease, in-phase coordination is still retained. In-phase coordination is pervasive in biological systems (Strogatz, Stewart, 1993), and thus it is little surprising that frameworks that view speaking as a biological behavior seek to trace in-phase coordination in language. In this vein, articulatory phonology has proposed that the intrinsically stable modes are the basis of syllable structure. For instance, the near-simultaneous articulator movement onset of consonant and vowel in a CV syllable (Löfqvist, Gracco, 1999) has been interpreted as arising from an underlying in-phase coordination of syllable onset and vowel. Due to the greater temporal extent of the vowel (vowel articulations are intrinsically slower than consonantal articulations), the impression of serial order arises. Codas have been associated with an anti-phase or eccentric coordination mode (Goldstein, Pouplier, 2014; Krakow, 1999; Pouplier, 2012b). A rather direct link has been drawn between the basic sta-

bility of in-phase coordination and linguistic preference, i.e. statistical asymmetries in cross-linguistic patterns. Typological CV preference, the order of acquisition of onsets before codas, onsets having a greater resistance to sound change compared to codas and moraic weightlessness of onsets have all been linked to the hallmark stability of in-phase coordination (Nam, 2007; Nam, Goldstein & Saltzman, 2009).

In this model, onset clusters require more complex coordination patterns than singleton onsets: they are characterized by two competing phase relationships which prevent the multiple consonants from being produced on top of each other. All onset consonants are coupled in-phase to the vowel, but to ensure their recoverability, they are coupled anti-phase to each other. This underlying coupling topology gives rise to the so-called 'c-center effect' (Browman, Goldstein, 1995; 2000). The c-center simply denotes the surface correlate of articulator timing that the competing underlying coupling graphs give rise to. Specifically, the c-center provides evidence for the onset being coordinated at some level as a single entity to the vowel – while the timing of each individual consonant to the vowel changes with increasing onset complexity, the timing of the onset as whole (the c-center) does not. For example, when going from /**k**a/ to /s**k**a/, the timing of the /k/ to the vowel will differ in the two syllables, but the timing of the *entire onset* /**k**a/ and /**sk**a/ to the vowel will be the same. For the data we will consider below, it is important to point out that the c-center effect conditions increasing overlap of the vowel-adjacent consonant with the vowel as onset complexity increases: the /k/ in /s**k**a/ is more overlapped with the vowel compared to the /k/ in /**k**a/. For recent evidence on this matter see Peters and Kleber (2014).

Over the past decade, the model has been tested for a range of languages (Brunner, Geng, Sotiropoulou & Gafos, 2014; Gafos, Hoole, Roon & Zeroual, 2010; Goldstein, Chitoran & Selkirk, 2007; Hermes, Mücke & Grice, 2013; Marin, 2013; Marin, Pouplier, 2010; Pouplier, Beňuš, 2011). Results have supported the general idea that syllable constituency is articulatorily expressed in the relative timing of the consonants to each other and to the vowel. At the same time it has become clear that the basic gestural syllable model is not able to fully capture onset-vowel timing; a number of exceptions have been reported which are not straightforwardly accommodated by the model.

One specific exception to the patterns predicted by the gestural syllable model has been discussed in the context of obstruent clusters. We will report the relevant studies here as a case in point for languages stabilizing a greater range of patterns than anticipated by physiologically based syllable models. In particular, we focus on the role of coarticulation resistance. Coarticulation resistance is a concept in speech motor control that describes the differential contextual flexibility of sounds (Bladon, Al-Bamerni, 1976; Farnetani, Recasens, 2010; Iskarous, Mooshammer, Hoole & Recasens, 2013; Recasens, 2002): sounds with stringent aerodynamic production requirements or a high degree of dorsal control will resist coarticulation with neighboring consonants and vowels, for example, sibilants and the palatal vowel /i/ are highly ranked on the coarticulation resistance scale. Labials on the

other hand, which require no lingual control at all, are among the least coarticulation resistant sounds: tongue shape during the production of a labial will be entirely dominated by the surrounding vowel context.

Complexity in the gestural model is defined as any deviation from the fundamental in-phase/anti-phase pattern and these deviations have often been motivated on perceptual grounds (Browman, Goldstein, 2000; Chitoran, Goldstein & Byrd, 2002). If coarticulation resistance interacts with syllable-level timing, this provides evidence that different gestural coordination patterns may be stabilized by languages on grounds of speech motor control, i.e. there may be trade-offs in speech motor control itself which lead to a contextual conditioned redefinition of preferred articulatory coordination patterns without necessarily implying a lesser stability thereof (Pouplier, 2012a). This in turn would invite us to take a broader view of what are deemed to be preferred articulatory patterns across languages.

3. *Onset-vowel timing and coarticulation resistance in Polish*

Marin (2013) reported for an articulography (EMA)[2] study on Romanian that consonant order interacts with the c-center effect. In particular, the onset clusters /sp-, sk-, sm-/ patterned as expected by the gestural model, showing an increase in consonant-vowel overlap between singleton and corresponding cluster condition (i.e., the /p/ in /**p**alə/ overlapped less with the vowel than the /p/ in /s**p**alə/). Stop-initial clusters /ps-, ks-, kt-, kn-/ failed to exhibit a c-center effect, there was no increase in C-V overlap between /**s**alt/ and /p**s**alm/. She attributed the lack of a c-center effect for the second cluster group to C1 being a stop, also discussing the possible role of frequency (Marin, 2011). Pastätter and Pouplier (2014; 2015) recently followed up on Marin's results on the basis of Polish, pursuing in particular the possibility that it is less manner of C1, but rather the coarticulation resistance of C2 that conditioned her results. Increasing consonant-vowel overlap with increasing onset complexity may not be a viable pattern for highly coarticulation resistant consonants such as sibilants.

Polish is a Slavic language featuring an unusual range of consonant clusters (Rochon, 2000). In particular, for a set of CC clusters the consonant members can appear in either order in both onset and coda, thereby allowing us to study the effect of order reversal in a systematic fashion within the same syllable position. By means of articulography (EMA) data from 6 Polish speakers were recorded by Pastätter and Pouplier (2014; 2015). The aim of the study was to test the hypothesis that increasing onset-vowel overlap conditioned by increasing onset complexity as predicted by the gestural syllable model is in an inverse relationship with the coarticulation resistance of the vowel-adjacent consonant (C_2). A highly coarticulation resistant

[2] Electromagnetic Articulography (EMA) is a recording technique that tracks the position of receiver coils attached to a speaker's articulators within a magnetic field (for an introductory overview see Stone, 2010).

consonant is expected to block increasing vowel-consonant overlap and no c-center effect will be observed. Sibilants are known to be among the most coarticulation resistant consonants both in terms of lingual and jaw position (Recasens, Espinosa, 2009), therefore the study in first instance focused on SCV and CSV onset clusters, where S stands for /s, ʃ/. Polish allows for a range of symmetrical SCV and CSV clusters all of which were included in our study: /ks-, sk-; mʃ-, ʃm-; ps-, sp-; pʃ-, ʃp-/; corresponding singleton conditions were also recorded (e.g., ***mʃ****alik* – ***ʃ****alik;* ***ʃm****ata* – ***m****ata*). As an index of onset-vowel overlap, the temporal distance between the vowel-adjacent consonant of the cluster and a constant reference measurement point (a following word-medial consonant) was calculated for singleton-cluster stimulus pairs. For instance, the temporal lag of /ʃ/ in ***ʃ****alik* to the word-internal /l/ was compared to to the same lag in ***mʃ****alik*. The gestural syllable model predicts a decrease in this distance measure as going from singleton to cluster. The zero centered cluster:singleton ratio served as dependent variable. On this ratio, values <0 indicate an increase in onset-vowel overlap between singleton and cluster condition. Results confirmed the predicted significant main effect of sibilant position (vowel adjacent vs. vowel remote) on onset-vowel timing (Figure 2). For the CSV condition, there was no change in onset-vowel overlap as onset complexity increased from SV to CSV. For SCV conditions, on the other hand, such an increase in overlap was observed between CV => SCV, in accordance with the gestural model (Figure 2)[3]. That is, SCV clusters conformed to the preferred onset-vowel coordination pattern, while CSV clusters did not.

Coarticulation resistance is also known to hold for the jaw. The content-frame model of speech production (MacNeilage, 1998) has also drawn a strong link between the biological givens of speech production and linguistic structure, in particular syllable phonotactics. The framework ascribes a central role to the jaw cycle and predicts that languages will preferably arrange their CV phonotactics in accordance with the jaw cycle (MacNeilage, Davis, 2000). Frame-content theory has also been applied to consonant phonotactics within the syllable, deducing the typical reversal of consonant order between onset and coda from an interaction of consonantal jaw position requirements and the jaw cycle: for instance, /l/ requires an open jaw position, hence the 'natural' consonant order from a jaw perspective is /bl/ in onset but /lb/ in coda (Redford, 1999). Sibilants are produced by tunneling a jet of air against the front teeth and thus require a high jaw position and allow for little context-conditioned variability. We now take a look at jaw position in the Pastätter & Pouplier data. Figure 3 shows representative data from one speaker. Here for the same clusters that have entered into Figure 2, we see averaged jaw position at the temporal constriction midpoint of each segment of the (C)CVC syllable separately for the two onset types. The singleton condition is on the right in each graph. The relevant point here is the relative position differences between C1 and C2 for the

[3] Some approaches have argued, mostly on theory-internal grounds, for a 'special' status of obstruent-obstruent onset clusters in Polish and, for SCV clusters, an extrasyllabic status of /s/. We note that s-initial clusters show canonical syllable organization in our articulatory data (see also Rochon, 2000).

two cluster types. SC clusters show a monotonic lowering of the jaw throughout the cluster into the vowel. For CS clusters, on the other hand, we see a higher jaw position during C2 than during C1, i.e. the sibilant enforces a reversal of the jaw cycle during the onset for CS but not for SC clusters[4].

Figure 2 - *Cluster: singleton onset-vowel overlap ratio (zero-centered) as a function of sibilant position in the cluster. Negative values mean increasing C_2-vowel overlap with increasing onset complexity, a value of zero means no change in overlap between singleton and cluster conditions*

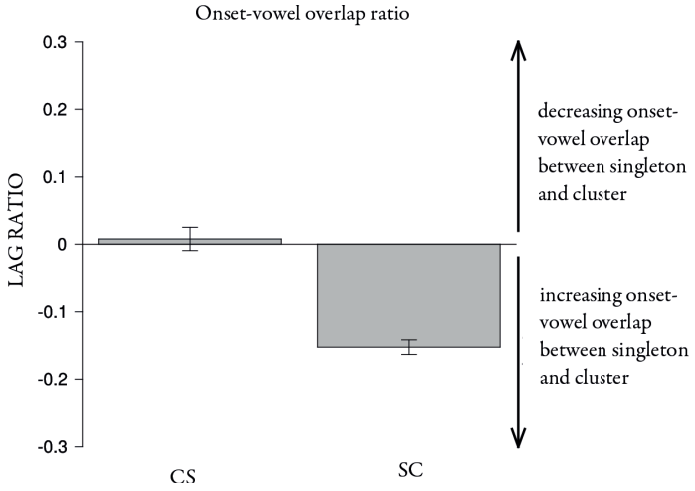
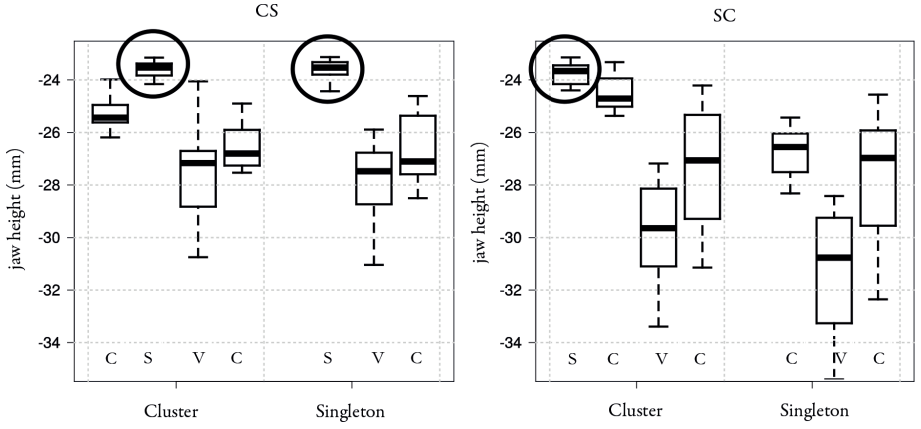


Figure 3 - *Average jaw position for each segment of the (C)CVC syllable for the two cluster types. CS clusters on the left, SC clusters on the right. Within each graph, the singleton condition is on the right, the cluster condition on the left. The sibilant is circled in each graph*

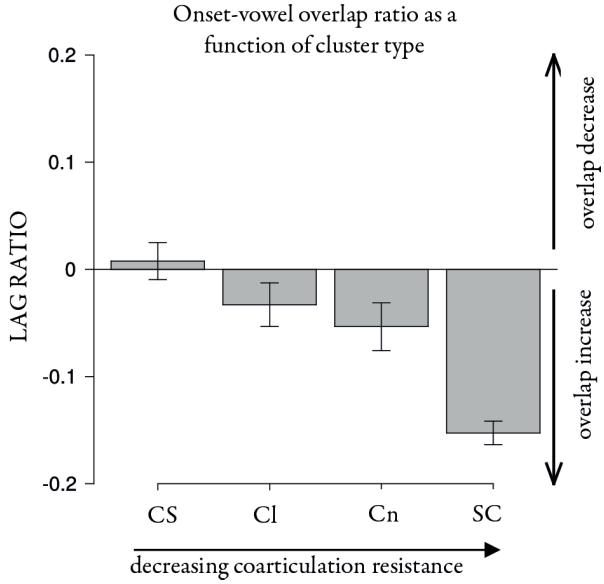


[4] Jaw coarticulation resistance may also explain why in the Marin data /kt/ patterned with /ks, ps/: While lingually among the lesser coarticulation resistant consonants, /t/ is for aerodynamic reasons quite stringent on jaw position requirements (Mooshammer, Hoole & Geumann, 2007).

Having identified differences in onset-vowel combination in SC vs. CS clusters which seem to be conditioned with differences in coarticulation resistance, we take an initial look at whether this would imply a difference in stability. In terms of token-to-token variability of the lag values, we find a significant difference in standard deviations SC: SD = .098; CS: SD = .134. A mixed model on these standard deviations was significant (dependent variable: standard deviations of onset-vowel overlap measure; fixed factor: Sibilant Position, random factors: Repetition, Cluster; $X^2[1] = 4.69$; $p < .05$). This suggests, against expectations, a lesser stability of CSV clusters.

The role of coarticulation resistance could be further substantiated by expanding the range of clusters beyond sibilants. Pastätter and Pouplier (2015) also included Cn (/kn-, pn-/) and Cl clusters (/kl-, pl-, ml-, vl-/). While Polish phonotactics does not allow us to systematically study the effect of order reversal for these clusters, they can be used to vary the coarticulation resistance of C_2. Results are given in Figure 4: Onset-vowel overlap increases linearly with decreasing coarticulation resistance of the vowel-adjacent consonant.

Figure 4 - *Cluster:singleton lag ratios as a function of coarticulation resistance of the vowel-adjacent consonant (decreasing from left to right). Negative values mean increasing C_2-vowel overlap with increasing onset complexity, a value of zero means no change in overlap between singleton and cluster conditions. Adapted from Pastätter and Pouplier (2015)*



Overall, the Polish data exhibit a systematic variation in onset-vowel timing as a function of manner of C_2, in accordance with the independently established coarticulation resistance hierarchy (Recasens, 1999; 2012; Recasens, Pallarès & Fontdevila, 1997). From the view of the gestural syllable model this means that a general, by hypothesis preferred pattern of temporal onset-vowel coordination may

be overridden by other, spatial factors such as coarticulation resistance. The significant difference in standard deviations evidence a greater token-to-token variability in lag ratios for the CS compared to SC conditions. More sophisticated probing of the production system in terms of flexibility under prosodic variation or in the speech error probability should render a more complete picture of a possible differential plasticity of these patterns.

4. *Conclusions*

The issue we have laid out at the beginning of this paper was how frameworks which have made rather strong claims about syllable structure being rooted in, and ultimately epiphenomenal to biologically preferred movement coordination patterns may negotiate unexpected diversity in movement coordination. We have used data from Polish to trace a particular instance in which presumably preferred production pattern (the c-center, expression of underlying in-phase coordination) can be overridden by another production constraint, coarticulation resistance. The Polish data show that syllable-level timing preferences interact with a spatial factor – coarticulation resistance. The data show some indication for a lesser stability, providing initial evidence that this comes at a cost. To the extent that further analyses confirm the asymmetrical token-to-token variability between CS and SC cluster, this would in turn reinforce biologically oriented models of the syllable which have made rather strong claims about the foundational role of preferred timing patterns for the syllable, and have implicitly defined complexity as any deviation from the preferred timing pattern.

The Polish data are, however, only a first step in gaining a deeper understanding of the full range of production patterns that we may find across languages. Goldstein et al. (Goldstein, Nam, Saltzman & Chitoran, 2009) have proposed, comparing syllable onset data from English and Georgian, to increase the degrees of freedom for onset clusters by recognizing closure formation and release of each consonant as independent gestures (Browman, 1994; Nam, 2007). For the syllable model this means that the closure and release gestures of each consonant can independently participate in coupling relations to each other and to the vowel, exponentially increasing the number of possible coupling graphs for onset clusters with a corresponding loss of predictive power. This idea mainly served to capture language-specific differences in the behavior of C_1 in a C_1C_2 cluster in terms of C-C and C_1-V timing. This idea cannot be expanded to the Polish data in the current context, since coarticulation resistance is predominantly a spatial effect.

For the current data, a question to ask is how we can model the interaction between coarticulation resistance and timing, and whether we expect coarticulation resistance to differ between languages. At first blush, the answer to the latter question may be 'no' for sounds that are produced in the same fashion, i.e. using the same articulator synergies. /l/ for instance, can be produced with quite different synergies in terms of dorsal configurations giving rise to the well-known 'clear' vs. 'dark' con-

tinuum, which links correspondingly to the degree of dorsal control to differences in coarticulatory resistance (Recasens, Fontdevila & Pallarès, 1996). Yet for sibilants there have to my knowledge been no reports of a differential coarticulation resistance across languages. Velars are quite instructive to consider in this context. Velars exert a high degree of control on the tongue dorsum and hence should be coarticulation resistant on that count, but the well-known effect of velar fronting shows that velars in fact do allow for an adjacent vowel to exert considerable influence on the velar's place of articulation (Öhman, 1967). Iskarous et al. (2012) present a novel analysis of how language-specific effects in coarticulation can be modelled, using allophony in Navajo velar fricatives as case study: they argue for a language-specific element in parameter blending in the gestural model. Parameter blending is an element of the task dynamic (Saltzman, Munhall, 1989) model designed to capture dominance relations in cases of conflicting demands on the same articulator, i.e. it effectively models coarticulation resistance (Fowler, Brancazio, 2000; Fowler, Saltzman, 1993). Iskarous et al. are the first to discuss how language-specific effects in coarticulation resistance can be captured in a principled fashion using the blending parameter and by recognizing the independence of constriction location and degree in this respect: languages may differ on whether blending averages both constriction location and degree of the conflicting articulations (Navajo), or whether averaging occurs in only one of the parameters (English: constriction location allows for averaging, constriction degree is entirely dominated by the consonant). While this account elegantly captures the known language-specific spatial effects in velar-vowel coarticulation, the Polish data go further in suggesting that there is an interaction between syllable-level timing and coarticulation resistance, i.e. between spatial and temporal parameters, blocking increasing onset-vowel overlap with increasing coarticulation resistance. Further modelling work will have to be carried out in order to understand how to appropriately capture the Polish data, and cross-linguistic work will help pinning down how languages may negotiate different spatial and temporal demands of the speech motor system.

The existing work on articulatory coordination cited in this paper suggests that both spatial and temporal aspects of spoken language may carry a language specific component, and the range patterns languages allow for is currently not well understood. The biologically oriented models of speech production have tremendously contributed to our understanding of speech motor control and spoken language as a behavioral system. There is however a general tension between the models and linguistic diversity which we encounter not only in terms of phonological inventories and phonological rules, but also in terms of the articulatory implementation of speech. It is only if we increase our knowledge of how exactly and under which circumstances languages may go beyond a presumably 'naturally given' pattern, that we can achieve a deeper understanding of how languages oscillate between being both a cultural and a biological behavior.

References

Beddor, P.S., Harnsberger, J.D. & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. In *Journal of Phonetics*, 30, 591-627.

Bladon, R.A.W., Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. In *Journal of Phonetics*, 4, 137-150.

Bombien, L., Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. In *Journal of the Acoustical Society of America*, 134 (1), 539-550.

Browman, C. (1994). Lip aperture and consonant releases. In Keating, P. (Ed.), *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, 331-353.

Browman, C., Goldstein, L. (1985). Dynamic modeling of phonetic structure. In Fromkin, V. (Ed.), *Phonetic Linguistics. Essays in Honor of Peter Ladefoged*. Orlando: Academic Press, 35-53.

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

Browman, C., Goldstein, L. (1995). Gestural syllable position effects in American English. In Bell-Berti, F., Raphael, L.J. (Eds.), *Producing Speech: Contemporary Issues.* Woodbury, NY: American Institute of Physics Press, 19-33.

Browman, C., Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. In *Bulletin de la Communication Parlée*, 5, 25-34.

Brunner, J., Geng, C., Sotiropoulou, S. & Gafos, A. (2014). Timing of German onset and word boundary clusters. In *Laboratory Phonology*, 5 (4), 403-454.

Butskhrikidze, M. (2002). *The Consonant Phonotactics of Georgian*. Utrecht: LOT Dissertation Series.

Chitoran, I., Goldstein, L. & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In Gussenhoven, C., Rietveld, T. & Warner, N. (Eds.), *Papers in Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 419-448.

Davidson, L. (2006). Phonotactics and articulatory coordination interact in phonology: evidence from non-native production. In *Cognitive Science*, 30 (5), 837-862.

de Jong, K. (2001). Rate-induced resyllabification revisited. In *Language and Speech*, 20, 197-216.

Dell, F., Elmedlaoui, M. (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer.

Evans, N., Levinson, S.C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. In *Behavioral and Brain Sciences*, 32, 429-492.

Farnetani, E., Recasens, D. (2010). Coarticulation and connected speech processes. In Hardcastle, W.J., Laver, J. & Gibbon, F.E. (Eds.), *The Handbook of Phonetic Sciences*, 2nd ed. Wiley-Blackwell, 316-352.

Fowler, C., Brancazio, L. (2000). Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel articulation. In *Language and Speech*, 43 (1), 1-41.

Fowler, C., Saltzman, E. (1993). Coordination and coarticulation in speech production. In *Language and Speech*, 36, 171-195.

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

Gafos, A., Goldstein, L. (2012). Articulatory representation and organization. In Cohn, A., Fougeron, C. & Huffman, M. (Eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Gafos, A., Hoole, P., Roon, K. & Zeroual, C. (2010). Variation in timing and phonological grammar in Moroccan Arabic clusters. In Fougeron, C., Kühnert, B., D'Imperio, M. & Vallée, N. (Eds.), *Laboratory Phonology 10. Variation, Detail and Representation*. Berlin: Mouton de Gruyter, 657-698.

Goldstein, L., Chitoran, I. & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVI International Congress of Phonetic Sciences, Saarbrücken*, 241-244.

Goldstein, L., Nam, H., Saltzman, E. & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In *Proceedings of the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers, Beijing, April 18-20, 2008*, CD-ROM.

Goldstein, L., Pouplier, M. (2014). The temporal organization of speech. In Goldrick, M., Ferreira, V. & Miozzo, M. (Eds.), *The Oxford Handbook of Language Production*. Oxford: Oxford University Press, 210-227.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E. & Byrd, D. (2007). Dynamic action units slip in speech production errors. In *Cognition*, 103 (3), 386-412.

Grimme, B., Fuchs, S., Perrier, P. & Schöner, G. (2011). Limb versus speech motor control: a conceptual review. In *Motor Control*, 15 (1), 5-33.

Haken, H., Kelso, J.A.S. & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. In *Biological Cybernetics* (51), 347-356.

Hermes, A., Mücke, D. & Grice, M. (2013). Gestural coordination of Italian word initial clusters - the case of 'impure s'. In *Phonology*, 30 (1), 1-25.

Hoole, P., Pouplier, M., Benus, S. & Bombien, L. (2013). Articulatory coordination in obstruent-sonorant clusters and syllabic consonants: data and modelling. In Spreafico, L., Vietti, A. (Eds.), *Rhotics. New Data and Perspectives*. Bolzano: Bolzano University Press, 81-97.

Iskarous, K., McDonough, J. & Whalen, D. (2012). A gestural account of the velar fricative in Navajo. In *Laboratory Phonology*, 3 (1), 195-210.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C., Salzman, E. & Whalen, D. (2013). The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. In *Journal of the Acoustical Society of America*, 134 (2), 1271-1282.

Kelso, J.A.S. (1986). Pattern formation in speech and limb movements involving many degrees of freedom. In Heuer, H., Fromm, C. (Eds.), *Generation and modulation of action patterns*. Berlin: Springer, 105-128.

Kelso, J.A.S., Saltzman, E.L. & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. In *Journal of Phonetics*, 14, 29-59.

Kisseberth, C. (2011). Conspiracies. In van Oostendorp, M., Ewen, C.J., Hume, E. & Rice, K. (Eds.), *The Blackwell Companion to Phonology*. Wiley-Blackwell, 1644-1665.

Kochetov, A., Pouplier, M. & Son, M. (2007). Cross-language differences in overlap and assimilation patterns in Korean and Russian. In *Proceedings of the XVI International Congress of Phonetic Sciences, Saarbrücken*, 1361-1364.

Krakow, R.A. (1999). Physiological organization of syllables: A review. In *Journal of Phonetics*, 27, 23-54.

Ladefoged, P. (1983a). The limits of biological explanations in phonetics. In *UCLA Working Papers in Phonetics*, 57, 1-10.

Ladefoged, P. (1983b). 'Out of chaos come order'; phsyical biological, and structural patterns in phonetics. In van den Broecke, M.P.R., Cohen, A. (Eds.), *Proceedings of the Xth International Congress of the Phonetic Sciences, Utrecht*. Dordrecht: Foris, 83-95.

Lindblom, B. (1983a). Can the models of evolutionary biology be applied to phonetic problems?. In van den Broecke, M., Cohen, A. (Eds.), *Proceedings of the Xth International Congress of the Phonetic Sciences, Utrecht.* Dordrecht: Foris, 67-81.

Lindblom, B. (1983b). Economy of speech gestures. In MacNeilage, P.F. (Ed.), *The Production of Speech.* New York: Springer, 217-246.

Lindblom, B., MacNeilage, P.F. & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In Butterworth, B., Comrie, B. & Dahl, B. (Eds.), *Explanations of Linguistic Universals*. The Hague: Mouton, 181-203.

Löfqvist, A., Gracco, V.L. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. In *Journal of the Acoustical Society of America*, 105 (3), 1864-1876.

MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. In *Behavioral and Brain Sciences*, 21, 499-546.

MacNeilage, P.F., Davis, B.L. (2000). On the origin of internal structure of words. In *Science*, 2888 (April), 527-531.

MacNeilage, P.F., Davis, B.L. (2001). Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. In *Current Opinion in Neurobiology*, 11, 696-700.

MacNeilage, P.F., Davis, B.L., Kinney, A. & Matyear, C. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. In *Child Development*, 71 (1), 153-163.

Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. In *Journal of the Acoustical Society of America*, 88 (3), 1286-1298.

Marin, S. (2011). *Organization of complex onsets in Romanian.* Paper presented at the Proceedings of the 9th International Seminar on Speech Production, Montréal, June 20-23, 2011.

Marin, S. (2013). Organization of Complex Onsets and Codas in Romanian: A gestural approach. In *Journal of Phonetics*, 41, 211-227.

Marin, S., Pouplier, M. (2010), Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. In *Motor Control*, 14 (3), 380-407.

Mooshammer, C., Hoole, P. & Geumann, A. (2007). Jaw and order. In *Language and Speech*, 50 (2), 145-176.

Nam, H. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In Cole, J., Hualde, J.I. (Eds.), *Papers in Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 483-506.

Nam, H., Goldstein, L. & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In Pellegrino, F., Marisco, E., Chitoran, I. & Coupé, C. (Eds.), *Approaches to Phonological complexity.* Berlin: Mouton de Gruyter, 299-328.

Nelson, W.L. (1983). Physical principles for economy of skilled movements. In *Biological Cybernetics*, 46, 135-147.

Öhman, S.E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. In *Journal of the Acoustical Society of America*, 39 (1), 151-168.

Öhman, S.E. (1967). Numerical model of coarticulation. In *Journal of the Acoustical Society of America*, 41, 310-320.

Ostry, D., Cooke, J.D. & Munhall, K. (1987). Velocity curves of human arm and speech movements. In *Experimental Brain Research*, 68, 37-46.

Pastätter, M., Pouplier, M. (2014). The temporal coordination of Polish onset and coda clusters containing sibilants. In *Proceedings of the 10th International Seminar on Speech Production, Cologne, Germany, 5-8 May 2014*, 312-315.

Pastätter, M., Pouplier, M. (2015). Onset-vowel timing as a function of coarticulation resistance: Evidence from articulatory data. *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*.

Peters, S., Kleber, F. (2014). Articulatory mechanisms underlying incremental compensatory vowel shortening in German. In *Proceedings of the 10th International Seminar on Speech Production, Cologne, Germany, 5-8 May 2014*, 316-319.

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. In *Journal of Phonetics*, 36, 114-140.

Pouplier, M. (2011). The atoms of phonological representations. In van Oostendorp, M., Ewen, C.J., Hume, E. & Rice, K. (Eds.), *The Blackwell Companion to Phonology*. Malden, MA: Wiley-Blackwell, 107-129.

Pouplier, M. (2012a). The gaits of speech: re-examining the role of articulatory effort in spoken language. In Solé, M.-J., Recasens, D. (Eds.), *Sound Change*. John Benjamins, 147-164.

Pouplier, M. (2012b). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In Fuchs, S., Weirich, M., Pape, D. & Perrier, P. (Eds.), *Speech Planning and Dynamics.* Berlin: Peter Lang, 63-96.

Pouplier, M., Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. In *Journal of Laboratory Phonology*, 2 (2), 243-273.

Recasens, D. (1999). Lingual coarticulation. In Hardcastle, W., Hewlett, N. (Eds.), *Coarticulation: theory, data, and techniques.* Cambridge: CUP, 80-104.

Recasens, D. (2002). An EMA study of VCV coarticulatory direction. In *Journal of the Acoustical Society of America*, 111, 2828-2841.

Recasens, D. (2012). A study of jaw coarticulatory resistance and aggressiveness for Catalan consonants and vowels. *Journal of the Acoustical Society of America*, 132 (1), 412-420.

Recasens, D., Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. In *Journal of the Acoustical Society of America*, 125 (4), 2288-2298.

Recasens, D., Fontdevila, J. & Pallarès, M.D. (1996). Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized /l/. In *Journal of Phonetics*, 24, 165-185.

Recasens, D., Pallarès, M.D. & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. In *Journal of the Acoustical Society of America*, 102 (1), 544-561.

Redford, M. (1999). *An articulatory basis for the syllable.* PhD dissertation, University of Texas, Austin.

Redford, M. (2001). Constrained emergence of universals and variation in syllable systems. In *Language and Speech*, 44 (1), 27-56.

Rochet-Capellan, A., Schwartz, J.-L. (2007). An articulatory basis for the labial-coronal effect: /pata/ seems a more stable articulatory pattern than /tapa/. In *Journal of the Acoustical Society of America*, 121 (6), 3740-3754.

Rochon, M. (2000). Optimality in complexity: The case of Polish consonant clusters. Berlin: Akademie Verlag.

Saltzman, E., Munhall, K.G. (1989). A dynamical approach to gestural patterning in speech production. In *Ecological Psychology*, 1, 333-382.

Stetson, R.H. (1951). *Motor phonetics: A study of speech movements in action*. Amsterdam: North-Holland.

Stone, M. (2010). Laboratory techniques for investigating speech articulation. In Hardcastle, W.J., Laver, J. & Gibbon, F.E. (Eds.), *The Handbook of Phonetic Sciences*, 2nd ed. Malden, MA: Wiley-Blackwell, 9-38.

Strogatz, S.H., Stewart, I. (1993). Coupled oscillators and biological synchronization. In *Scientific American*, Dec., 102-109.

Turvey, M.T. (1990). Coordination. In *American Psychologist*, 45 (8), 938-953.

Vallée, N., Rossato, S. & Rousset, I. (2009). Favoured syllabic patterns in the world's languages and sensorimotor constraints. In Pellegrino, F., Marisco, E., Chitoran, I. & Coupé, C. (Eds.), *Approaches to Phonological Complexity.* Berlin: Mouton de Gruyter, 111-139.

Warner, N., Brenner, D., Schertz, J., Carnie, A., Fisher, M. & Hammond, M. (2015). The aerodynamic puzzle of nasalized fricatives: Aerodynamic and perceptual evidence from Scottish Gaelic. In *Laboratory Phonology*, 6 (2), 197-241.

DOI: 10.17469/O2101AISV000004

LORENZO FILIPPONIO, SONIA CAZZORLA[1]

The vowels of Bari. A comparison between local dialect and regional Italian

In the mid-70s, the stressed vowel system of Barese displayed 7 phonemes, with an endangered opposition between mid-high and mid-low vowels, which occurred only in open syllable in non-metaphonetic contexts. The corresponding regional Italian already had 5 phonemes, mid-high and mid-low vowels being allophones of /e/ und /o/ respectively in open and closed syllables. According to our analysis, based on data collected in 2014 and grouped by gender, age and residence (inhabitants of Bari vs. migrants), the opposition is now completely disappeared. Moreover, innovative speakers show a generalized lowering (higher F1 values) of regional Italian vowels. Young females, the group supposed to be more sensitive to innovation, exhibit the same pattern even in Barese.

1. *Stressed vowels in the dialect and in the Regional Italian of Bari*

The Pugliese dialect spoken in Bari (henceforth Barese), capital city of Apulia (327.000 inhabitants), belongs to the south-eastern Italo-Romance varieties, whose stressed vowel system – based on the so-called common-Romance system – displays a series of processes affecting vowels in open syllables (Loporcaro, 2011b: 136) which differentiate the vowel quality in open and closed syllables, like raising of lower-mid vowels or diphthongization[2]: for instance, the outcomes of common-Romance /e/ in open and closed syllable are [aj] and [e] in the dialect of Bisceglie (45 km east from Bari): **cre$do* > [ˈkrajtə] '(I) believe' vs. **stel$la* > [ˈsted:ə] 'star' (see Cocola, 1925; De Gregorio, 1939); [aj] and [ɛ] in the dialect of Altamura respectively (50 km southeast from Bari, henceforth Altamurano): **ve$de* > [vajt] '(he/she) sees' vs. **len$gua* > [lɛɲ:] 'tongue' (see Loporcaro, 1988: 33-34; 2011b: 136). According to Valente (1975: 15-17), urban Barese shows no longer diphthongization in open syllables, the diphthongs (which are nowadays still attested in the hinterland) having been monophthongized (see Stehl, 1988: 703). The only diphthongs remaining are due to metaphony on lower-mid vowels, whereas the higher-mid vowels raise in the same context.

[1] This paper has been written jointly by the two authors. For academic purposes, LF bears responsibility for §§ 1, 3.1, 4; SC for §§ 2, 3.2. While remaining solely responsible for any weaknesses or inaccuracies to be found here, both authors would like to thank Michele Loporcaro, Stephan Schmid and two anonymous reviewers for their helpful comments.

[2] Depending on the dialect, these processes can be sensitive to sentence-phonetics or applied at word level with restructuring of the underlying representation (Loporcaro, 2011b: 136).

Table 1 - *Stressed vowels system of Barese (Valente, 1975: 16)*

	-V#	ī > i	ē,ĭ > e	ĕ > ɛ	ă > a	ŏ > ɔ	ō,ŭ > o	ū > u
σ[	-a, -e, -o	i	ɛ	e	a	o	ɔ	u
	-ī, -ŭ	i	i	i(ə)	a	we, e	u	u
σ]	-a, -e, -o	i	ɛ	ɛ	a	ɔ	ɔ	u
	-ī, -ŭ	i	i	i(ə)	a	wɛ, ɛ	u	u

Moreover, Valente indicates that the oppositions /ɛ/ ~ /e/ and /ɔ/ ~ /o/, which occur only in open syllable in non-metaphonetic contexts, are going to be neutralized. This is supposedly due to a restrictive interpretation of the Italian patterns by the upper class (1975: 16). In any case, the interplay between local dialect and Regional Italian (henceforth RIt) has an important role in the evolution of Barese, and consequently in the other Pugliese dialects, due to Bari being the epicenter of linguistic innovations for Northern and Central Apulia (Stehl, 1988: 703).

According to Canepari (1983[2]: 173), Pugliese RIt has 5 vocalic phonemes: vowel trapezium is characterized by two stable high vowels [i] and [u] and three floating areas, two in the middle, front and back, on the high-low axis, and one central low, where fronted [æ] is the realization of /a/ in open syllable. This sound is according to Valente absent in the city of Bari (see Table 1 above), but quite common in the surrounding areas (1975: 42). As in the corresponding dialects, the different pronunciation (high and low) of RIt stressed mid vowels depends on syllable structure, as shown by Loporcaro (1988: 205-207) for the RIt of Altamura. The dialectal pattern ([+tense] stressed vowel in open syllable, [-tense] in closed syllable) has a clear substrate influence on RIt. However, the growing influence of RIt among younger speakers has caused a change in the vowel system of Altamurano: the set of the speakers of the young generation had yet at the end of the 80s only five phonemes /ʊ o a e ɪ/, with [ɔ] and [ɛ] remaining only as allophones of /o/ and /e/ in closed syllables.

Thus, the process shown for Barese in the 70s was completed in Altamurano about ten years later.

The goal of this study is to give an account of Barese stressed vowel system 40 years after Valente's description, in order to verify whether the tendencies presented in his work (without considering here his typical mid-70s sociolinguistic statement) led to a complete neutralization of the unstable oppositions, as in Altamurano. Here, we carried out a phonetic analysis based on new data, being aware that while the AVIS and CLIPS corpora provide data for the experimental analysis of many Italo-Romance dialects (see Clemente et al., 2006), there is still a significant lack of experimental research on Barese.

2. *Materials and methods*

We have chosen 2 groups, each composed of 4 Barese native speakers belonging to the working class. The first group is formed by speakers who have never left Bari (henceforth: resident group), the second one is formed by migrants who left Bari for Germany and the German-speaking part of Switzerland in the late 80s/early 90s. The two groups

have a homogeneous composition, both represented by two adult speakers and two young speakers, male and female respectively[3]. This configuration enables us to sort the speakers by residence (residents vs. migrants) and by age (adult vs. young), whereas the results are always kept separated by gender because of the differences in phonation between females and males due to physiological reasons (Peterson, Barney, 1952; Ferrero et al., 1995). The low number of informants makes this analysis strictly qualitative. A broader empirical basis is a purpose for further research.

We have recorded two different kinds of corpora: For RIt, each informant had to read a list of 70 words, 10 per each stressed vowel of Standard Italian (/i e ɛ a ɔ o u/), which allows lower-mid and higher-mid stressed vowels both in open and in closed syllable (e.g. /ˈbɛne/ 'well, good (adverb)', /ˈsɛtːe/ 'seven', /ˈrete/ 'net', /ˈstelːa/ 'star'); for Barese, each informant had to translate from Italian into dialect a list of 80 words containing all Protoromance stressed vowels in open and closed syllable in different contexts (paroxyones, proparoxytones, metaphonetic, non-metaphonetic, etc.).

Digital recordings were made using a Zoom H4N portable recorder. The instrumental analysis was carried out using the Praat software: all the stressed vowels were selected and their F1 and F2 values – measured on the mid-point – were extracted with the script *Vokalanalyse* by means of a textgrid file. The results are displayed by ellipse plots created by the script *Ellipsenplotter*.

3. *Results*

3.1 Preliminary remarks about Barese

For the sake of simplicity, we have left out of our experimental analysis of Barese data all the stressed vowels affected by metaphony and all proparoxytones, which are briefly discussed in the next paragraphs. For the experimental analysis, 52 words (at least 6 per vowel) remained.

3.1.1 Metaphony

Metaphony due to final -ī and -ŭ is no longer an allophonic rule, since every final unstressed vowel has been weakened to *-ə* (see below §3.1.2). Thus, stem alternations in nouns deriving from Latin third class (**-e/*-i*, sg./plur.: non-metaphonetic (sg.)/metaph. (pl.) context) or in adjectival paradigms (**-e/*-i* and **-u/*-i/*-a/*-e*, m.sg./m.pl./f.sg./f. pl.: metaphonetic (m.)/non-metaph. (f.) context) can only be analyzed synchronically in terms of lexicalized allomorphy. Some uncertainties are shown by resident speakers: for instance, three of them seem to be unable to process the allomorphy for gender distinction in the adjectival paradigm of *CŎCT-U/-I/-A/-E ('cooked m.sg./m.pl./f.sg./f. pl.'), accepting [ˈkɔtːə] as generalized form, perhaps under the influence of Italian *còtto/-i/-a/-e*. The adult male speaker, which can be considered the most conservative of the

[3] In the resident group, males are 55 and 30 years old and females 54 and 25; in the migrants group, males are 49 and 28 years old and females 46 and 19 respectively: that means that the migrant young speakers have been partially or totally schooled in German.

group according to age and gender criteria (see Labov, 1990: 205), is the only able to provide the expected outcomes [ˈkwɛtːə] for m.sg. and m.pl. (metaphonetic context) and [ˈkɔtːə] for f.sg. and f.pl. (non-metaphonetic context, see Table 1 above). In the migrants group, the young female speaker, which should be the most innovative by the same parameters, overgeneralizes the diphthong (which means that the diphthong is still present in her variety)[4]. It is noteworthy that migrant speakers, with the partial exception of the young female, display always the expected dialectal forms[5]: living abroad has probably preserved their dialect from the innovations due to the daily contact with the RIt spoken nowadays in Bari (see §3.2.2).

3.1.2 Proparoxytones (and rhythm pattern)

Our results confirm that stressed vowels in proparoxytones are treated in the same way as those in closed syllable, as pointed out by Carosella (2005: 69-79) and Loporcaro (2011b: 136 [note 44]) for southern Italo-Romance dialects: for instance, DOMĪNĪCA > [dəˈmɛnəkə] 'Sunday' (It. *doménica*); PĔCŎRA > [ˈpɛkərə]/[ˈpɛgərə] 'sheep' (It. *pècora*). This could be interpreted as a collateral effect of a compensative rhythm pattern (in sense of Bertinetto, Bertini, 2010; see Filipponio, 2012: 298-307): the stressed syllable gets stronger and attracts the weakened unstressed syllables causing by compression the shortening of the stressed vowel and a close contact with the following consonant, which skips to the coda position of the stressed syllable[6]. One can also observe that both the reduction of the unstressed vowels (which is common to the whole Mid-Southern Italoromance area influenced by Naples, with some lateral remnants of older conditions) and the stressed vowel quality distinction in open and closed syllables are conform to the same rhythm pattern. Thus, Barese, like other modern Southern Italo-Romance dialects (e.g. Neapolitan), should be classified as a variety with an at least partially compensated rhythmical structure.

---

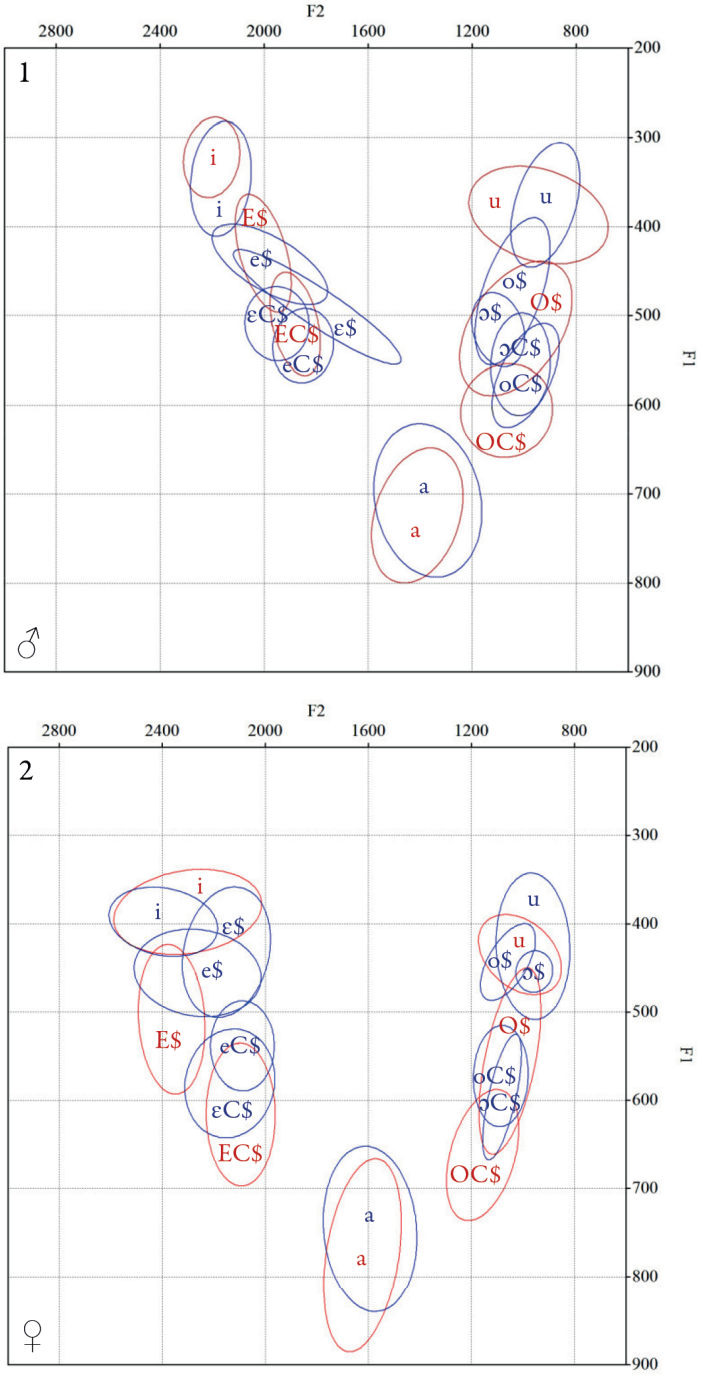
[4] According to Valente (1975: 17), in Barese [we] and [wɛ] are reduced to [e], [ɛ] except after [k] and labial consonants, whereas the stronger tendency is to reduce the metaphonetic outcomes of Protoromance *ò o u* to [u] (testified in our data by the gender distinction [ˈbuːnə] (m.)/[ˈboːnə] (f.) displayed by migrant speakers as result of the paradigm *BŎN-U/-I/-A/-E 'good'; even in this case, resident speakers provided incoherent answers, see below Note 5).

[5] Other examples: *PĬLU 'hair', resident female speakers [ˈpeːlə] (It. *pélo*), all other speakers [ˈpiːlə] (expected metaphonetic form); *TĒCTU 'roof', resident speakers (with exception of the young male) [ˈtɛtːə] (It. *tétto*, RIt. *tètto*, see above §1), all other speakers [ˈtitːə] (expected metaphonetic form).

[6] This explanation can be considered as an application at the word-level of Vennemann's law of *attractiveness of rhyme* (1988: 61). The same law is maybe responsible for the coda position of plosives in *muta cum liquida* clusters, testified by the fact that stressed vowels before this cluster display the same quality of those in closed syllable, like in the dialect of Bisceglie (see above §1) *PA$LU > [ˈpɔːlə] 'pole, post' vs. *PLAT$JA > [ˈcatːsə] 'square' and LA$TRO > [ˈlatrə] 'thief'. Given this picture, Loporcaro (2011a: 104) maintains that south-eastern Italo-Romance varieties have kept the Late Latin heterosyllabicity of *muta cum liquida* clusters (see Filipponio, 2014 for a discussion) and that proparoxytones in this dialects never underwent open syllable lengthening (2011b: 136 [note 44]), typical for the Romance central area. Theoretically, the closing of stressed syllables before *muta cum liquida* clusters and in proparoxytones as well as the changes of stressed vowels quality could have took place *after* a more recent change in rhythm pattern, which could have become again similar to the Late Latin one: but there are no proofs for such an explanation, and the continuity hypothesis remains more economical.

3.2 Barese vs. RIt

Figures 1-2 - *Ellipses of Barese (blue) vs. RIt (red) stressed vowels formant values of all informants: males (above) and females (below)*



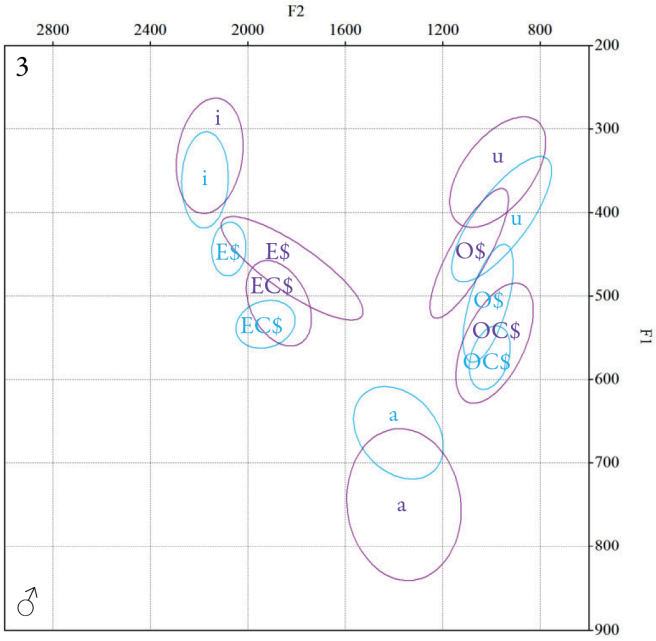
The first diagrams (Figs. 1-2) show the ellipse plots of the stressed vowels pronounced by male and female speakers both in Barese and in RIt. The formant values of Barese outcomes of Protoromance mid vowels are kept separate according to the syllable structure[7]. It is clear that the oppositions /ɛ/ ~ /e/ and /ɔ/ ~ /o/ found by Valente (even though endangered) are generally suppressed. The syllable structure is the only parameter differentiating the low/high degree of mid stressed vowels, as shown by Loporcaro for the youngest generation of Altamurano speakers in the late 80s. This allows us to keep the Protoromance ancestors of mid vowels unspecified in the next figures, and to relate only to the syllable structure, as shown in Figs. 1-2 for RIt[8].

There are also two remarkable differences between male and female speakers. Females overlap the higher-mid and high Barese vowels and lower the articulation of RIt mid and low vowels in comparison to Barese, as displayed by higher F1 values. Males show a similar tendency only in back vowels, and only in a slighter way.

Let us consider next the data grouped by age and residence. Each parameter is presented by two pairs of graphs, the first ones showing the Barese data, and the second ones the RIt data each for male and female speakers.

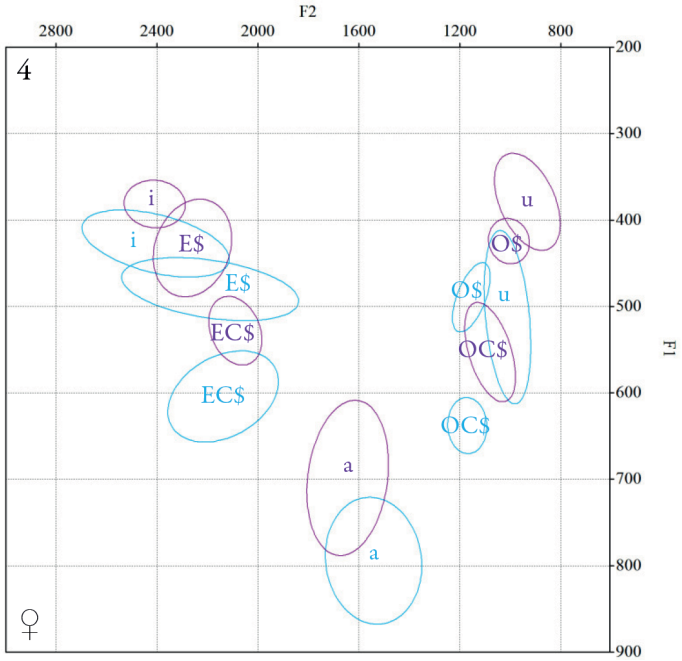
3.2.1 Age

Figures 3-4 - *Formant values of young (turquoise) vs. adult (purple) Barese speakers: males and females (next page)*

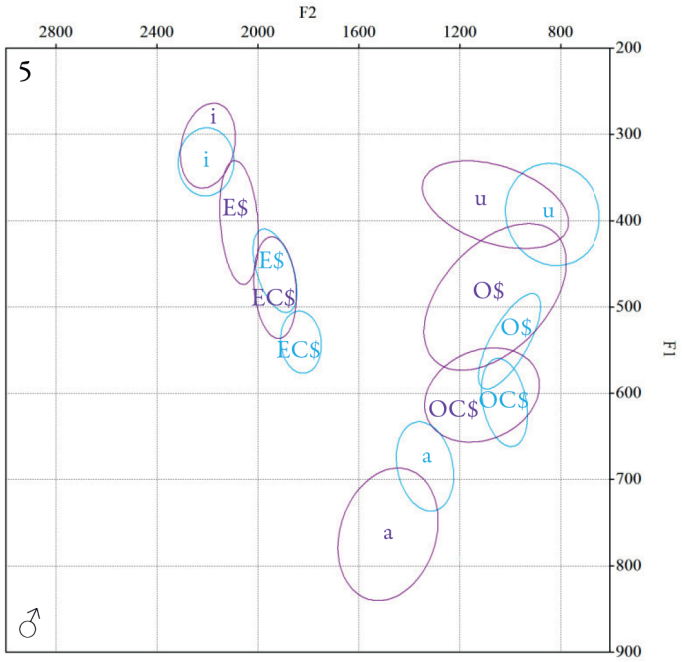


[7] e$/o$ = Protoromance higher-mid vowel in open syllable; eC$/oC$ = Protoromance higher-mid vowel in closed syllable; ɛ$/ɔ$ = Protoromance lower-mid vowel in open syllable; ɛC$/ɔC$ = Protoromance lower-mid vowel in closed syllable.

[8] E$/O$ = mid vowel in open syllable; EC$/OC$ = mid vowel in closed syllable.

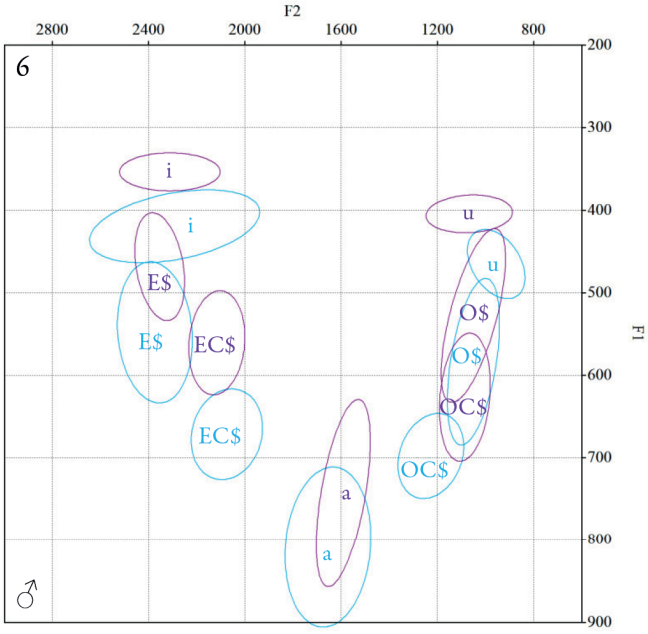


Figures 5-6 - *The same as Figures 3-4 (see previous page) for RIt speakers*



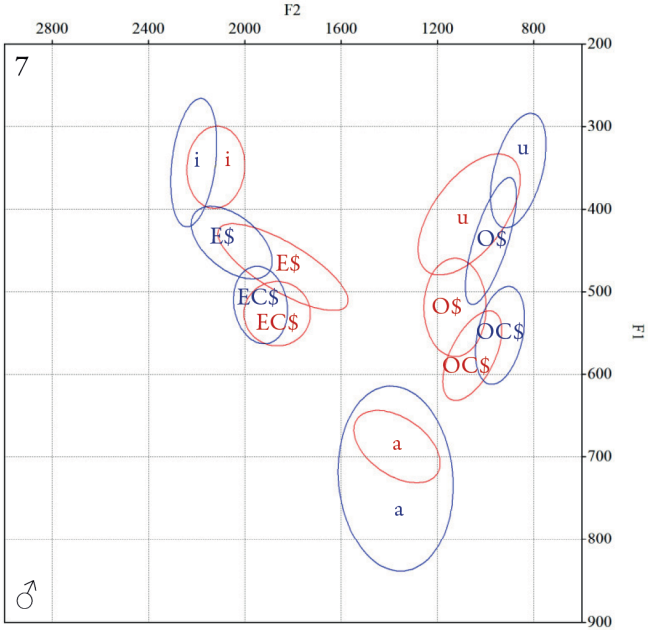
Young female speakers show higher F1 values, i.e. a generalized lowering of all vowels both in Barese and in RIt. Young male speakers display the same pattern only for

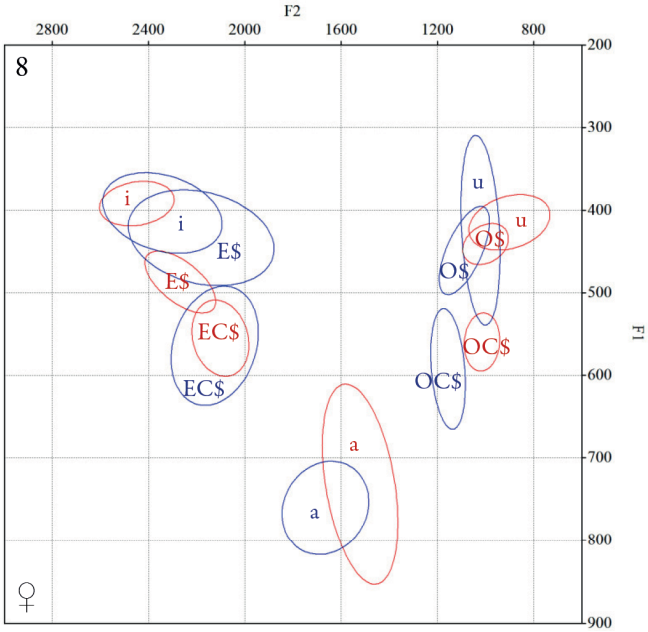
Barese [u] and [o] and for RIt [e], [ɛ] and [o], whereas F1 values of [a] are lower than those of adult speakers.



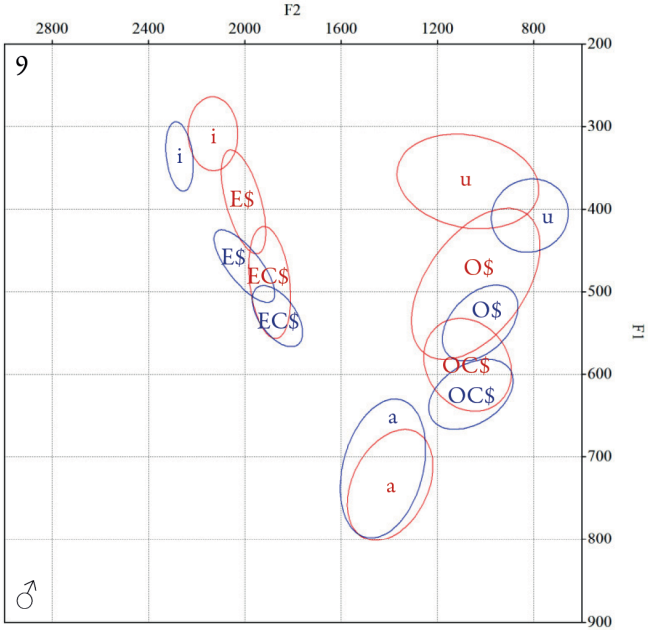
3.2.2 Residence

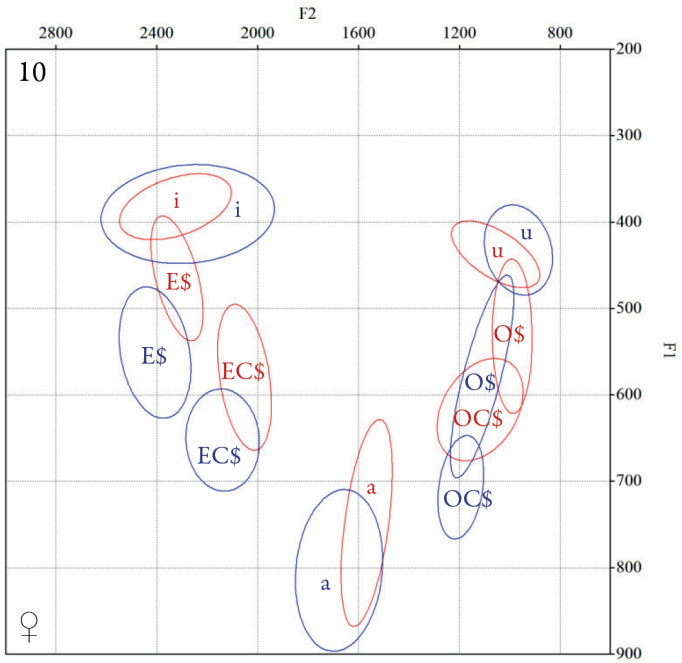
Figures 7-8 - *Formant values of resident (red) vs. migrant (blue) Barese speakers: males and females (next page)*





Figures 9-10 - *The same as Figures 7-8 (see previous page) for RIt speakers*





Place of residence seems to be a stronger factor in RIt than in Barese. In the latter case (Figs. 7-8), male migrants display a centralized trapezium in comparison to residents, with remarkable differences of the F1 values of [e], [u] and [o], whereas females show a more homogeneous picture, except for the absence of overlap between [i] and [e] of migrant speakers. In the case of RIt (Figs. 9-10), all migrant speakers display generalized higher F1 values of mid vowels. Furthermore, females of this group show higher F2 values of front mid vowels and of [a].

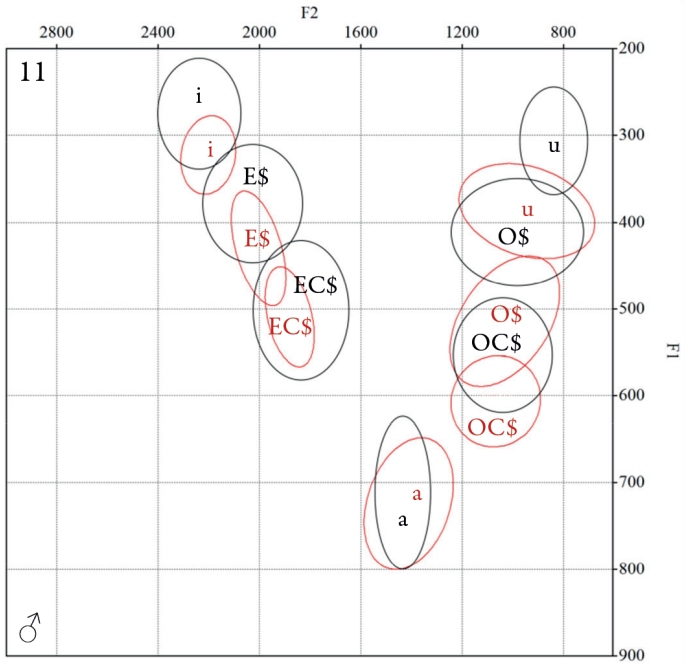
4. *Conclusions*

As we have seen in §3.2, the deletion process of the oppositions /ɛ/ ~ /e/ and /ɔ/ ~ /o/ is completed. Young and adult Barese speakers display a system which is similar to the one of young Altamurano speakers in the 80s. Given the asymmetric structure of Barese stressed vowels system in the 70s (see Table 1), there would be no need to ascribe this process to the influence of RIt. However, at least a conspiracy of internal (structural) and external (growing influence of RIt, which has strengthened the tendency) factors seems to be plausible.

A slight fronting of [a] appears in migrant females RIt (see §3.2.2, Figure 10) as well as in adult males RIt (Figure 5) and in migrant females Barese (Figure 8). This haphazard situation could be due to the low number of informants: nonetheless, dividing Barese F2 values of [a] into two sets based on the syllable structure leads to discover an unexpected (given the data in Table 1) palatalization in open syllable common to all migrant speakers. The phenomenon is more relevant in male

(medians of F2 values of /a/: 1632Hz in open syllable, 1174Hz in closed syllable) than in female speakers (F2: 1845Hz in open syllable, 1583 in closed syllable). By crossing the blue ellipses displayed in Figures 7 and 9 (males) and 8 and 10 (females) one could notice that Barese and RIt stressed vowel system of migrant speakers are generally different: this may be due to the fact that speakers living abroad are less exposed to regional varieties of their native tongue which may cause interferences with their dialect, as proposed for the explanation of the stability of allomorphy in noun and adjectival paradigms (see above §3.1.1).

Figure 11 - *Formant values of RIt male speakers from Bari (red) compared with Standard Italian male speakers (black) analyzed by Albano Leoni and Maturi*



Concerning other sociophonetic factors, we have seen that young females tend to lower the articulation of all stressed vowels (higher F1 values, see Figures 4 and 6) both in Barese and RIt, and that females have generally lower vowels in RIt than in Barese (Figure 2). If we consider [+female] and [+young] as sociophonetic features which are more sensitive to innovation, lowering could be seen as an innovative articulatory pattern. By taking into account all informants, RIt has generally lower vowels than Barese (see Figures 1-2)[9] and even male speakers show RIt lower vowels if compared with the masculine formant values based on the analysis of 12 regional news anchors from Northern, Central and Southern Italy provided by Albano

[9] Maybe a confirmation of our sociolinguistic hypothesis, if we assume that innovation is actually produced in RIt more than in Barese.

Leoni and Maturi (2002[3]: 106), as shown in Figure 11. The lowering of RIt vowels by migrant speakers (in a higher fashion than residents, see Figures 7-10) could suggest that this tendency is active since at least 30 years ago[10].

In summary, our data show that the structural changes in progress 40 years ago are now completed. Moreover, the innovative speakers display now a new articulatory pattern in RIt that is affecting the Barese vowels of younger female speakers.

Bibliography

Albano Leoni, F., Maturi, P. (2002[3]). *Manuale di fonetica*. Roma: Carocci.

Bertinetto, P.M., Bertini, C. (2010). Towards a unified predictive model of natural language rhythm. In Russo, M. (Ed.), *Prosodic Universals. Comparative studies in rhythmic modeling and rhythm typology*. Roma: Aracne, 43-77.

Canepari, L. (1983[2]). *Italiano standard e pronunce regionali*. Padova: CLEUP.

Carosella, M. (2005). *Sistemi vocalici tonici nell'area garganica settentrionale fra tensioni diatopiche e dinamiche variazionali*. Roma: Edizioni Nuova Cultura.

Clemente, G., Savy, R. & Calamai, S. (2006). Sistemi vocalici in diatopia. In Savy, R., Crocco, C. (Eds.). *Analisi prosodica. Teorie, modelli e sistemi di annotazione. Atti del II Congresso Nazionale AISV (Salerno, 30 novembre - 2 dicembre 2005)*. Torriana: EDK, 439-460.

Cocola, F. (1925). *Vocabolario dialettale biscegliese-italiano*. Bisceglie: Carmastro.

De Gregorio, I. (1939). Contributo alla conoscenza del dialetto di Bisceglie (Bari). In *L'Italia Dialettale*, 15, 31-52.

Ferrero, F., Magno Caldognetto, E. & Cosi, P. (1995). Le vocali al femminile. In Marcato, G. (Ed.), *Donna e linguaggio*. Padova: CLEUP, 413-436.

Filipponio, L. (2012). *La struttura di parola dei dialetti della valle del Reno. Profilo storico e analisi sperimentale*. Sala Bolognese: Forni.

Filipponio, L. (2014). Il dibattito sull'eterosillabismo di alcuni nessi consonantici in latino tardo e le sue implicazioni strutturali. In Molinelli, P., Cuzzolin, P. & Fedriani, C. (Eds.), *Latin Vulgaire – Latin Tardif X. Actes du Xe colloque international sur le latin vulgaire et tardif (Bergamo, 5-9 settembre 2012)*, vol. I. Bergamo: Sestante Edizioni, 45-62.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. In *Language Variation and Change*, 2, 205-254.

Loporcaro, M. (1988). *Grammatica storica del dialetto di Altamura*. Pisa: Giardini.

Loporcaro, M. (2011a). Syllable, segment and prosody. In Maiden, M., Smith, J.C. & Ledgeway, A. (Eds.), *The Cambridge History of Romance Languages, vol. 1: Structures*. Cambridge: CUP, 51-108.

Loporcaro, M. (2011b). Phonological processes. In Maiden, M., Smith, J.C. & Ledgeway, A. (Eds.), *The Cambridge History of Romance Languages, vol. 1: Structures*. Cambridge: CUP, 109-154.

[10] And/or perhaps not limited to RIt of Bari: in order to verify that, further investigations about other RIt varieties are obviously needed.

Peterson, G.E., Barney, H.L. (1952). Control methods used in a study of the vowels. In *Journal of the Acoustical Society of America*, 24 (2), 175-184.

Stehl, Th. (1988). Italienisch: Areallinguistik XI. Apulien und Salento. In Holtus, G., Metzeltin, M. & Schmitt, Ch. (Eds.), *Lexikon der Romanistischen Linguistik. Band IV: Italienisch, Korsisch, Sardisch*. Tübingen: Niemeyer, 695-716.

Valente, V. (1975). *Profilo dei dialetti italiani. Puglia*. Pisa: Pacini.

Vennemann, Th. (1988). *Preference laws for syllable structure and the explanation of sound change*. Berlin/New York/Amsterdam: Mouton de Gruyter.

DOI: 10.17469/O2101AISV000005

CINZIA AVESANI, VINCENZO GALATÀ, MARIO VAYRA, CATHERINE T. BEST, BRUNO DI BIASE, OTTAVIA TORDINI, GRAZIANO TISATO

Italian roots in Australian soil: coronal obstruents in native dialect speech of Italian-Australians from two areas of Veneto[1]

We will discuss the maintenance of the heritage dialect coronal fricatives in the speech of Italian-Australian trilinguals (dialect/Italian/English) originating from North Veneto, Italy, as compared to the variability found in the productions of comparable Italian-Australian trilinguals originating from Central Veneto. Results on coronal fricatives' distribution based on narrow phonetic trancriptions and on their acoustic characteristics based on spectral moments analysis show that the immigrants have generally maintained the fine-grained features of their dialect. After more than five decades of residence in Australia, traces of interlinguistic influences exerted by L3-English are evident in one speaker only. We consider both internal (linguistic) and external (sociolinguistic) factors for this difference in maintenance of first language speech features between immigrants from two geographical areas of the same province of origin in Italy (Veneto).

1. *Introduction*

We live in an age of globalization that is seeing movements of large groups of people across vast distances. These immigrants must adapt to their new home across multiple dimensions. As they do, their cultural and linguistic origins take root in new soil, providing a foundation for their transplanted lives. Immigrant groups moving to a new country with a different dominant language tend to maintain their heritage language(s) as part of their cultural identity, while at the same time they must acquire and use the language of the host country for everyday communication. As bilingual or multilingual speakers, they hold a complex repertoire of linguistic forms and structures at their disposal, including the phonemes and pronunciation features of each of their languages. They draw on all elements of their linguistic repertoires in a selective way, as dictated by their language experience, conversational intentions, social activities and by the needs of the speech participants in specific conversational contexts. Our interest here lies in maintenance of the phonological and phonetic details of the specific varieties of language spoken by immigrants from a given region. Specifically, we address the extent to which fine-grained details of

[1] Authorship note: The project was designed and conducted by the first five authors. Tordini completed the transcription. Tisato provided tools for recording and access to materials in the *Multimedia Atlas of Veneto Dialects*.

their speech in their mother tongue may continue to provide unique markers of their linguistic origins.

In the linguistic contact situation typical of immigration settings, this is an important but little-studied question: How do people negotiate between their home language(s) and their new language even at the level of the fine-grained features of their speech in each? Do they maintain the detailed speech characteristics in their native dialect even across decades of speaking and hearing the language of their adoptive country? Or do those heritage language phonetic features dissolve over time as a result of living in the new setting, far away from the heritage language community in the old country? And how might the answers to these questions depend on use of the heritage language(s) and opportunities for speaking with other immigrants from their home region, and/or on other sociolinguistic factors?

Studies on bilingualism and second language acquisition indicate that the speaker's first language (L1) and second language (L2) can interact at the phonetic and phonological levels in several possible ways. One well documented pattern is that the L1 influences production of L2 phonemes and contrasts, consistent with the classic notion that the L1 acts as a sieve or filter on spoken production of later-learned languages (Trubeckoj, 1971, orig. ed. 1938). A considerable body of work has shown that the first acquired language (L1) influences the phonetic production and perception of a later acquired language even after many years of continued use of the L2 (e.g. Flege, 1991; Sebastián-Gallés, Soto-Faraco, 1999), and that the extent of such influence depends on the age of acquisition of L2 (e.g. Flege, Munro & MacKay, 1995). A second type of observed effect, however, is a bidirectional linguistic influence: in addition to interference from L1 to L2, learning an L2 can eventually induce changes in the production of L1. Such a bidirectional interaction can lead to the production of phones that differ from the native productions of both languages, as has been shown by a series of studies on production of stop consonant voicing as indexed by Voice Onset Time (e.g. Flege, Eefting, 1987; Sancier, Fowler, 1997; Antoniou, Best, Tyler & Kroos, 2011), and on vowel production (Chang, 2012). A third attested pattern of inter-language influence is that, independently of age of acquisition, if an L2 becomes the dominant language it may be "freed" from at least some of the effects exerted by L1 and instead exert a unidirectional influence on production of the non-dominant L1 (Flege, Mackay & Piske, 2002) and also on perception (Antoniou, Best, Tyler & Kroos, 2011; Antoniou, Tyler & Best, 2012). The final possibility is that no interaction occurs between the two languages, and that the productions of bilingual speakers are equivalent to those of monolingual speakers of each language, at least on certain specific measures of phonetic performance (e.g., Antoniou, Best, Tyler & Kroos, 2010).

Interlinguistic influences in speech production can also be modulated according to the sociolinguistic factors that determine the use of a given language. As well, they can be conditioned by the internal dynamics of the linguistic repertoires[2] of the

[2] According to the sociolinguistic literature, "repertoire" refers to the set of language varieties proper of a given linguistic community. The varieties that form a "linguistic repertoire" can belong to the same

speakers. A case in point is represented by bilingual speakers who learned as children an L2 that co-existed in a diglossic relation with the L1 in their community, and then later, as adults, learned a third language. Examples of such a case are those Italians who left their country in the post-World War 2 years (mid 40s-early 60s). According to the 1951 census, 65% of Italian population spoke only dialect and 35% also spoke Italian; ten years later, in the 1961 census, these figures had changed respectively to 43% and 57%.[3] Migrants who left Italy in those years had the local dialect of their village/town acquired from birth as their L1; the variety of Standard Italian (SI) of their territory, if spoken at all, was their "early" L2. It was typically learned in primary schools from about age 6 or, more infrequently, learned before then, at home, after the dialect. The local dialect was the language of everyday communication, exclusively oral, while Italian was the high variety, used for writing and, if spoken at all (see above figures), used in formal situations. Considering that dialect and Italian are structurally related to each other and yet are also perceived as clearly delimited from each other, they qualify as co-existing varieties standing in a diglossic relation in the linguistic repertoire of those speakers (type A repertoires; see Auer, 2005)[4]. The official language of the country where they moved and established was their "late" L3.

Complying with the conventions of Italian language research, we use the term "dialect" to indicate independent language systems and not geographical or social varieties of the national language. Italian (Italo-Romance) dialects refer to the local continuations of the Latin languages spoken across the Italian peninsula ever since the loss of Latin; even though they display different degrees of structural distance and mutual intelligibility, they are all sister languages of Italian, as they evolved in parallel with the Florentine dialect from which SI developed (e.g., Maiden, Parry, 1997; Cerruti, 2011; Dal Negro, Vietti, 2011).

In this paper we address the question of maintenance, attrition/drift or convergence of specific L1 phonetic features in two subgroups of Italian speakers who migrated to the greater Sydney area of Australia in the mid-1900s. Immigration from Italy to Australia started well before this time[5], but it was after the Second

language or to different languages (monolingual, bilingual or multilingual repertoires).

[3] Note however that these percentages are indirect estimates based on the data on schooling. People who did not attend the elementary schools were considered dialect monolinguals. Reliable statistic data are available only from 1974, when polls reported the percentage of people who declared to speak dialect, Italian or both and in which circumstances (cf. De Mauro, 1972 - ed. orig. 1963; D'Agostino, 2007).

[4] The contact and dynamics between the different Italian dialects and Italian as the base language in Italy have been extensively studied by sociolinguists, who distinguish among dialectalization of Italian, italianization of dialect, koineisation (dialect "mixing" that gives rise to a new variety), hybridisation, and discuss the different repertoires that emerge in such a complex dynamics (Berruto, 1989; 2005; Cerruti, Regis, 2011; see also Samarin, 1971).

[5] According to Campolo (2009), the first Italian immigrants to Australia, in 1770, were two sailors of Italian descent and an Italian convict aboard the *Endeavour*, Captain Cook's ship. Later, others Italian settlers followed as free citizens. The first large wave of immigration started in the second half of the

World War that mass immigration waves from different regions in Italy[6] first took place, induced by the difficult economic and social conditions in Italy and favored by good diplomatic and commercial relations between Australia and Italy[7] (e.g. Bettoni, Rubino, 1996; Campolo, 2009; Gallina, 2011; Rubino, 2014). We specifically focus on this group of first generation emigrants as they are aged (they show the highest mean age for resident groups born outside Australia) but still reasonably accessible: they can offer a unique window into mechanisms of language change in general and speech change in particular, *within* a given generation.

Almost all Italians in the first large wave of migration to Australia were late adolescents or young adults of the same generation at the moment of their departure toward the host country. Moreover, most were sequential dialect-Italian bilinguals. In the new homeland they came into contact with a third new language, which they needed to learn for their survival and social life in Australia. English became their L3, most often learned spontaneously after their arrival in Australia. There, they came into contact with other Italian language varieties, too, such as dialects from different regions in Italy that are structurally distant enough from their own dialect as to be mutually unintelligible, as well as the varieties of SI spoken in those regions that carried different phonetic, phonological and morphosyntactic features from those of their own variety of SI. Even more, speakers of a local variety of a given dialect could easily come into contact with speakers of other local varieties of the same dialectal system (for example, emigrants from North Veneto and from Central Veneto). Therefore, once in Australia, they found themselves in contact not only with English as a new language, but also with many varieties of Italian languages (local dialects and different forms of SI) spoken by the other Italian immigrants who formed a vast and variegated community in their adoptive country. Therefore, their multilingual repertoire includes the local dialect, the variety of Italian spoken by their community in Italy, and Australian English.

Given such a complex contact linguistic situation, and in light of the dynamics of dialect-Standard convergence of Italian repertoires in Italy (Auer 2005; 2011; Berruto, 2005; Cerruti, Regis, 2011) we can sketch some possible scenarios of interactions among the phonological and phonetic aspects of the linguistic repertoires available to those Italian speakers. These follow from the four types of interlanguage influences summarised above for bilinguals and L2 learners:

1. *Conservation of the phonetic/phonological features* of the linguistic heritage: immigrant groups tend to maintain the features of their linguistic heritage, crystalizing the state of the languages at the time they migrated and conserving archa-

1800s. At the end of that century, the first group of Italian settlers established themselves in North Queensland.

[6] Primarily from Sicily, Calabria, Abruzzo, Campania, Veneto, Friuli.

[7] In 1951 a bilateral agreement for assisted immigration was signed between Australia and Italy, based on which 20,000 immigrants were allowed to relocate to the country each year. The agreement was suspended in 1952 but it was resumed in 1954 (for details, see Campolo, 2009). From 1947 to 1961 more than 200,000 Italians reached Australia and settled there, contributing to more than 20% of the total immigration in that time span.

ic features more than the language(s) of the homeland[8] (Bartoli, 1925; 1945). Neither dialect nor Italian drift toward the other, as in the situation of diglossia (if any) they experienced in the late 40s/early 50s in Italy before leaving for Australia. In addition, English as L3 will not exert any influence on dialect or Italian.

2. *Interlanguage interaction*: the phonetic/phonological features of the heritage languages are subject to a bidirectional influence with respect to one another and with respect to English. This scenario opens up to four further possibilities:
   a. in a sociolinguistic situation in which speakers of different dialects need to communicate to one another within the Italian community, the structural distance of the home dialects can impede communication in their own L1. Italian becomes the vehicular language among the different groups of immigrants, and their L1-dialect can drift towards L2-Italian in a process of vertical advergence (Auer, 2005; Berruto, 2005; Cerruti, Regis, 2011), in which the dialect looses (some of) its features and becomes Italianized. An alternative is that a dialect-Standard continuum is formed (cf. type C diaglossia, Auer 2005) and that a new Australian version of Standard Italian will develop that takes on the typical phonetic features of the different dialectal substrata of the local Italian community.
   b. A related possibility is that speakers of a local dialect of a given region in Italy, once coming into contact with other speakers from the same region, may loose the most specific features of their local dialects converging toward a regional dialectal koiné.
   c. A third possibility is that the local dialect as L1 and Italian as L2 are kept in an attenuated form of diglossia (Auer, 2005) in which both varieties are structurally and attitudinally kept apart and used in different situations, but at the same time they maintain their own prestige (the former as a marker of regional identity, the latter as the official language of the Italian community). With respect to L3-English then, we can ask: (i) whether dialect as L1 will show more resistance to attrition than Italian as L2; (ii) whether both L1 and L2 drift towards English; and (iii) if they do, whether they show the same type of drift.
   d. Finally, English may become the dominant L3 after years of extensive use, exerting a strong influence on both dialect and Italian, while at the same time its production remains unaffected by the two heritage languages.

[8] According to the linguistic norms of Spatial Linguistics, a colony or "area seriore" is the area where a given linguistic variety arrived later with respect to the area to which it is traditionally connected. It tends to conserve archaic features more than the "area anteriore", i.e. the homeland (Bartoli, 1925; 1945).

2. *Focusing on Veneto immigrants and their dialect*

We explore the preceding possibilities by focusing on Italian immigrants to Australia who originated from Veneto, a region in northeast Italy. The dialect spoken in that region, named Veneto after the region, is better described as a macro-dialect (a dia-system according to Trumper, 1977) showing intra-dialect differences among its five subsystems (Mioni, Trumper, 1977; Zamboni, 1974; 1988). We selected four speakers originating from Northeastern Veneto (centered on the towns of Belluno, Treviso, Feltre) and Central Veneto (centered on the towns of Padova, Vicenza, Rovigo) subsystems. From now on we will refer to these subsystems respectively as NeVen and CVen. Of our speakers, two originate from the area centered on Belluno (BL) and two from the area centered on Rovigo (RO).

The two systems have been selected because they show interesting differences in the phonetic properties of the class of coronal obstruent consonants. These consonants are especially useful for examining interlanguage effects in immigrant groups, as they show a wide range of variation in fine-grained details across languages and regional accents and, relatedly, tend to undergo phonetic shifts over time and distance (language change).

According to Zamboni (1974; 1988), the NeVen system presents the following class of coronal obstruents: /t, d[9], θ, ð, s, z, ʧ, ʤ/; the CVen system has /t, d, θ, s, z, ʧ, ʤ/; in both systems the coronal stops are dental, as in SI. In this paper we will focus only on the voiceless fricatives /θ, s, ʧ/. Examples of NeVen voiceless fricatives include: (inter)dental /θ/ as in [ˈ**θ**ento] (“one hundred”, SI [ˈ**ʧ**ento]), alveolar /s/ [ˈ**s**ento] (“I hear”, SI [ˈ**s**ento]) and alveo-palatal /ʧ/ as in [ˈ**ʧ**ave] (“key”, SI [ˈ**kj**ave]). In the southern part of the CVen system, specifically in the area which is centered around the town of Rovigo (Polesine), /θ/ has further evolved in a dento-alveolar fricative [s̪] (Trumper, 1972: 16). Examples of CVen fricatives include: /s̪/ as in [ˈ**s̪**egoe] (“onions”, SI [**ʧ**iˈpol:e]), /s/ as in [ˈ**s**aleze] (“willow”, SI [ˈ**s**aliʧe]); /ʧ/ as in [ˈ**ʧ**ɔð̞i] (“nails”, SI [ˈ**kj**ɔdi]).

Note that the dental affricates found in SI ([ts], [dz]) do not occur in either of these Veneto dialect subsystems (Trumper, 1972; Zamboni, 1974). This absence is explained by the diachronic development of the phonology of Italo-Romance dialects from Latin. Palatalization and in certain cases affrication of consonants that were followed by the palatal glide /j/ in Proto-Romance occurred at an early date: it started probably in century I, AD, in any case not later than century II (Tekavčić, 1980: 114 ff), and affected all Romance varieties. This process was followed by a later, non Pan-Romance palatalization (and affrication) process affecting velar stop consonants /k/, /g/ immediately followed by a front vowel (this latter process was active not before century V: Tekavcic ibid.; see also Maiden & Parry, 1995: 48 ff). This is one of the most intricate chapters in the development of the Italo-Romance

[9] /d/ gets lenited in intervocalic position in both NeVen and CVen systems. But while the dialectological literature refers to the lenited allophone as a dental fricative [ð], in our data /d/ has two lenited allophones: a dental approximant [ð̞] in most of the cases (110/132) and a dental fricative [ð] in a minority of cases (22/132).

dialects' sound system from Latin. In general, palatalization of Latin /k/ produced voiceless apical affricate /ts/ (with further developments) in Northern Italian dialects, including Veneto, and palatal affricate /ʧ/ in the Central and Southern ones. As for Latin /g/, it produced the voiced affricate /dz/ (with further developments) in Northern Italian dialects, affricate /ʤ/ in Tuscany and Central Italian dialects, and the voiced palatal fricative /j/ in most of Southern dialects (see Tekavcic, 1970: 187 ff for a detailed discussion). In both systems Latin /k, t/ + /l/ clusters evolved in the voiceless affricate /ʧ/.

Comparing the phonetic properties and the distribution of the coronal fricatives of NeVen and CVen to one another and to those occurring in SI and English, the other languages of our trilingual Italian-Australian speakers, similarities and differences emerge. The NeVen phonological system shares with English the voiceless [θ] interdental fricative, while SI lacks it (so does CVen). Neither English nor SI (nor NeVen) has the lamino-denti-alveolar fricative [s̪] that is present in the Rovigo area of the CVen system. On the other hand, SI has dental affricates that do not occur in any of Veneto's regional subsystems.

The differences in the phonological/phonetic properties of coronal obstruents in Veneto dialects with respect to Italian and English will be the lens through which we will observe how the linguistic repertoires managed by the immigrant speakers maintain, loose or change their phonetic features in a multi-contact situation. To accomplish this, we conducted phonetic-acoustic analyses of words containing these target voiceless coronal fricatives in two Italian-Australian first generation (immigrant) speakers from each of the two Veneto subregions.

3. *Target words for the corpus*

In order to verify the presence and the phonetic characteristics of those coronal fricatives, we designed a corpus that, among other consonants, included the specific target consonants we are interested in: for dialect, /θ, s̪, s, ʧ/ and for Italian, /s, ʃ, ʧ/. As for the dialect, we drew baseline comparative data from the literature with the data available in the *Multimedia Atlas of Veneto Dialects* (Atlante Multimediale dei Dialetti Veneti: AMDV; Tisato, Barbierato, Ferrieri, Gentili & Vigolo, 2013) in order to ascertain whether, where and in which items the interdental and the lamino-denti-alveolar fricatives are attested in the present-day Veneto dialects as spoken in Italy. The AMDV contains Jaber and Jud's *Sprach und Sachatlas Italiens und der Südschweiz* data (Atlante Italo-Svizzero, AIS, 1921) collected for the Veneto region in 1921 by Scheuermeier–(Perco, Vigolo & Sanga, 2011) and new data collected in 2009 by the authors of the AMDV. For the 1921 data, only graphic forms are available, while for the 2009 data also phonetic transcriptions in IPA and audio recordings are available. Relying on the results of this search in AMDV and AIS, we compiled a list of words containing in their dialectal realization one of the target consonants. For the sake of future comparability, most of the items we selected for this project were present in the AMDV. As far as possible, the list was balanced for

position of the consonant in the word (word-initial, word-medial, word-final) and phonetic context. The Latin derivation was also taken into consideration, in order to select only those items having a similar and comparable realization both in their dialectal as well as in their Italian form. The final and revised list of dialectal items used in the production task (see §4) consisted of 64 depictable words (see examples in Table 1), matched with a picture representing it and later used to elicit the target item[10].

In order to verify whether the dialect exerts any influence on the participants' spoken Italian, out of the 64 words that constitute our dialectal corpus, we selected a subset of 46 to be produced in Italian for the Italian production task (as described in §4).

Table 1 - *Examples of words used and their respective pronunciations*

Depictable word	*NeVen - Belluno*	*CVen - Rovigo*	*SI*
clogs	[ˈθɔkoj]	[ˈʂɔkoj]	[ˈtsɔkoli]
halter	[kaˈveθa]	[kaˈveʂa]	[kaˈvets:a]
trowel	[kaˈθɔla]	[kaˈʂoja]	[kaˈts:wɔla]
nail	[ˈʧɔð̩o]	[ˈʧɔð̩o]	[ˈkjɔdo]

4. *The Italian Roots corpus recordings*

To create our trilingual speech corpus, spontaneous speech samples in dialect, Italian and English were elicited and recorded by means of a MatLab recording tool (*Sync*Rec by G. Tisato). Audio was captured through a Shure model SM10A-CN headset microphone connected to an external Edirol UA-25 EX sound card. Each interview was recorded at 96 kHz 24bit-mono, ranging in duration from 1:45' to 2:45'. The recordings took place in a quiet location at the participants' home or at MARCS Institute (University of Western Sydney).

In order to tune the participants into the appropriate language mode (Grosjean, 1998) during the collection of the three speech samples (i.e. dialect, Italian and English), we adopted a specific interview protocol, as follows:

Starting with dialect, the initial phase was designed to perceptually attune the informants to their local dialect. This was achieved by having the speaker listen to excerpts of autobiographic tales (from AMDV) spoken by monolingual Veneto speakers living in the same areas in Italy where the informants were born and lived until they emigrated to Australia. The informant was asked to listen to these stories, while trying to bring back his/her memories of the past. A guided interview was then conducted in Veneto dialect. At this stage, one of the two interviewers

[10] We preferred using pictures instead of orthographic or verbal stimuli in order to avoid providing hints to the participants about the word and its pronunciation.

asked questions in the participant's own dialect: the resulting bouts of conversation produced by the participant related to facts, events, stories or memories from their childhood in Italy, proverbs, nursery rhymes. The whole interview was kept at an informal level. The participants were allowed to express themselves freely and without time constraints.

Then, after a short practice session, the 64 target images were presented to the speaker in random order on a 15" monitor and asked him/her to name and describe them. In order to stimulate the participants' verbal productivity, the interviewer engaged them in conversation asking for descriptive details, but never directly asking for or suggesting the target word: this allowed us to elicit more tokens for each target item.

After a short break, the same procedure was repeated in Italian. In the conversation held in Italian, both interviewers interacted with the informant. The questions they asked addressed aspects of the speaker's linguistic competence in Italian, such as how and when they learned Italian, in which situations he/she prefers to use Italian or dialect, as well as their attitude towards each of them. Then, the informant was presented with the 46 items to be named and described in Italian.

The interview ended with a short conversation in English focusing on the following issues: i) which of the languages (dialect, Italian, English) he/she considered to be his/her "dominant" language; ii) the reasons of his/her switches between languages during the interview, if any; iii) if he/she has ever consciously suppressed his/her dialectal accent when speaking Italian or English; iv) if he/she ever consciously maintained or even exaggerated his/her dialectal accent.

5. *Participants*

Based on sociolinguistic information collected by means of a questionnaire, for the investigation reported here we selected two speakers each for the NeVen and the CVen areas (one female and one male, respectively). Participants CZ and GP were born in the province of Belluno (BL, NeVen; Bellunese variety); AM and JF were born in the province of Rovigo (RO, CVen; Rodigino variety). These speakers are also matched between BL and RO with respect to: age and local dialect as L1; number of years of experience of English, as indexed by length of residence (LOR) in Australia (range = 50-57 years). Moreover, the age they started to acquire Italian as L2 coincides for all of them with the beginning of primary school in Italy at age 6. The age of arrival in Australia (AoA) is lower for the females than for the males, and also represents the age they began to learn English as L3 (AoA Eng). Further detailed socio-linguistic information about our informants can be found in Table 2.

Table 2 - *Informants' sociolinguistic information: Age = age at time of recording; LOR = Length of Residence in Australia; AoA Eng = Age or Arrival and onset of Acquisition of English*

Veneto origins	*ID*	*Sex*	*Age (yr)*	*AoA Eng*	*LOR*	*L1*	*L2*	*L3*	*Highest education achieved*	*Profession*
RO	AM	F	64	14	50	dialect (Rodigino)	Italian	English (formal instruction)	university	judge
RO	JF	M	81	26	55	dialect (Rodigino, Vicentino)	Italian & Ferrarese dialect	English (formal instruction)	primary school	storekeeper
BL	CZ	F	74	17	57	dialect (Bellunese)	Italian	English (spontaneous immersion)	primary school	housewife
BL	GP	M	82	29	53	dialect (Bellunese-Cadorino)	Italian	English (spontaneous immersion)	primary school	craftsman

6. *Data preparation and acoustic analysis of coronal fricatives*

All target words were segmented and IPA labeled in Praat. Four different textgrid tiers were created for the full recording. They included:

a. segmentation and orthographic transcription in Italian at sentence level;
b. segmentation and narrow IPA transcription at word level (for dialect);
c. segmentation and narrow IPA transcription at phone level of the target consonants and of the preceding and/or following phonetic context (for dialect). As for affricates, we separately segmented and transcribed the closure and frication phases;
d. coding of the target consonant and word position: for the target consonants, we coded their manner of articulation and their position within the word (initial or medial).

The acoustic analysis centered on the set of voiceless coronal fricatives. The onset of the fricative was defined as the first appearance of aperiodic noise on the waveform (the point at which the number of zero crossings rapidly increased) simultaneously accompanied by high-frequency "noisy" energy on the spectrogram. The offset for voiceless fricatives was defined as the first zero-crossing of the periodic waveform of the following vowel (Jongman, Wayland & Wong, 2000; Li, Edward & Beckman, 2009). In order to ensure transcription accuracy, the preliminary segmentation and coding by one of the authors were checked by two other authors (from among CA, VG, MV).

Fricatives are sounds produced by making a tight constriction in the vocal tract (with an area of the order of 0.1-2 cm), which gives rise to an airstream forming "a turbulent jet" that provides a noise source (Shadle, 2010: 50). Previous research on languages that have a place-of-articulation contrast for fricatives, such as English or Italian, have focused on the spectral properties of the frication noise in order to differentiate them by place of articulation (e.g. Hughes, Halle, 1956; Stevens, 1960; Heinz, Stevens, 1961). The spectral content of a fricative sound is determined by shape and size of the oral cavity in front of the constriction. For example, a /ʃ/ is characterized by a longer front cavity than /s/, both because it is produced with a more posterior place of articulation and because of the typical lip protrusion (in the context of unrounded vowels) that further extends the front cavity. Such an increase in the length of the front resonance cavity produces a low-energy spectrum for /ʃ/ and a higher-energy spectrum for /s/.

In more recent years, a common way to acoustically analyze fricative sounds has been to consider their power spectrum as a probability distribution and to calculate its spectral center of gravity (also mean or centroid frequency), its standard deviation (or variance), its skewness and its kurtosis[11]. These are known as *spectral moments* and are referred to as M1, M2, M3, M4 respectively (Forrest, Weismer, Milenkovic & Dougall, 1988; Jongman et al., 2000; Harrington, 2010).

The Center of Gravity (CoG) provides information regarding where, on average, the energy is concentrated, and correlates negatively with the length of the front cavity resonance. Standard Deviation (SDev) is a measure of the diffuseness of the spectrum around the CoG. Skewness (Skew) refers to the distribution's asymmetry: a value of zero indicates a symmetrical distribution around the mean, a positive value indicates that the right tail of the distribution extends further than the left tail, while a negative value indicates that the left tail of the distribution extends further than the right one. Skewness provides information on the spectral tilt: a positive skewness suggests a negative tilt with a concentration of energy in the lower frequencies, while a negative skewness is associated with a positive tilt and a predominance of energy in the higher frequencies. The fourth spectral moment, Kurtosis (Kurt), indicates the peakedness of the distribution: a positive value indicates a relatively high peakedness (the higher the value, the more peaked the distribution, containing one or a small number of relatively sharp peaks), while a negative one indicates a relatively flat distribution. Positive kurtosis thus suggests a clearly defined spectrum with well-resolved peaks, while negative kurtosis indicates a flat spectrum without clearly defined peaks.

The spectral moments metric thus incorporates both local (spectral peak) and more global (spectral shape) information (Jongman et al., 2000: 1253). We adopt the spectral moment analysis in order to classify the dialectal productions of the coronal fricatives of our speakers, as that set of measures can be sensitive enough to show acoustic cues indexing fine-grained articulatory differences, such as those

[11] Another important acoustic cue to fricative place of articulation is the onset frequency of F2 in the following vowel (Li et al., 2009).

between the interdental versus the lamino-denti-alveolar fricatives described for Belluno and Rovigo dialects, respectively. CoG and Skew may be useful in differentiating fricatives that have a different place of articulation; as CoG negatively correlates with the length of front resonating cavity, it roughly describes where the constriction is made relatively to the length of the oral cavity (Li et al, 2009: 112): the lower the CoG value the more posterior the place of articulation, i.e., the larger the front cavity. Skew also correlates to a place-of-articulation distinction: a positive value would indicate concentration of energy in the lower frequencies below the mean value and hence it would point to a more posterior place of articulation. SDev and Kurt distinguish a compact and peaky spectral shape, respectively, from a diffuse and flat one, and thus help in differentiating fricatives based on their tongue posture, specifically apical vs laminal fricatives.

So, while CoG and Skew distinguish [θ]-[s]-[ʃ] and help in identifying the place of articulation of the Rovigo [s̪], SDev and Kurt should further distinguish the non-sibilant [θ] from the sibilant [s]-[ʃ] and possibly help characterize the tongue posture of Rovigo's [s̪].

Prior to analysis, each file was down-sampled to 48 kHz 16-bit. In order to compute the four spectral moments we adopted the time-averaging technique proposed by Shadle (2012)[12]. Adapting a Praat script by Di Canio (2013)[13] that implements Shadle's suggestion, before the analysis a 300 Hz low pass cut-off filter was applied to all the recordings to remove any F0-related influence. Then CoG, SDev, Skew and Kurt were computed over the central 80% of the fricative segment's duration[14] using 5 DFTs with an analysis window set to 10ms.

8. *Results*

We report here the results of the dialectal target word productions of the 4 selected speakers, specifically concentrating on the set of words containing unvoiced coronal fricatives of the NeVen (Bellunese) and CVen (Rodigino) dialect systems. In this set we include also the postalveolar affricate [ʧ], of which we will acoustically analyse only the fricative release phase, [ʃ].

We considered 501 tokens, 210 produced by the Belluno speakers, 291 by those from Rovigo. The number of collected tokens in the two groups is unbalanced due to the eliciting procedure, in which we gave complete freedom to our participants to describe and comment on the object representing the target word. The conse-

[12] Specifically, each fricative segment is first divided into many shorter intervals; each of these is windowed and a Discrete Fourier Transform (DFT) is computed for each. The obtained DFTs are then averaged. Shadle's assumption is that for fricatives, "the signal properties are stationary during the long interval, and therefore the short windows represent independent samples of the same random process" (Shadle, 2012: 515).

[13] http://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives.praat

[14] In order to limit coarticulatory effects from the preceding and following segments we kept Di Canio's script settings and discarded 10% of the segment duration at its onset and 10% at its offset.

quence of allowing this more naturalistic approach in eliciting the data is that for each speaker the number of repetitions of a target word could range from 0 (if he/she did not know or recall the word in dialect) to a dozen repetitions.

8.1 Distribution of target coronal fricatives: Latin root words

Table 3 shows the percent distribution of the target coronals, per our fine-grained IPA transcriptions of them, as a function of their Latin derivation.

Table 3 - *Distribution of voiceless coronal fricatives, as produced in all word position by the four speakers*

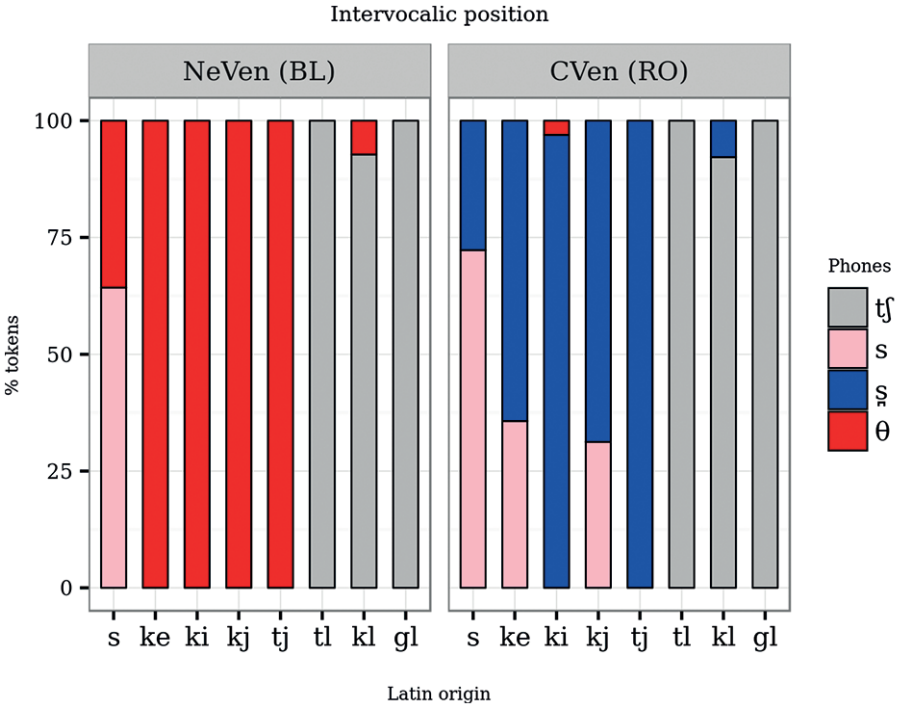
	NeVen system (Belluno)				*CVen system (Rovigo)*			
Latin Root	ʧ	s	θ	ʂ	ʧ	s	θ	ʂ
s	0	84% (43/50)	16% (7/50)	0	0	75% (58/77)	0	25% (19/77)
tj	0	0	100% (10/10)	0	0	0	0	100% (31/31)
ke	0	0	100% (28/28)	0	0	36% (10/28)	0	64% (18/28)
ki	0	0	100% (31/31)	0	0	5% (2/39)	3% (1/39)	92% (36/39)
kj	0	0	100% (16/16)	0	0	25% (5/20)	0	75% (15/20)
tl	100% (3/3)	0	0	0	100% (11/11)	0	0	0
kl	93% (64/69)	0	7% (5/69)	0	92% (72/78)	0	0	8% (6/78)
gl	100% (3/3)	0	0	0	100% (6/6)	0	0	0

The NeVen system has three voiceless coronal fricatives in its phonemic inventory: dental [θ], alveolar [s] and postalveolar [ʧ]. All were produced consistently by our BL speakers in the relevant dialect target words. [s] was our BL speakers' realization of the voiceless coronal fricative for 84% of the cases of words with /s/ in Latin. In the remaining 16% of Latin /s/ derivations, the BL speakers' realization was [θ], however, this was restricted to just one specific lexical item: Latin "soccus" > BL ['θɔk], SI "zoccolo" ['dzɔk:ɔlo] (sing.) (*clog*). By comparison, for dialect words derived from Latin stops that would have been palatalized due to a following front vowel or glide, /t, k/+ /i, j, e/, the BL speakers' realization is [θ]; there are no exceptions. BL productions did include the affricate [tʃ] but only for derivations from Latin words that had contained stop + liquid clusters /t, k, g/ + /l/. No cases of the lamino-denti-alveolar voiceless fricative [s̪] (described for Rodigino dialect: RO), were attested in our BL speakers' productions.

For the voiceless coronal fricatives produced by our speakers from CVen (RO), a different picture emerged. On the one hand, consistent with the speakers from Belluno, derivations from Latin /t, k, g/ + /l/ clusters were homogeneously realized in RO dialect as [ʧ] (as in Belluno, only /kl/ has a different output in a minority of cases: [s̪], in 8% of cases). On the other hand, deviating from the BL situation, derivations from Latin palatalized stop consonants resulted in more variable output in RO speech. In the majority of cases, the RO fricative produced is phonetically transcribed as the lamino-denti-alveolar [s̪] (/ke/ = 64%; /ki/ = 92%; /kj/ = 75%). But in the remaining cases, the realizations are transcribed as "plain" [s] or [θ] (the latter only in 1 case). Dialect words with Latin /s/ derivations are also produced variably by our RO speakers, including both [s] and [s̪] realizations.

The difference between the NeVen (BL speakers) and CVen (RO speakers) systems can be better appreciated in Fig. 1, which includes only the target voiceless coronals followed by a vowel (respectively 192 and 285 occurrences).

Figure 1 - *Percent distribution of voiceless coronal fricatives (y axis) followed by vowel as a function of their Latin derivation (x axis). For BL, n =192; for RO, n = 285*



In CVen (RO speakers), Latin /s/ gives rise to [s] in the majority of cases (75%) and [s̪] in the remaining (25%). Latin /tj/ is realized as [s̪] in all cases; /kj/ as [s̪] in 75% of cases, as [s] in the remaining 25%. Latin /ke/ and /ki/derivations are realized as [s̪] in the majority of cases (64% and 92%) and as [s] in a minority of cases (36%, 5%). In this system, then, the majority of derivations from original Latin palatalized

stops are realized as [s̪] while a minority of derivations from original Latin /s/ are realized as [s̪]. Conversely, the majority of derivations from Latin /s/ and the minority of derivations from Latin palatalized stops converge to [s].

On the contrary, in NeVen the Latin palatalization of stops in the context of front vowels results consistently in BL realizations as [θ]. The more variable outputs in the Rovigo system cannot be entirely explained in terms of a lexically induced variation: while the two RO speakers produce "*cherry*" (late Lat. "ceresia") as [sa'reza] in dialect instead of the expected [s̪a'reza], in other cases both speakers alternated between [s] and [s̪] realizations for the same lexical item.

8.2 Acoustic characterization: spectral moments

For the acoustic analysis we selected a subset of fricative tokens based on the quality of the following vowel: only those fricatives that occurred before /a/, /e, ɛ/ and /o, ɔ/ were included in the analysis, in order to balance the anticipatory coarticulatory influence of the following vowel. Means and standard errors of the 4 spectral moments for each speaker of the NeVen (Belluno) and CVen (Rovigo) systems are displayed in Figure 2[15].

The acoustic analyses are aimed to verify: 1) whether Veneto speakers conserved the consonants of their L1-dialect; 2) whether the two dialects have drifted to one another converging horizontally to form a Veneto koiné; 3) whether they have drifted toward SI in a process of vertical advergence; 4) whether they have drifted toward English.

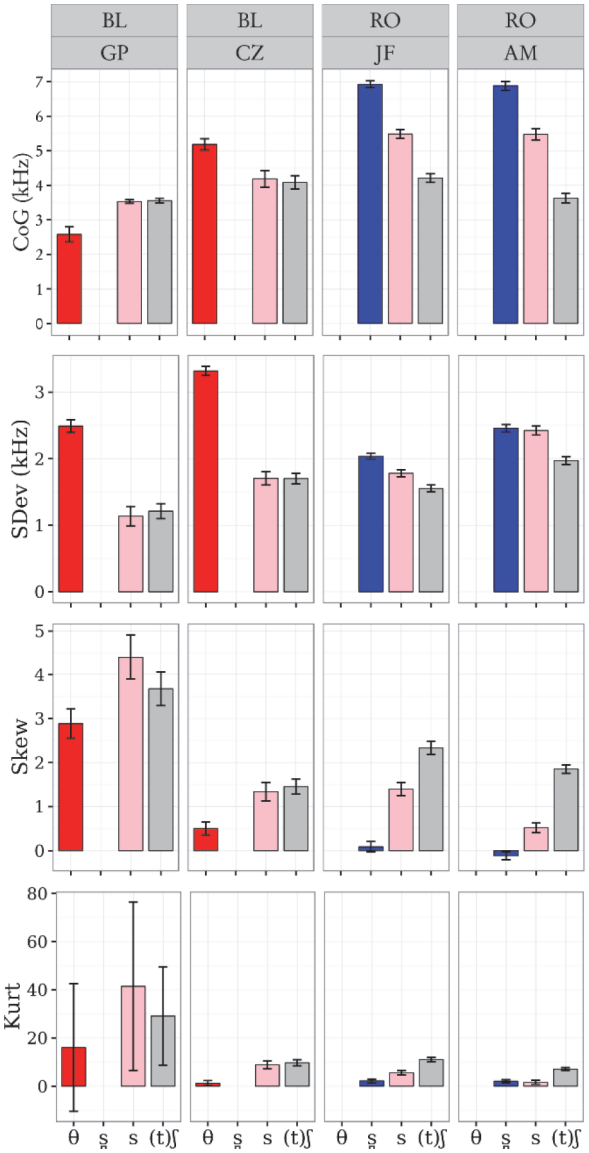
In the first case, comparing the NeVen to the CVen system, we expect that only the fricatives that are unique to each system, i.e. [θ] and [s̪], will be significantly different from each other, while the others will not be affected as they are shared consonants. Within each system, every fricative will be different from the others.

In the second case, [θ] and [s̪] will not be significantly different from each other at least on some aspect of their spectrum: if they have merged to a common place of articulation, they will have the same CoG and Skew; if they have merged to a consonant produced with the same tongue shape, they will have the same SDev and Skew; if they have completely merged, there will be no difference on all spectral moments. Within each system, [s] and the fricative release phase of [ʧ] will be different from the others (from now on, for simplicity, we will refer to it only as [ʃ]).

In the third case, [θ] and [s̪], will drift to the closest fricative consonant [s] that is present both in their own dialect and in SI: if this is the case, both in Belluno and in Rovigo the dental fricatives will equal the alveolars. As we assume that the other fricatives will not show significant difference between each other, we expect no difference between Belluno and Rovigo on any fricative. Within each system, [s] will be different from [ʃ].

[15] The spectral moment analysis has been applied only to the fricative release phase [ʃ] of the postalveolar affricate [ʧ].

Figure 2 - *Mean and standard error of the spectral moments for the four speakers*



In the fourth case, we expect different patterns of results in the two systems: for Belluno, we expect significant differences in the spectral characteristics of [θ] as compared to [s] and [ʃ], because the contact with English would reinforce the maintenance in the dialect of [θ], a consonant shared by Belluno and English. For Rovigo, we expect no difference between the dento-alveolar [s̪] and the alveolar [s], a fricative shared with English (and SI), while [s] would be significantly different from [ʃ].

To test those hypotheses, we performed two sets of statistical analyses. In the first we applied a Repeated Measure Model from the Generalized Linear Models provided by the JMP platform. Gender (male, female), dialect (Belluno, Rovigo) and Fricative (dental, alveolar, post-alveolar) were set as fixed factors, Subjects were included as random effect. Each of the four acoustic measures CoG, SDev, Skew and Kurt were the dependent variables.

In the second, we aimed at checking individual differences in the pattern of contact-induced drifts and we performed four 2-way Repeated Measure ANOVAs for each speaker. The within-subjects factor was type of Fricative (for Belluno: [θ], [s], and [ʃ]; for Rovigo: [s̪], [s], and [ʃ]) and the between-subject factor was type of Following Vowel. The dependent variable was each of the four acoustic measures CoG, SDev, Skew and Kurt.

8.2.1 *Statistical analysis*

The results of the first set of analyses for all spectral moments indicate that Fricative is significant and the interaction between Fricative and dialect is significant only for CoG, SDev and Skew. Gender and dialect are never significant (details are shown in Table 4).

The post-hoc tests (Student's *t,* alpha level 0.05) indicate that the difference in CoG between dental, alveolar and post-alveolar fricatives both in Belluno and Rovigo is non significant (while every fricative is different from the others within the Rovigo system, no difference is attested among the set of fricatives in Belluno). For SDev, both Belluno dentals and alveolars are significantly different from Rovigo ones (dentals: BL = 2929.34, RO = 2251.6; alveolars: BL = 1418.4, RO = 2095.3). Within each system, dentals are different from alveolars and post-alveolars in Belluno; all of them are different from each other in Rovigo. As for Skew, only dentals are significantly different from each other in Belluno vs Rovigo (BL= 1.696; RO= 0.022). Within each system, dentals are different from alveolars and post-alveolars in Belluno; all of them are different from each other in Rovigo. For Kurt, none of the fricatives shows significant differences between Belluno and Rovigo (see Table 5 for a summary).

Table 4 - *ANOVA values on the four acoustic parameters of the fricative spectrum*

Factors	*CoG*	*SDev*	*Skew*	*Kurt*
Gender	n.s.	n.s.	n.s.	n.s.
Dialect	n.s	n.s.	n.s.	n.s.
Fricative	F(2,476)= 105.201***	F(2,476)= 227.351***	F(2,476) = 65.915***	F(2,476)=4.615*
Fric*Dia	F(2,476)= 85.777***	F(2,476) = 88.613***	F(2,476) = 13.925***	n.s.

***p <0.001; **p <0.01; *p <0.05

Table 5 - *Post-hoc tests on the four acoustic parameters of the fricative spectrum*

Fricatives	*Differences*			
Dentals (D), Alveolars (A), PostAlveolars (P)	*CoG*	*SDev*	*Skew*	*Kurt*
BL vs RO	none	D, A	D	none
BL	none	D#(A, PA)	D#(A,PA)	none
RO	D#A#PA	D#A#PA	D#A#PA	none

The data on Skew support the conservation of features (first) hypothesis: the two systems differ only for the dental fricatives (interdental in Belluno vs. dento-alveolar in Rovigo) on this measure, while they do not differ about the remaining consonants. Dentals have a different place of articulation in Belluno as compared to Rovigo, while alveolars and postaveolars do not differ in the two systems. Looking at each system separately, though, the hypothesis is fully supported only by Rovigo, where each fricative is significantly different from each other, while in Belluno the dental fricative differs from the other consonants, but the alveolar equals the postalveolar. Neither the second (Veneto koinè) nor the third hypothesis (advergence to Italian) appears to be confirmed by the data of the four spectral moments, based on the post-hoc comparisons between BL and RO and within each system combined. Neither is the fourth hypothesis, a drift toward English, supported by the data of Rovigo, which is the crucial case for evaluating this hypothesis. Belluno data cannot inform about this hypothesis, as the extended contact with English would reinforce the *maintenance* of the dental fricative in the NeVen system (BL), as it is a shared consonant.

Table 6 - *ANOVA values for each acoustic parameter of the fricative spectrum*

Speaker	*Factors*	*Center of Gravity*	*Standard Deviation*	*Skewness*	*Kurtosis*
GP	Fricative	$F(2,92) = 10.284^{***}$	$F(2,92) = 45.608^{***}$	$F(2,92) = 3.319^{*}$	n.s.
(BL)	Vowel	n.s.	n.s.	n.s.	n.s.
	Inter.	n.s.	n.s.	n.s.	n.s.
CZ	Fricative	$F(2,97) = 13.81^{***}$	$F(2,97) = 149.167^{***}$	$F(2,97) = 14.549^{***}$	$F(2,97) = 17.263^{***}$
(BL)	Vowel	$F(2,97) = 7.463^{***}$	$F(2,97) = 3.232^{*}$	$F(2,97) = 18.683^{***}$	$F(2,97) = 10.3^{***}$
	Inter.	n.s.	$F(4,93) = 2.636^{*}$	n.s.	$F(4,93) = 3.168^{*}$
JF	Fricative	$F(2,140) = 143.881^{***}$	$F(2,140) = 22.547^{***}$	$F(2,140) = 71.505^{***}$	$F(2,140) = 23.811^{***}$
(RO)	Vowel	$F(2,140) = 7.006^{**}$	n.s.	$F(2,140) = 8.233^{**}$	n.s.
	Inter.	n.s.	$F(4,136) = 4.904^{***}$	$F(4,136) = 3.103^{*}$	$F(4,136) = 7.726^{***}$
AM	Fricative	$F(2,135) = 157.438^{***}$	$F(2,135) = 19.32^{***}$	$F(2,135) = 108.501^{***}$	$F(2,135) = 17.237^{***}$
(RO)	Vowel	$F(2,135) = 13.574^{***}$	n.s.	n.s.	n.s.
	Inter.	n.s.	$F(4,131) = 19.32^{*}$	n.s.	n.s.

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

As the post-hoc tests clearly reveal a different picture for the two Veneto dialect subsystems, we proceed with the second set of analyses performed on each speaker. In reporting the results summarized in Table 6, we will comment about the articulatory configurations of each fricative consonant that can be inferred from the spectral moments data (for a summary, see Table 7).

Looking at the results we observe that the main effect of Fricative is significantly different for every spectral moment and every speaker, with the only exception being for GP's Kurtosis.

Starting with the Belluno system and the female speaker CZ, [θ] is significantly different from [s] and [ʃ] for all spectral moments, while [s] and [ʃ] do not differ from each other (Tukey-Kramer HSD, alpha level 0.05). CoG is higher for [θ] than for [s] and [ʃ] (Hz 5198,53 vs 4215,11 and 4086,12), indicating a concentration of energy in the higher frequencies, which supports a more advanced place of articulation for [θ] than for [s] and [ʃ]. Skew is lower for [θ] than for [s] and [ʃ] (0.513 vs 1.346 and 1.452), indicating a quasi symmetrical distribution of energy, while the values for the other two fricatives indicate a concentration of energy in the lower frequencies below the mean value. As skewness negatively correlates with the length of the front resonating cavity, these values suggest a more advanced place of articulation for [θ]. SDev is higher for [θ] than for [s] and [ʃ] (Hz: 3345.86 vs 1702.22 and 1701.90) and Kurt is lower (1.218 vs 9.758 and 9.162), both indicating a more diffuse spectrum that correlates with a more laminal articulation of [θ].

For GP, the male speaker from Belluno, the spectral moments that correlate with place of articulation distinctions give contradictory results: CoG values are lower for [θ] than for [s] and [ʃ] (Hz 2580.45 vs 3470.45 and 3556.75), unexpectedly suggesting a *less* advanced place of articulation for [θ], while Skew values are lower for [θ] than for [s] (2.883 vs 4.398), more in line with expectations, indicating a more advanced place of articulation for [θ] than for [s]. SDev, also as expected, shows higher Hz values for [θ] that are compatible with a more diffuse spectrum and a more laminal articulation of the dental fricative ([θ]: 2490.94; [s]: 1150.06; [ʃ]: 1212.34).

As for the Rovigo system, all target fricatives produced by the male speaker JF are significantly different from each other for all spectral moments. Progressively lower values of CoG and progressively higher values of Skew indicate that in the front-back dimension [s̪] has a more advanced place of articulation than [s], and also that [s] is more advanced than [ʃ] (CoG - [s̪]: 6922.34; [s]: 5532.48; [ʃ]: 4214.49. Skew - [s̪]: 0.100 [s]: 1.359; [ʃ]: 2.309). Moreover, progressively higher values of SDev and progressively lower values of Kurt indicate a more diffuse and more flat spectrum in the sequence [s̪] > [s] > [ʃ], compatible with a more laminal articulation for [s̪] than [s], and for [s] than [ʃ] (SDev - [s̪]: 2041.02; [s]: 1783.63; [ʃ]: 1561.33. Kurt - [s̪]: 2.214 [s]: 5.245; [ʃ]: 10.916).

For the female Rovigo speaker AM, [s̪], [s] and [ʃ] are significantly different from each other for CoG and Skew, with [s̪] as the most fronted fricative of the series, again according to the sequence [s̪] > [s] > [ʃ] (CoG - [s̪]: 6880.83; [s]:

5477.04; [ʃ]: 3628.33. Skew - [s̪]: -0.117; [s]: 0.521; [ʃ]: 1.852). As for SDev and Kurt, [s̪] and [s] are significantly different from [ʃ] but not from each other, indicating a spectrum that is more diffuse and less peaky for [s̪] than [ʃ] but equally diffuse and flat for [s̪] and [s] (SDev - [s̪]: 2459.1; [s]: 2424.44; [ʃ]: 1967.14. Kurt - [s̪]: 2.090; [s]: 1.643; [ʃ]: 7.135).

Table 7 - *Articulatory interpretation of spectral moments for each speaker*

	CZ	GP	AM	JF
CoG ⇒ : more advanced place	[θ] ⇒ {[s], [ʃ]}	{[s], [ʃ]} ⇒ [θ]	[s̪] ⇒ [s] ⇒ [ʃ]	[s̪] ⇒ [s] ⇒ [ʃ]
SDev > : more laminal	[θ] > {[s], [ʃ]}	[θ] > [s] > [ʃ]	{[s̪], [s]} > [ʃ]	[s̪] > [s] > [ʃ]
Skew ⇒ : more advanced place	[θ] ⇒ {[s], [ʃ]}	[θ] ⇒ {[s], [ʃ]}	[s̪] ⇒ [s] ⇒ [ʃ]	[s̪] ⇒ [s] ⇒ [ʃ]
Kurt > : more laminal	[θ] >{[s], [ʃ]}	n.s.	{[s̪], [s]} > [ʃ]	[s̪] > [s] > [ʃ]

9. *Discussion and conclusions*

The data collected and analysed so far allow us to draw a sociophonetic picture of how some members of an Italian-Australian community have negotiated the varieties that form its linguistic repertoire after 50 or more years since their arrival in Australia. Our focus was on their local dialect as L1, and we asked: 1) whether they have maintained the dialect-specific variety of the Veneto area from which they originated; 2) whether, due to the contact with other Veneto speakers, they have formed a Veneto koiné loosing some of the phonetic features of their local dialect; 3) whether they have converged toward SI as the vehicular language of the whole Italian community; and 4) whether their dialect shows signs of phonetic attrition due to the extensive contact with English.

The distribution of the target coronals as a function of their Latin derivation allows us to appreciate to what extent the actual pronunciation of a given word in our speakers conforms to the norms of the different Veneto systems. In turn, the spectral moments analyses offer a way to characterize how the coronals are differentiated acoustically (and articulatorily).

There are two fricatives that act as identity marks of each of the two Veneto dialectal varieties we considered: the voiceless interdental fricative [θ] differentiates the NeVen from the other Veneto consonantal systems; [s̪] differentiates the southern part of CVen from the other Veneto dialects. Those two consonants will be the pivots of our discussion.

The distribution data that are based on our narrow phonetic transcriptions show that in NeVen (Belluno) [θ] is maintained by both speakers in all the words in which it is expected based on their Latin derivation, with no exceptions. There are no cases in which [θ] drifts toward the lamino-denti-alveolar fricative [s̪] that characterizes the dialectal system of CVen (Rovigo) speakers, nor to [s] of SI and English, nor to the dental and alveopalatal affricates that characterize the corresponding words

in SI. Both BL speakers behave identically in these regards, independently of their gender and their age of arrival in Australia: recall that the female speaker arrived at age 17, 11 years younger than the male speaker, and at the time of the recording she had 57 years of residency in the country as compared to 53 years of the male speaker.

Also for the CVen Rovigo speakers, the data show that none of them produced any [θ], confirming that the speakers have not shifted toward the NeVen system (nor toward English interdental fricatives). However, they show more variable renditions as compared to the Bellunesi. The words that etymologically have /kj, ke/ in Latin can present both [s̪], as expected in Rovigo's dialect, and [s]. More importantly, while in some cases such variability can be reduced to the specific rendition of a given lexical item (i.e. a given word is always produced by both speakers with [s] instead of [s̪]), in a few other cases [s̪] or [s] may randomly occur in the same word (e.g.: [ˈlas̪]/[ˈlaso], *lace*; [ˈs̪egoe]/[ˈsegoe], *onions*). The latter type of more variable rendition, however, is restricted to only the female speaker AM.

From the distribution data it appears that all speakers of each dialectal variety have conserved the distinctive fricatives of their systems and the acoustic analyses confirmed that the landmark consonants [θ] and [s̪] are distinguished by different values of Skewness, which indicate that Belluno's [θ] has a more advanced place of articulation than Rovigo's [s̪], as expected from earlier descriptive work summarized in our introduction.

Moreover, all speakers but one are consistent in the production of a given fricative in all repetitions of the same lexical item. Only the variability observed for the Rovigo speaker AM, admittedly limited, could be a reflection of linguistic interference. We exclude the possible account that the exchange of [s] for [s̻] could be attributed to her uncertainty in recalling those specific words in dialect, as they are common words. Instead, it is the phonetic similarity between [s̻] and [s] that could be the initial opportunity for a possible attrition of this distinction. Only for AM in fact (but not for the other Rovigo speaker JF) do the two sounds appear to have the same tongue posture, as indexed by SDev and Kurt, which are not significantly different from each other. Lamino-denti-aveolar [s̻] maintains a different place of articulation, as indexed by the values of CoG and Skew, with [s̻] being more advanced than the alveolar [s]. The two places are so close to one another in the articulation space, though, that sharing the same tongue posture could be the trigger for an eventual complete merger of the two sounds, and certainly could increase variability in production of the two consonants in our CVen (RO) speakers. The attrition could be exerted both by English and Standard Italian, as both lack [s̻] while in both of them [s] is present. Based on AM's sociolinguisic data, the latter appears to be the most plausible of the two possible sources of attrition: AM arrived to Australia as the youngest of the group of our four informants and had her education in English from high school through university. The other RO speaker (JF) arrived later in the country, received all of his education in Italy and does not show any uncertainty in the articulation of any fricative: each is significantly differentiated from the others on the basis of all spectral moments. [s̪], [s] and [ʃ] have a progressively

less forward place of articulation and a less laminal tongue posture, as we expect in the productions of monolingual native RO speakers.

Also the two Belluno speakers differ from each other as for the acoustic characterization of the fricatives' spectra. CZ, the female speaker, systematically distinguishes [θ] from the other coronal fricatives: the analysis shows that her [θ] has an articulation that is significantly more front and more laminal than the other two fricatives. However, [s] and [ʃ] are not differentiated by any spectral moment. Note that this result is at odds with the narrow phonetic transcriptions: perceptually, none of the transcribers ever had any doubt about how to categorize any instance of [s] and [ʃ].

The male speaker GP (BL) produces the three fricatives with a significant difference in the diffuseness of the spectrum, with values of SDev that indicate a more laminal articulation for [θ] and progressively less so for [s] and [ʃ]. As for the other Belluno speaker, [s] and [ʃ] are not distinguished for CoG and SDev, but, contrary to CZ and contrary to our expectations, CoG values indicate a less advanced place of articulation for [θ] than for [s] and [ʃ]. However, Skew and SDev values are consistent with a more advanced place for the dental fricative than the others.

Since for GP too there are no doubts on the nature of these consonants on a perceptual basis, we infer that the spectral moments analysis is not sufficient in itself to properly characterize the fricatives of such speakers: further acoustic measures are needed such as, for example, the F2 onset of the following vowel. Future and more extended acoustic analyses on other Veneto emigrants will tell us whether the results of the Belluno speakers are idiosyncratic or shared by other speakers. We aim at widening the database of our analysis to include on the one hand more Veneto-Australians, and on the other hand to compare their productions with those of monolingual Veneto speakers in Italy living in the same areas or villages from where the emigrants originated, and of the same generation. These extensions will allow a more thorough characterization of NeVen and CVen dialects and of their resilience in the speech of Australian-Italians.

At the moment, the occurrence, distribution and acoustic features of the identity-mark-fricatives that uniquely characterize the Belluno and Rovigo systems, confirm the first of our hypotheses. All of our speakers maintained in their L1-dialect both [θ] and [s̪] after more than 5 decades of residence in Australia. Interlinguistic influence exerted by L3-English is evident in one speaker only, the one who had more formal exposure to it from a younger age and over a longer period of time: she shows the signs of an initial drift that could induce a merge between the lamino-denti-alveolar and the alveolar fricative. In the other speakers, L3-English apparently does not interfere with the production of the target fricatives, even though our broader recordings of their spontaneous conversations show frequent code switches to English.

No traces of a Veneto koiné formation are found in any of the speakers, nor traces of Italianisation of dialect, neither in their productions of the coronal fricatives nor in the lexicon of their spontaneous speech. Dialect as L1 is preserved and main-

tained in the context of an attenuated form of diglossia with respect to L2-Italian in Australia. In the sociolinguistic questionnaires as well as in the final interviews, our speakers report speaking dialect with family, relatives and friends who emigrated from their same areas, while Italian is used as the language of socialization with other Italians who emigrated from different regions.

The sociolinguistic picture suggests that the local L1-dialect will quickly decay in the competence of the speakers as the number of people with whom it can be spoken diminishes, and/or as the situations in which the individual must operate in English increases (in our study here, AM of the RO group). A related question that is still open and will be addressed in our future work is whether L2-Italian will show the same type of resistance to L3-English or will show more drift toward it.

Acknowledgments

We acknowledge financial support: to V. Galatà in 2011 from the Italian National Research Council (STM Grant, prot. n. 53653, 19/07/2011) and in 2012 from MARCS Labs, University of Western Sydney (Australia) for a visiting research fellow grant; to C. Avesani and M. Vayra in 2013 for a Visiting Fellowship from the University of Western Sydney (IRIS grant). Special thanks go to Giacomo Ferrieri for his precious help in checking the IPA transcriptions. Finally, we would like to thank all the participants for their collaboration and participation without whom this project would not have been possible.

References

Antoniou, M., Best, C.T., Tyler, M.D. & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. In *Journal of Phonetics*, 38, 640-653.

Antoniou, M., Best, C.T., Tyler, M.D. & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. In *Journal of Phonetics*, 39, 558-570.

Antoniou, M., Tyler, M.D. & Best, C.T. (2012). Two ways to listen: Do bilinguals perceive stop voicing differently according to language mode? In *Journal of Phonetics*, 40, 582-594.

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.

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.

Bartoli, M. (1925). *Introduzione alla neolinguistica: principi – scopi – metodi*. Genève: Olschki.

Bartoli, M. (1945). *Saggi di linguistica spaziale*. Torino: Rosenberg & Sellier.

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.

Bettoni, C. (1985b). Italian language attrition: a Sydney case-study. In Clyne, M. *Australia, meeting place of languages*. Camberra: Australian National University Press, 63-79.

Bettoni, C., Rubino, A. (1996). *Emigrazione e comportamento linguistico. Un'indagine sul trilinguismo dei siciliani e dei veneti in Australia*. Galatina: Congedo Editore.

Campolo, C. (2009). L'italiano in Australia. In *Italiano LinguaDue*, 1, 128-141.

Cerruti, M., Regis, R. (2011). Standardization patterns and Dialect/standard convergence: A northwestern Italian perspective. In *Language in Society*, 43 (1), 83-111.

Cerutti, M. (2011). Regional varieties of Italian in the linguistic repertoire. In *International Journal of the Sociology of Language*, 210, 9-28.

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

D'Agostino, M. (2007). *Sociolinguistcia dell'Italia contenporanea*, Bologna: il Mulino.

Dal Negro, S., Vietti, A. (2011). Italian and Italo-Romance Dialects. In *International Journal of the Sociology of Language*, 210, 71-92.

De Mauro, T. (1972[3]). *Storia linguistca dell'Italia unita*. Roma-Bari: Laterza (orig. ed. 1963).

Flege, J.E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. In *Journal of the Acoustical Society of America*, 89 (1), 395-411.

Flege, J.E., Eefting, W. (1987). Cross-language switching in stop consonant production and perception by Dutch speakers of English. In *Speech Communication*, 6, 185-202.

Flege, J.E., Mackay, I.R. & Piske, T. (2002). Assessing bilingual dominance. In *Applied Psycholinguistics*, 23, 567-598.

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

Gallina, F. (2011). Australia e Nuova Zelanda. In Vedovelli, M. (Ed.), *Storia linguistica dell'emigrazione italiana nel mondo*. Roma: Carocci, 429-475.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. In *Bilingualism: Language and CoGnition*, 1 (2), 131-149.

Harrington, J. (2010). Acoustic Phonetics. In Hardcasle, W., Laver, J. & Gibbon, F. (Eds.), *The Handbook of Phonetic Sciences*. Chichester: Wiley-Blackwell, 81-129.

Heinz, J.M., Stevens, K.N. (1961). On the Properties of Voiceless Fricative Consonants. In *The Journal of the Acoustical Society of America*, 33 (5), 589-596.

Hughes, G.W., Halle, M. (1956). Spectral Properties of Fricative Consonants. In *The Journal of the Acoustical Society of America*, 28 (2), 303-310.

Jaber, K., Jud, K. (1928-1940). *Sprach- und Sachatlas Italiens und der Siidschweiz*. Zofingen: Ringier.

Jongman, A., Wayland, R. & Wong, S. (2000). Acoustic characteristics of English fricatives. In *The Journal of the Acoustical Society of America*, 108 (3), 1252-1263.

Li, F., Edwards, J. & Beckman, M. (2009), Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. In *Journal of Phonetics*, 37, 111-124.

Maiden, M., Parry, M. (Eds.) (1997). *The Dialects of Italy*. London: Routledge.

Maniwa, K., Jongman, A. & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. In *Journal of the Acoustical Society of America*, 125, 3962-3973.

Mioni, A., Trumper, J. (1977). Per un'analisi del continuum linguistico veneto. In Simone, R., Ruggiero, G. (Eds.), *Aspetti sociolinguistici dell'Italia contemporanea*, Atti dell'VII Congresso Internazionale di Studi. Bressanone, 329-372.

Perco, D., Vigolo, M.T. & Sanga, G. (Eds.) (2011). Paul Scheuermeier - Il Veneto dei contadini 1921-1932. Costabissara: Angelo Colla Editore.

Rubino, A. (2014). L'italiano in Australia tra lingua immigrata e lingua seconda. In De Meo, S., D'Agostino, M., Iannaccaro, G. & Spreafico, L. (Eds.), *Varietà di contesti di apprendimento linguistico*, Studi AitLA 1. Milano: AitLA, 241-261.

Samarin, W.J. (1971). Salient and substantive pidginization. In Hymes, D. (Ed.), *Pidginization and creolization of languages*, Cambridge: Cambride University Press, 117-140.

Sancier, M.L., Fowler, C.A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. In *Journal of Phonetics*, 27 (4), 421-436.

Sebastián-Gallés, N., Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. In *Cognition*, 72 (2), 111-123.

Shadle, C.H. (2010[2]). The Aerodynamics of Speech. In Hardcasle, W., Laver, J. & Gibbon, F. (Eds.), *The Handbook of Phonetic Sciences*. Chichester: Wiley-Blackwell, 39-80.

Shadle, C.H. (2012). On the acoustics and aerodynamics of fricatives. In Cohn, A.C., Fougeron, C. & Huffman, M. (Eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press, 511-526.

Stevens, P. (1960). Spectra of fricative noise in human speech. In *Language and Speech*, 3 (1), 32-49.

Tekavčić, P. (1970[2]). *Grammatica Storica dell'italiano*, vol. I. Bologna: il Mulino.

Tisato, G., Barbierato, P., Ferrieri, G., Gentili, C. & Vigolo, M.T. (2013). AMDV Atlante Multimediale dei Dialetti Veneti, *Atti del IX Convegno dell'Associazione Italiana Scienze della Voce*. Roma: Bulzoni, 1-18.

Trubeckoj, N.S. (1971). *Fondamenti di fonologia*. Torino: Einaudi (ed. orig. 1938).

Trumper, J. (1972). *Il gruppo dialettale padovano-polesano. La sua unità, le sue ramificazioni.* Badia Polesine: Rebellato Editore.

Trumper, J. (1977). Ricostruzione nell'Italia settentrionale: sistemi consonantici. Considerazioni sociolinguistiche nella diacronia. In *Prolblemi Della Ricostruzione Linguistica*, 259-310.

Zamboni, A. (1974). *Veneto*. Pisa: Pacini Editore.

Zamboni, A. (1988). 270. Italienisch: Areallinguistik IV a) Venezien. In Holtus, G., Metzeltin, M. & Schmitt, C. (Eds.), *Lexicon der Romanistischen Linguistik*, vol. IV. Tubingen: Niemeyer Verlag, 517-538.

DOI: 10.17469/O2101AISV000006

FRANCESCA PINTO

High vowels elision in Japanese. A perceptive approach

The present study will focus on the perception by the native speakers of elided high vowels in Japanese and of deleted sounds in spontaneous speech. Study of mother tongues perception of the said phenomenon may serve as a capstone toward the acceptance of elided vowels in Japanese. It is a common belief among scholars that vowel elision may interfere with words comprehension, for the missing vowel would create a high rate of ambiguity, especially when dealing with minimal pairs. For this reason, the present study aims to demonstrate that words are easily understood despite vowel elision, through a perceptive test based on the recognition of 11 minimal pairs. Results revealed that vowel elision does not cause much trouble in the identification of the target words.

1. *Introduction*

Japanese high vowel devoicing is a well-known and studied phenomenon, examined by many native and non-native scholars[1] since the early stages of literature on Japanese linguistics.

High vowel devoicing is a phenomenon for which vowels are pronounced without the vibration of the vocal folds and where oral organ configuration is more similar to that of whispering: they are hence voiceless. A voiceless vowel can be easily distinguished from a fully voiced vowel for a series of acoustic peculiarities: it does not display any trace of the voice bar at the bottom of the spectrogram, it shows a sensitive increase of both F1 and F2, formants look less neat and sharp, pitch is absent, signal is aperiodic, there should be relatively little or no turbulent airflow during articulation[2], and overall the phone length is shorter. Vowel devoicing is more likely to occur in specific phonotactic contexts, specifically between two voiceless consonants or between a voiceless consonant and a pause. Factors like speech rate may also have a strong influence.

Under the same circumstances, high vowel elision may occur as well. Acoustically, a vowel is said to be elided when there is the complete absence of the phone with all its characteristics and therefore cannot be retraced. Depending on a higher speech rate, on the quality of the consonantal environment, and on the lexicon recurrence, this latter phenomenon may have an even higher degree of occurrence as compared to vowel devoicing. In fact, it is quite evident that some specific contexts require vowel elision more than fully voiced or devoiced vowels.

[1] See, for example, Akamatsu (1997), Maekawa (1993), Shibatani (1990), Vance (2008), Yuen (2000).
[2] This is needed to distinguish devoicing from frication, Yuen (2000: 532).

The reason for the alternation among vowel devoicing or elision is partly due to the ease of pronouncing certain consonantal clusters instead of others. A cluster (resulting from the elision of an interconsonantal vowel) composed, e.g., of a palatal fricative plus an alveolar occlusive is easier to pronounce than one composed of two occlusives. Additionally, word recurrence plays an important role in determining vowel elision regularity. Words like *desu* 'to be' or *shita* '(verbal morpheme for past tense)' are very frequent in Japanese and their realization with high vowel elision has gradually become a common habit of pronunciation which conflicts with the written interface of those words.

Notwithstanding neat phonetic evidences that prove vowel elision, traditionally only vowel devoicing is accepted by scholars in this field, especially by native ones[3]. It is possible to assume that the reason for that is mainly due to the influence exerted by Japanese writing system. Apart from Chinese ideograms and Latin characters, Japanese language is represented by *kana*, a syllabic writing system where each symbol stands for a mora composed of a vowel or a consonant plus a vowel, i.e. an open syllable. The only exceptions to this are represented by the symbol <n>, the only consonant that can occur alone and form a mora, thus closing the syllable, and by the symbol <tsu> conventionally used in a smaller size to mark intense consonants, which makes the preceding syllable closed as well.

As a consequence, it is not rare that native speakers experience difficulty in discriminating acoustic differences between vowel devoicing and elision. In other words, Japanese *kana* has influenced the perception of Japanese phonology and phonotactics so much that even scholars are led to think that only open syllables are possible – with the said exceptions – and hardly acknowledge vowel elision and consonantal clusters that derive from this process. Therefore, since speakers' intuition is not reliable in this case, a direct observation of acoustic data is needed.

The aim of the present study is to understand to what degree Japanese mother tongues' intuition, notoriously influenced by the writing system, is in contrast with the acoustic and perceptual reality. Investigating perception represents a key point to the full understanding of this research topic and of the possible drift of the language. This research may also constitute an original contribution to the research in this field, since perceptive tests are still not very popular in this branch of research.

Unfortunately experimental analyses on speaker perception are still not very popular in this field. On the other hand, investigating the native speakers' way of interpreting speech is an efficient method to confirm results from acoustic analysis, especially when interacting with phenomena that may be suspected of being in the middle of an ongoing linguistic change. What will be exposed here is partially examined in another study, dealing with the analysis of acoustic and

[3] See, for example, Akamatsu (1997: 50).

perceptive aspects of vowel reduction in Japanese, which has been conducted by Yuen[4].

Yuen starts from the acoustic analysis of six minimal pairs pronounced by eight native speakers – the large majority of the analysed vowels were elided – and successively used the same tokens as the basis for a perceptive test, which he conducted on forty native speakers. The experiment conveyed by Yuen hence originates from a different analysis basis, not only due to the number of tokens and subjects taken into account, but also for the fact that he considers also devoiced – in addition to elided – realizations of high vowels. On the other hand, in the present study only elided vowel's role will be analysed, since it is believed that devoiced vowels affect word recognition, despite being reduced. Notwithstanding this difference, it is particularly relevant that the author got experimental results that are similar to those illustrated in the present paper: even though most of the devoiceable vowels are deleted, the listeners did not seem to have problems recovering the underlying vowel identity.

Apart from the recognition of word including elided vowels, the present paper aims to demonstrate that vowel elision can be considered as an innovative feature in Japanese, also on the base of relevant acoustic analyses on diachronic and diatopic variants[5]. In fact, it is possible to postulate that an elided vowel may have consequences on Japanese's phonological system, mainly dealing with its palatalization correlation. As results will show, the correct understanding of the proposed stimuli occurs through coarticulation phenomena on the preceding consonant. It is well known that coarticulation is crucial in phone recognition and in the present experiment it plays a key role in determining the correct identification of the proposed stimuli, even though it occurs at different degrees according to the nature of the consonant. This process constitutes a matter of importance, because phonetic traces deriving from elided vowels allow the correct transmission of the message informative content with no comprehension inconvenience. The said phenomena, being a constant feature of consonants followed by front vowel, would enrich the already existent correlation, opposing consonant followed by non-front vowels to the palatalized ones.

1.1 Expected results

Expected results will show that the listener will be able to correctly identify most of the proposed words, notwithstanding vowel elision. The reason for this assumption relies on the fact that phonetic traces still remain in the coarticulation of the preceding consonant. These traces will constitute the necessary hint for the right comprehension of the proposed stimuli.

For example, if we take the pair *kisai* and *kusai*, we can expect that the speaker will decode the proposed input due to the different pronunciation manner of

[4] Yuen (2000).

[5] See, for example, Byun (2011), Frellesvig (2010), Maekawa (1993), Shibatani (1990).

/k/, [kʲ] in the first case and [k] in the second, coherently with the palatalization correlation already present in this language.

2. *Methodology and collected data*

2.1 Subjects

Fifty Japanese native speakers with no apparent auditory problems participated as volunteers. Other than a privacy and a safety agreement form, subjects were asked to fill in an anonymous form dealing with information about their personal and linguistic history such as: sex, age, birthplace, living place, number of years spent in the living place, mother's birthplace, father's birthplace, and highest education level obtained.

The information required was needed because it is not rare that subject's pronunciation may vary greatly according to sociolinguistic factors. In particular, diatopic related information was needed in order to understand the degree of influence of the dialect on missing vowels discrimination. It is a well-known fact[6] that the phone reduction phenomena may vary according to the spoken language geographic variety. For example, Japanese speakers from the Kansai area[7] are more used to the rounded high back vowel, which is more audible and neater than its unrounded counterpart, typical of the Standard Japanese.

On the other hand, information related to the education level was needed in order to evaluate the degree of awareness of the first language speakers towards their native language. In fact, higher educated speakers are considered to have a more accurate knowledge of language mechanisms.

Table 1 - *Speakers dialectal background related data*

	Birthplace	*Mother's birthplace*	*Father's birthplace*
Hokkaidō	0	0	0
Tōhoku	2	2	2
Kantō	35	24	24
Chūbu	9	8	8
Kansai	1	1	0
Chūgoku	2	1	1
Shikoku	0	0	0
Kyūshū/Okinawa	0	0	0
Other	1	0	0

[6] Maekawa (1993).

[7] Kansai area includes Kyōto, Ōsaka and Kobe dialects, among others.

[8] In the following table data related to mother and father birthplace were reported only when they coincide with those of the speakers' birthplace. In this way it will be easier to verify the dialectal influence of the parents' birthplace on the subjects.

Table 2 - *Personal data related to the 50 speakers*

Sex		*Average age*	*Years spent in Tōkyō*	*Education level*[9]			
M	F			H	B	M	P
46	4	24,64	16,33	13	28	6	3

2.2 Stimuli and task

The test is based on a series of 11 minimal pairs (i.e. 22 lexical items). Selected words are extracted from an interview expressly created for this research purpose. A native speaker with no apparent pronunciation problem or strong accent was asked to pronounce each word with the elided pronunciation of the vowels involved. Words were then carefully checked one by one with Praat to make sure that each target vowel was elided. Words were chosen according to the vowel quality and to the consonantal environment. Furthermore, particular attention was paid in order to have a similar speech rate and not to combine rare and common words so that the listeners would not be influenced by the lexicon occurrence rate. It was verified that all the subjects were familiar with the chosen words.

Pairs including meaning differences caused by /i/ and /u/ in the same position within the word were selected so that vowels occurred both in low pitched and in unaccented morae. Due to a palatalization process occurred in the 16th century[10], many voiceless consonants shifted place or manner of articulation or both in front of high vowels, producing what are now complementary distributed allophones, e.g.:

/t/+/i/→ [tɕ], vs. /t/+/u/→ [ʦ] ([t] otherwise);

/h/+/i/→ [ç], vs. /h/+/u/→ [ɸ] ([h] otherwise)[11].

This is the reason why only [k] and [ɕ] are used here, for they are the only voiceless consonants that can contrastively occur before both high vowels (note that, while [ɕi] is still a result of the palatalization of /s/+/i/, it is nevertheless contrasted with [ɕu] coming from the palatalization of the cluster /si/ followed by /u/). Considering this, and in order to match minimal pairs in the most convenient way, all the other voiceless consonants have been excluded. As for the following consonant an attempt was made to vary the possible phonotactic contexts.

Selected words are the following:

[9] H stands for High school; B stands for Bachelor degree; M stands for Master degree; P stands for PhD.

[10] Frellesvig (2010: 322).

[11] Japanese is among the few languages that allow palatalization phenomena in front of a back vowel. See Bath (1978: 53).

Table 3 - *Selected pairs*

Word	Transcription	Meaning	Pair	Transcription	Meaning
Shishoku	/ɕiɕokɯ/	Sampling food	*Shushoku*	/ɕɯɕokɯ/	Staple food
Kishin	/kʲiɕin/	Donation	*Kushin*	/kɯɕin/	Effort
Ishitsu	/iɕitsɯ/	Heterogeneous	*Ishutsu*	/iɕɯtsɯ/	Export
Kakite	/kakʲite/	Writer	*Kakute*	/kakɯte/	Thus
Yukisaki	/jukʲisakʲi/	Destination	*Yukusaki*	/jukɯsakʲi/	Future
Kikai	/kʲikaj/	Machine	*Kukai*	/kɯkaj/	Recital
Kisai	/kʲisaj/	Mention	*Kusai*	/kɯsaj/	Smelly
Kichō	/kʲiteo:/	Precious	*Kuchō*	/kɯteo:/	Tone
Shikan	/ɕikan/	Official	*Shukan*	/ɕɯkan/	Subjective
Shitai	/ɕitaj/	Corpse	*Shutai*	/ɕɯtaj/	Subject
Shichō	/ɕiteo:/	Mayor	*Shuchō*	/ɕɯteo:/	Opinion

For the aim of this test subjects were asked to listen to each of the 22 words and to choose between two provided alternatives forming the minimal pair, namely the word containing /i/ and the one containing /u/. Overall, subjects needed one to three hearings per each word. Total time needed to complete the test was less than three minutes.

2.3 Analysis of data

Words pronounced by the native speaker within a sentence have been cut and then proposed to listeners without the help of a phonetic or a semantic context. This process was needed so that the listener would have not been influenced by any other element that might have interfered with words identification.

For word segmentation the Praat software, version 5.3.49, has been used. The test has been developed using the Max Runtime software, version 6.1.7. Metadata was installed into the said software and selected words were reproduced randomly. The said software was programmed in order to obtain a yes/no output for each audio segment, corresponding to the correct or incorrect guess of the word including the elided vowel. Test answers were automatically collected in a database and statistically analysed using the correction for guessing formula[12], a method that clarifies the link between the score on the test and the subject's real understanding.

The test was run with the support of Sony stereo headphones MDR-XD100 and was conducted in a soundproof room.

[12] Chiorri (2011: §4.7)

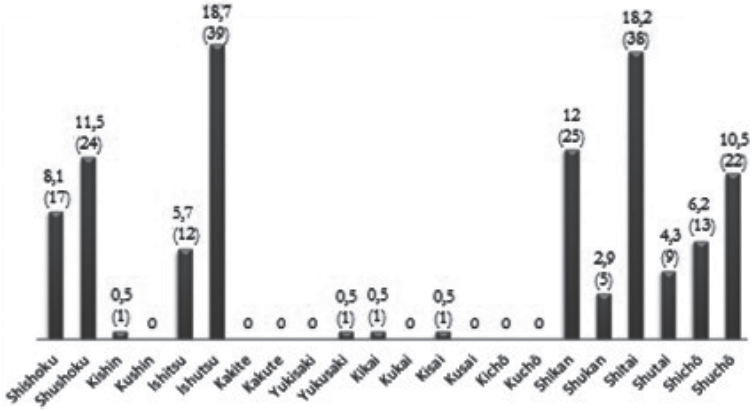
3. *Test results*

Here follow tables and graphs that illustrate results from the perception test.

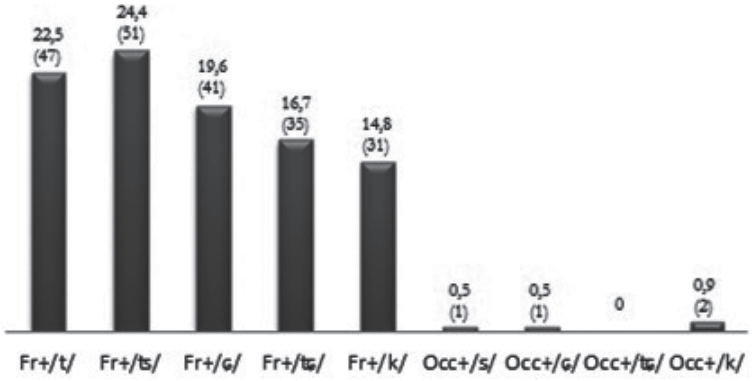
Table 4 - *General results*

Tot. amount of proposed stimuli	*Tot. amount of identified words*	*Tot. amount of mistaken words*
1100	891	209
100%	81%	19%

Graph 1 - *Word related error rate plus error total amount in brackets*



Graph 2 - *Phonotactic context related error rate plus error total amount in brackets*



Results from the test based on Japanese minimal pairs seem to confirm the above mentioned hypothesis according to which native speakers can correctly discriminate words involving high vowel elision.

The total amount of proposed stimuli was 1100 and the total number of errors committed was 209. Using the correction for guessing formula and considering that

test responses might have been the result of a random guess, it is possible to state that the total amount of correctly guessed words is 62%.

As graphs 1 and 2 show, the number of errors differs highly depending on the phonetic context, but the question is which part of the context is more influential. It seems that the timbre of the vowel involved does not have a significant influence on the error rate for the number of errors related to /i/ is 108 and 101 related to /u/. Nor does the following consonant have a strong influence on the error production. Not surprisingly, it is the preceding consonant that exerts a most considerable impact on the occurrence, which is due to some very clear phonetic reasons. This is quite evident when comparing error rate related to words where an elided vowel is preceded by a fricative to those preceded by an occlusive, for coarticulation reasons that will be explained later; while in the first case the number of errors is 205, in the latter case it is only four.

Another point of interest is represented by the eight out of twenty-two words that were never mistaken, i.e. *kushin*, *kakite*, *kakute*, *yukisaki*, *kukai*, *kusai*, *kichō*, *kuchō*, which are the ones that involve [k], regardless of the following vowel or consonant. Even the mistaken words where [k] is the preceding consonant were misunderstood only once per each word. On the other hand, ambiguity caused by [ɕ] is much more relevant, even though the number of errors differs significantly depending on the word. The lowest number of errors, 6, is related to the word *shukan* 'manager' while the highest, 39, goes with the word *ishutsu* 'export': using the correction for guessing formula and assuming that the maximum error rate here is 50%, for the subjects only had two alternatives, the error rate of the word *shukan* is 76%, while the error rate of ishutsu is 56%. There are no single words mistaken by all participants.

It seems quite evident at this stage that coarticulation caused by elided vowels has a different effect depending on the quality of the preceding consonant. Stops seem to better preserve coarticulation – a clear example of this is represented by aspiration that usually follows the phone burst when combined with /i/ – making words identification much easier. On the contrary, coarticulation on fricatives mostly disperses and only a few traces remain: this is the main reason why a higher degree of ambiguity occurs in coincidence with these phones.

For what concerns subjects' related data, in this test the average number of errors is about four per speaker, with a minimum of two errors and a maximum of seven. The majority of the test subjects were from Tōkyō and surroundings, so it is possible to say that this sample may represent a good mirror of the perception of this phenomenon in the standard variety of Japanese (which correspond to the dialect of the capital). Dialectal differences do not seem to have a great influence on perception as answers given by speakers from other regions are totally comparable to those given by Tōkyō speakers.

4. *Discussion*

Results from this test are quite interesting for two main reasons: they constitute new support material for the hypothesis according to which Japanese phonology may eventually include elided vowels and consonantal clusters that would derive from it and they can contribute to the advancement of the state of the art of missing phones recognition in speech.

These results show that elided vowels are well tolerated perceptually. In general, we can state that notwithstanding differences in the degree of identification of words including vowel elision, missing phones do not consistently obstacle word recognition and do not cause any relevant trouble to the message transmission. We should not forget that in spontaneous speech and in everyday language the number, type of errors, and ambiguities would be even lower, due to the role of semantic context, of sentence intonation and of all the other mechanisms involved in oral communication, whereas in the present study words where proposed in isolation. Moreover, in spontaneous speech phones elision is quite frequent, especially when the speech is inaccurate and at a high rate. For these reasons, speakers already have the habit to interpret fragmented speech.

It is possible to state that when speaking, people are more interested in the mere pragmatic aspects of the language and the message transmission is a priority over the speech hyperarticulation. It frequently happens that in a linguistic act, especially when inaccurate, there is an apparent high number of imperfections and gaps: it is not rare that speakers disregard the speech form and orthoepy and prefer to focus more on other aspects of the communication. In this way elisions or abbreviations often take place, being them of whole words or small fragments. When there is a good mutual comprehension, speakers will also benefit from an articulatory and mental saving, without drawing upon an orthoepic and accurate speech. In other words, during a speech act we do not hyperarticulate sounds, but we rely upon conversational redundancy. This allows us to economize on articulatory effort when we can count on listener's cooperation.

The dialectic balance among interlocutors finds its basis on the principle of pragmatic cooperation principle[13]. According to this principle, complicity and participation are established among people involved in a conversation, so that orthoepy no longer represents the most important parameter. Speakers, in fact, capitalise on their lexical competence and activate a meaning recognition mechanism – when the discourse is pertinent – in order to trigger phonetic perception. Phone elision is certainly able to generate ambiguities, especially as related to minimal pairs, in languages with a high rate of homonymy. The context then, plays a fundamental role in determining the semantic field of a conversation in case of ambiguities. Hence, the listener will not be lead to identify the missing phone with the wrong vowel, but will understand without many efforts the pronounced words, even though an inaccurate pronunciation or a high speech rate might have provoked vowel elision.

[13] Grice (1975: 45).

4.1 The role of coarticulation in speech recognition

As mentioned above, one of the most important perplexities raised by scholars in this field deals with speech recognition, or to put it into other words, the impossibility to correctly identify words including vowel elision[14]. The test presented above showed that stimuli recognition, especially when detached from the context, occurs mostly due to vowel coarticulation – even when not pronounced – on the preceding consonants, representing somehow the last trace of its presence.

Results from the perceptive test on mother tongue speaker's perception showed that in the vast majority of cases subjects are able to discern minimal pairs of words where the main obstacle was represented by the elision of the vowel involved. The given explanation was that subjects focused their attention to the acoustic environment of the elided vowel – mostly on the preceding consonant – to clearly label each proposed stimulus. In other words, we can state that the effect of coarticulation on the preceding consonant might have been the main feature for the subjects to give the correct answer. This process takes place because a linguistic chain is not made by single sounds, separated among them, pronounced and perceived in isolation. We can state instead that sounds tie up[15].

Hence, in the missing phone recognition process and, more specifically, when dealing with Japanese high vowels, coarticulation plays a very important role[16]. In fact, as stated above and as Kawakami[17] notes, vowel identification, even though elided, occurs at any rate through coarticulation clues exerted by /i, u/: more specifically, it ensues in a more advanced position in the first case and in a more retracted one in the second case. We can expect that the said mechanism takes place also in conjunction with a devoiced vowel, as well as with a fully voiced one. The fact that this may take place also coincidentally with an elided vowel is something that raised experts' doubts and perplexities, whose most consistent objection is that, summing up what Kondo[18] says, if a vowel no longer exists, it shall not cause coarticulation. Even though we can state that a missing sound may no longer provoke an influence on the phonotactic context, not being part of it anymore, we should not forget that vowel elision is the result of a process for which a vowel undergoes many reductions going from a status of full sonority to its complete elimination. Note that during this process the subsequent consonant can retain the coarticulation effects, thus leading to the split of the allophones.

What has been said can give elucidations about some scholars doubts – among them Akamatsu – according to whom the recognition of words including vowel elision might be highly undermined by this phenomenon and the ambiguity degree would be even higher when dealing with minimal pairs. In fact, if we consider the coarticulation impact, word related phonetic image would be certainly differ-

[14] See for example Akamatsu (1997: 50).
[15] Nusbaum, DeGroot (1990: 296).
[16] Labrune (2006: 51).
[17] Kawakami cited by Vance (1987: 54).
[18] Kondo (1997: 14).

ent and this is quite evident if we try and give a narrow phonetic transcription of the words considered here. For example, if we take the pair *kishi* 'shore' /kiɕi/ and *kushi* 'comb' /kɯɕi/, in case of vowel elision we would not have, as many scholars say, a pronunciation like [kɕi] in both cases, but [kʲɕi] in the first case and [kɕi] in the second. To put it into other words, a slightly more advanced pronunciation of the occlusive is the key that would grant speech recognition without many efforts. As Vance states[19], the phoneme /k/ in *kishi* is realized with the allophone expected in front of /i/, while the /k/ in *kushi* with the allophone expected in front of /ɯ/.

It shall be noted that at the present stage of Japanese, apart from the exception of /n/ and of geminates (in case we interpret them as clusters of identical consonants), its phonological system only allows the presence of a vowel preceding or following a consonant. Therefore, if we pronounce the phoneme /tɕ/ it will necessarily follow /i/; in the same way, after the phoneme /ɸ/ there must be /ɯ/[20].

As mentioned above, Japanese phonotactics is mainly composed of CV sequences and it goes without saying that the presence of an elided vowel would lead to relevant modifications of the simply panorama of its phonotactic rules. We may postulate, in fact, an increase of possible combinations of consonants and vowels and above all a wider number and types of consonantal clusters. This process is absolutely coherent with the present system and with the naturalness of the linguistic types. Even though this is not the place where to exhaustively discuss this topic, we can briefly postulate the creation on the following syllable structures: (1) VC; (2) CVC; (3) CCV; (4) CyVC; (5) CCVN; (6) CCVQ; (7); CCVR.

4.2 Possible effects of high vowels elision on the palatalization correlation

It the introductory paragraph it was mentioned that we are probably observing a mutation process in act, regarding the acceptance and the inclusion of the phonological consequences that would arise from vowel elision, primarily the increase of the quantity and the complexity of consonantal clusters.

It was said that coarticulation played a fundamental role in stimuli recognition. Coarticulation may sometimes induce cases of allophony, most of the time due to the influence of adjacent sounds in the speech chain: in some cases new forms prevail and become phonologically stable. A further step is when the two allophones split into two phonemes because of the elision of the conditioning context, such as the vowel elision under consideration here.

It is well known that allophones participate in speech recognition in that they anticipate the following phone, for example, as briefly seen above, [ç, tɕ, dʑ] precede [i], as much as [ɸ, ʦ, ʣ] are connected with [ɯ], being the allophones of /h, t, d/. Since reduced vowels cause coarticulation on the preceding consonants and being it a steady phenomenon, it is possible to state that they may originate new allophones. If we take a closer look at the sequence [ki], taken into account for the experiment

[19] Vance (2008: 209).

[20] Exceptions deriving from loanwords allowing other combinations are not taken into account here.

presented here, we will see that it contemplates an aspiration after the VOT which is intrinsic to its pronunciation, making this sound tend to palatalization. The said characteristic allows this sound to be clearly distinguished from occurrences where /k/ is followed by other vowels[21]. Moreover, aspirated realization of /k/ in front of /i/ is a constant and regular phenomenon: it is not an irregular or a sporadic circumstance, but it has a peculiar stability. Even from a perceptive point of view, as the test results show, the identification of the said articulation when not followed by a vocalic phone takes place through the acoustical recognition of the frication accompanying the consonant, making values similar to those of a palatal consonant. Because of this recognition process, subjects will overtake the potential conditioning caused by the image of the writing system that a speaker may take as a datum point.

Given the identification degree, it is possible to postulate that in this language the palatalization correlation is enlarging, with the inclusion of new palatalized consonants. In other words, we can hypothesize an increase of the correlation through the phonologization of the allophone [kʲ], which is already existent in this language. Supposing that high vowels elision might become a more stable reality, it can lead to the consequence of the phonologization of the palatalized allophone of /k/. In support of this hypothesis, it should be reminded that at the present stage of this language the palatalized variant of /k/ already appears in front of /a, u, o/ and in front of /e/ in case of loanwords. This sound, hence, is not only existent in Japanese, but is also part of the speakers' phonetic and articulatory habits.

It is a matter of fact that languages are subjected to constant changes which are most of the time coherent with its system, in other words, they follow their drift. Languages are economical systems where nothing is superfluous: when changes occur they have the function to balance what other changes has made unstable. As Lazzeroni says[22], since every system tends to its balance, when a language experiences some irregularities, balance will be restored through other changes. It is clear that the unbalanced situation caused by vowel elision may be successfully rebalanced through the enlargement of palatalization correlation, other than a new way of structuring morae and syllables. When a phonological system already displays that correlation, there are higher chances that it will be enriched through consonantal phonemes not yet integrated in it. The author also illustrates[23] that the presence itself of an opposition leads the system to extend that distinctive feature to more phonemes, through a new modification.

At this point, it is possible to presume that the influence of the elided vowels on the preceding consonant may lead to the creation of new minimal pairs, through new palatalized consonants that would enhance the already prolific palatalization correlation in this language. At the moment these palatalized consonants shall be considered as context determined allophonic variations, since within speech from

[21] A clear example of this, pertaining to the couple *kishi* – *kushi*, is exposed in the previous paragraph.
[22] Lazzeroni (1998: 50).
[23] Lazzeroni (1998: 46).

the same speaker, realizations of the same lexemes occur both with elided or devoiced vowel, according to the diaphasic variance and, consequently, to the speech rate. Hence, we must suppose that, from a paradigmatic point of view, these vowels are phonologically present for the speakers. However, in case vowel elision would predominate in the course of time in a consistent group of speakers, we should hypothesize that a phonologization process of the palatalized allophones of these consonants might have been established. The said process would lead to an enrichment of the already existent palatalization correlation in Japanese phonology. Even though these consonants are not phonologized yet, these allophones constitute a good prerequisite to hypothesize that this phonologization process may eventually take place in Japanese.

Acknowledgements

The author would like to express her gratefulness to Prof. Claudia Ciancaglini, Prof. Andrea Paoloni, Prof. Artemij Keidan, Prof. Shinoda Koichi, Prof. Maekawa Kikuo, Prof. Timothy J. Vance and Prof. Kibe Nobuko for their precious comments and suggestions. Many thanks also to Prof. Ozawa Naoko for her assistance with the Japanese based test and to Dr. Massimiliano Todisco from the Fondazione Ugo Bordoni that helped with the engineering part of the experiment. The author is also grateful to the fifty subjects that voluntarily participated in this research.

References

Akamatsu, T. (1997). *Japanese Phonetics: Theory and Practice*. Munchen, Newcastle.

Bhat, S.D.N. (1978). A General Study of Palatalization. In Greenberg, H. (Ed.), *Universals of Human Language*, vol. 2. Stanford, California: Stanford University Press, 47-92.

Byun, H. (2011). High Vowel Devoicing in the Keihanshin Area of Japan. In *The Journal of the Acoustical Society of Japan*, vol. 15 n. 2, 23-37.

Chiorri, C. (2011). *Teoria e tecnica psicometrica. Costruire un test psicologico*. McGraw-Hill Companies S.r.l., Publishing Group Italia.

Dupoux, E. et al. (1999). Epenthetic Vowels in Japanese: a Perceptual Illusion? In *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1568-1578.

Frellesvig, B. (2010). *A History of the Japanese Language*. Cambridge, University Press.

Grice, H.P. (1975). Logic and Conversation. In Cole, P., Morgan, J. (Eds.), *Syntax and Semantics*, vol. 3. New York: Academic Press, 41-58.

Kondo, M. (1997). *Mechanism of Vowel Devoicing in Japanese*, unpublished PhD thesis.

Jakobson, R. (1973). *Essais de linguistique générale*. Ruwet, N. (Ed.), Paris: Les éditions de minuit.

Labrune, L. (2006). *La phonologie du japonais.* Leuven, Paris: Peeters.

Ladefoged, P., Johnson, K. (2010). *A Course in Phonetics*, 6th edition. Cengage Learning.

Ladefoged, P., Broadbent, D.E. (1957). Information Conveyed by Vowels. In *The Journal of the Acoustical Society of America*, vol. 29, n. 1, 98-104.

Maekawa, K. (1993). Boin no museika. In Miyaji, S. et al. (Eds.), *Kōza nihongo to nihongo kyōiku, 2 Nihongo no onsei – on'in*, vol. 1. Tōkyō: Meiji Shoin, 135-153.

Nusbaum, H.C., Francis, A.L. & Luks, T. (1997). Speech Perception: A Special Mechanism or a Specialized Cognitive Process? In *CLS* 33, vol. 2, 415-434.

Shibatani, M. (1990). *The Languages of Japan*. N.Y.: Cambridge University Press.

Vance, T.J. (1987). *An Introduction to Japanese Phonology*. Albany, N.Y.: State University of New York Press.

Vance, T.J. (2008). *The Sounds of Japanese*. Cambridge: University Press.

Vygotsky, L.S. (1980). *Pensiero e Linguaggio*. Massucco Costa, A. (Ed.), Firenze: Giunti.

Yuen, C.L.K. (2000). The perception of Japanese Devoiced Vowels. In *CLS* 36, 531-547.

DOI: 10.17469/O2101AISV000007

CAMILLA BERNARDASCI

Aspetti quantitativi del vocalismo tonico del dialetto di Piandelagotti

The present study investigates some quantitative aspects of the tonic vocalism of the western Emilian dialect of Piandelagotti based on a corpus of data collected in 2008. It considers the relation between tonic vowel and post-tonic consonant lengths in three different syllabic structures: in open syllable and in vowel lengthening contexts, the vocalic length is (nearly) double the consonantal length, whereas in closed syllables they tend to be equivalent. Further investigation of the postonic contexts in closed syllables revealed differences in the respective V/C rating. Thanks to the result obtained in the second part of the analysis dedicated to the distinctiveness of vowel quantity (V/V: 0.55), it is possible to place the Piandelagotti dialect under the Gallo-italic dialects.

1. *Introduzione*

Piandelagotti è una località dell'Appennino modenese in cui si parla un dialetto alto-frignanese. Secondo la classificazione di Pellegrini (1977) Piandelagotti si trova in area emiliana occidentale sul confine che divide il territorio a nord (o a occidente) e il territorio a sud (o a oriente) del fiume Panaro (cfr. Foresti, 1988: 575). Il piandelagottese è quindi un dialetto emiliano occidentale di tipo conservativo, come trapela dal vocalismo atono finale, che vede mantenuta, perlomeno in posizione prepausale, l'opposizione tra -/a/ e -/ə/ (cfr. Loporcaro et al., 2007: 71; Loporcaro, 2014: 171, 190). La sua posizione isolata a ridosso del fascio di isoglosse La Spezia-Rimini o, secondo Pellegrini (ad es. 1992: 285), Carrara-Fano, ha fatto sì che suscitasse l'interesse di diversi studiosi. Già all'inizio del secolo scorso Giuseppe Malagoli si è dedicato allo studio della distribuzione dell'articolo maschile singolare a Piandelagotti (cfr. Malagoli, 1910-1913); in tempi più recenti gli interessi si sono orientati sul livello strutturale fonetico-fonologico: Loporcaro et al. (2007) si sono occupati, in particolare, del vocalismo atono finale, mentre Biasini (2010) ha studiato il sistema fonetico e fonologico di questa varietà; Bernardasci (2013b) ha condotto un'analisi fonetica di alcuni aspetti del sistema vocalico e consonantico.

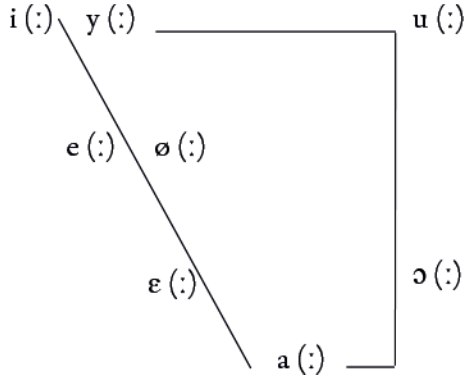
Dalla bibliografia sul piandelagottese emergono i seguenti tratti pertinenti al vocalismo tonico: si ha differenza di trattamento tra vocali in sillaba aperta e vocali in sillaba chiusa e si registra l'insorgere di una nuova quantità vocalica distintiva. Nel vocalismo atono sono frequenti le sincopi delle vocali interne in posizione postonica. Per quel che riguarda i principali fatti del consonantismo, a Piandelagotti si ha regolarmente lenizione delle consonanti scempie intervocaliche; la degeminazione è sistematica in posizione protonica e ha raggiunto uno stadio molto avanzato an-

che in postonia (come emergerà dai risultati relativi al rapporto tra vocale tonica in sillaba chiusa e consonante postonica, per cui cfr. § 3.2).

Dato questo quadro di partenza, lo scopo del presente contributo è quello di verificare empiricamente alcuni tratti cruciali del vocalismo tonico del dialetto piandelagottese: si calcoleranno i rapporti tra la durata della vocale tonica e della consonante postonica e si verificherà il livello di distintività della quantità vocalica.

Sarà quindi utile, a questo proposito, illustrare le principali caratteristiche del sistema vocalico tonico della varietà in questione sia in ottica sincronica, sia in ottica diacronica.

Figura 1 - *Vocalismo tonico piandelagottese in sincronia*



In sincronia (cfr. figura 1) il dialetto di Piandelagotti conta otto vocali toniche articolate su quattro gradi di apertura. La lunghezza della vocale è sensibile alla struttura sillabica: si hanno vocali lunghe in sillaba aperta e vocali brevi in sillaba chiusa. Inoltre, provocano allungamento vocalico i contesti consonantici di sonorante seguita da ostruente (per cui cfr. oltre, § 3.1). Sono brevi le vocali toniche dei proparossitoni.

In diacronia il sistema vocalico piandelagottese si è evoluto come illustrato nella tabella 1, dove si distinguono gli esiti in sillaba aperta (σ[) e in sillaba chiusa (σ]):

Tabella 1 - *Vocalismo tonico piandelagottese in diacronia*

	Ī	Ĭ, Ē	Ĕ, AE	A	Ŏ	Ō, Ŭ	Ū	AU
protorom.	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	/ɔ̯/
σ[	/iː/	/ɛː/	/eː/	/aː/	/øː/	/uː/	/yː/	/ɔː/
σ]	/i/	/ø/	/ɛ/	/a/	/ɔ/	/u/	/y/	

Oltre che delle differenze di durata vocalica, la struttura sillabica è responsabile anche del diverso trattamento timbrico. In sillaba aperta, infatti, nelle vocali medie del ramo palatale si registrano inversioni timbriche, per cui da lat. ĭ, ē > [ɛː] ([ˈnɛːʊa] 'neve') e da lat. ĕ > [eː] ([ˈɟeːza] 'chiesa'); in sillaba chiusa si registrano, rispettiva-

mente, [ø] e [ɛ]: pdg. [ˈkwøʃtə] 'questo' e [ˈtɛrˑa] 'terra'. Non presentano caratteristiche particolari le vocali che continuano lat. ī e ă ([ˈviːta] 'vita', [ˈviɲˑa] 'vigna', [ˈtaːna] 'tana', [lat(ˑ)] 'latte'), mentre palatalizzazioni sono attestate per il ramo velare: /y(ː)/ < ū ([ˈbryːʒa] 'brucia', [saˈlyt] 'saluto'), /øː/ < ŏ (ma solo in sillaba aperta, [ˈkøːgə] 'cuoco', mentre in sillaba chiusa si ha regolarmente /ɔ/, [ˈkɔʃˑa] 'coscia'). I continuatori di ō, ŭ sono andati ad occupare il timbro lasciato libero in seguito alla palatalizzazione dell'esito di ū: [ˈsuːlə] 'sole', [ˈkup(ˑ)a] 'coppa'.

Come già accennato, in posizione atona finale in piandelagottese è mantenuta l'opposizione tra -/a/ e -/ə/: le vocali finali non basse non hanno quindi (ancora)[1] subito sistematicamente l'apocope, ma soltanto una neutralizzazione, in posizione prepausale, nella vocale centrale schwa[2]. In posizione interna questa vocale può dileguarsi: [e ˈbratːʃə] 'il braccio', [ke ˈbratːʃ ki e ɱ ˈfa um ˈmą̣lʌ] 'questo braccio qui mi fa un male...' (cfr. Loporcaro et al., 2007: 72)[3].

2. *Dati e metodo*

2.1 Dati

I dati sui quali si basa l'analisi effettuata per questo contributo sono stati raccolti a Piandelagotti in occasione di un'escursione dialettologica organizzata nel 2008 da Michele Loporcaro e Stephan Schmid con un gruppo di studenti dell'università di Zurigo, dei quali l'autrice di questo contributo non faceva (ancora) parte[4]. Le registrazioni sono state realizzate in ambienti poco rumorosi (ma, purtroppo, non insonorizzati) con un registratore digitale DAT SONY TCD-D e con un microfono SONY ECM 959A. I ricercatori hanno sottoposto ai parlanti dei questionari con stimoli in italiano da tradurre in dialetto. Purtroppo all'epoca non si è chiesto agli informatori di inserire le parole bersaglio in frasi cornice (procedura che avrebbe garantito lo stesso contesto prosodico per ogni parola escussa), per cui non si hanno contesti omogenei o equivalenti per ogni realizzazione. Il materiale audio è stato trasferito su computer e salvato in formato .wav con una frequenza di campiona-

[1] Si può infatti supporre, insieme a Loporcaro et al. (2007: 57), che «anche i dialetti settentrionali siano passati, in uno stadio predocumentario, per una fase in cui le vocali atone finali poi cancellatesi fossero dapprima soltanto centralizzate».

[2] «La bibliografia dialettologica fra fine Ottocento e primi del Novecento segnala tre sole varietà emiliane rispondenti a queste caratteristiche: i dialetti di Sillano e Gombitelli, *enclaves* in Lucchesia descritte da Pieri (1893a, b), e il dialetto alto-frignanese di Piandelagotti, in provincia di Modena» (Loporcaro et al., 2007: 58).

[3] I dati raccolti da Loporcaro et al. (2007) mostrano che, in realtà, a differenza di quanto registrato all'inizio del secolo scorso da Malagoli, «nel piandelagottese odierno [ə] può essere opzionalmente cancellato anche nella pronuncia di parole in isolamento» (*Ibidem*: 72).

[4] Desidero ringraziare gli informatori piandelagottesi che hanno contribuito a rendere possibile questo studio, Ferdinando Lunardi, Lina Manattini ed Emiliano Mazzoni. Per i ragguagli e i consigli ricevuti nel corso del lavoro ringrazio Lorenzo Filipponio, Marie-José Kolly (in particolare per l'analisi statistica), Michele Loporcaro, Davide Mantilla, Dieter Studer-Joho e Stephan Schmid.

mento di 44100 Hz ed è attualmente conservato all'Archivio fonografico dell'università di Zurigo.

Il *corpus* sul quale si basa la presente analisi è costituito da una selezione di registrazioni di tre parlanti, due uomini e una donna, nati tra il 1928 e il 1978, come indicato nella tabella 2.

Tabella 2 - *Parlanti analizzati e numero di occorrenze*

Sigla	*Sesso*	*Data di nascita*	*Numero di occorrenze*
FL	m	*1944	240 entrate: 99 ossitoni, 141 parossitoni
EM	m	*1978	184 entrate: 67 ossitoni, 117 parossitoni
LM	f	*1928	226 entrate: 95 ossitoni, 131 parossitoni

Nel *corpus* analizzato sono confluite in totale 650 entrate (261 parole ossitone e 389 parole parossitone). Le voci sono state elicitate perlopiù in posizione isolata, senza l'ausilio di frasi cornice. Oltre che della struttura di parola, di quella sillabica, e, naturalmente, dei timbri vocalici, nella selezione delle occorrenze da integrare nel *corpus* si è tenuto conto dei contesti consonantici postonici (per cui, tra le consonanti che chiudono la sillaba tonica si hanno 148 casi di ostruente sorda, 130 casi di ostruente sonora, 184 casi di sonorante e 100 casi costituiti da nessi di consonante e ostruente). Le voci con vocale tonica in sillaba aperta sono 88. Purtroppo in occasione delle inchieste sul campo non è stato sottoposto ai parlanti un questionario dedicato alla verifica della quantità vocalica distintiva, motivo per cui per queste analisi si è quindi stati costretti ad allestire un sotto-*corpus* costituito da 39 coppie (semi)-minime: ad es. [di] 'giorno' ~ [diː] 'dito', [pø] 'può' ~ [bøː] 'buoi'.

2.2 Metodo

Le registrazioni sono state segmentate ed etichettate manualmente in *Praat* (versione 5.4.04): si sono isolate le vocali toniche (facendo riferimento all'andamento di F1 per segnarne il confine d'inizio e all'andamento di F2 per stabilirne la fine) e le consonanti postoniche. La misurazione e l'esportazione delle durate vocaliche e consonantiche sono state effettuate automaticamente grazie a uno *script* di *Praat*. I calcoli dei rapporti tra la durata della vocale tonica e la durata della consonante postonica (V(ː)/C), così come quelli tra vocali brevi e lunghe (V/Vː) e tra consonanti dopo vocale lunga e consonanti dopo vocale breve ((Vː)C/(V)C) in contesto di coppia (semi-)minima sono stati effettuati con *Excel*. Le verifiche statistiche dei risultati sono state messe a punto con il programma R (R core Team, 2013), in particolare con R package *lme4* (Bates, Maechler, 2009) e R package *multcomp* (Hothorn et al., 2008). I dati sono stati analizzati sulla base di modelli lineari a effetti misti (*linear mixed effect models* (LMEs)), i quali permettono di fare delle predizioni tenendo conto sia degli effetti fissi (nel nostro caso, ad esempio, il parametro 'struttura sillabica'), sia di quelli casuali (ad esempio il parametro 'parlante') (cfr. Kliegl et al., 2011): eventuali effetti sono stati testati confrontando un *full model* nel quale il fattore in questione – ad esempio, nel §

3.1, la struttura sillabica – era codificata come un effetto e un *reduced model* nel quale il fattore era escluso. I *p-values* sono stati ottenuti confrontando i due modelli sulla base di test standard per verificare la probabilità di un rapporto.

3. *Risultati*

3.1 Rapporto tra vocale tonica e consonante postonica

Sulla base delle misurazioni delle durate vocaliche (toniche) e consonantiche (postoniche) delle entrate comprese nel *corpus* si sono calcolati, in primo luogo, i rapporti di durata (V(ː)/C) differenziando tre contesti: quello di sillaba aperta, quello di sillaba chiusa e quello di sillaba chiusa in contesto di sonorante + C che provoca allungamento vocalico. Rientrano nel sottogruppo del contesto di sillaba chiusa le sillabe toniche chiuse originarie (es. [ˈrɛʃˑta] 'resta') e quelle insorte negli ossitoni e nei parossitoni secondari in seguito a sincopi (es. [ˈtrødˑʒə] 'tredici').

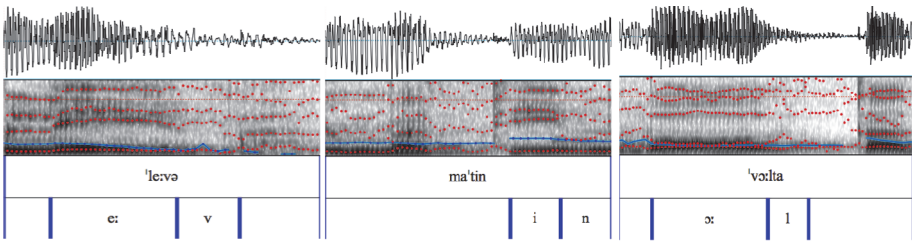
Come noto, l'allungamento della vocale tonica in un determinato contesto consonantico è un fenomeno diffuso nei dialetti emiliani. Il lizzanese (cfr. Malagoli, 1930: 134, 143; Uguzzoni, 1975; Filipponio, 2012: 72-73, 86), il frignanese (Uguzzoni, 1971: 124) e il bolognese (Coco, 1970: 7, 17) presentano vocali lunghe se il nesso consonantico postonico è costituito da *liquida* + C: lizz. [ˈbaːrba] 'barba', frign. [ˈseːrća] 'cerchia', bol. [ˈeːrba] 'erba'. Più limitata è generalmente la presenza di vocali lunghe prima di *nasale* + C: nei dialetti emiliani, infatti, si registrano «tre tipologie di comportamento della vocale davanti al nesso *nasale* + C: quella lizzanese, con vocale sempre lunga (Malagoli, 1930: 142); quella mediofrignanese, con vocale lunga davanti a *nasale* + C sorda e breve davanti a *nasale* + C sonora (Uguzzoni, 1975: 60-61); quella del bolognese, con vocale sempre breve (Coco, 1970: 61, n. 115)» (Filipponio, 2012: 199). A Piandelagotti si registra una situazione analoga a quella del medio Frignano e del medio Appennino bolognese, per cui si ha vocale lunga soltanto dopo *nasale* + C sorda, mentre negli altri contesti di nasale + C la vocale è breve: pdg. [ˈseːmper] 'sempre' ma [kwãnd] 'quando'.

I risultati del calcolo del rapporto tra vocale tonica e consonante postonica per i tre contesti considerati sono riportati nella tabella 3 dove, nella terza riga – oltre al valore medio del rapporto V(ː)/C – sono indicati, tra parentesi, la deviazione standard e il numero delle occorrenze computate (questo vale anche per le tabelle 4 e 5); nella quarta riga sono riportati degli esempi (ai quali corrispondono gli spettrogrammi delle figure 2, 3 e 4) e nell'ultima riga della tabella si è calcolato il coefficiente di correlazione tra la durata della vocale tonica e della consonante postonica.

Tabella 3 - *Risultati dei rapporti tra vocale tonica e consonante postonica (V(ː)/C) in sillaba aperta, in sillaba chiusa e in contesto di allungamento vocalico*

sillaba aperta (σ[)	*sillaba chiusa (σ])*	*sillaba chiusa in contesto di allungamento vocalico*
Vː/C	V/C	Vː/C
1.99 (± 0.76) (263)	1.02 (±0.79) (287)	2.39 (±1.09) (100)
[ˈleːva] ‘[mi] alzo’	[maˈtin] ‘mattino’	[ˈvɔːlta] ‘volta’
r = 0.3869905	r = 0.1412782	r = 0.2864141

Figure 2, 3 e 4 - *Spettrogrammi per l’illustrazione del rapporto tra vocale tonica e consonante postonica (parlante: FL)*



[ˈleːvə] ‘[mi] alzo’ [eː] 171 ms, [v] 85 ms

[maˈtin] ‘mattino’ [i] 67 ms, [n] 68 ms

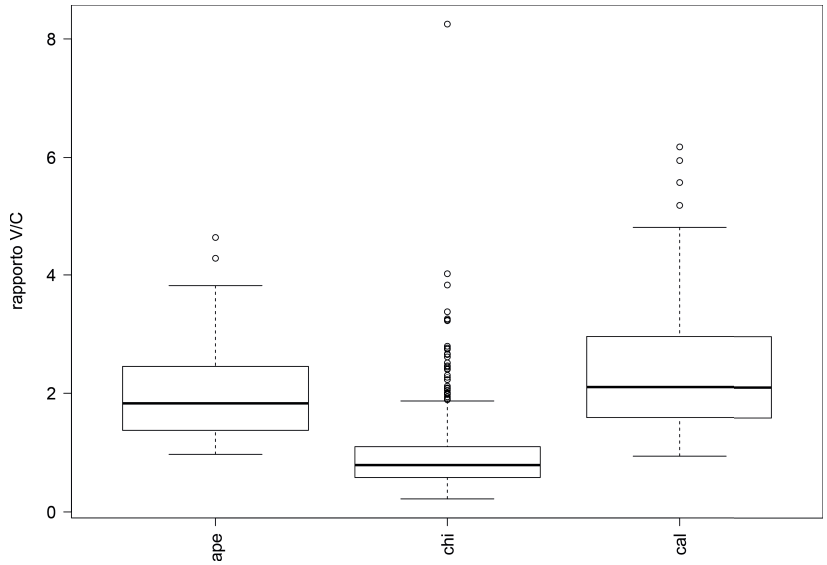
[ˈvɔːlta] ‘volta’ [ɔː] 202 ms, [l] 71 ms

Dai rapporti di durata emerge chiaramente che le vocali in sillaba aperta durano il doppio delle consonanti postoniche (Vː/C 1.99), mentre in sillaba chiusa la durata della vocale tonica tende ad equivalere a quella della consonante postonica (V/C 1.02)[5]. Sempre in sillaba chiusa la vocale tonica si allunga quando è seguita da una sonorante preconsonantica (secondo le modalità descritte sopra): in questi casi il rapporto Vː/C è di 2.39 (il rapporto è a favore della vocale soprattutto nei casi in cui la seconda consonante è un’ostruente sorda, come ad es. in [biˈfuːlk] ‘bifolco’). Da questi dati emerge che la struttura soggiacente del piandelagottese è VC: è la durata della vocale tonica ad essere dirimente, mentre la lunghezza consonantica, anche in posizione postonica, non presenta forti oscillazioni di durata.

I rapporti V(ː)/C risultano evidenti dai box-plots riportati nella figura 5, dove si evidenziano i diversi contesti sillabici: sillaba aperta (ape), sillaba chiusa (chi) e sillaba chiusa in contesto di allungamento vocalico (cal).

[5] In questo caso l’indice di correlazione tra la durata della vocale tonica e della consonante postonica è basso: questo risultato può dipendere dal fatto che non si sono (ancora) disaggregati i diversi contesti consonantici, per cui cfr. oltre, § 3.2.

Figura 5 - *Box-plots che rappresentano il rapporto V(:)/C in sillaba aperta (ape), in sillaba chiusa (chi) e in contesto di allungamento vocalico (cal)*



Le differenze tra i rapporti risultano essere significativi (p = 9.3e-05***)[6]. Le mediane dei due contesti di sillaba aperta e di allungamento vocalico sono molto vicine, mentre quella dei contesti di sillaba chiusa ha un valore inferiore[7]. L'alto numero di *outliers* per le sillabe chiuse si spiega sulla base del fatto che i dati comprendono tutti i contesti consonantici che chiudono la sillaba senza ulteriori disaggregazioni (per cui cfr. oltre § 3.2).

3.2 Rapporto tra vocale tonica e consonante postonica in sillaba chiusa

Dopo aver calcolato i valori dei rapporti tra la vocale tonica e la sillaba postonica per le diverse strutture sillabiche ci si concentra ora sulle sillabe chiuse, escludendo i contesti in cui si hanno vocali toniche foneticamente lunghe. Operando una distinzione tra i contesti consonantici ostruenti (sorde e sonore) e sonoranti si può osservare come la

[6] Per quanto concerne il rapporto V(:)/C si sono svolte delle analisi preliminari per testare se i fattori 'parlante' e 'struttura di parola' mostrassero effetti significativi: dato che i risultati di questi test non rispecchiano differenze statisticamente rilevanti (p = 0.02023* per il fattore 'parlante' e p = 0.03191* per il fattore 'struttura di parola') nelle analisi principali non si sono disaggregati i dati per queste due variabili.

[7] Le differenze tra le diverse strutture sillabiche sono state poi testate sulla base di confronti multipli (funzione *glht* di R dal R package *multcomp*, per cui cfr. Hothorn et al., 2008), dai quali è emerso che ci sono differenze significative tra le sillabe aperte (ape) e le vocali lunghe in contesto di allungamento (cal) (p = 0.0072**) così come tra le sillabe chiuse (chi) e le vocali lunghe per allungamento (cal) (p < 0.001***). I risultati del test per le sillabe chiuse (chi) e per quelle aperte (ape) mostrano che i due contesti non differiscono significativamente (p = 0.2042). Per l'interpretazione di quest'ultimo risultato bisogna tenere conto dell'eterogeneità dei contesti consonantici che chiudono la sillaba, la quale si rispecchia nell'alto numero di *outliers* presente nel box-plot della figura 5. È quindi comprensibile che le differenze tra le sillabe aperte e quelle chiuse non siano statisticamente significative.

durata della vocale tonica sia inversamente proporzionale alla forza della consonante postonica[8]. Si vedano a questo proposito i dati riportati nella tabella 4.

Tabella 4 - *Risultati dei rapporti tra vocale tonica e consonante postonica (V/C) disaggregati sulla base del contesto consonantico postonico*

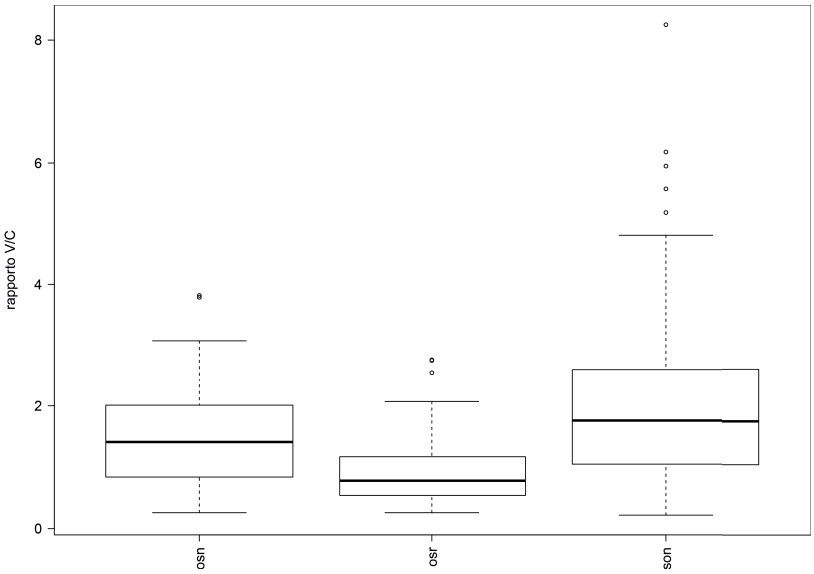
ostruenti		*sonoranti*
sorde	*sonore*	
V/C	V/C	V/C
0.68 (± 0.29) (106)	1.02 (± 0.59) (62)	1.33 (±1.04) (119)

La media dei rapporti tra le vocali toniche e le consonanti più forti (nello specifico le ostruenti sorde, V/C 0.68) è decisamente a favore delle consonanti, quella tra le vocali e le consonanti meno forti (le ostruenti sonore, V/C 1.02) indica che la durata della vocale e quella della consonante grossomodo si equivalgono, mentre la media dei rapporti tra le vocali e le consonanti deboli (le sonoranti, V/C 1.33) pende a favore della vocale (l'alto valore della deviazione standard nel caso delle sonoranti postoniche può essere ricondotto alle oscillazioni nelle durate della nasale bilabiale [m], che viene spesso pronunciata come geminata dai parlanti)[9]. La rappresentazione grafica di questi rapporti si trova nella figura 6, dove si distinguono i contesti di ostruente sorda (osr), di ostruente sonora (osn) e di sonorante (son).

[8] Si fa rifermento all'universale fonetico di forza consonantica così come definito da Vennemann (1988: 8): «The speech sounds of language system have **phonetic correlates**, which may differ according to their syntagmatic and prosodic contexts. These correlates may be projected on a phonetic parameter of degree of deviation from unimpeded (voiced) air flow, called **Universal Consonantal Strength**. This projection induces an ordering among the correlates» (grassetti nel testo). Si adotta, qui, la scala di forza consonantica (crescente) proposta, ad esempio, da Vennemann (1988: 9): vocali basse, vocali medie, vocali alte, vibranti, laterali, nasali, fricative sonore, fricative sorde, occlusive sonore, occlusive sorde. Proposte alternative per rendere conto della scala di forza consonantica si trovano, ad esempio, in Vogel (1982: 89-102) (dove si confrontono e discutono diverse gerarchie di forza segmentale in relazione a più lingue per giungere a una gerarchia di forza universale) e in Nespor (1993: 43-71).

[9] Il fatto che la nasale bilabiale sia pronunciata dai parlanti piandelagottesi come geminata non sorprende: «both gemination after stress in proparoxytones (of whatever consonant) and categorical gemination of Latin -m- (with merger with -mm-) are general features across Northern Italo-Romance, as can be seen from inspection of ancient texts, from data from modern dialects still retaining, at least in part, geminates, and, for the rest of modern Northern Romance dialects, in which degemination is nowadays complete, from syllable-related processes» (Loporcaro, 2014: 176).

Figura 6 - *Box-plots che rappresentano il rapporto V/C in tre contesti consonantici: ostruente sorda (osr), ostruente sonora (osn) e sonorante (son)*



Riprendendo esempi dal *corpus* analizzato, si può infatti notare che in [tʃik·] 'piccolo' il rapporto tra la [i] tonica e l'occlusiva velare sorda, una consonante forte, rispecchia una maggior lunghezza della consonante rispetto alla durata della vocale tonica (parlante EM, durata della vocale tonica: 113 ms, durata della consonante postonica 258 ms, rapporto V/C 0.44), mentre la relazione sintagmatica tra la vocale tonica e l'occlusiva sonora di ['vødva] 'vedova' si avvicina a 1 (parlante FL, durata della vocale tonica: 101 ms, durata della consonante postonica 112 ms, rapporto V/C 0.90). Il rapporto tra la vocale tonica e la sonorante postonica in [kwøl] 'quello' mostra come la vocale possa anche durare il doppio rispetto alla consonante (parlante LM, durata della vocale tonica: 82 ms, durata della consonante postonica 41 ms, rapporto V/C 2).

Dal quadro qui presentato non sembrerebbe esserci quindi traccia di allungamento consonantico secondario: questo si sarebbe potuto ipotizzare per le ostruenti sorde postoniche, ma non per le altre consonanti, il che porta a pensare che il rapporto medio di 0.69 sia di fatto dovuto a una proprietà intrinseca del segmento (la forza della consonante) e non a un fenomeno di allungamento consonantico secondario.

3.3 Quantità vocalica distintiva

Dopo aver descritto i rapporti tra le vocali toniche e le consonanti postoniche piandelagottesi si analizza ora la distintività della quantità vocalica con l'intento di apportarne evidenza empirica. Come noto, la quantità vocalica, che era distin-

tiva in latino e che si è generalmente persa nel passaggio alle lingue romanze[10], ha nuovamente assunto valore distintivo in diverse varietà settentrionali in seguito alla diffusa degeminazione postonica[11].

Il sotto-*corpus* messo a punto per il calcolo dei rapporti tra le vocali lunghe e le vocali brevi in contesto di coppia (semi-)minima conta 78 entrate (39 coppie). I valori dei rapporti tra vocale breve e vocale lunga (V/V:) e tra consonante preceduta da vocale lunga e consonante preceduta da vocale breve ((V:)C/(V)C) sono riassunti nella tabella 5[12].

Tabella 5 - *Risultati dei rapporti tra V/V: e tra (V:)C/(V)C in 39 coppie (semiminime)*

V/V:	*(V:)C/(V)C*
0.55 (± 0.28) (39)	0.76 (± 0.59) (25)

I risultati mostrano che a Piandelagotti la quantità vocalica è ancora saldamente distintiva: le vocali lunghe presentano, in media, una durata quasi doppia rispetto alle vocali brevi. Il rapporto tra consonanti dopo vocale breve e dopo vocale lunga, invece, registra un valore mediamente maggiore, il che rispecchia, come noto, differenze meno marcate tra le lunghezze dei due gruppi di consonanti.

La quantità vocalica rappresenta quindi una realtà nel sistema fonologico dei parlanti. A sostegno di questa distintività vi sono altri tratti strutturali, quali la degeminazione protonica e la presenza di distintività vocalica anche negli ossitoni uscenti in vocale[13], quindi in voci che non presentano una coda consonantica che possa influenzare la durata della vocale tonica (pdg. [di] 'giorno' ~ [di:] 'dito', [pø] 'può' ~ [bø:] 'buoi')[14]. La parziale presenza, in postonia, di consonanti (sonoranti) non brevi quanto le ostruenti non inficia la distintività della quantità vocalica in sede tonica. Questo è quanto anche affermato da Filipponio (2012: 46-47) sulla base dei risultati emersi dallo studio sui dialetti dell'Appennino bolognese:

> Il quadro mostra come, anche di fronte a una situazione di non piena degeminazione [in postonia, CB], l'interdipendenza strutturale di alcuni fenomeni (degeminazione

[10] «Sembra [...] prudente postulare il verificarsi del collasso della quantità vocalica latina, a *tutti* i livelli della lingua, attorno al V secolo d.C. (Loporcaro, 2011: 58), *terminus post quem* peraltro compatibile con la ricostruzione di tutti i processi fonologici protogalloitalici e galloitalici» (Filipponio, 2012: 32, corsivi nel testo).
[11] Cfr., ad es., Lüdtke (1956), Filipponio (2012: 22-67), Loporcaro (2014).
[12] Si ricordi che quanto minore è il valore del rapporto V/V:, tanto maggiore è la realizzazione fonetica della quantità vocalica; più tale rapporto si avvicina a 1 e meno marcata è la realizzazione della quantità vocalica.
[13] Questo contesto, infatti, come noto, permette di accertare la distintività della quantità vocalica. Cfr. Martinet (1956: 75 e 1975: 205, da cui si cita): «[C]e trait [la geminazione nei parossitoni con vocale tonica breve, CB] n'est jamais distinctif, puisqu'il accompagne automatiquement la brièveté de la voyelle précédente, et que c'est cette brièveté qui est le trait pertinent, puisqu'il est des positions, la finale tonique absolue par exemple, où la voyelle brève s'oppose à la non brève sans que la question se pose de géminer une consonne suivante».
[14] In piandelagottese gli ossitoni con vocale lunga sono (quasi) sempre secondari, insorti in seguito all'apocope della sillaba finale (cfr. Biasini, 2010: 17).

protonica, coppie minime di ossitoni di sillaba libera) permetta di individuare la pertinenza fonologica del vocalismo tonico.

La distintività della quantità vocalica quantificata in questa sede permette quindi di collocare, anche se con prudenza (cfr. paragrafo conclusivo), il piandelagottese all'interno del quadro dei dialetti gallo-italici (cfr. cartina 1). Gli studi che hanno verificato la saldezza della quantità vocalica in questi dialetti sono infatti numerosi (si ricordino, tra gli altri, quelli di Uguzzoni, Busà, 1995; Loporcaro et al., 2005; Filipponio, 2012): se ne riassumono alcuni nella tabella 6.

Cartina 1 - *Quantità vocalica: località indagate*

Tabella 6 - *Rapporti tra V/V: e tra (V:)C/(V)C in diversi dialetti gallo-italici*

Località	*V/V:*	*(V:)C/(V)C*	
Crocette	0.45	0.95	Uguzzoni, Busà (1995: 14-15)
Benedello	0.51	1.03	Uguzzoni, Busà (1995: 10)
Piandelagotti	0.55	0.76	
Lizzano	0.55	0.77	Filipponio (2012: 246-47)
S. Vigilio di Marebbe	0.64	1.11	Bernardasci (2013a: 135)[16]
Soglio	0.85	0.64	Loporcaro et al. (2005: 608)

I dati per i dialetti di Crocette (Pavullo nel Frignano) e Benedello (prov. di Modena) mostrano chiaramente un'opposizione fonologica tra vocali lunghe e vocali brevi in contesto di coppia minima; la lunghezza consonantica, per contro, non ha un ruolo fonologicamente determinante: le durate delle consonanti, infatti, sono tendenzialmente equivalenti sia dopo vocale lunga, sia dopo vocale breve (cfr. Uguzzoni, Busà 1995). I rapporti qui calcolati per Piandelagotti sono pressoché identici a quelli ottenuti da Filipponio (2012: 246-47) per Lizzano. Alla marcata distintività della quantità vocalica si accompagna un rapporto meno netto per le consonanti postoniche: «[q]uesto scarto di durata tra C postoniche sembra insufficiente per poter parlare di resistenza alla geminazione postonica in lizzanese» (Filipponio, 2012: 243)[16]. Lo stesso vale, come si è detto, per lo statuto della degeminazione postonica in piandelagottese. Ai dati lizzanesi e piandelagottesi si allineano anche i risultati della varietà di S. Vigilio di Marebbe, un dialetto ladino centrale che presenta saldamente quantità vocalica distintiva (cfr. Bernardasci, 2013a).

Diversa è invece la situazione di Soglio: nella località bregagliotta il rapporto vocalico calcolato da Loporcaro et al. (2005) è maggiore di quello consonantico, il che significa che le durate di tutte le vocali toniche sono tendenzialmente omogenee e che le consonanti geminate postoniche non si sono (ancora) degeminate (mentre

[15] Il rapporto consonantico non è presente nella pubblicazione citata: esso è stato calcolato in seguito, per quest'occasione, sulla base dei dati riportati a p. 138.

[16] I rapporti consonantici lizzanesi presentati da Loporcaro et al. (2006: 511) divergono notevolmente da quelli riportati in Filipponio (2012): nel primo caso si è calcolato un rapporto C/C(:) di 0.52 per i bisillabi parossitoni, mentre nel secondo lo stesso rapporto è, come si è già detto, di 0.77. La differenza tra i due rapportti porta a due interpretazioni diverse del ruolo fonetico della lunghezza consonantica: dai dati di Loporcaro et al. risulta che le consonanti postoniche geminate sono ancora ben salde, mentre dai suoi dati Filipponio conclude che la lunghezza consonantica è subordinata alla durata vocalica. Tale divergenza nei risultati è riconducibile, come sostenuto anche da Filipponio (2012: 246), a sostanziali differenze nell'elicitazione dei dati: nel primo studio «l'analisi acustica è stata svolta quasi esclusivamente sulle parole elicitate in isolamento» (Loporcaro et al., 2006: 494), mentre nel secondo caso si è adottato un sistema di elicitazione sintagmatica utilizzando frasi di senso compiuto (cfr. anche Loporcaro, 2014: 191).

in protonia la degeminazione è sistematica): [ˈkaʃːɐ] 'cassa', [kaˈʃɔtːɐ] 'cassetta' (dati citati da Loporcaro 2014: 177). Inoltre, mancano coppie minime di ossitoni in sillaba aperta che si differenzino esclusivamente per il tratto di durata della vocale. Sulla base dei dati del dialetto di Soglio non si può quindi attribuire distintività fonologica alla quantità vocalica. Soglio rappresenta infatti il

> polo di massima conservatività lungo una scala che ordina, in gradini successivi, gli altri tipi di dialetti [dell'Italia settentrionale e costituisce una sorta di] [...] sistema 'proto-italoromanzo settentrionale', in quanto permette di osservare *in vivo* la situazione da ricostruirsi per la fase medievale in cui la degeminazione non s'era ancora imposta e l'instaurazione della quantità vocalica distintiva era ancora di là da venire (Loporcaro et al., 2005: 597).

Il pregio dei risultati dei lavori riassunti nella tabella 6 è senza dubbio quello di essere scaturiti da un'analisi sperimentale basata su dati raccolti sul campo. Il fatto che, com'è inevitabile quando le inchieste sono svolte da ricercatori diversi a distanza di anni, non tutti i risultati siano stati ottenuti seguendo la stessa metodologia (cfr. ad es. n. 17) suggerisce prudenza nel loro confronto. Nel caso specifico del presente contributo, bisogna tenere presente il contesto in cui i dati sui quali si basa l'analisi acustica sono stati elicitati: la conseguenza dell'assenza dell'utilizzo di frasi cornice o di frasi di senso compiuto è che le parole bersaglio sono state pronunciate in un contesto prosodico marcato (in isolamento) con uno stile di eloquio enfatico. Una volta considerato e problematizzato questo aspetto metodologico i risultati piandelagottesi possono venire integrati nel quadro degli studi sulla distintività della quantità vocalica in area gallo-italica.

Bibliografia

Bates, D.M., Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes, R package version 1.1-7.

Bernardasci, C. (2013a). La quantità vocalica distintiva in marebbano. Un'analisi sperimentale. In *Ladinia*, 37, 125-146.

Bernardasci, C. (2013b). Un'analisi fonetica di alcuni aspetti del sistema vocalico e consonantico del dialetto di Piandelagotti (MO), tesi di master, Università di Zurigo.

Biasini, C. (2010). Fonetica e fonologia del dialetto di Piandelagotti sull'Appennino modenese, tesi di licenza, Università di Zurigo.

Coco, F. (1970). *Il dialetto di Bologna. Fonetica storica e analisi strutturale*. Bologna: Forni.

Filipponio, L. (2012). *La struttura di parola dei dialetti della Valle del Reno*. Bologna: Forni.

Foresti, F. (1988). Emilia-Romagna. In Holtus, G., Metzeltin, M. & Schmitt, C. (Eds.), *Lexikon der Romanistischen Linguistik (LRL)*, IV. Tübingen: Niemeyer Verlag, 569-593.

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. In *Biometrical Journal*, 50 (3), 346-363.

Kliegl, R., Wei, P., Dambacher, M., Yan, M. & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. In *Frontiers in Psychology*, 1, 1-12.

Loporcaro, M. (2011). Syllable, segment and prosody. In Maiden, M., Smith, J.C. & Ledgeway, A. (Eds.). *The Cambridge History of the Romance Languages*. Cambridge: Cambridge University Press, 50-108.

Loporcaro, M. (2014). Vowel Length from Latin to Romance. Oxford: Oxford University Press.

Loporcaro, M., Paciaroni, T. & Schmid, S. (2005). Consonanti geminate in un dialetto lombardo alpino. In Cosi, P. (Ed.), *Misura dei parametri: aspetti tecnologici ed implicazioni nei modelli linguistici.* AISV 2004, 1° Convegno Nazionale (Padova, 2-4 dicembre 2004). Brescia: EDK, 597-618.

Loporcaro, M., Delucchi, R., Nocchi, N., Paciaroni, T. & Schmid, S. (2006). La durata consonantica nel dialetto di Lizzano in Belvedere (Bologna). In Savy, R., Crocco, C. (Eds.), *Analisi prosodica. Teorie, modelli e sistemi di annotazione.* AISV 2005, 2° Convegno Nazionale (Salerno, 30 novembre - 2 dicembre 2005). Torriana (RN): EDK, 491-517.

Loporcaro, M., Delucchi, R., Nocchi, N., Paciaroni, T. & Schmid, S. (2007). Schwa finali sull'Appennino emiliano: Il vocalismo del dialetto di Piandelagotti. In Giordani, V., Bruseghini, V. & Cosi, P. (Eds.), *Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche.* AISV 2006, 3° Convegno Nazionale (Trento, 29-30 novembre - 1 dicembre 2006). Brescia: EDK, 57-76.

Lüdtke, H. (1956). *Die strukturelle Entwicklung des romanischen Vokalismus*, Romanisches Seminar an der Universität Bonn, Bonn.

Maddieson, I. (1997). Phonetic Universals. In Hardcastle, W.J., Laver, J. (Eds.), *The Handbook of Phonetic Sciences*. Oxford - Cambridge: Blackwell, 619-639.

Malagoli, G. (1910-1913). L'articolo maschile singolare nel dialetto di Piandelagotti (Modena). In *Archivio Glottologico Italiano*, 17, 250-254.

Malagoli, G. (1930). Fonologia del dialetto di Lizzano in Belvedere (Appennino Bolognese). In *L'Italia Dialettale*, 6, 125-196.

Martinet, A. (1956). *La description phonologique, avec application au parler franco-provençal d'Hauteville (Savoie)*. Genève - Paris: Droz - Minard.

Martinet, A. (1975). *Évolution des langues et reconstruction*. Paris: Presses Universitaires de France.

Nespor, M. (1993). *Fonologia*. Bologna: il Mulino.

Pellegrini, G.B. (1977).*Carta dei Dialetti d'Italia*. Pisa: Pacini.

Pellegrini, G.B. (1992). Il «cisalpino» e l'italo-romanzo. In *Archivio Glottologico Italiano*, 77, 272-296.

Pieri, S. (1983a). Il dialetto gallo-romano di Gombitelli nella provincia di Lucca. In *Archivio Glottologico Italiano*, 13, 309-328.

Pieri, S. (1983b). Il dialetto gallo-romano di Sillano. In *Archivio Glottologico Italiano*, 13, 329-354.

R Core Team (2013). R. A language and environment for statistical computing, version 3.0.1. http://www.R-project.org.

Uguzzoni, A. (1971). Quantità fonetica e quantità fonologica nell'area dialettale frignanese. In *L'Italia Dialettale*, 34, 115-136.

Uguzzoni, A. (1975). Appunti sulla evoluzione del sistema vocalico in un dialetto frignanese. In *L'Italia Dialettale*, 38, 47-76.

Uguzzoni, A., Busà, M.G. (1995). Correlati acustici della opposizione di quantità vocalica in area emiliana. In *Rivista Italiana di Dialettologia*, 19, 7-39.

Vennemann, T. (1988). *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin: Mouton de Gruyter.

Vogel, I. (1982). *La sillaba come unità fonologica*. Bologna: Zanichelli.

DOI: 10.17469/O2101AISV000008

ALESSANDRO VIETTI, LORENZO SPREAFICO, VINCENZO GALATÀ[1]

Sull'allofonia di /ʀ/ in tirolese: metodo e dati da un'analisi articolatoria ultrasonografica preliminare

Moving from traditional dialectological literature and inspired by contemporary research on rhotics, in this paper we present preliminary data on the distribution and the articulation of /ʀ/ in Tyrolean, an under-researched South Bavarian dialect. Two speakers produced a comprehensive selection of Tyrolean words containing /ʀ/. Overall, they uttered up to five different uvular rhotics: [χ, ʁ, ʁ̞, ɾ, ʀ]. We found mild tendencies in the allophonic distribution of the variants, but systematic differences in their lingual configurations: trills are produced steep lowering the tongue tip; vocalizations are markedly lowered and retracted, with the tongue dorsum flat; approximants are retracted.

1. *Introduzione*

L'allofonia di /ʀ/ costituisce uno dei temi privilegiati nell'annoso dibattito sulla problematica interfaccia tra fonetica e fonologia (Simpson, 1998; Wiese, 2000; Wiese, 2001; Scobbie, Sebregts, 2010). In tal senso, le lingue germaniche – e in particolare quelle occidentali – da tempo costituiscono un ambito probante per i ricercatori. Per quanto riguarda i dialetti alto tedeschi, i molti tentativi fatti per offrire una compiuta descrizione di tutte le varianti di /ʀ/ attestate, così come gli sforzi tesi a rendere conto della loro distribuzione contestuale, non si sono rivelati risolutivi. Per questo motivo nel nostro contributo affronteremo la tematica con nuovi dati e ulteriori preliminari riflessioni sull'allofonia di /ʀ/ in tirolese. La nostra proposta è alternativa alle – o quantomeno integrativa delle – precedenti perché contempla anche la dimensione articolatoria. Nella letteratura dialettologica disponibile, basata essenzialmente su indagini uditive impressionistiche, il sistema delle rotiche del tirolese è descritto come composto da due sole varianti uvulari, una fricativa e una polivibrante, raramente compresenti come allofoni nella medesima varietà di dialetto (Klein, Schmitt, 1965; Tonelli, 2001). Se possibile, le informazioni sulla distribuzione allofonica delle due varianti nelle varietà che lo prevedano sono ancora più scarse. Lo spoglio dei dati in Klein, Schmitt (1965) consente di inferire una generale preferenza per la polivibrante in attacco di sillaba, e per la fricativa - sorda o sonora – in coda di sillaba. Tuttavia, studi più recenti – basati su parlanti bilingui di tirolese *e* italiano – hanno riscontrato una maggiore variabilità interna al

[1] Sebbene il lavoro qui presentato sia frutto della collaborazione tra i tre autori, ai soli fini concorsuali il § 1 è da attribuire ad AV e LS; il § 2.4 a VG; il § 2.5 a LS e VG; il § 3.1 ad AV; il § 3.2 a LS; i restanti §§ 2.1, 2.2, 2.3 e 4 congiuntamente a tutti e tre gli autori.

parlante, specialmente nelle rese uvulari di /ʀ/[2]. Secondo Spreafico, Vietti (2013) in particolare, l'articolazione primaria delle rotiche uvulari sarebbe caratterizzata dall'innalzamento e/o l'ammassamento del post-dorso della lingua, mentre tratti secondari come la faringalizzazione e/o l'innalzamento dell'apice risulterebbero essere responsabili della presenza di altri allofoni uvulari quali approssimanti, monovibranti e persino vocalizzazioni.

Al fine di verificare con maggiore accuratezza se e quali varianti uvulari di /ʀ/ siano presenti nelle produzioni di parlanti tirolesi; se queste coesistano o meno nello spazio fonologico di singoli parlanti; e, infine, se e quali siano le proprietà articolatorie linguali condivise dalle varianti, abbiamo predisposto una ricerca pilota finalizzata allo studio dell'allofonia articolatoria di /ʀ/ impiegando dati ecografici.

Descriviamo nei prossimi paragrafi le fasi di raccolta e di preparazione dei dati seguite nella ricerca che, data la sua fruttuosità, intendiamo sviluppare in un più ampio progetto che miri a indagare l'implementazione dell'allofonia in parlanti monolingui e bilingui.

2. *Materiali e metodi*

2.1 Materiali

Poiché era nostro obiettivo documentare l'allofonia di /ʀ/, il primo passaggio è stato rappresentato dalla definizione di un compito escussivo che permettesse l'elicitazione di quanti più, se non tutti, i foni rotici del dialetto tirolese. Pertanto, valutata l'impossibilità dell'operazione qualora si fossero raccolti solo dati spontanei, perché pochi sarebbero stati i dati utili rispetto a quelli raccolti e analizzati, abbiamo optato per un compito guidato di lettura di parole. Dopo un attento spoglio di tutti i lemmari cartacei e informatizzati attualmente disponibili per il dialetto tirolese meridionale (Schatz, 1956; Haller, Lanthaler, 2004; Oschpele, 2006), siamo giunti alla definizione di una lista di 80 parole contenenti /ʀ/ in attacco e in coda di sillaba (sia atona che tonica, ove possibile), in posizione iniziale assoluta, mediana e finale di parola (Appendice 1). Al fine di limitare il carico di lavoro per gli informanti, probante di per sé dato l'impiego di un casco stabilizzatore (cfr. *infra*), nel redigere la lista abbiamo limitato i segmenti adiacenti a /ʀ/ alle sole vocali (V) /a, a:, i, i:, o, o:/, e alle sole consonanti (C) /t, d, k, g/ per /ʀ/ in attacco di sillaba (CRV) e in coda di sillaba (VRC)[3]. Per /ʀ/ in posizione di coda abbiamo considerato anche parole con /ʀ/ + nasale o liquida.

[2] L'alta variabilità individuale nella produzione di rotiche uvulari è confermata anche dalla ricca letteratura in materia: cfr. il volume a cura di Spreafico, Vietti (2013), *Rhotics. New data and perspectives*. Bozen-Bolzano: BU Press, in particolare i contributi di Van de Velde, Tops & van Hout e di Sankoff, Blondeau; e il contributo di Ulbrich, Ulbrich (2007), in particolare le osservazioni sul tedesco d'Austria.

[3] Ove possibile, abbiamo preferito parole contenenti più di una /r/ al fine di mantenere, come detto, basso il numero di parole utilizzate.

Per essere sicuri che le parole scelte fossero riconoscibili e pronunciate correttamente dai partecipanti all'esperimento, la lista di parole è stata sottoposta al vaglio di due parlanti nativi di tirolese estranei alla ricerca, che l'hanno approvata: le parole segnalate come problematiche sono state sostituite con altre da loro suggerite.

2.2 Partecipanti

Al fine di rispondere alle domande di ricerca abbiamo anzitutto reclutato due partecipanti di sesso femminile (MRL ed EVK), entrambe di 33 anni, nate e residenti nell'area di Merano (Bolzano) e parlanti native di dialetto tirolese. Entrambi i soggetti hanno riferito di non avere deficit linguistici e di possedere competenze linguistiche native sia per il tedesco che per l'italiano standard. Per entrambe è stato tuttavia verificato, tramite la somministrazione di un questionario relativo agli usi del repertorio linguistico individuale e alla rete sociale, quale fosse il tasso di incidenza delle due lingue nell'uso quotidiano. Abbiamo così osservato che il tirolese è lingua fortemente dominante per EVK e blandamente dominante per MRL.

2.3 Somministrazione e acquisizione dei dati

Per ciascuna delle partecipanti l'acquisizione dei dati ha avuto luogo in cabina silente e in un'unica sessione della durata compresa tra i 40 e i 50 minuti.

La lista di parole, randomizzata e intervallata da distrattori[4], è stata somministrata due volte allo scopo di ottenere almeno due ripetizioni dello stesso target. Le parole sono state presentate singolarmente su uno schermo da 21" con un font di almeno 48 punti. L'istruzione fornita ai partecipanti era di leggere in dialetto tirolese le parole che di volta in volta apparivano a schermo.

Il segnale sonoro è stato catturato da un microfono Sennheiser ME2, posto ad una distanza di ca. 15 cm dal soggetto, collegato a un registratore B1 Marantz PMD660 con funzioni di amplificatore del segnale.

I dati ultrasonografici sono stati acquisiti attraverso la sonda Ultrasonix C9-5/10 operante a 5MHz e collegata al sistema ecografico Ultrasonix SonicTablet dotato di applicativo Exam software v. 5.7.3. La sonda, del tipo microconvesso, è stata posizionata e bloccata sotto il mento del partecipante tramite un elmetto di stabilizzazione (Scobbie, Wrench & Van der Linden, 2008).

Il segnale acustico e il segnale ultrasonografico sono quindi stati sincronizzati e registrati in tempo reale su una postazione Desktop PC grazie al software Articulate Assistant Advanced v. 2.15 (AAA) della Articulate Instruments Ltd (2014) in grado di gestire i parametri di configurazione del sistema a ultrasuoni. Il segnale ultrasonografico è stato acquisito a una frequenza di 91Hz e 95Hz con un campo visivo di

[4] Ovvero, delle pause per il partecipante in cui gli viene chiesto di deglutire dell'acqua così da migliorare la visibilità della superficie della lingua tramite ultrasonografia e permettere anche la tracciatura del profilo del palato (cfr. infra).

127° e 134° (profondità 80mm) per MRL ed EVK, rispettivamente[5]. La frequenza del segnale ecografico è adeguata per il campionamento dei foni rotici uvulari, inclusi i mono- e i polivibranti. Il segnale sonoro è stato campionato 22050Hz 16-bit mono.

2.4 Preparazione dati acustici

Ad acquisizione dati ultimata, il solo segnale sonoro è stato esportato da AAA in formato *.wav e successivamente importato in PRAAT (Boersma, Weenink, 2014) per la fase di segmentazione e annotazione delle parole e delle porzioni di interesse. L'operazione di segmentazione e annotazione è stata eseguita da uno degli autori in due fasi distinte, una automatica e una manuale.

Nella la fase automatica, con l'ausilio di uno script di PRAAT, ciascuna parola presente nei singoli file *.wav è stata segmentata e annotata in modo codificato e automatico all'interno di un TextGrid a diversi livelli:

- sul Tier *prompt* è stata annotata la parola;
- sul Tier *key* sono state codificate le informazioni che riguardano la struttura della sillaba, la posizione di /ʀ/, il tipo e la posizione della sillaba all'interno della parola, l'elemento che precede e che segue /ʀ/, durata e posizione dell'accento.

Nella fase manuale, i singoli segmenti sono stati perfezionati in termini di allineamento dei confini temporali grazie all'ispezione visiva della forma d'onda e del sonogramma in associazione al feedback uditivo-percettivo. In questa fase tutti i segmenti di /ʀ/ sono stati identificati, segmentati e annotati a un ulteriore livello sul Tier *label* in riferimento a:

- luogo di articolazione (PoA, *place of articulation*): per il campione in esame erano presenti solo target di tipo uvulare;
- modo di articolazione (MoA, *manner of articulation*): polivibrante (o trill [ʀ]), monovibrante (o tap [ɾ]), approssimante [ʁ̞], fricativo [χ, ʁ]. Sono stati considerati e annotati anche vocalizzazioni e cancellazioni[6];
- sonorità (±voice): presenza o assenza di vibrazione delle corde vocali.

Eventuali casi dubbi o problematici di /ʀ/ sono stati segnalati come "undef", riportando possibili note o commenti sul Tier *notes*.

[5] Tali parametri sono frutto della calibrazione della strumentazione a inizio registrazione con lo scopo di ottenere la migliore risoluzione possibile del dato ultrasonografico. La configurazione adottata è rimasta invariata sino alla fine della sessione di registrazione.

[6] Sono state considerate come vocalizzazioni quelle realizzazioni fonetiche di /r/ in forma dittongata con una visibile e percepibile modificazione dell'andamento formantico e in forma monottongata con realizzazione di un vocoide più centralizzato; al contrario, sono state segnalate come cancellazioni le porzioni terminali relative al segnale del nucleo sillabico in cui percettivamente e visivamente, con riferimento alle traiettorie formantiche, non erano rilevabili modificazioni qualitative attribuibili o riconducibili alla presenza di una /r/.

2.5 Preparazione dati articolatori

Le annotazioni effettuate nel dominio acustico sono quindi state importate in AAA per la tracciatura dei profili linguali relativi alle porzioni di interesse: oltre ai riferimenti temporali per ciascuna porzione annotata, sono stati esportati anche i rispettivi valori annotati nei Tier *prompt*, *label*, *key* e *notes* popolando i corrispondenti campi presenti nell'interfaccia di AAA. Due degli autori hanno tracciato i profili linguali per ciascun segmento di /R/ utilizzando l'algoritmo di tracciatura implementato in AAA e intervenendo in modo manuale nei casi in cui era necessario affinare o perfezionare il profilo proposto. L'altro autore ha verificato e all'occorrenza corretto i profili tracciati prima dell'esportazione dei dati da sottoporre ad analisi.

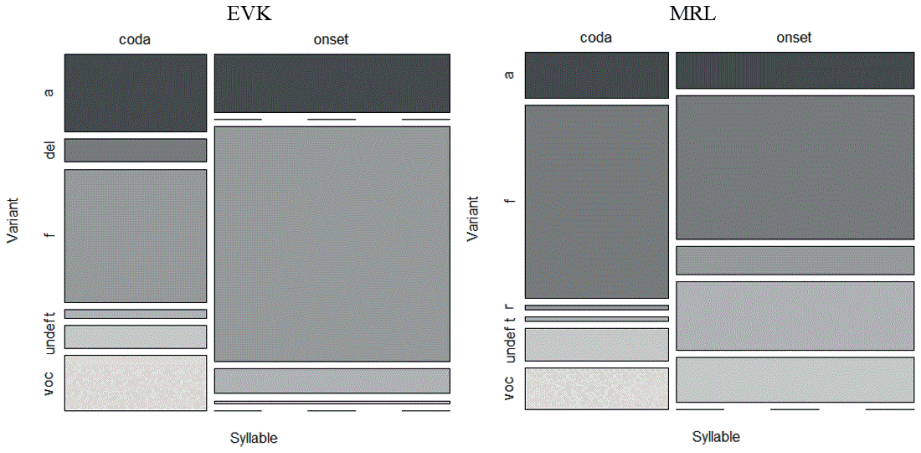
Nella fase di esportazione dei dati articolatori, i profili linguali sono stati estratti nel punto medio di ciascun intervallo etichettato come /R/.

3. *Analisi e discussione dei dati*

3.1 Distribuzione delle varianti

I diagrammi a mosaico riportati in Figura 1 permettono di osservare sinteticamente la distribuzione e il peso delle diverse rotiche uvulari nei diversi macro-contesti sillabici. I diagrammi mostrano come entrambe le parlanti usino tanto allofoni vocalici quanto allofoni consonantici di /R/.

Figura 1 - *Diagramma a mosaico riportante la distribuzione delle varianti di /R/ uvulare per EVK ed MRL, rispettivamente, raggruppate per posizione di sillaba (onset e coda) e MoA (a = approssimante, f = fricativo, t = tap, r= trill, voc = vocalizzazione, del = cancellazione)*



Per quanto concerne questi ultimi, EVK articola [ʁ̞, χ, ʁ, ɾ][7], mentre MRL articola [ʁ̞, χ, ʁ, ɾ, ʀ]. Non emergono forti correlazioni tra le varianti e i contesti sillabici, tuttavia si rilevano alcune tendenze, ovvero:

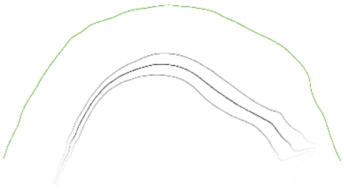
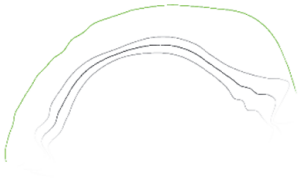
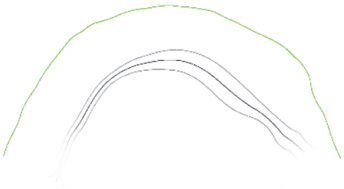
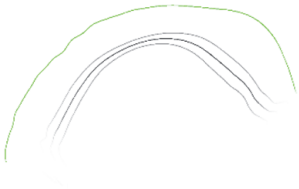
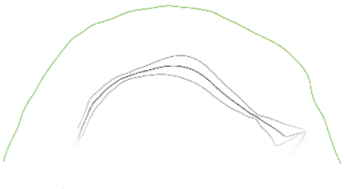
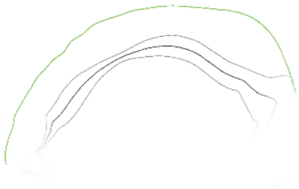
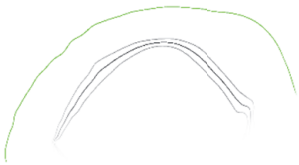
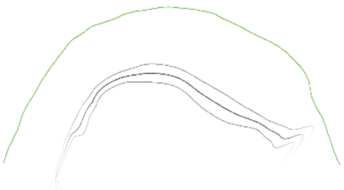
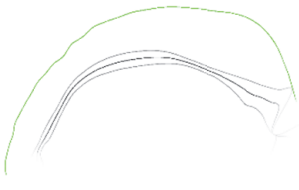
– per entrambe le parlanti la variante più frequente è [χ], sia in testa che in coda di sillaba;
– [ʁ̞] è preferita in testa di sillaba, soprattutto nei contesti intervocalici;
– le vocalizzazioni e le cancellazioni di /ʀ/ si hanno solo in coda di sillaba;
– [ʀ] emerge in testa di sillaba.

3.2 Articolazione delle varianti

L'analisi è basata sul confronto delle varianti di /ʀ/ indipendentemente dal contesto fonetico in cui fossero inserite. Nonostante emergano differenze di varianti, gli schemi articolatori sono relativamente stabili. L'indagine impressionistica dei profili linguali di /ʀ/ estratti in corrispondenza del punto mediano mostra, nonostante le differenze acusticamente percepibili, configurazioni ricorrenti. Risultano infatti somiglianze di forma e posizione indipendentemente da fenomeni di coarticolazione (con la rilevante eccezione di /ʀ/ + nasale o liquida). I profili linguali risultano in tal senso simili a quelli già individuati in Spreafico, Vietti (2013) e in Scobbie, Sebgrets (2011) per le rotiche uvulari del nederlandese. I profili risultano simili anche tra le due parlanti se analizzati alla luce di una comparazione dei profili medi estratti secondo la proposta in Davidson (2006). In entrambi i casi si rileva un profilo linguale convesso, con il posterodorso sollevato verso il palato e l'anterodorso e la porzione di lamina visualizzata rivolte verso il basso (esclusa la già notata eccezione di /ʀ/ + liquida o nasale, specialmente in MRL). EVK presenta però una modesta variazione allofonica che caratterizza la radice e il dorso linguale, con le rotiche vocalizzate a presentare maggiore abbassamento e arretramento, le monovibranti a mostrare innalzamento e avanzamento, quasi a sovrapporsi ai profili per le fricative. Emerge inoltre un lieve avvallamento nel predorso indicativo forse di una articolazione secondaria. La variazione allofonica emerge appena più marcatamente in MRL, laddove il posterodorso risulta più retratto e meno convesso per le vocalizzazioni, più retratto per le approssimanti e sensibilmente più abbassato per le monovibranti. Significativamente, anche alla luce del dibattito sulla salienza percettiva delle rotiche, le produzioni di MRL, per quanto esclusivamente uvulari, mostrano un lieve innalzamento dell'apice linguale tanto nelle vocalizzazioni e approssimanti, quanto nelle monovibranti. Tale tratto è assolutamente assente nelle polivibranti, caratterizzate anche da una maggiore pendenza di postdorso e predorso e, in generale, da minore variabilità soggiacente.

[7] Come proposto in Demolin et al. (submitted), adottiamo qui il simbolo [ɾ] per riferirci alla monovibrante uvulare.

Tabella 1 - *Profili linguali medi (con deviazione standard) della realizzazione di /ʀ/ uvulare per EVK ed MRL, rispettivamente (punta della lingua sulla destra)*

	EVK	MRL
Approssimante [ʁ]		
Fricativo [χ, ʁ]		
Tap [ɾ]		
Trill [ʀ]	*assente*	
Vocalizzazione		

4. *Conclusione sul metodo e i dati*

Questo breve studio pilota su un gruppo ristrettissimo di parlanti ha permesso di verificare la bontà della proposta elaborata per un'indagine più rappresentativa dell'allofonia di /ʀ/ in tirolese che includesse anche informazioni sulla dimensione articolatoria.

La categorizzazione delle varianti su base uditivo-percettiva e coadiuvata dall'ispezione visiva della forma d'onda oltre che del sonogramma, conferma l'elevata variabilità individuale nella produzione della classe delle rotiche. L'analisi distribuzionale mostra inoltre come, nonostante talune regolarità (es.: la variante meno marcata è la fricativa, poli- e monovibranti sono più frequenti in testa di sillaba; le cancellazioni in coda di sillaba), non emergano correlazioni forti. L'analisi ultrasonografica mostra infine come parametri articolatori possano essere utilmente considerati per meglio comprendere differenze nelle rese delle varianti, in particolare quelli riassunti nella tabella che segue e ricavabili dalle considerazioni sui profili linguali di EVK e MRL espressi sopra.

Tabella 2 - *Possibili combinazioni di parametri articolatori/linguali per la caratterizzazione delle varianti di /ʀ/*

	[ɾ, ʀ]	[χ, ʁ,]	[ɻ]
Ritrazione radice	–	–	+
Abbassamento lamina	+	–	–
Innalzamento dorso	+	+	–

Ciò illustra una volta di più come il dato articolatorio non vada considerato semplicemente come complementare all'analisi acustica, ma sia portatore di informazioni rilevanti anche sul piano della teorizzazione fonologica. Con riferimento alla tecnica e al fenomeno in esame, l'ultrasonografia risulta perciò particolarmente adeguata allo studio delle rotiche uvulari e alla variabilità allofonica in ragione dell'ampia risoluzione spaziale della superficie linguale e della elevata risoluzione temporale che permette di catturare gesti e transizioni estremamente rapide.

I dati qui esposti rappresentano solo un'analisi preliminare e non hanno, al momento, la pretesa di esaurire il dibattito che interessa il variegato fenomeno della /ʀ/. Sono infatti in corso analisi e approfondimenti con ulteriori soggetti che ci forniranno elementi utili a meglio definire lo status di /ʀ/ in questa varietà dialettale ancora poco studiata.

Ringraziamenti

Ringraziamo i curatori e due revisori anonimi per gli utili commenti. La ricerca è stata finanziata dalla Provincia Autonoma di Bolzano - Alto Adige, Ripartizione allo studio, Università e ricerca scientifica 2013-16, "Sociofonetica articolatoria di parlanti bilingui: il potenziale della tecnica di Ultrasound Tongue Imaging".

Bibliografia

Articulate Instruments Ltd (2014). *Articulate Assistant Advanced User Guide: Version 2.15*. Edinburgh, UK: Articulate Instruments Ltd.

Boersma, P., Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from http://www.praat.org.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imagining using smoothing spline analysis of variance. In *JASA*, 120 (1), 407-415.

Demolin, D., Franchetto, B. & Fausto, C. (submitted). Pharyngeal taps in Kuikuro. In *Journal of the International Phonetic Association*.

Haller, H., Lanthaler, F. (2004). Passeier Wörterbuch - Psairer Werterpuach. Wörter - Ausdrücke - Beispiele. Verlag Passeier: St. Martin in Passeier.

Klein, K., Schmitt, L. (1969). *Tirolischer Sprachatlas*. Tyrolia-Verlag.

Oschpele (2006), Oschpele - das Südtiroler Dialekt-Wörterbuch. http://www.oschpele.ritten.org.

Schatz, J. (1956). Wörterbuch der Tiroler Mundarten. Wagner: Innsbruck.

Scobbie, J., Sebregts, K. (2011). Acoustic, articulatory and phonological perspectives on rhoticity and /r/ in Dutch. In Folli R., Ulbrich, C. (Eds.), *Interfaces in linguistics: new research perspectives*. OUP, 257-277.

Scobbie, J., Wrench, A. & van der Linden, M. (2008). Headprobe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In *Proceedings of the 8th International Seminar on Speech Production*, 373-376.

Simpson, A. (1998). Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung, Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel. In *Arbeitsberichte*, 4, 1-233.

Spreafico, L., Vietti, A. (2013). On rhotics in a in a bilingual community: A preliminary UTI research. In Spreafico L., Vietti A. (Eds.), *Rhotics. New data and perspectives*. Bozen-Bolzano: BU Press, 57-77.

Spreafico, L., Vietti, A. (2013). *Rhotics. New data and perspectives*, Bozen-Bolzano: BU Press.

Tonelli, L. (2002). *Regionale Umgangssprachen*. Unipress.

Ulbrich, C., Ulbrich, H. (2007). The Realisation of /r/ in Swiss German and Austrian German. In *ICPhs*, 1761-1764.

Wiese, R. (2000). *The Phonology of German*. Oxford: OUP.

Wiese, R. (2001). The unity and variation of German /r/. In *Etudes et Travaux*, 4, 11-26.

Wrench, A., Scobbie, J. (2006). Spatio-temporal inaccuracies of video-based ultrasound images of the tongue. In *ISSP7*, 451-458.

Appendice 1: stimoli

Audraanen; Barbara; bartl; bevor; birkn; bohrn; diir; dofiir; kameradrohne; eurokrise; dorfmodrotz; horoskop; dschiro; fabrizirn; febraar; fischer; ohrn; fora; fuori;

gartn; gekroblt; getriller; getroogn; girtl; halodri; hoangart; hoor; integral; integrirn; iraze; jangger; karg; kearn; kotzngrint; kraaln; kraxn; kriibeskraabes; krippe; kroogn; kropfn; laar; lattern; magari; maschgra; maustroppl; neutral; nudltriebl; ort; paarl; paragrof; kolgrov; puschtra; raadl; rappl; rearl; riibl; rittn; rodrigo; roggn; roudl; sakrisch; schlutzkropfn; schmårrn; schneakralle; siirig; stirl; stork; tirk; töldra; traaget; tratzn; triib; tritt; troogn; troppl; vorarlberg; weiberer; wirt.

DOI: 10.17469/O2101AISV000009

ROSALBA NODARI

Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi

This paper investigates the role of internal and external factors triggering aspiration of voiceless stops in the Regional Italian spoken by a group of adolescents from central Calabria. The results showed that variation in VOT duration has both phonetic and extra-linguistic explanations. The findings showed that aspiration is generally favoured in velar and post-stress stops. At the same time, individual speakers show fine-grained differences in the realization of aspiration as a function of their social characteristics. In particular, a significant effect of the attitude towards schooling, with males with a negative attitude showing significantly longer VOT, was found.

1. *Introduzione*

Ciò che si propone in questo lavoro è cercare di fornire un quadro interpretativo di un fenomeno linguistico presente in alcuni dialetti meridionali estremi e nei corrispettivi italiani regionali – e cioè l'aspirazione delle occlusive sorde – che non sia limitato alla sola descrizione diatopica e all'assegnazione del fenomeno a uno specifico codice linguistico (e cioè dialetto o italiano), bensì tenga conto della competenza multipla (Sornicola, 1977) dei parlanti, consideri la variazione come elemento centrale di quel *continuum* di varietà (Berruto, 1993) che caratterizza il repertorio linguistico italiano, metta in relazione i fenomeni fini della variazione fonetica ai significati sociali da essi veicolati (Foulkes, Docherty, 2006), e analizzi l'identità sociolinguistica per come essa è percepita dai parlanti stessi (Eckert, 2008; Preston, 1999). Per fare ciò sarà necessario innanzitutto fornire una dettagliata descrizione del fenomeno dal punto di vista fonetico; in seconda battuta si metteranno in relazione le variabili fonetiche con la dimensione sociale, mostrando come ci sia un'interrelazione complessa, frutto delle competenze multiple e degli usi variabili dei parlanti.

Il presente studio fa parte di un progetto più ampio che si incentra sulla varietà di italiano regionale parlato a Lamezia Terme e nei paesi circostanti. Si è scelto innanzitutto di lavorare in un ambiente urbano, tenendo in considerazione come in una situazione più complessa quale è quella cittadina i processi di standardizzazione non possono essere considerati in maniera statica, bensì vanno considerati nella loro interrelazione con variabili extralinguistiche quali l'età, il sesso dei parlanti, i quartieri di residenza o la classe sociale (Sobrero, 1998). Si è poi preferito conside-

rare la varietà parlata dagli adolescenti, giacché è proprio in questa fascia d'età che normalmente si ritrova il picco massimo dell'innovazione fonologica, soprattutto in relazione alla costruzione delle identità personali (Eckert, 1988; Tagliamonte, D'Arcy, 2009). Per verificare ciò si è scelto di analizzare il parlato di un campione di parlanti giovani di due istituti scolastici di Lamezia Terme (un liceo classico e un istituto tecnico commerciale), accedendo così più facilmente a soggetti provenienti da diversi ambienti sociali, di diverse fasce d'età e di entrambi i sessi. In questo contributo saranno analizzati solamente i materiali di parlato letto provenienti dagli studenti del liceo classico.

2. *L'aspirazione delle occlusive sorde*

Le occlusive sorde aspirate sono delle consonanti il cui gesto di apertura glottidale comincia durante la fase della tenuta e raggiunge il suo massimo grado di apertura all'incirca durante la fase del rilascio. La conseguenza di questo ritardo nel raggiungimento della fase di massima apertura della glottide è che un'occlusiva sorda aspirata richiede un periodo fisiologico di chiusura completa della glottide, successivo all'apertura del diaframma, che permetta l'articolazione dei suoni sonori successivi (Ladefoged, Maddieson, 1996).

Ladefoged e Maddieson (1996: 69) evidenziano come si possano avere due principali significati associati al termine aspirazione. Da un lato, se si considera l'aspirazione come il punto finale del *continuum* delle aperture glottidali si porrà l'attenzione soltanto sul meccanismo laringeo; d'altro canto se si interpreta l'aspirazione come un periodo di assenza di sonorità dopo il rilascio di un'occlusiva, prima che le corde vocali inizino a vibrare, si porrà l'attenzione solo sugli aspetti del coordinamento temporale. Più corretto è invece considerare il fenomeno come un'interazione tra più fattori: come visto, la differenza fra un'occlusiva sorda non aspirata e la sua controparte aspirata non è solo la coordinazione temporale fra raggiungimento della massima apertura glottidale e rilascio dell'occlusiva, ma sta anche nell'ampiezza maggiore dell'apertura glottidale stessa, definita da Catford (1988: 59) di tipo *breathy,* e in diverse attività muscolari a livello tiroaritenoideo. Una proposta recente che fa tesoro di questo insieme di conoscenze considera infatti l'aspirazione come il risultato della coordinazione relativa tra chiusura glottidale e quella sovraglottidale (Nance, Stuart-Smith, 2013).

Il corrispettivo acustico principale del fenomeno è il Voice Onset Time (VOT), la relazione temporale che intercorre tra il rilascio della consonante occlusiva e l'inizio delle pulsazioni glottidali necessarie per l'articolazione di un suono sonoro. Una sua maggiore durata è in rapporto diretto con una maggiore apertura glottidale e con il tempo necessario affinché la glottide possa tornare ad assumere una conformazione adeguata per la realizzazione di un suono sonoro (Cho, Jun & Ladefoged, 2002; Nance, Stuart-Smith, 2013). Si suole riferire le variazioni del VOT nelle diverse lingue a partire da tre categorie principali:

- occlusive con VOT negativo (*voicing lead VOT)*, vale a dire le occlusive sonore, nelle quali le vibrazioni delle pliche vocali cominciano prima del rilascio;
- occlusive con VOT positivo breve (*short lag VOT*), per cui le vibrazioni del suono sonoro successivo cominciano subito dopo il rilascio;
- occlusive con VOT positivo lungo (*long lag VOT)*, per cui intercorre un intervallo di tempo maggiore fra il rilascio della consonante e la vibrazione delle pliche vocali, e cioè occlusive sorde aspirate.

2.1 Le occlusive aspirate sorde in Calabria

L'aspirazione delle occlusive sorde, sebbene assente nella varietà standard di italiano, è un fenomeno ben attestato in alcuni dialetti meridionali e nei corrispettivi italiani regionali, soprattutto nella penisola salentina e in Calabria (Canepari, 1986; Telmon, 1993; Fanciullo, Librandi, 2002), dove è di natura allofonica.

Per quanto riguarda la presenza dell'aspirazione delle occlusive sorde nei dialetti parlati in Calabria, i primi accenni sono già nella "Grammatica Storica della Lingua Italiana e dei suoi dialetti", laddove Rohlfs notava che

> in alcuni dialetti della provincia di Cosenza (territorio della Sila), si osserva una chiara aspirazione della *-t-* – per esempio (Spezzano Grande, Aprigliano, San Giovanni in Fiore) *lavathu, acithu, sintuthu* 'sentito' – fenomeno non limitato però alla posizione intervocalica (per esempio anche *fatthu*, *vienthu* 'vento') (Rohlfs, 1966: 277).

Più dettagliato è il riferimento all'aspirazione di Falcone (1976), secondo cui le occlusive sorde aspirate "almeno nella provincia di Reggio [sono] un fatto pressoché endemico nella fascia più conservativa del sistema socio-linguistico" (Falcone, 1976: 42). Nel dialetto reggino il fenomeno riguarderebbe l'intera classe delle occlusive sorde (bilabiali, dentali, palatali e alveolari) se precedute da una nasale, da una vibrante o se geminate. In quest'area vi sarebbero inoltre condizionamenti prosodico-accentuali a governare l'intensità dell'aspirazione, più o meno forte a seconda della posizione del nesso o della geminata rispetto all'accento di sillaba (Falcone, 1976: 41-45).

Considerate nel loro insieme, le annotazioni riportate in Rohlfs e Falcone non permettono di inquadrare bene il fenomeno né dal punto di vista geografico (per Rohlfs infatti l'aspirazione sembra riguardare l'area della Presila cosentina, mentre Falcone la riporta come fenomeno peculiare dell'area reggina) né da un punto di vista fonetico: come nota Sorianello (1996) non sono infatti chiare né le consonanti coinvolte (per Rohlfs solo l'occlusiva alveolare, per Falcone l'intera serie delle occlusive), né i contesti fonetici (per Rohlfs anche intervocalico, per Falcone le occlusive sorde quando geminate, post-nasali o post-rotiche).

Sorianello (1996) contiene uno studio acustico del parlato di locutori cosentini; l'autrice mette a confronto dialetto e italiano regionale, mostrando come, sia nel dialetto cosentino sia nella varietà regionale di italiano, le occlusive sorde geminate o post-rotiche sono articolate in maniera differente rispetto alle altre varietà regionali di italiano, mostrando una durata maggiore del VOT. In seguito, il fenomeno è stato

considerato da Stevens e Hajek (2010), i quali, confrontando le durate del VOT di diversi italiani regionali, mostrano come l'italiano regionale parlato a Catanzaro si differenzi massimamente solo nella classe delle occlusive alveolari geminate.

Considerando il complesso panorama sociolinguistico dell'Italia contemporanea e le innovazioni prodotte soprattutto dalle giovani generazioni nell'uso originale dei diversi codici nell'ambito di una nuova competenza multilingue che va oltre la tradizionale nozione di diglossia (Sobrero, 1992; Auer, 2005), risulta importante analizzare con attenzione l'uso che dell'aspirazione fanno i parlanti nell'interazione quotidiana e la percezione che hanno di questo tratto rispetto alla costruzione delle identità linguistica e culturale. Questo studio si propone perciò mettere in luce quali siano i contesti sociolinguistici favorevoli o sfavorevoli alla produzione dell'aspirazione, nel parlato di soggetti giovani con un diverso orientamento verso la scuola, età e sesso. L'importanza del VOT come variabile sociolinguistica capace di indicare l'esistenza di dinamiche di variazione sociolinguistica e di veicolare informazioni sullo status (reale e percepito) dei parlanti è stata messa in luce non solo nelle situazioni di contatto linguistico (Heselwood, McChrystal, 1999; Nagy, Kochetov, 2013) ma anche per varie lingue, ad esempio Scobbie (2006), Docherty et al. (2011) per le varietà di inglese parlato in Scozia, Nance e Stuart-Smith (2014) per il gaelico, Takada e Tomimori (2006) per il giapponese.

3. *Metodologia*

3.1. Materiali

I dati analizzati in questo lavoro derivano da una serie di registrazioni effettuate dall'autrice fra il febbraio e il maggio 2014 presso il liceo Classico F. Fiorentino di Lamezia Terme (CZ). Ai parlanti è stato chiesto di leggere una lista di frasi in italiano appositamente preparata, cercando di mantenere una velocità di eloquio costante, nella maniera più naturale possibile, immaginando di rivolgersi a un amico. La lista constava di 156 frasi di identica lunghezza (8 sillabe fonetiche) e struttura intonativa uniforme. Ogni frase conteneva al proprio interno una parola target contenente, in una posizione determinata, una delle occlusive sorde dei tre diversi luoghi di costrizione (/p t k/). L'occlusiva poteva essere una geminata, oppure una scempia preceduta da suono rotico o nasale[1]. Si è inoltre inserito il contesto post-laterale[2], segnalato da Canepari (1986) come possibile contesto di aspirazione, ma finora mai analizzato negli studi precedentemente citati. Si è incluso anche il contesto di occlusiva sorda scempia intervocalica come contesto di controllo. Le parole target pote-

[1] Nei dialetti parlati nell'area centrale le occlusive sorde precedute da nasale non subiscono infatti processi di sonorizzazione, come avviene ad esempio nella zona calabrese settentrionale (Falcone, 1976; Trumper, 1997; Loporcaro, 2009).

[2] Nei dialetti parlati nell'area centrale si possono riscontrare velarizzazione o dileguo ([ˈautu], [ˈatu] 'alto') o rotacismo se la laterale è seguita da labiale o velare ([ˈgurpi] 'volpe'); il fenomeno della velarizzazione risulta essere di natura variabile (Romito, Tarasi & Renzelli, 2010), ma in ogni caso non attecchisce nell'italiano regionale.

vano essere parossitone (con occlusiva sorda in posizione pretonica o postonica), o proparossitone (con occlusiva sorda in posizione postonica). Le occlusive potevano essere seguite da /i/, /u/, o /a/, in rari casi da /e/ e da /o/.

3.2 Soggetti

All'esperimento hanno partecipato 24 studenti del liceo Classico di entrambi i sessi, di cui 12 studenti di 2° ginnasio (15 anni ca.) e 12 studenti di 2° liceo (17 anni ca.). Sulla base delle interviste condotte prima della lettura del corpus di frasi, i parlanti sono stati poi suddivisi in base a quello che d'ora in poi verrà definito come 'orientamento verso la scuola'. Oltre a rispecchiare il reale andamento scolastico degli studenti, si è pensato di segnalare con questo indice il rapporto che, in generale, i ragazzi avevano verso la scuola[3]. Con orientamento positivo sono stati indicati quegli studenti che avevano scelto di propria volontà il liceo da seguire, che miravano a degli ottimi risultati e che avevano intenzione di proseguire gli studi; con orientamento negativo sono stati indicati quegli studenti, molto spesso ripetenti di qualche anno, che non avevano scelto di propria iniziativa la scuola da seguire, con risultati sotto la sufficienza e che soprattutto non credevano in ciò che la scuola offriva loro, preferendo orientarsi per il futuro verso il mondo lavorativo; con intermedi sono stati indicati invece quegli studenti, molto spesso titubanti, i quali non avevano chiaro il loro futuro di studi e che avevano un atteggiamento ambivalente nei confronti dell'istituzione scolastica.

3.3 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).

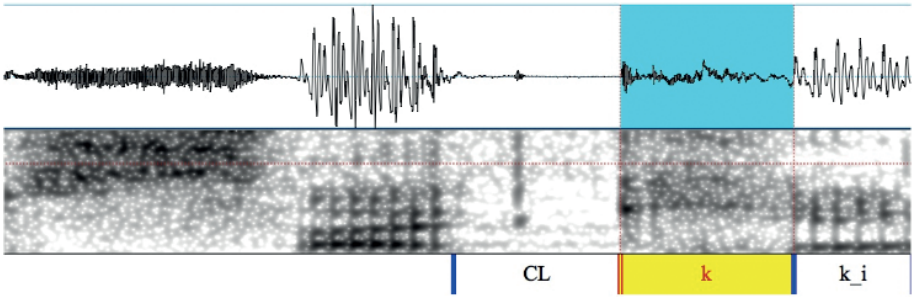
3.4 Analisi

Le frasi sono state analizzate e annotate su Praat (Boersma, Weinink, 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 che lo spettrogramma. Per l'annotazione del VOT si è scelto di segnalare l'inizio includendo il *burst*; nei casi di *burst* multipli si è adottato il criterio di Cho e Ladefoged (1999), misurando il VOT dall'ultimo *burst* presente sullo spettrogramma. I casi in cui il *burst* non era chiaramente visibile sono stati scartati. Per definire

[3] L'assegnazione dei soggetti alla variabile 'orientamento verso la scuola' non è stata costruita seguendo un indice rigido con assegnazione di punteggi, bensì è stata guidata dalla valutazione dell'autrice stessa. Attraverso la valutazione di risposte a domande somministrate a tutti i soggetti ("Chi ti ha indirizzato nella scelta di questa scuola", "Ti piace questa scuola?", "Ti piacerebbe continuare a studiare dopo il liceo e se sì, cosa vorresti studiare?" "Qual è la tua media scolastica?") si sono quindi suddivisi i soggetti nelle tre categorie di "positivi", "negativi" e "intermedi".

la fine dell'intervallo del VOT si è scelto di considerare la definizione data da Lisker e Abramson (1964) cioè l'intervallo che intercorre tra il rilascio di un'occlusione e l'attacco delle vibrazioni glottidali: di conseguenza l'attacco della vocale seguente è stato marcato nel momento in cui si riscontrava sulla forma d'onda l'attacco delle vibrazioni della vocale.

Figura 1 - *Schermata di Praat con esempio di segmentazione* (*parola 'sacchi,'* [sak^{h}:i])
CL = closure (fase di tenuta), k = VOT della consonante /k/, k_i = vocale /i/ successiva a /k/



Attraverso uno script di Praat è stato possibile estrarre automaticamente le durate in millisecondi del VOT di 2839 occlusive.

La durata del VOT è stata testata in relazione a diverse variabili, alcune prettamente fonetiche e altre di natura sociolinguistica. Le variabili prese in analisi nel seguente contributo sono: luogo di articolazione dell'occlusiva, contesto fonetico, posizione dell'accento, timbro della vocale seguente (limitando l'analisi ai soli casi di C/i/, C/u/ e C/a/), età, sesso del parlante, orientamento verso la scuola. È stata eseguita un'analisi della varianza, seguita da tests post-hoc quando necessario.

4. *I risultati*

4.1 Variabili fonetiche

Di seguito si riportano i risultati dell'analisi delle variazioni del VOT in funzione delle variabili fonetiche, e cioè luogo di articolazione della consonante, contesto di occorrenza, posizione dell'accento e timbro della vocale seguente.

Rispetto agli effetti principali, si riscontra nei dati un chiaro effetto del luogo di articolazione ($F(2, 2723)=195.522$, $p<0.05$), per cui, in linea con le aspettative (Cho, Ladefoged, 1999), /k/ presenta valori di VOT maggiori (media=25 ms. dev. st.=9.7 in contesto intervocalico, media=50 ms. dev.st.=16.7 in contesti di aspirazione) rispetto a /t/ (media=19 ms. dev.st.=6.9 in contesto intervocalico, media=34 ms. dev.st.=14.0 in contesto di aspirazione) e quest'ultima rispetto a /p/ (media=14 ms. dev.st.=5.1 in contesto intervocalico, media=27 ms. dev.st.=12.2 in contesti di aspirazione).

È stato inoltre riscontrato un effetto del contesto di occorrenza ($F(4, 2723)=51.302$, $p<0.05$). Il contesto intervocalico, il quale non dovrebbe favorire

l'aspirazione, mostra infatti durate del VOT inferiori (media=20 ms. dev.st.=9.1) rispetto ai contesti di aspirazione, e cioè il contesto di geminazione (media=37 ms. dev.st.=17.5), dopo un suono rotico (media=38 ms. dev.st.=16.9), dopo un suono nasale (media=39 ms. dev.st.=18.6), e dopo laterale (media=36 ms. dev.st.=17.0). I contesti di geminazione, post-rotica, post-nasale e post-laterale si comportano invece in maniera omogenea, con differenze non statisticamente significative.

Anche la posizione dell'accento contribuisce a determinare le variazioni nella durata del VOT delle occlusive sorde nei contesti di aspirazione ($F(2, 2326)=9.377$, $p<0.05$), con durate maggiori per le occlusive postoniche proparossitone (media=40 ms. dev.st.=16.9) e parossitone (media=38 ms. dev.st.=17.8) rispetto alle occlusive pretoniche parossitone (media=35.9 ms. dev.st.=17.6). I tests post-hoc mostrano che tutti questi contrasti sono significativi ($p<0.05$ in tutti i casi).

In linea con le aspettative (Morris, McCrea & Herring, 2008), infine, è stata trovata una differenza significativa nella durata del VOT causata dal timbro della vocale seguente ($F(2, 2326)=232.190$ $p<0.05$). Le durate maggiori si riscontrano per le occlusive sorde seguite da vocali alte /i/ (media=43 ms. dev.st.=19.4) e /u/ (media=43 ms. dev.st.=16.5) in opposizione alle occlusive seguite da /a/ (media=31 ms. dev.st.=13.2; $p<0.05$), mentre non sono significativi i contrasti tra /i/ e /u/.

Tra le interazioni, sono risultate significative quelle tra posizione dell'accento e contesto di occorrenza, tra posizione dell'accento e timbro vocalico seguente e tra luogo articolatorio e timbro della vocale seguente. Non risultano invece significative le interazioni fra luogo articolatorio e posizione dell'accento, tra contesto di occorrenza e timbro della vocale seguente e tra luogo articolatorio e contesto di occorrenza.

Per quanto riguarda l'interazione non significativa tra luogo articolatorio e contesto di occorrenza, l'effetto del luogo di articolazione risulta agire allo stesso modo sia nei contesti di aspirazione (geminazione, post-rotica, post-nasale, post-laterale) sia nel contesto intervocalico, che non favorisce l'aspirazione. I tests post-hoc mostrano infatti che, sia nei contesti di aspirazione che nel caso delle scempie intervocaliche, tutti i contrasti fra i tre luoghi di articolazione sono significativi, ossia il VOT di /k/ è significativamente più lungo di quello di /t/ e di /p/ mentre il VOT di /t/ è significativamente più lungo di quello di /p/ ($p<0.05$ in tutti i casi). Questo significa dunque che anche l'aspirazione della varietà calabrese qui studiata obbedisce al ben noto effetto fisiologico della ritrazione del punto di costrizione sulla durata del VOT delle occlusive (Cho, Ladefoged, 1999), senza apparenti deviazioni rispetto alla tendenza generale valida per tutte le occlusive, anche quelle considerate non aspirate.

Per quanto riguarda l'interazione tra la posizione dell'accento e il contesto di occorrenza ($F(8, 2723)=2.961$, $p<0.05$), le postoniche (parossitone e proparossitone) sono caratterizzate da VOT più lungo di quello delle pretoniche solo in due contesti di aspirazione su quattro, e cioè in caso di geminazione (media=40 ms. per le postoniche, media=35 ms. per le pretoniche) e in occlusiva post-rotica (media=41 ms. per le postoniche, media=34 ms. per le pretoniche); negli altri due con-

testi (post-nasale e post-laterale) le differenze tra le condizioni accentuali non sono altrettanto nette, sebbene le proparossitone postoniche mantengano le durate maggiori (media=42 ms. per il contesto di postnasale, media=39 ms. per il contesto di post-laterale). Nel contesto che non favorisce l'aspirazione (occlusiva sorda scempia intervocalica) le differenze tra condizioni accentuali non sono altrettanto significative (media=19 ms. dev.st.=7.6 per le parossitone postoniche, media=21 ms. dev. st.=10.8 per le parossitone pretoniche, media=20 dev.st.=7.7 per le proparossitone postoniche). Questo dunque sembra indicare che le differenze legate alla posizione accentuale non sono altrettanto forti e generalizzate nel lessico come quelle legate al luogo articolatorio, ma dipendono in ampia misura dalla cooccorrenza di altre caratteristiche contestuali (nella fattispecie, dalla lunghezza della consonante, e dalla presenza di una rotica precedente).

L'interazione tra posizione dell'accento e timbro vocalico seguente ($F(4, 2723)=848.250$), $p<0.05$) mostra che le occlusive sorde seguite da vocali chiuse /i/ e /u/ hanno durate maggiori rispetto a quelle seguite da /a/ quando pretoniche in parole parossitone (media=41 ms. dev.st.=19.8 per /i/, media=41 ms. dev.st.=16.8 per /u/, media=27 ms. dev.st.=11.9 per /a/) e postoniche in parole proparossitone (media=44 ms. dev.st.=18.3 per /i/, media=48 ms. dev.st.= 14.4 per /u/, media=35 ms. dev.st.=14.3 per /a/), ma non nei contesti postonici parossitoni, nei quali si ritrova una durata maggiore solo per /i/ (media=40 ms. dev.st.=20.0). Anche questo dato sembra indicare che le differenze legate alla posizione accentuale non sono altrettanto forti e generalizzate nel lessico come quelle legate al luogo articolatorio, ma dipendono in ampia misura dalla cooccorrenza di altre caratteristiche contestuali (nella fattispecie, dalla presenza di una vocale alta seguente).

Infine, l'interazione tra luogo articolatorio e timbro vocalico seguente ($F(4, 2723)=27.475$, $p<0.05$) mostra come i valori più alti siano quelli delle occlusive velari seguite da /i/ (media=59 ms. dev.st.=16.9) e /u/ (media=54 ms. dev.st.=13.9) mentre i più bassi siano quelli delle bilabiali seguite da /a/ (media=23 ms. dev. st.=10.9). La tendenza per cui le occlusive velari, quando seguite da vocali alte /i/ e /u/, sono concordi nel mostrare le durate maggiori è un dato abbastanza atteso, essendo stato riscontrato anche in altre lingue come l'inglese, il francese o il cinese (Rochet, Yanmei, 1991; Neary, Rochet, 1994).

In conclusione, dai dati qui presentati si desume che tutti i fattori presi in esame concorrono nel modulare la durata del VOT, ma non tutti con lo stesso peso. Infatti, il fattore più pesante sembra essere il luogo di articolazione delle occlusive, seguito dalle differenze legate al timbro della vocale adiacente (con /i/ e /u/ che favoriscono l'allungamento del VOT); meno chiari, e comunque dipendenti dall'interazione con gli altri fattori, sono gli effetti del contesto di occorrenza (se si esclude la netta bipartizione tra scempie intervocaliche e tutti gli altri contesti analizzati) e gli effetti delle variazioni prosodico-accentuali (se si eccettua una generale preferenza per l'allungamento del VOT in sillaba atona dopo un accento lessicale). Dato che tutte le tendenze riscontrate sono coerenti con le predizioni basate sui meccanismi di interazione tra movimenti glottidali e dinamiche sovraglottidali (Nance, Stuart-

Smith, 2013), si può concludere che l'aspirazione delle occlusive sorde nella varietà calabrese di Lamezia Terme risponde a principi generali articolatori e aerodinamici in misura tendenzialmente pari alle variazioni del VOT generalmente riscontrate per qualsiasi classe di occlusiva. Gli allungamenti del VOT nel parlato dei giovani lametini analizzati, infatti, pur essendo quantitativamente maggiori di quelli riscontrabili in altre varietà di italiano regionale, non appaiono sostanzialmente diversi, dal punto di vista qualitativo, dalle variazioni predicibili in base ai principi generali dell'interazione tra movimenti degli articolatori orali e laringali e della loro caratterizzazione aerodinamica.

4.2 Variabili sociolinguistiche

Di seguito si riportano i risultati dell'analisi delle variazioni nelle durate del VOT in funzione delle variabili sociolinguistiche, e cioè età, sesso e orientamento verso la scuola.

Rispetto agli effetti principali non è stato riscontrato un effetto statisticamente significativo del fattore età (quindicenni: media=40 ms. dev.st.=16 vs. diciottenni: media=36 ms. dev.st.=17).

Non è stato riscontrato neanche un effetto significativo del fattore sesso (maschi: media=38 ms. dev.st.18 vs. femmine: media=36 ms. dev.st.=16).

Infine, è stato invece trovato un effetto del fattore orientamento verso la scuola ($F(2, 2407)=36.967$, $p<0.05$) con valori più alti per i soggetti con orientamento negativo (media=38 ms. dev.st.=16) rispetto ai soggetti con orientamento positivo (media=30 ms. dev.st.=13) e intermedio (media=30 ms. dev.st.=13).

Tra le interazioni, sono risultate significative quelle tra sesso e orientamento verso la scuola, ed età e orientamento verso la scuola. Non risulta invece significativa l'interazione fra sesso ed età.

L'interazione tra il sesso e l'orientamento verso la scuola ($F(1, 2407)=11.937$, $p<0.05$) mostra come i valori più alti siano riscontrati tra i soggetti di sesso maschile con orientamento negativo (media=41, dev.st.=15), di contro ai valori più bassi mostrati dai soggetti di sesso femminile con orientamento positivo (media=30 ms. dev.st.=13.3).

Per quanto riguarda l'interazione tra l'età e l'orientamento verso la scuola ($F(1, 2407)=7.320$, $p<0.05$), i valori più alti si ritrovano tra gli studenti più giovani con orientamento negativo (media=39 ms. dev.st.=16) rispetto ai meno giovani con orientamento positivo (media=30 ms. dev.st.=14).

Da questo quadro, si deduce che non è sufficiente guardare solo il sesso o l'età dei parlanti, bensì sia l'interazione tra più fattori a essere correlata con l'aspirazione; il rapporto trovato tra l'orientamento verso la scuola e le variabili di sesso ed età mette in luce come agli estremi del *continuum* ci siano da una parte studentesse più grandi con orientamento positivo verso la scuola, il cui VOT è più breve, e dall'altra studenti di sesso maschile più giovani, con orientamento negativo verso la scuola, che invece rappresentano il nucleo di parlanti con durate più lunghe.

5. *Discussione*

5.1 Variabili fonetiche

Per quanto riguarda il luogo di articolazione, i risultati ottenuti sono coerenti con le aspettative nel mostrare durate maggiori per le velari, seguite da alveolari e bilabiali; il volume della cavità dietro il punto di costrizione influisce sulla durata del VOT, nella misura in cui occlusive posteriori avranno un VOT comparativamente più lungo di occlusive anteriori. Nelle occlusive posteriori, infatti, il volume dell'area sovraglottidale è più ridotto e conseguentemente causa un aumento di pressione, la quale impiegherà un tempo maggiore per dissiparsi e per permettere di raggiungere adeguati valori di pressione transglottidale per la vibrazione delle pliche vocali (Cho, Ladefoged, 1999). L'effetto viene messo anche in relazione alla maggiore estensione del contatto, per cui da un'area di contatto più estesa consegue un rilascio più lento, e al movimento degli articolatori, per cui nelle bilabiali le labbra riescono a separarsi velocemente l'una dall'altra, laddove per le velari il post-dorso impiega più tempo per distaccarsi dal velo palatino, necessitando così un tempo maggiore per raggiungere la fase di apertura. I dati sono invece in controtendenza rispetto ai dati di Sorianello (1996) sul dialetto cosentino, nel quale le durate di velare e alveolare in contesti di aspirazione non differiscono in maniera statisticamente significativa; sono invece paragonabili ai dati, presenti sempre nello stesso lavoro, riguardanti l'italiano regionale prodotto da locutori cosentini, il quale mostra la stessa gerarchia qui riscontrata.

I contesti di aspirazione, e cioè occlusiva sorda geminata, occlusiva sorda post-rotica, post-nasale e post-laterale mostrano delle durate simili statisticamente non rilevanti. Per quanto il fenomeno dell'aspirazione sia di origine dialettale, l'italiano regionale mostra una generalizzazione per cui esso si presenta anche nei contesti che in dialetto possono avere come esito una occlusiva scempia intervocalica, come i casi di post-laterale: per quanto l'interferenza fra i due codici esista, il fenomeno sembra avere quindi una propria vitalità autonoma nell'italiano regionale.

I dati riguardanti la posizione dell'accento mostrano come viga in generale una tendenza per cui la durata del VOT è generalmente più lunga nelle occlusive postoniche; ciò vale tanto per le parole proparossitone che per le parossitone, in controtendenza con la gerarchia proposta da Falcone (1976) ma in accordo con Sorianello (1996). In ogni caso si è notato come le differenze legate alla posizione accentuale non siano altrettanto generalizzate quanto quelle legate al luogo articolatorio, ma dipendono in ampia misura dalla cooccorrenza di altre caratteristiche contestuali (nella fattispecie, dalla lunghezza della consonante, dalla presenza di una rotica precedente e di una vocale alta seguente).

Per quanto riguarda il timbro della vocale seguente, si nota come esso concorra nell'influenzare la durata del VOT. Il fenomeno è in accordo con quanto riportato in bibliografia, per cui la durata del VOT tende ad aumentare davanti a vocali alte e chiuse e a diminuire davanti a vocali basse e aperte (Ohala, 1981; Ohala, 1983; Morris, McCrea & Herring, 2008). Le ragioni sarebbero da ricondurre al fatto che le vocali alte e chiuse offrono una maggiore resistenza alla fuoriuscita dell'aria dalla

cavità orale, ritardando quindi il raggiungimento di una pressione transglottidale adeguata per la vibrazione: un picco più basso del flusso d'aria per le vocali alte fa sì che l'abbassamento della pressione nell'area sovraglottidale necessario per creare la differenza di pressione transglottidale sia raggiunto più tardi.

5.2 Variabili sociolinguistiche

I risultati presentati permettono di offrire un quadro del fenomeno dell'aspirazione che renda conto anche della sua natura variabile e dell'utilizzo che di essa fanno i parlanti. Si è visto come il comportamento dei soggetti non sia assolutamente uniforme, e anzi solo attraverso una selezione di variabili sociolinguistiche non selezionate a priori ma situate nel contesto scolastico, come può essere ad esempio il fattore 'orientamento verso la scuola', sia possibile osservarne la loro natura variabile.

Particolarmente importante risulta la variabile relativa all'orientamento verso la scuola, che fornisce preziose indicazioni sulla distribuzione del fenomeno dell'aspirazione rispetto alla collocazione sociolinguistica e socio-culturale dei parlanti. Le differenze di genere prese singolarmente non spiegano la variabilità nelle realizzazioni dell'aspirazione, ma modulano invece l'effetto dell'orientamento verso la scuola. Infatti non sono semplicemente i parlanti di sesso maschile ad avere durate maggiori, ma i maschi con orientamento verso la scuola negativo. Anche per le femmine l'orientamento è direttamente correlato alla variabile del VOT: più l'orientamento è positivo più il VOT sarà breve. Si può quindi ipotizzare che la manipolazione del fenomeno dell'aspirazione sia considerata a tutti gli effetti un tratto marcato, dotato di prestigio 'coperto' (Labov, 1963; Trudgill, 1972) e significativamente correlato alle attitudini verso la scuola.

Considerazioni simili possono essere fatte per la variabile età. Come per la variabile sesso, la variabile età non risulta significativa, a meno che non sia vista nella sua correlazione con la variabile 'orientamento verso la scuola' e si ipotizzi che i ragazzi che si accostano alla maggiore età comincino a essere maggiormente consapevoli di ciò che è giudicato positivamente nel 'mercato linguistico' (Bourdieu, 1992; Chambers, 2003). Si può pensare come l'avvicinarsi dell'ingresso nel mondo universitario spinga i ragazzi con orientamento positivo verso la scuola – e quindi con un maggiore interesse a proseguire gli studi - a liberarsi di ciò che ai loro occhi sembra ancorato al codice dialettale giacché non necessario per il proprio avanzamento personale e professionale. Superata la fase critica in cui l'adolescente ha bisogno del gruppo dei pari per trovare la propria legittimazione, il processo di costruzione dell'identità comincia a farsi via via più personale e privato. Secondo Kerswill (1996), durante la preadolescenza, i bambini dai 6 ai 12 anni scelgono una norma linguistica che è differente da quella dei 'grandi', orientandosi verso la rete sociale del gruppo dei pari[4]; i ragazzi oramai adolescenti (dai 12 ai 17 anni) cominciano invece

[4] Valga come esempio il caso citato da Kerswill (1996), che mostra come i preadolescenti di Milton Keynes siano concordi nel mostrare una regola fonologica non-standard caratteristica dei parlanti dell'Inghilterra sudorientale, (il *Roland-rolling split*, in altre parole l'abbassamento della vocale che

ad acquisire (ma non per forza a sposare) una conoscenza di ciò che sono le norme degli adulti, assieme a una maggiore consapevolezza delle regole dello *style-shifting*. Nel caso degli adolescenti calabresi è possibile osservare il picco di consapevolezza che si realizza proprio in questa finestra temporale, notando lo scarto che avviene tra i quindicenni e i loro compagni più grandi, già maggiorenni. Del resto ogni parlante "crea i sistemi del suo comportamento verbale in modo tale che essi somiglino a quelli del gruppo o dei gruppi con i quali di volta in volta potrà volere essere identificato" (Le Page, Tabouret Keller, 1985: 181), e questo può portarci a credere che ciò abbia un riflesso sul suo comportamento linguistico vero e proprio.

6. *Conclusioni e sviluppi futuri*

I risultati mostrano come le occlusive sorde presenti nella varietà di italiano parlata a Lamezia Terme presentano degli allungamenti significativi del VOT in determinati contesti di occorrenza (occlusive sorde geminate, post-rotiche, post-nasali o post-laterali). Sebbene differenti quantitativamente dalle altre varietà regionali di italiano, le occlusive sorde dell'italiano regionale calabrese sono comunque in linea con le predizioni basate sui principi generali dell'interazione tra movimenti degli articolatori orali e laringali e della loro caratterizzazione aerodinamica.

Si è inoltre mostrato come il fenomeno dell'aspirazione, esaminato attraverso l'analisi delle durate del VOT, sia soggetto a una variazione fonetica di tipo fine da mettere direttamente in relazione con variabili di tipo sociolinguistico, quali possono essere il sesso, l'età e l'orientamento verso le istituzioni scolastiche.

Negli studi futuri si analizzerà prima di tutto l'interazione tra variabili sociolinguistiche e variabili fonetiche, per avere un quadro più dettagliato del comportamento linguistico dei diversi gruppi sociali presi in esame. Inoltre si amplierà l'analisi ai dati raccolti nella seconda scuola e si cercherà di vedere se vi sono comportamenti contrastanti fra gli studenti del liceo classico e gli studenti dell'istituto tecnico commerciale. L'analisi sociolinguistica verrà inoltre ampliata grazie all'analisi delle produzioni di parlato spontaneo, già raccolte sul campo, che permetteranno di confermare, o eventualmente smentire, i risultati ottenuti dalla lettura delle frasi.

Verranno inoltre considerati i dati relativi alle durate della fase di occlusione, da testare come eventuale parametro che permetta di normalizzare tenendo sotto controllo la velocità di eloquio, e alle modalità di fonazione della vocale seguente[5], per

precede una liquida /l/ solo in confine di morfema, es. r[ʊə]lling 'rolling' vs R[äë]land 'Roland') anche quando questa non è presente nel parlato dei propri genitori.

[5] La speciale coordinazione temporale richiesta per produrre una consonante aspirata può interferire anche con la modalità di fonazione della vocale seguente (Löfqvist, 1992). È infatti probabile che nel momento di attacco della vocale la glottide non abbia ancora raggiunto una posizione di massima chiusura: si avrà quindi un breve intervallo durante il quale si ha vibrazione delle pliche vocali assieme a un certa apertura glottidale che porta a una modalità di fonazione non modale (Jessen, 1999: 109).Per questo motivo sono in fase di programmazione delle analisi che tengano conto di H1-H2 (Esposito, 2010) come indice di modalità di fonazione di tipo *breathy* della vocale seguente una occlusiva sorda aspirata.

verificare eventuali effetti coarticolatori riscontrati in altre lingue nel cui inventario sono presenti delle occlusive sorde aspirate.

Bibliografia

Auer, P. (2005). Europe's sociolinguistic unity, or: a typology of European dialect/standard constellations. In Delbecque, N., van der Auwera, J. & Geeraerts, D. (Eds.), *Perspectives on Variation. Sociolinguistic, historical, comparative*. Berlin: Mouton de Gruyter, 7-42.

Berruto, G. (1993). Le varietà del repertorio. In Sobrero, A.A. (Ed.), *Introduzione all'italiano contemporaneo, la variazione e gli usi*. Roma - Bari: Laterza, 3-36.

Boersma, P., Weenink, D. (2015). *Praat: doing phonetics by computer [Computer program]*, Version 5.4.09, retrieved 1 June 2015 from http://www.praat.org.

Bourdieu, P. (1992). *Language and Symbolic Power*. Cambridge: Polity Press.

Canepari, L. (1986). *Italiano standard e pronunce regionali*. Padova: CLEUP.

Catford, J.C. (1988). *A practical introduction to phonetics*. Oxford: Clarenton Press.

Chambers, J.K. (2003). *Sociolinguistic Theory*, second edition. Oxford: Blackwell.

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

Cho, T., Jun, S. & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. In *Journal of Phonetics*, 30, 193-228.

Docherty, G., Watt, D., LLamas, C., Hall, D. & Nycz, J. (2011). Variation in Voice Onset Time along the Scottish-English border. In Lee, W., Zee, E. (Eds.), *Proceedings of the 17th ICPhS, Hong Kong*. 591-594.

Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. In *Language in Society*, 17 (2), 183-207.

Eckert, P. (2008). Variation and the indexical field. In *Journal of Sociolinguistics*, 12, 453-476.

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.

Falcone, G. (1976). *Calabria*. Pisa: Pacini.

Fanciullo, F., Librandi, R. (2002). La Calabria. In Clivio, G. et al. (Eds.), *I dialetti italiani. Storia, struttura, uso.* Torino: UTET, 793-833.

Foulkes, P., Docherty, G. (2006). The social life of phonetics and phonology. In *Journal of Phonetics*, 34, 409-438.

Heselwood, B., McChrystal, L. (1999). The effect of age-group and place of L1 acquisition on the realisation of Panjabi stop consonants in Bradford: an acoustic sociophonetic study. In *Leeds Working Papers in Linguistics & Phonetics*, 7, 49-68.

Jessen, M. (1999). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamin.

Kerswill, P. (1996). Children, adolescents and language change. In *Language Variation and Change*, 8, 177-202.

Labov, W. (1963). The social motivation of a sound change. In *Word*, 19, 273-309.

Ladefoged, P., Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Le Page, R., Tabouret-Keller, A. (1985). *Acts of Identity: creole-based approaches to language and ethnicity*. Cambridge: University Press.

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. (1992). Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants. In *Language & Speech*, 35, 15-28.

Loporcaro, M. (2009). *Profilo linguistico dei dialetti italiani*. Bari: Laterza.

Morris, R.J., McCrea, C.R. & Herring, K.D. (2008). Voice onset time differences between adult males and females: isolated syllables. In *Journal of Phonetics*, 36, 308-317.

Nagy, N., Kochetov, A. (2013). Voice onset time across the generations: a cross- linguistic study of contact-induced change. In Siemund, P., Gogolin, I., Schulz, M.E. & Davydova, J. (Eds.), *Multilingualism and language diversity in urban areas: acquisition, identities, space, education*. Amsterdam: John Benjamins, 19-38.

Nance, C., Stuart-Smith, J. (2013). Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. In *Journal of the International Phonetic Association*, 43 (2), 129-152.

Neary, T., Rochet, B.L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. In *Journal of the International Phonetic Association*, 24 (1), 1-18.

Ohala, J.J. (1981). Articulatory constraints on the cognitive representation of speech. In Myers, T., Laver, J. & Anderson, J. (Eds.), *The cognitive representation of speech*. Amsterdam: North Holland, 111-122.

Ohala, J.J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P.F. (Ed.), *The production of speech*. New York: Springer-Verlag, 189-216.

Preston, D.R. (1999). A language attitude approach to the perception of regional variety. In Preston, D.R. (Ed.), *Handbook of perceptual dialectology*. Amsterdam: John Benjamins, 359-373.

Rochet, B.L., Yanmei, F. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. In *Canadian Acoustics*, 19 (4), 105-106.

Rohlfs, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino: Einaudi.

Romito, L., Tarasi, A. & Renzelli M. (2011). Verso una ridefinizione dei confini linguistici dialettali: uno studio fonetico-fonologico in alcuni centri della Calabria. In Gili Fivela, B., Stella, A., Garrapa, L. & Grimaldi, M. (Eds.), *Proceedings del VII Convegno AISV*. Roma: Bulzoni, 37-47.

Scobbie, J. (2006). Flexibility in the face of incompatible English VOT systems. In Goldstein, L., Whalen, H. & Best, C. (Eds.), *Laboratory Phonology VIII: varieties of phonological competence.* Berlin: Mouton de Gruyter, 367-92.

Sobrero, A.A. (1992). *Il dialetto nella conversazione. Ricerche di dialettologia pragmatica.* Roma: Galatina.

Sobrero, A.A. (1988). Villages and towns in Salento: the way code-switching switches. In Dittmar, N. & Schlobinski, P. (Eds.), *The sociolinguistics of urban vernaculars: case studies and their evaluation*. Berlin: Mouton De Gruyter, 207-216.

Sorianello, P. (1996). Indici fonetici delle occlusive sorde nel cosentino. In *Rivista Italiana di Dialettologia*, 20, 123-159.

Sornicola, R. (1977). *La competenza multipla: un'analisi micro-sociolinguistica*. Napoli: Liguori.

Stevens, M., Hajek, J. (2010). Post-aspiration in standard Italian: some first cross-regional acoustic evidence. In *Proceedings of Interspeech, Makuhari, Japan*, 1557-1560.

Tagliamonte, S., D'Arcy, A. (2004). Peaks beyond phonology: adolescence, incrementation, and language change. In *Language*, 85 (1), 58-108.

Takada, M., Tomimori, N. (2006). The relationship between VOT in initial voiced plosives and the phenomenon of word-medial plosives in Nigata and Shikoku. In Kawaguchi, Y., Zaima, S. & Takagaki, T. (Eds.), *Spoken Language Corpus and Linguistic Informatics.* Amsterdam: John Benjamin, 365-379.

Telmon, T. (1993). Varietà regionali. In Sobrero, A.A. (Ed.), *Introduzione all'italiano contemporaneo, la variazione e gli usi*. Roma - Bari: Laterza, 93-149.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. In *Language in Society*, 1 (2), 175-195.

Trumper, J. (1997). Calabria and Southern Basilicata. In Maiden, M., Parry, M. (Eds.), *The dialects of Italy*. London: Routledge, 355-364.

DOI: 10.17469/O2101AISV000010

VALENTINA PAGGINI, SILVIA CALAMAI[1]

L'anafonesi in Toscana. Il contributo degli archivi sonori del passato

The paper provides a geolinguistic analysis of the so called 'anaphonesis of second type' (Castellani, 1952), a phonetics phenomenon that concerns the changing of high mid vowels to [i] and [u] before nasal + velar, e.g. ['loŋgu] > ['luŋgo], 'long'; ['leŋgua] > ['liŋgua], 'tongue'; ['veŋko] > ['viŋko], 'I win'. The analysis is based on the inspection of *Carta dei Dialetti Italiani* (CDI) sound archive, one of the most important geolinguistic entreprises of the last century. The surveys undertaken in Tuscany have been recovered, digitized, catalogued, and published for scientific use by *Grammo-foni. Le soffitte della voce (Gra.fo)*, a research project financed by Regione Toscana, PAR FAS 2007-13. The sound archive of CDI includes recordings collected in the 60s and 70s by scholars involved in a nation-wide research project on Italian dialects. It originally implied a survey on 500 voices and the collection of a dialectal version of the *Parable of the prodigal son* in every Italian municipality; in any event, all the localities investigated were at that time represented by several informants with fieldwork sessions recorded on tape. Today, these recordings are among the oldest available materials on Tuscan dialects. The geolinguistic picture emerging from the auditory inspection of the data seems to collide with Castellani's description; in this period, Tuscany already appears to be almost completely anaphonetical and the absence of the phenomenon is limited to the southern areas as the countryside around the cities of Siena and Arezzo.

Introduzione

Discipline come la sociologia, l'antropologia o la storia orale si avvalgono già da tempo dei documenti contenuti negli archivi sonori del passato utilizzandoli come fonti per avviare riflessioni inedite su argomenti talvolta poco studiati, ovvero per aprire nuove possibili strade di ricerca. In ambito linguistico, gli archivi sonori possono contribuire a meglio comprendere le dinamiche geolinguistiche e sociolinguistiche di fenomeni a volte non troppo problematizzati dalla letteratura specialistica. Può essere questo il caso del noto fenomeno di fonetica storica etichettato da Arrigo Castellani (1952) come 'anafonesi di secondo tipo', oggetto del presente studio, incentrato sulla diffusione delle forme anafonetiche nelle parlate della Toscana all'altezza degli anni Sessanta e Settanta del Novecento. La base dei dati è rappresenta-

[1] Il lavoro rappresenta l'elaborazione di una tesi di laurea magistrale in Glottologia e Linguistica generale (a.a. 2013-2014), Dipartimento di Scienze della Formazione, Scienze Umane e della Comunicazione Interculturale di Arezzo, Università degli Studi di Siena. Per fini meramente accademici, a VP devono essere attribuiti i §§ 1, 2, 3, a SC l'Introduzione e il § 4. Le autrici ringraziano due revisori anonimi per le utilissime osservazioni critiche ricevute.

ta dalle 462 inchieste toscane relative all'archivio sonoro della *Carta dei Dialetti Italiani*, recuperato, digitalizzato e messo a disposizione degli studiosi dal progetto finanziato dalla regione Toscana *Grammo-foni. Le soffitte della voce (Gra.fo)*. Il quadro geolinguistico che emerge dall'analisi va ad integrare l'analisi dialettologica di Castellani (1980) e Giannelli (2000), soprattutto per quanto concerne la persistenza di forme non anafonetiche nella Toscana meridionale.

Il lavoro è così strutturato: in § 1 viene presentato il fenomeno nei suoi aspetti fonetici e dialettologici, in § 2 sono descritti l'impresa della *Carta dei dialetti italiani* nel suo complesso e il segmento del questionario spogliato per lo studio, nonché alcune criticità nell'interpretazione dei risultati; il § 3 riporta l'analisi e il quadro geolinguistico risultante. Chiude il lavoro una riflessione metodologica incentrata sul riuso dei materiali di precedenti imprese di ricerca. Il progetto *Gra.fo* ha messo a disposizione degli studiosi una mole rilevante di documenti linguistici del passato, che possono e debbono essere rimessi in discussione dalla comunità scientifica, non solo dei linguisti. Tale riutilizzo si porta dietro questioni etiche e legali di non poco conto: anche la comunità dei linguisti è dunque chiamata a interrogarsi sui modi di accesso e di riuso dei dati raccolti in passato da gruppi di ricerca e da singoli, dati che in mancanza di una sensibilizzazione e di una azione capillari rischiano di essere perduti per sempre.

1. *L'anafonesi in Toscana*

Se è vero che si deve ad Arrigo Castellani l'utilizzo del termine *anafonesi* per indicare il fenomeno linguistico fra quelli "che servono a connotare il fiorentino"[2] (Castellani, 1952: 21), è anche vero che la prima attestazione della parola compare in ambito letterario (Barbato in stampa); è infatti del 1946 l'uso di anafonesi con il significato di 'riavvicinamento alla forma originale, etimologica di una parola' (LEI) per mano di Carlo Emilio Gadda che in *Quer pasticciaccio brutto de via Merulana* scrive a proposito del cognome veneto *Menegazzi* che a Roma viene storpiato in *Menicacci* o *Menecacci*: "Sui loro labbri stupendi quel nome veneto risaliva l'etimo, puntava contro corrente, cioè contro l'erosione operata dagli anni. L'anafonesi trivellava il deflusso col perforante vigore d'un'anguilla"[3], a indicare la continua insistenza dei romani nel pronunciare male il cognome tipico di un dialetto molto lontano dal loro.

Ci troviamo, quindi, di fronte a un gioco linguistico dove al termine *anafonesi* viene attribuito un significato diverso rispetto a quello etimologico di 'innalzamento tonale' che poi Castellani ha assegnato alla parola rifacendosi al termine greco (Castellani, 1952) e con cui Gadda si diverte a stupire e confondere il lettore in linea con l'estrema poliedricità linguistica dei suoi romanzi. Non le *Meraviglie d'I-*

[2] Cfr. De Mauro ([1963] 2011: 305), Loporcaro ([2009] 2013: 117), Calamai (2011).

[3] *Quer pasticciaccio brutto de via Merulana*, prima di vedere la sua pubblicazione come romanzo nel 1957, esce, a partire dal 1946 (gennaio-febbraio, anno VIII, n. 1), nella rivista *Letteratura*, periodico bimestrale di letteratura contemporanea diretta da Alessandro Bonsanti. La citazione è a pagina 80.

talia dunque, come cita erroneamente l'estensore della voce *anafonesi* nel *Grande dizionario della lingua italiana* (GDLI), facendo risalire l'uso gaddiano del termine al 1939. I dizionari che nel tempo hanno registrato il significato e gli usi del termine sono nel complesso concordi nel definire la parola come grecismo medico, "esercizio del canto per dilatare i polmoni e fortificar gli organi della voce"[4]. Nonostante nel corso dei secoli siano state fornite diverse definizioni di *anafonesi*, la storia linguistica della parola merita ancora qualche ulteriore approfondimento, come del resto rilevano Cortelazzo (2000: 174) e Barbato (in stampa).

L'*anafonesi* è un fenomeno di fonetica storica per il quale si verifica un cambiamento di suono di due vocali in posizione tonica che vengono a trovarsi in determinati contesti consonantici: la /e/, proveniente da Ē o da Ĭ latine, si chiude nella vocale palatale /i/, e la /o/, proveniente da Ō o da Ŭ latine, si chiude in /u/[5]. L'anafonesi, caratteristica della sola Toscana, anzi "anticamente neppure di tutta la Toscana, ma solo di fiorentino, pratese, pistoiese, lucchese, pisano e volterrano (e non dei dialetti centrali e orientali: senese, aretino e cortonese)" (Serianni 1998: 51), è un fenomeno di natura contestuale, dipendente dai nessi consonantici che seguono la vocale in questione. Quella che Castellani ha definito "di primo tipo" (1980: 85) vede la presenza di *i* al posto di *e* dinanzi a *l* palatale [ʎ] o *n* palatale [ɲ] provenienti rispettivamente da -LJ- e -NJ-; infatti dalla base latina FAMĬLIA(M) abbiamo avuto inizialmente la forma *faméglia*, con passaggio vocalico e palatalizzazione regolari, che si è diffusa in tutti i volgari d'Italia tranne che a Firenze e nella Toscana occidentale, dove, per la vitalità dell'anafonesi, si è affermata la forma *famiglia*, poi confluita nell'italiano (Patota, 2002: 62-63). Lo stesso vale per i seguenti esempi: CĬLIUM > *ciglio*, GRAMĬNEA(M) > *gramigna*, CONSĬLIUM > *consiglio*, MATRĬNIA > *matrigna*. Il fatto che la nasale palatale debba provenire dal nesso -NJ- e non da un originario -GN- è una condizione necessaria per l'attivazione dell'anafonesi, poiché altrimenti parole come LĬGNU(M) e DĬGNU(M) avrebbero dato **ligno* e **digno*. Proprio il periodo storico in cui sarebbe avvenuta la palatalizzazione dei nessi latini -LJ- e -NJ- (tra il I e il II secolo), "ma prima che il nesso -gn- si palatalizzasse in [ɲ], ovvero prima dell'emergere delle forme *legno* e *degno* [...] (IV secolo)" (Calamai, 2011: 58), ha permesso a Castellani (1980: 86-87) di circoscrivere l'anafonesi più antica tra il III e il IV secolo.

L'anafonesi etichettata da Castellani "di secondo tipo" riguarda il passaggio di /e/ > /i/ e di /o/ > /u/ in presenza di *n* velare, cioè seguita da occlusiva velare sorda [k] o sonora [g]; è questo il caso di *lingua* che, proveniente dalla base latina LĬNGUA(M), ha avuto un'iniziale e regolare evoluzione in *léngua* – attestata nei

[4] *Dizionario Universale critico enciclopedico della lingua italiana* dell'abate D'Alberti di Villanova, 1797. Si veda anche il LEI (*Lessico etimologico italiano*) e il *Vocabolario Universale della lingua italiana* nell'edizione del 1845.

[5] In realtà, non tutti gli studiosi sono d'accordo sul fatto che si sia avuto innalzamento: le forme anafonetiche potrebbero essere anche forme conservative. L'anafonesi si trova descritta a livello storico-linguistico in Castellani (1952), Serianni (1995; 1998) e Patota (2002). La motivazione del fenomeno è tuttora ignota: tentativi di spiegazione sono in Rohlfs (1966), Castellani (1980), Tuttle (1991), Castellani (1992), Calamai-Celata (2012), Calamai (in stampa) e Barbato (in stampa).

volgari italiani fuori di Toscana – che poi, "a causa della /e/ trovatasi in contesto anafonetico, si è trasformata in *lingua*" (Patota, 2002: 64). Altri esempi sono: VĬNCO > *vinco*, IŬNCU > *giunco*, CĬNGO > *cingo*, STRĬNGO > *stringo*, FŬNGO > *fungo*. Salvo il caso di *giunco*, però, l'anafonesi non si attiva in presenza del nesso [oŋk] (abbiamo infatti TRŬNCU > *tr*[o]*nco* e SPELŬNCA > *spel*[o]*nca*) forse a causa della "maggiore forza articolatoria della velare sorda rispetto alla velare sonora, in grado di indebolire il potere di 'chiusura' esercitato dalla nasale sulla vocale precedente" (Castellani, 1980: 78-79). Castellani data l'anafonesi di secondo tipo tra la fine del IV secolo e l'inizio del V, sottolineando come in Toscana, a partire dal III secolo, siano esistite basi con /e/ ed /o/ per tutte quelle voci in cui più tardi si trovano /i/ ed /u/, tendenza che sarebbe rimasta vitale almeno per tutto il V secolo (1980: 86-87). Quest'affermazione, però, è stata recentemente rivista e contestata da Barbato (in stampa) il quale, avvalendosi anche di esempi provenienti dalle altre lingue romanze, anticiperebbe l'anafonesi dinanzi a velare rispetto a quella davanti a [ʎ] e [ɲ].

Anche se, come vedremo meglio in §3.2, in merito ad alcune questioni che solleva il fenomeno linguistico dell'anafonesi non tutti gli studiosi si trovano d'accordo, è assolutamente certo il suo carattere toscano, specialmente fiorentino: l'anafonesi differenzia quindi il toscano da molti dei dialetti italiani – che semmai conoscono fenomeni metafonetici – e il fatto "che l'italiano abbia *consiglio, lingua, fungo*, dimostra inequivocabilmente che esso coincide in buona parte col fiorentino letterario del Trecento" (Patota, 2002: 65). D'altro canto, il fiorentino corrente mantiene ancora oggi questa sua tendenza di fonetica storica, per cui il dialetto "mal sopporta /e o/ in posizione anafonetica e trasforma in /-ɛngo/ il suffisso /-engo/ presente in toponimi e in antroponimi" (Giannelli, [1976] 2000: 25)[6].

2. *La Carta dei Dialetti Italiani*

La nostra base di dati proviene dall'archivio sonoro della *Carta dei Dialetti Italiani* (CDI), una fra le imprese geolinguistiche più importanti condotte nel corso del Novecento.

2.1 La storia e il recupero della CDI

Ideata e fortemente voluta da Oronzo Parlangèli – membro del Comitato centrale insieme a Carlo Battisti e Giovan Battista Pellegrini – la *Carta dei Dialetti Italiani* nasce con l'obiettivo di fotografare la situazione dei dialetti d'Italia all'altezza degli anni Sessanta e Settanta. Per farlo, vengono presto costituiti in ogni regione dei gruppi di ricerca formati da giovani studenti di linguistica, ricercatori e dialettologi che dovranno intervistare almeno un abitante per ogni comune italiano[7] attenendosi alle

[6] Cfr Larson (2010: 1518).

[7] Queste erano le intenzioni iniziali di Parlangèli; strada facendo, il progetto ben presto si modifica restringendo il suo raggio d'azione.

domande di un questionario appositamente realizzato e registrando ogni inchiesta su nastro magnetico. Uno degli aspetti più importanti della *Carta* è proprio l'utilizzo costante del registratore, novità assoluta nel panorama italiano delle ricerche sul campo (cfr. Telmon, 1999: 29): "è poi grande merito del Parlangèli di avere introdotto l'obbligatorietà da parte dei raccoglitori della registrazione su nastro che costituisce una delle novità più rilevanti della raccolta poiché essa si traduce nella costituzione di un grande archivio delle nostre parlate popolari assai prossime alla dissoluzione se non altro nell'aspetto più autentico e microcosmico del patuà" (Pellegrini, 1977: 12, n. 4).

Nel primo numero del *Bollettino della Carta dei Dialetti Italiani* (BCDI) del 1966, contenente "istruzioni per l'attività del «Gruppo di ricerca per la dialettologia italiana» e relazioni sul lavoro svolto dai collaboratori" leggiamo:

> Il 15 maggio 1965, gli studiosi riunitisi a Bari, in occasione del II convegno per la preparazione della *Carta dei dialetti italiani: rilievi regionali* (CDI), deliberano di costituire l'impresa scientifica della CDI, in parte già finanziata dal Consiglio nazionale delle Ricerche (CNR), in «Gruppo di ricerche per la dialettologia italiana» che ha per suo primo compito la redazione della *Carta dei dialetti italiani*.

Il regolamento dell'organizzazione e della struttura interna della CDI, suddiviso in 51 punti, continua in maniera piuttosto dettagliata con informazioni riguardanti anche la parte amministrativa e prettamente burocratica dei lavori, come le modalità con cui si riuniscono periodicamente le assemblee dei vari Comitati regionali (CR), costituitisi in ogni regione e a "cui spetta il compito di organizzare e condurre a termine i rilievi dialettali nell'ambito del territorio di propria competenza" (BCDI, 1966: 5), o come le puntuali indicazioni sulla retribuzione che ogni ricercatore avrà a fronte del lavoro svolto, con precisazioni e dettagli anche in merito alle spese di viaggio e all'assicurazione (BCDI, 1966: 11-12, 15-22).

In origine il progetto risulta molto ambizioso, ma alla morte di Parlangèli, avvenuta improvvisamente nel 1969, l'impresa si arena e la conduzione del Comitato Centrale passa in mano al CNR di Padova nella persona di Manlio Cortelazzo. Il comitato regionale toscano è affidato a Gabriella Giacomelli dell'Università degli Studi di Firenze "la quale, negli anni, si adopera presso l'Accademia di Arti, Scienze e Lettere la Colombaria [...] e presso Manlio Cortelazzo per poter portare avanti i lavori in maniera capillare ed efficace" (Calamai, Bertinetto, 2012: 341). Il lavoro della studiosa, nonostante diverse difficoltà legate al passaggio gestionale, ha nel tempo dato i suoi frutti: il progetto *Gra.fo*, finanziato dalla Regione Toscana e condotto congiuntamente dalla Scuola Normale Superiore di Pisa e dall'Università di Siena, ha così potuto recuperare a Firenze delle inchieste ben conservate insieme ai verbali curatissimi rivisti da Luciano Giannelli (cfr. Calamai, Bertinetto 2012: 345 e http://grafo.sns.it/grafo). La distribuzione geografica delle inchieste svolte in Toscana non è omogenea: alcune province sono capillarmente indagate, come Siena e Firenze, mentre altre, come Arezzo, Lucca e soprattutto Livorno sono più scoperte; questo perché Gabriella Giacomelli poteva contare su un gruppo di ricerca

distribuito soprattutto tra le province di Siena e Firenze (Calamai, Bertinetto, 2012: 340-347)[8].

Dall'ascolto delle interviste e dallo spoglio dei verbali possiamo ricavare il profilo socio-culturale degli informatori. La scelta della fonte è demandata al ricercatore, che di solito riesce ad intervistarne almeno un paio per inchiesta, per la maggior parte contadini o artigiani anziani, maschi, dal basso profilo scolastico (non si va oltre la quinta elementare), mai (o quasi mai) allontanati dal luogo di residenza. Talvolta, essendo spesso l'inchiesta registrata nell'abitazione privata dell'informatore, intervengono nelle risposte al questionario anche altri componenti della famiglia, e così l'intervista si estende anche alle donne. Più raramente, per ottenere anche utili confronti delle risposte, il ricercatore decide di intervistare soggetti più giovani, dal profilo socio-culturale leggermente più elevato: si tratta di persone che hanno ottenuto il diploma e al momento dell'inchiesta lavoravano come impiegati, partecipando attivamente a momenti di aggregazione sociale che il comune offre loro (teatri, cinema, circoli ricreativi).

2.2 I settori del questionario spogliati: precisazioni e criticità

La parte del questionario ispezionata è la seguente:

29. io vinco, tu vinci

30. io tingo, tu tingi

31. lingua, -e

[8] I punti d'inchiesta sono così distribuiti: Arezzo, con 10 comuni (Ambra di Bucine, Anghiari, Bibbiena, Castiglion Fiorentino, Cavriglia, Foiano della Chiana, Frassineto, Laterina, Loro Ciuffenna-Malva e San Giovanni Valdarno); Siena, con 31 (Chiusdino, Asciano, Abbadia San Salvatore, Buonconvento, Casole d'Elsa, Castiglion d'Orcia, Cetona, Chianciano Terme, Castellina in Chianti, Castelnuovo Berardenga, Costalpino San Galganello, Colle Val d'Elsa, Gaiole in Chianti, Monteroni d'Arbia-Corsano, Montepulciano, Murlo, Montalcino, Monteriggioni, Piancastagnaio, Poggibonsi, Pienza, Radicondoli Belforte, Radicofani, Rosìa, Rapolano Terme, San Casciano dei Bagni, San Casciano d'Asso, San Gimignano, San Quirico d'Orcia, Torrita e Sarteano); Lucca, con i comuni di Capannori, Massarosa e Montecarlo; Pisa, con Bientina, Casciana Terme, San Miniato e Volterra; Livorno, con Campo nell'Elba e Castagneto Carducci; Grosseto, con 25 comuni (Arcidosso, Castellazzara, Capalbio, Castiglion della Pescaia, Campagnatico, Civitella Paganico, Castel del Piano, Follonica, Gavorrano, Monte Argentario, Manciano, Magliano in Toscana, Montieri, Monte Rotondo Marittimo, Massa Marittima, Pitigliano, Roccalbegna, Roccastrada, Scansano, Scarlino, Seggiano, Semproniano, Santa Fiora, Sorano e Cinigiano); Firenze, che al tempo inglobava anche la provincia di Prato, con 45 comuni (Barberino del Mugello, Barberino Val d'Elsa, Cerreto Guidi Bassa, Cantagallo, Campi Bisenzio, Contea, Calenzano, Capraia Limite, Carmignano, Certaldo, Castelfiorentino, Empoli, Fucecchio, Firenzuola Castelvecchio, Fiesole, Figline Valdarno, Gambassi, San Godenzo, Grassina, Impruneta, La Briglia, Lastra a Signa, Montelupo Fiorentino, Montemurlo, Montaione, Marradi, Montespertoli, Poggio a Caiano, San Piero a Sieve, Pelago, Pontassieve, Panzano, Prato, Paterno, Rufina, Scarperia, Scandicci, Signa, Sesto Fiorentino, Tosi, Tavernelle Val di Pesa, Vicchio, San Casciano Val di Pesa, Vinci e Vernio); Pistoia con i comuni di Buriano, Calamecca di Piteglio, Lamporecchio, Le Piastre, Mazza e Cozzile, Momigno, Montale, Marliana, Monsummano Terme, Nievole, Pieve a Nievole, Posola, San Michele Agliana, San Rocco di Larciano e Serravalle Pistoiese.

55. stringere (inf., ind. pres., part.)

298. unghia

369. lungo, lunghi; lunga, lunghe;

371. mungere (inf., ind. pres.)

373. ungere (inf., ind. pres.)

[BCDI 1966]

Ai singoli informatori viene chiesto di tradurre i termini corrispondenti alle 8 domande dall'italiano standard al proprio dialetto.

Abbiamo visto che le domande del questionario si basano su un disegno schematico perlopiù fisso e sono sempre mediate dal ricercatore, che a volte non si risparmia frasi che condizionano il comportamento linguistico della fonte. Spesso infatti, ci troviamo di fronte a dei casi in cui non possiamo dirci certi del livello di competenza del parlante. A questo proposito, è emblematica l'inchiesta svolta a Castiglion Fiorentino (AR):

R[9]: [...] vinco

I: quande gioco, io vinco sempre

R: vinco o no? O un'altra parola: venco?

I: venco

R: ecco! dovete dirmi le più antiche eh

I: ah sì?

R: eh sì, sennò

I[2]: quande gioco io venco sempre

Come notiamo, è il ricercatore a pronunciare per primo il termine *venco*, (con mancata anafonesi) che solo dopo viene detto – ripetuto – anche dall'intervistato. In questo modo, come rileva anche l'antropologo Bruno Pianta (1980), "si trova solo quello che si cerca", e si crea un contesto artificioso dove è la fonte ad adattarsi al mondo del ricercatore, quando dovrebbe accadere il contrario. L'utilizzo di un questionario piuttosto rigido, da sottoporre agli informatori in tempi relativamente stretti, non favorisce la spontaneità della produzione linguistica: a volte si ha l'impressione che le risposte siano suggerite, talvolta 'estorte', senza rispettare i tempi (e soprattutto i silenzi) dell'informatore, in special modo quando questi si trova a dover rispondere a domande di tipo morfologico o sintattico.

Un altro esempio viene dall'inchiesta registrata a Piancastagnaio, piccolo comune in provincia di Siena, dove la metodologia messa a punto dal ricercatore nel porre la domanda alla sua fonte crea talmente tanta confusione e incomprensione che fa quasi sorridere l'ascoltatore:

[9] Indichiamo con R il ricercatore e con I l'intervistato.

R: dunque se lui vince, io?

I: perdo

R: no ehm, lei dice io vinco e tu?

I: e tu perdi, e te perdi

R: no, ma se anche lui vince?

I: io ho vinto e te hai perso

[...]

R: sì ma in questo momento, ora? hai vinto prima, e ora?

I: hai perso

I dati così ricavati sono difficilmente valutabili dal punto di vista della loro affidabilità (Sanga, 1991: 173): in altre parole, spesso non è possibile stabilire il livello di competenza (se attiva o passiva) posseduto dal parlante quando fornisce una certa risposta. Non è quindi difficile mettere in luce l'estrema astrattezza della metodologia utilizzata per ottenere le risposte, in un contesto comunicativo artificioso, dove si pretende dal testimone l'accesso ad una riflessione metalinguistica che, non solo non gli compete, ma che "senza un'abitudine precostituita è difficile avere" (Carpitelli, Iannaccaro, 1995).

3. *Analisi dei dati*

Come abbiamo già accennato nel §1, il fenomeno dell'anafonesi è tipico di un'area molto ristretta della nostra penisola: Castellani (1980), Serianni (1995 e 1998), Patota (2002) e Calamai (2011) sono concordi nel concentrare l'anafonesi a Firenze, nelle zone di Volterra e San Gimignano e nella Toscana occidentale (Prato, Pisa, Pistoia e Lucca), lasciando fuori i dialetti centrali e orientali dell'aretino, del senese e del cortonese. Questa situazione linguistica si rispecchia solo in parte con quella emersa dallo spoglio dei dati dell'archivio della CDI, come mostrano le due carte della regione (vd. Figg. 1 e 2): l'una costruita sulla base delle considerazioni[10] di Arrigo Castellani sui territori toscani non anafonetici, l'altra creata con i risultati ottenuti dall'ascolto delle inchieste, per un totale di 1136 risposte spogliate.

[10] Le analisi linguistiche di Arrigo Castellani (1952 e 1980) si basano principalmente su testi antichi ma, da attento linguista qual era, non tralascia dati più recenti, come quelli dell'AIS e quelli riportati nei vocabolari dialettali degli anni '50.

Figura 1 - *Territori non anafonetici secondo la ricostruzione di Arrigo Castellani (1952 e 1980)*

Si è deciso di colorare di azzurro le zone non anafonetiche della regione, differenziando la provincia senese da quella aretina e grossetana con un tono più chiaro, poiché Castellani afferma (1952) che la provincia di Siena partecipa solo sporadicamente a fenomeni anafonetici (p. 25). Nella figura 2 i punti d'inchiesta della CDI segnati con un piccolo cerchio azzurro corrispondono ai comuni in cui rimangono delle forme non anafonetiche: procedendo da nord a sud, per la provincia di Arezzo registriamo Castiglion Fiorentino, Frassineto, Foiano della Chiana e Anghiari; per Siena abbiamo i comuni di Castellina in Chianti, Asciano, Murlo, Montalcino, San Quirico d'Orcia, Montepulciano, Chianciano Terme, Abbadia San Salvatore e infine Piancastagnaio alle pendici del Monte Amiata, insieme all'unico comune della provincia di Grosseto, Pitiglia.

Man mano che procediamo verso sud il quadro geolinguistico esplicitato dalle due carte tende a modificarsi; se infatti, per la Toscana nord-occidentale la situazione anafonetica delineata da Castellani coincide con quella disegnata dalle inchieste della *Carta dei Dialetti Italiani*, per le province sud-orientali assistiamo a modifiche piuttosto consistenti. Vediamole più in dettaglio.

Figura 2 - *Territori non anafonetici secondo le inchieste della Carta dei Dialetti Italiani*

Per quanto riguarda la Toscana occidentale, nelle province di Lucca, Pistoia, Pisa e Livorno ci troviamo di fronte a un quadro atteso: il territorio risulta infatti interamente anafonetico sia nella ricostruzione di Castellani che in quella emersa dall'analisi delle risposte della *Carta*[11]. Nell'area grossetana solo un comune, Pitigliano, si distingue dal resto del territorio anafonetico: il parlante dichiara infatti che lingua nel suo dialetto si dice ancora *l[e]ngua*. Questo dato è spiegabile alla luce delle varianti non anafonetiche registrate anche in altre zone del Monte Amiata vicino a Pitigliano: ad esempio a Piancastagnaio, in provincia di Siena, abbiamo *t[e]ngo* e *str[e]ngo* come competenze attive del nostro parlante; d'altronde Giannelli ([1976] 2000) considera non anafonetico il dialetto amiatino che, parlato sia a Pitigliano che a Piancastagnaio, egli classifica come una parlata "ben distinta da tutti gli altri dialetti toscani per caratteristiche di chiara provenienza centro-meridionale e con tracce di un sistema fonologico ben diverso da quello fiorentino e senese a livello di dialetto rustico" (p. 105).

Per quanto riguarda l'anafonesi nella Toscana orientale e centrale le due cartine ci mostrano una diversa situazione geolinguistica. Dalle inchieste della CDI, infatti, sembrerebbe che anche il territorio aretino e senese, almeno a quest'altezza cronologica, sia già quasi del tutto anafonetico, a differenza della situazione delineata in Castellani (1952 e 1980). I termini senza chiusura anafonetica – *l[e]ngua, v[e]nco,*

[11] In merito all'anafonesi in area tosco-occidentale si veda l'interpretazione, più controversa, di Franceschini (1991); nonché le forme riportate da Rohlfs (1966: 73) successivamente smentite da Castellani (1980: 82) grazie ai dati provenienti dall'AIS e a quelli contenuti nel *Vocabolario Cateriniano* (Mattarucco, 2008).

l[o]ngo, str[e]ngo e *t[e]ngo* – che spesso gli informatori definiscono "all'antica", "usi contadineschi antiquati" o "parole che prima si dicevano" – rimangono in forme residuali nel contado attorno alle province di Arezzo e Siena. Nello specifico: nella Val di Chiana aretina – tra Castiglion Fiorentino, Frassineto e Foiano della Chiana – si registrano *v[e]nco, t[e]ngo, l[e]ngua, str[e]ngo, [o]nghia, l[o]ngo, m[o]ngo* e *[o]ngo*; nella campagna senese, tra Montalcino e il confine con il Lazio e l'Umbria, e alle pendici del Monte Amiata con il comune di Piancastagnaio troviamo *v[e]nco, t[e]ngo, l[e]ngua, str[e]ngo* e *l[o]ngo*. Ad Anghiari (Arezzo), uno dei comuni dove l'anafonesi sembra ancora non essersi diffusa del tutto, il parlante intervistato arriva quasi a ostentare il suo dialetto, in un atteggiamento volto a "fare colpo" sul ricercatore, come a voler dire che in quel borgo toscano è custodita la vera e antica parlata locale. L'informatore di questo comune, infatti, fornisce sempre risposte non anafonetiche, *l[e]ngua/l[e]ngue*, *str[e]ngo*, *[o]nghia*, *l[o]ngo/l[o]nghi* e *[o]ngono*: è tuttavia arduo stabilire se siano parte della competenza attiva ovvero frutto di una sorta di 'esibizione dialettale'. Nella zona dell'aretino, a riprova delle considerazioni di Nocentini (1989), il modello di Firenze si è già diffuso in quasi tutto il territorio: "ad Arezzo il fenomeno della mancata anafonesi è in regresso e dalla varietà alta si sono introdotte molte parole col fonetismo fiorentino" (p. 38).

Dal punto di vista geolinguistico sembra dunque che le forme non anafonetiche si siano ritirate nell'estremo sud della regione e quasi sempre al confine con i territori non toscani dell'Umbria e del Lazio, quindi con parlate di per sé non anafonetiche: da questo punto di vista tutte e tre le province – Arezzo, Siena e Grosseto – risultano abbastanza coerenti.

4. *Usare e riusare i dati di ricerca*

Questo esercizio di analisi ha mostrato i pregi e i limiti di una ricerca che faccia uso di precedenti dati scientifici, raccolti per scopi in parte diversi, e resi accessibili agli studiosi attraverso un progetto di recupero e valorizzazione degli archivi sonori del passato. Naturalmente il ricco giacimento della *Carta dei Dialetti Italiani* si presta a un ventaglio molto ampio di analisi fonetiche, morfologiche, lessicali, in parte anche sintattiche, oltre che a una riflessione sulle metodologie dell'inchiesta sul campo. Un siffatto esercizio di analisi consente anche di avviare una riflessione più ampia, incentrata sui temi dell'archiviazione, dell'accessibilità e del riuso dei dati di ricerca – temi che peraltro sono al centro di *Horizon* 2020 e di diverse comunità di ricerca in varie parti del mondo. Il numero monografico di *Science* 331 (2011) pone al centro proprio il cosiddetto *data deluge*. Ora che la massa dei dati a disposizione sta aumentando in maniera piuttosto massiccia, per molte discipline diventa cruciale stabilire quali dati archiviare e quali scartare, come accedere ai dati archiviati, come riusarli in maniera efficiente, sostenibile, eticamente corretta. Questa esigenza è stata fortemente sentita tra i fisici, che hanno dato vita, nel 2009 a un gruppo di lavoro chiamato 'Data Preservation in High Energy Physics" (DPHEP). Appare complesso e per certi versi contraddittorio il dibattito nell'ambito delle scienze so-

ciali (King, 2011). Il panorama anglosassone e francese risulta essere estremamente stimolante a questo proposito, nell'ambito della ricerca sia quantitativa che qualitativa. Il tema della conservazione degli archivi costruiti dai ricercatori è del resto un tema trasversale che riguarda ricercatori, archivisti, documentalisti, istituzioni ed enti di ricerca. Ci sono reti (ad esempio il *réseau Quetelet*: www.reseau-quetelet.cnrs.fr) che si occupano specificamente di conservare le inchieste condotte con questionari; più recentemente sono apparsi gruppi (tra cui *beQuali*, www.bequali.fr) specializzati proprio nella raccolta, nella digitalizzazione e nella diffusione dei materiali legati alle inchieste qualitative. Nonostante l'estrema diversità dei metodi utilizzati, siffatte iniziative possono essere di grande aiuto sia nel mostrare la pluralità dei metodi e delle procedure nella ricerca qualitativa, sia – crucialmente – nella messa a disposizione di modalità di ricerca plurime utili per i novizi che si avvicinano al mestiere di ricercatore. I materiali di corredo, la documentazione dell'inchiesta, i supporti video e/o sonori possono rappresentare dei validi supporti pedagogici e dunque fornire un efficace accompagnamento alla ricerca (per esempio, nella comprensione delle condizioni reali concrete in cui si realizza un'inchiesta). Non sono rare le ricerche incentrate proprio sul riuso (e, conseguentemente, sulla messa a disposizione) di inchieste precedenti, che pongono al centro proprio i temi dell'analisi secondaria, della rivisitazione, della ricontestualizzazione storica di materiali precedentemente raccolti. Gli archivi sonori contenuti dentro *Gra.fo* sono a disposizione anche per questo.

Bibliografia

Alberti Di Villanova, F. (1797). *Dizionario universale critico enciclopedico della lingua italiana dell'abate D'Alberti di Villanuova*, tomo primo. Lucca: Marescandoli.

Andreini, A., Clemente, P. (Eds.) (2007). *I custodi delle voci. Archivi orali in Toscana: primo censimento*. Firenze: Regione Toscana.

Barbato, M. (in stampa). Anafonesi Latina e anafonesi romanza. In *Actes du XXVII^e Congrès International de Linguistique et Philologie romanes*, Nancy, 15-20 luglio 2013.

Bcdi (1966). *Bollettino della Carta dei Dialetti Italiani*, 1.

Calamai, S. (2011). Anafonesi. In Simone, R. (diretta da), *Enciclopedia dell'Italiano.* Roma: Istituto dell'Enciclopedia Italiana, 58-59.

Calamai, S., Bertinetto, P.M. (2012). Per il recupero della *Carta dei Dialetti Italiani.* In Telmon, T., Raimondi, G. & Revelli, L. (Eds.), *Coesistenze linguistiche nell'Italia pre- e postunitaria, Atti del XLV Congresso internazionale di studi della Società di Linguistica Italiana (Aosta/Bard/Torino 26-28 settembre 2011)*, 2 voll. Roma: Bulzoni, 335-356.

Calamai, S. (in stampa). The phonetics of anaphonesis between production and perception. In Gibson, M., Jil, J. (Eds.), *Romance Phonetics and Phonology*. Oxford: Oxford University Press.

Carpitelli, E., Iannaccaro, G., (1995). Dall'impressione al metodo: per una ridefinizione del momento escussivo. In *Dialetti e Lingue nazionali. Atti del XVII Congresso Internazionale della Società di Linguistica Italiana*. Roma: Bulzoni, 99-120.

Castellani, A. (Ed.) (1952). *Nuovi testi fiorentini del Dugento*, 2 voll. Firenze: Sansoni.

Castellani, A. (1980). Sulla formazione del tipo fonetico italiano. In *Saggi di linguistica e filologia italiana e romanza*, vol. I. Roma: Salerno Editrice, 73-122.

Cortelazzo, M. (1969). *Avviamento critico allo studio della dialettologia italiana, I, Problemi e metodi*. Pisa: Pacini.

Cortelazzo, M.A. (2000). *Italiano d'oggi*. Padova: Esedra.

De Mauro, T. [1963] (2011). *Storia linguistica dell'Italia unita*. Roma-Bari: Laterza.

Franceschini, F. (1991). Note sull'anafonesi in Toscana occidentale. In *Tra Rinascimento e strutture attuali. Saggi di Linguistica Italiana*, vol. II. 259-272.

Giannelli, L. [1976] (2000). *Toscana*. Pisa: Pacini.

King, G. (2011). *Ensuring the Data-Rich Future of the Social Sciences*. In *Sciences*, 331, 719-721.

Larson, P. (2010). Fonologia. In Salvi, G., Renzi, L. (Eds.), *Grammatica dell'italiano antico*, vol. II. Bologna: il Mulino, 1515-1678.

Loporcaro, M. [2009] (2012). *Profilo linguistico dei dialetti italiani*. Roma-Bari: Laterza.

Mattarucco, G. (2008) (Ed.). *Girolamo Gigli*, *Vocabolario cateriniano*. Firenze: Accademia della Crusca.

Nocentini, A. (1989). *Il vocabolario aretino di Francesco Redi. Con un Profilo del dialetto aretino*. Firenze: E.L.I.T.E.

Parlangèli, O. (1967). L'inchiesta dialettale. In Falcone, G., Mancarella, G.B. (1972) (Eds.), *Scritti di dialettologia*. Galatina: Congedo, 11-18.

Parlangèli, O. (1970). Il programma del gruppo di ricerca per la dialettologia italiana. In *I dialetti dell'Italia mediana con particolare riguardo alla regione umbra, Atti del V Convegno di Studi Umbri (Gubbio, 28 maggio-1 giugno 1967)*. Perugia, Facoltà di Lettere e Filosofia, 449-461.

Patota, G. (2002). *Nuovi lineamenti di grammatica storica dell'italiano*. Bologna: il Mulino.

Pfister, M. (1986) (Ed.). *LEI. Lessico Etimologico Italiano*, fascicolo 14°, vol. II. Wiesbaden: Dr. Ludwig Reichert Verlag, 1053-54.

Pianta, B. (1980). La ricerca "a prova d'idiota". In *"La ricerca folklorica"*, 2, 109-111.

Pellegrini, G.B. (1977). Carta dei Dialetti d'Italia. In Cortelazzo, M. (Ed.), *Profilo dei dialetti italiani, Consiglio Nazionale delle Ricerche-Centro di Studio per la Dialettologia Italiana*. Pisa: Pacini.

Rohlfs, G. (1966). *Grammatica storica della lingua italiana e dei suoi dialetti*, vol. I, *Fonetica*. Torino: Einaudi.

Sanga, G. (1991). I metodi della ricerca sul campo. In *Rivista Italiana di Dialettologia*, 15, 165-181.

Serianni, L. (1995). *Appunti di grammatica storica italiana*. Roma: Bulzoni.

Serianni, L. (1998). *Lezioni di grammatica storica italiana*, nuova ed. Roma: Bulzoni.

Telmon, T. (1999). Intervento. In *Aa.Vv., Archivi sonori. Atti dei seminari di Vercelli (22 gennaio 1993), Bologna (22-23 settembre 1994), Milano (7 marzo 1995)*, Roma: Ministero per i Beni e le Attività culturali (Pubblicazioni degli Archivi di Stato, 53), 28-34.

Tuttle, E. (1991). Nasalization in Northern Italy: sillabic constraints and strenght scales as developmental parameters. In *Rivista di linguistica*, 3, 23-92.

Sitografia

http://grafo.sns.it/grafo (Accessed 28.10.15).

www.reseau-quetelet.cnrs.fr (Accessed 28.10.15).

www.bequali.fr (Accessed 28.10.15).

DOI: 10.17469/O2101AISV000011

LUCIANO ROMITO, MARIA ASSUNTA CIARDULLO, ANDREA TARASI

Analisi acustica delle occlusive sorde aspirate del dialetto di San Giovanni in Fiore (CS)

The goal of this study is the acoustical analysis of aspirated voiceless stops /t k/ preceded by /n/ produced by the speakers of San Giovanni in Fiore, a small town of Calabria. Even though dialectological studies include this village in the second dialectal area of the Calabrian region, aspirated voiceless stops realized by speakers of San Giovanni in Fiore contrast with this idea. In fact, its dialect comes to be more similar with the phonetic features (i.e. aspiration) of the areas of Catanzaro and Reggio Calabria. After the measurement of the durations of the stops elicited, the analytic section investigates the phenomenon of aspiration that forces us to reconsider traditional isoglosses in order to put San Giovanni in Fiore in its proper dialectological place.

1. *Introduzione*

La definizione dei confini linguistici di un territorio ha da sempre costituito il motore di avvio di qualsiasi ricerca che puntasse alla descrizione delle differenze tra le varietà esistenti in una data area. Dunque, in dialettologia si rivelano cruciali sia il concetto di isoglossa (cfr. Dubois, 1979; Cardona, 1988) che si fonda proprio sull'individuazione di confini che attestano la presenza o l'assenza di fenomeni linguistici usati come caratterizzanti sia quello di corridoio (cfr. Trumper, 1997) che individua zone disomogenee in cui si condividono fenomeni linguistici (fonetico-fonologici, morfologici, sintattici e lessicali) appartenenti ad aree contigue.

L'eterogeneità dialettologica della Calabria ha da sempre attirato l'attenzione di numerosi studiosi e di conseguenza vari sono stati i tentativi di ripartire la regione calabrese in specifiche aree dialettali attraverso l'utilizzo di isoglosse o corridoi. La prima ripartizione dialettologica del territorio calabrese risale a Rohlfs (1962), il quale suddivise l'area in due sezioni facendo riferimento a differenze di sostrato (latino: Calabria settentrionale vs. greco: Calabria centrale e meridionale); Falcone (1976), invece, partendo dal sistema vocalico tonico di derivazione latina, descrisse per la Calabria tre zone dialettologiche: una zona meridionale, una centrale e una settentrionale. La zona centrale, a sua volta, venne ulteriormente suddivisa in due gruppi per via di differenze proprie del sistema vocalico: la sottosezione A avente vocalismo 'siciliano' e la sottosezione B avente una situazione di compromesso tra le condizioni 'siciliane' e 'napoletane'. Infine, un'articolazione più dettagliata è stata proposta da Trumper (1997) che, mediante diversi parametri linguistici, individua tre isoglosse con cui ripartisce la regione in quattro aree dialettali.

Figura 1 - *aree dialettali secondo Trumper 1997: 360.*

La seconda sezione dialettale individuata da Trumper (Op. cit.) include il paese di San Giovanni in Fiore[1], posto nella sfera amministrativa della provincia di Cosenza, il cui dialetto è oggetto di studio della presente ricerca. Questo dialetto esibisce delle peculiarità in contrasto con le caratteristiche dei dialetti inclusi nella seconda area. Nel dialetto di San Giovanni in Fiore le consonanti occlusive sorde /p t k c/ diventano aspirate quando precedute dalle consonanti nasali /n m/ o dalla vibrante /r/ oppure quando si manifestano come geminate. La presenza di queste consonanti a San Giovanni in Fiore rappresenta un fatto 'anomalo' e molto interessante, perché tali foni sono tipici della provincia catanzarese e reggina, cioè della terza e quarta area dialettale (cfr. Trumper, 1997), vale a dire di quelle aree con sostrato greco (cfr. Rohlfs, 1962).

2. *Ipotesi della ricerca*

La ricerca si concentra sullo studio delle occlusive sorde /t k/ precedute da nasale /n/, quindi nt-, -nt-, nk- e -nk-.

L'ipotesi di partenza è la dimostrazione della presenza di consonanti aspirate a San Giovanni in Fiore, nonostante tali foni costituiscano un tratto tipico delle aree linguistiche con sostrato greco (cfr. Rohlfs, 1962; Falcone, 1976; Marotta, Sorianello, 1992).

La scelta di queste due consonanti sorde aspirate è motivata dal fatto che esse, articolatoriamente, sono situate agli antipodi dell'asse linguale. La conformazione articolatoria che caratterizza questi due foni comporta differenze intrinseche negli esiti del VOT.

3. *Stato dell'arte del VOT*

L'inventario fonetico-fonologico dei suoni della lingua italiana standard è oltremodo nutrito ma quello concernente i dialetti parlati nel perimetro nazionale è ancor più vasto. Diversi studiosi si sono già occupati di fenomeni linguistici produttivi e attivi nel

[1] San Giovanni in Fiore, con una concentrazione demografica che sfiora i 20 mila abitanti, è il più antico e ampio centro presente sull'altopiano silano.

dialetto sangiovannese (cfr. Mele, Schmid, 2009; Mele, 2009) ma, ad eccezione di brevi cenni di contorno, l'aspirazione risulta essere scarsamente indagata. L'aspirazione è un periodo di sordità che si riverbera sul ritardo dell'attacco della sonorità del fono successivo, quasi sempre vocalico: quanto detto coincide con la definizione acustica di VOT (*Voice Onset Time*) (Lisker, Abramson, 1964). Questi studiosi definiscono il VOT, mettendo in relazione il meccanismo sovraglottidale con l'attività laringale, asserendo che

> VOT, the interval between the release of the stop and the onset of phonation as shown in spectrograms, was the simplest single measure we could find in the acoustic signal of the timing of laryngeal adjustements[2].

In uno studio successivo, i due linguisti sottolineano che il VOT è effettivamente:

> the single most effective measure for classifying stops into different phonetic categories with respect to voicing[3].

Un'ulteriore definizione, sempre di natura acustica, proposta da Lisker ed Abramson (1965), specifica che:

> this aspiration phase is accompanied by considerable attenuation of the first formant, an effect presumably ascribed to the presence of the tracheal tube below the open glottis.

La definizione acustica del VOT si è rivelata utile (cfr. Nance, Stuart-Smith, 2013) per discernere altri due fenomeni: pre-aspirazione[4] e post-aspirazione[5]. Poche lingue usano questi due parametri in opposizione fonologica e la stessa pre-aspirazione è attestata in un numero molto esiguo di lingue, eccezion fatta per l'Islandese (cfr. Thrainsson, 1978), la lingua parlata nelle isole Far Oer (cfr. Helgason, 2003), il dialetto di Siena (cfr. Stevens, Hajek, 2007) e poche altre varietà.

L'ultima proposta acustica esplicativa concernente il VOT è stata avanzata da Kessinger e Blumstein (1997), le quali descrivono il VOT come:

> an acoustic property that is associated with the feature voicing [and] is typically described as having three ranges of values: pre-voiced, short lag, and long lag VOT.

Il VOT non è stato specificato solo su base acustica, ma anche su criteri di tipo articolatorio, neurolinguistico e concernente la linguistica applicata. La definizione articola-

[2] Lisker, L., Abramson, A.S. (1972). Voice-Timing Perception in Spanish Word-Initial Stops. In *83rd Meeting of the Acoustical Society of America*. Buffalo, N.Y., 15.

[3] Lisker, L., Abramson, A.S. (1971). Distinctive features and laryngeal control. In *Language*, 47, 776-785.

[4] «[...] is referred to as the time period between the offset of normal voicing and the stop closure (Laver, 1994: 150). This definition conceptualizes pre-aspirated stops in terms of the timing coordination of glottal abduction and supra-glottal closure for the stop».

[5] «can similarly be thought of in terms of the timing coordination of glottal and supra-glottal closures ...during the closure for a voiceless stop, the vocal folds are abducted. If the closure is released before the vocal folds move together for a following voiced segment, this results in an audible puff of turbulent, noisy air, or aspiration» (cfr. Nance, Stuart-Smith, 2013).

toria[6] del VOT trova nella ricerca condotta da Cho & Ladefoged (1999) un paradigma essenziale per le ricerche linguistiche e, soprattutto, per la sezione analitica del presente studio. La loro ricerca si concentra sull'analisi articolatoria del VOT, definendolo:

> [...] the interval between cursors placed at the onset of release (the final release, if there was more than one) and the onset of the first complete vibration of the vocal folds as indicated on the waveform.

I due linguisti ritengono che il luogo di articolazione di una consonante occlusiva sorda e il relativo VOT a essa conseguente siano correlati in quanto l'aspirazione che avviene in seguito ad un'occlusione velare ha un valore maggiore rispetto al VOT che segue le occlusive la cui stretta è prossima alle labbra, vale a dire quelle dentali e bilabiali. Di recente, è stato parimenti condotto uno studio che analizza i correlati articolatori dei contrasti laringali (Fuchs, 2005).

Infine, diversi studi di linguistica applicata hanno dimostrato che il VOT è considerato come elemento distintivo nell'apprendimento delle lingue straniere di ogni ordine (L2, LS, L3, etc.) (cfr. Zlatin, 1974; Lieberman, Blumstein, 1988; Volaitis, Miller, 1992; Pind, 1999; Baum, Ryan, 1993; Pind, 1995; Whiteside, Marshall, 2001; Whiteside, Dobbin & Henry, 2003).

Anche la neurolinguistica dedica al VOT una copiosa letteratura che pone il VOT come parametro essenziale per attestare o meno la presenza di sospetta disprassia verbale e di altri disturbi logopedici (cfr. Lundeborg, Nordin, Zeipel-Stjerna & McAllister, 2014). Va in ultimo ricordato che, oltre al VOT, vi sono potenziali correlati acustici dell'aspirazione e in generale del tratto [*spread glottis*] come la differenza nelle armoniche (H1-H2) o la perturbazione di F0 che costituiranno potenziali fulcri per ricerche future.

4. *Materiali e metodi*

Il corpus di riferimento di questa ricerca contiene locutori del dialetto sangiovannese raggruppati in tre fasce di età coincidenti con quelle utilizzate da Labov (1972) attualizzate per ovvi motivi storici: la prima fascia di età raggruppa informatori che hanno un'età compresa tra i venti e i ventinove anni, la seconda riunisce parlanti di età compresa tra i trenta e i cinquantanove anni e il terzo gruppo include locutori che presentano un'età maggiore di sessant'anni. Per ogni fascia di età è stato registrato un numero equidistribuito di voci maschili e femminili: per la prima fascia sono state registrate sette voci mentre per la seconda e per la terza fascia sono state acquisite cinque voci; ai fini del presente studio, le voci complessive, oggetto d'indagine sono state diciassette.

Allo scopo di verificare l'ipotesi di partenza del presente lavoro (cfr. § 2), il campione sangiovannese è stato confrontato con un gruppo di controllo, stratificato solo

[6] Si rimanda allo studio di Keating, A. (1984), Phonetic and phonological representation of stop consonant voicing. In *Language* vol. 60 (2), 286-318, il quale classifica il ritardo breve e lungo nella seguente modalità: attribuisce il termine ritardo breve ai valori positivi di VOT quando sono compresi tra 20 e 35 ms, mentre con ritardo lungo definisce quei ritardi di sonorità che ricoprono intervalli temporali più alti.

per genere, composto da parlanti del dialetto cosentino. Il confronto che si opera col gruppo di controllo, che seguirà lo stesso iter metodologico del campione sangiovannese, è interessante perché le occlusive del cosentino, negli stessi contesti, presentano assimilazione progressiva con sonorizzazione delle consonanti stesse. Risulterà quindi evidente la peculiarità del dialetto sangiovannese che presenta un processo fonologico opposto (cfr. Falcone, 1976; Marotta, Sorianello, 1992).

Il corpus è anche stratificato diafasicamente: parlato letto e parlato semi-spontaneo. Il parlato letto, su cui si effettuano le analisi di questa ricerca, è realizzato attraverso la lettura di 14 frasi e di 20 parole che includono le due consonanti nel contesto sopra specificato. Pertanto, le frasi contengono al loro interno delle "parole bersaglio", cioè parole che racchiudono le due consonanti nel contesto menzionato. La lista di parole, invece, è organizzata in parossitoni e proparossitoni. Questa organizzazione consente di constatare se la posizione dell'accento influenza o meno la durata dell'aspirazione.

La lista di parole e di frasi sono state presentate ai locutori in forma scritta e agli stessi è stato domandato di tradurre il materiale dall'italiano al dialetto. Dalla lista di parole sono state analizzate 303 produzioni, mentre dalle parole in isolamento 300. Metodologicamente, in questa ricerca per VOT si intende l'intervallo compreso tra la comparsa del *burst* dell'occlusiva e il primo semiciclo periodico della vocale seguente sullo spettrogramma (cfr. Martinez, 1998; 2002; 2007; Hrycyna et al., 2011). È stata adottata questa procedura, perché riteniamo sia meno problematica e più precisa.

Il parlato semi-spontaneo[7] è stato elicitato attraverso lo strumento tipico della ricerca qualitativa, vale a dire attraverso un'intervista libera (o semi-strutturata) in cui è stato chiesto al parlante di raccontare un fatto autobiografico oppure legato alle tradizioni popolari al fine di rendere l'eloquio più spontaneo possibile. L'intervista è proposta come fase antecedente rispetto a quella della lettura delle liste di frasi e di parole, in maniera tale da agevolare la produzione di un parlato più naturale possibile.

Tutte le registrazioni sono state etichettate e segmentate attraverso il software PRAAT versione 5.3.08. In merito all'etichettatura dei file sonori è stato seguito un protocollo interno al Laboratorio di Fonetica. Lo stesso programma è stato utilizzato per definire la durata del VOT delle consonanti occlusive sorde. La durata del VOT è stata determinata attraverso uno script costruito *ad hoc*.

5. *Analisi dei dati*

Prima di addentrarci nell'analisi particolareggiata dei dati ottenuti, ci preme sottolineare che non sono state registrate significative differenze tra la produzione di parlato prodotto dalle diverse classi d'età o dai generi del campione sangiovannese:

[7] In questo caso, il parlato semi-spontaneo viene utilizzato solo come metro di conferma dell'effettiva presenza delle consonanti aspirate a San Giovanni in Fiore. Sarà nelle nostre intenzioni future condurre uno studio volto al confronto tra la durata delle occlusive sorde aspirate del parlato letto e semi-spontaneo.

Figura 2 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [k] letto nella lista di frasi e calcolato per ciascuna fascia di età che compone il corpus*

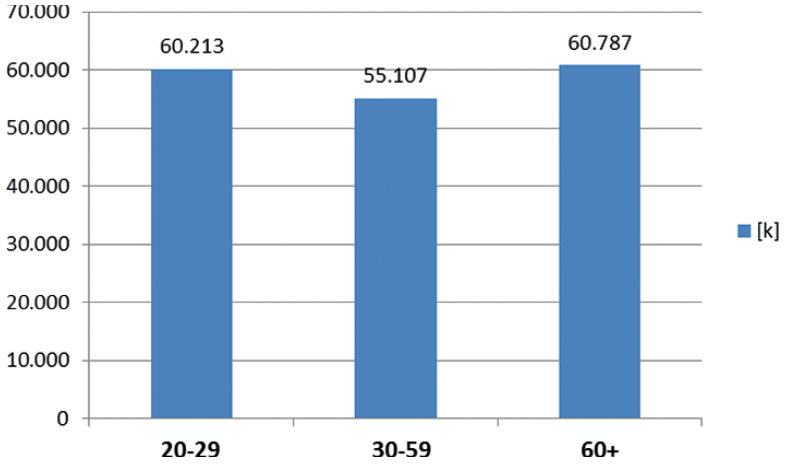


Figura 3 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [t] letto nella lista di frasi e calcolato per ciascuna fascia di età che compone il corpus*

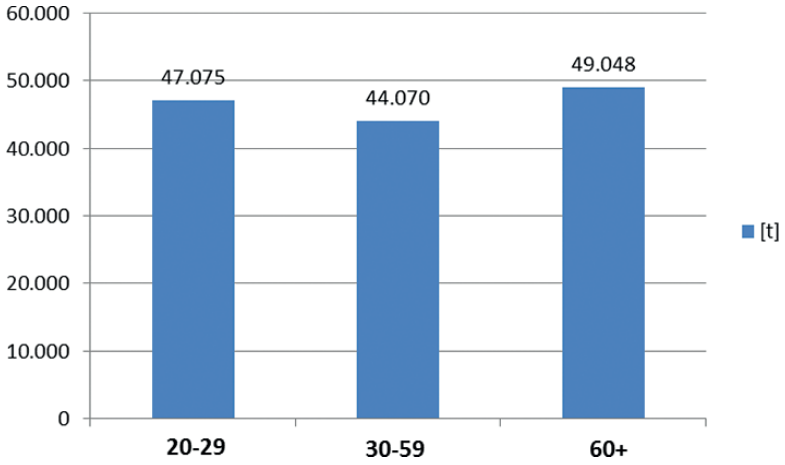


Figura 4 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [k] letto nella lista di frasi e calcolato per genere che compone il corpus*

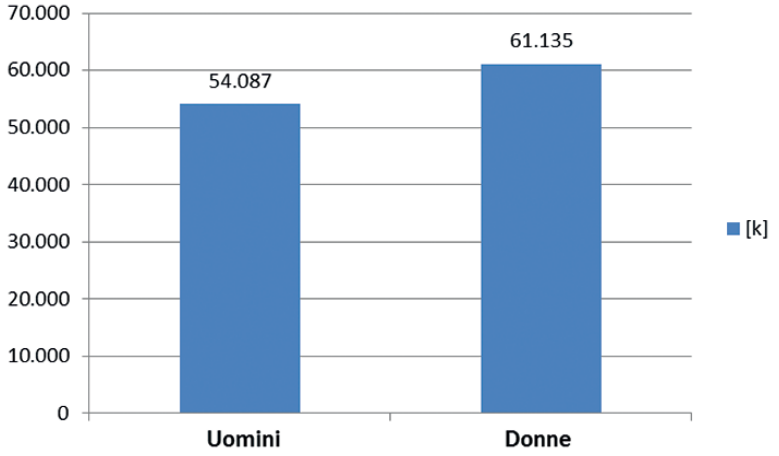
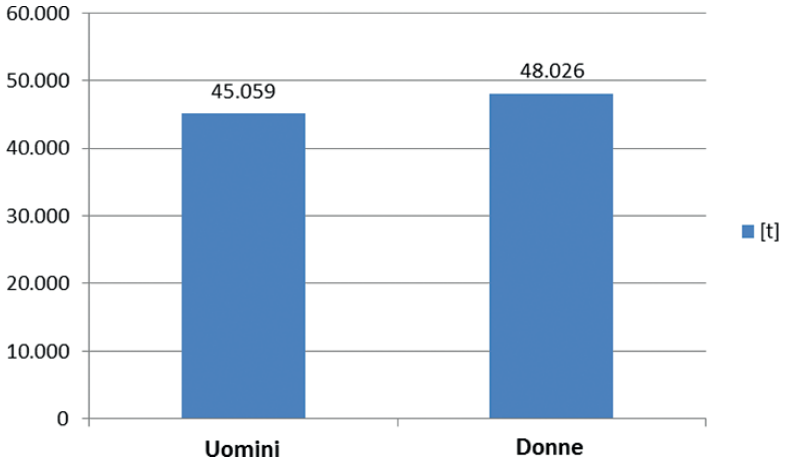


Figura 5 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [t] letto nella lista di frasi e calcolato per genere che compone il corpus*



Riportiamo di seguito, invece, i valori medi del VOT, suddiviso per fasce di età e genere, relativi alla lista di parole:

Figura 6 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [k] letto nella lista di parole e calcolato per ciascuna fascia di età che compone il corpus*

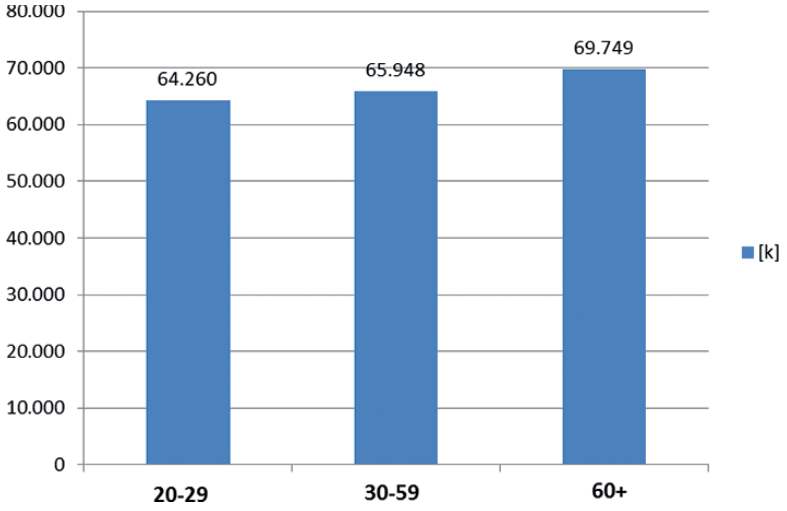


Figura 7 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [t] letto nella lista di parole e calcolato per ciascuna fascia di età che compone il corpus*

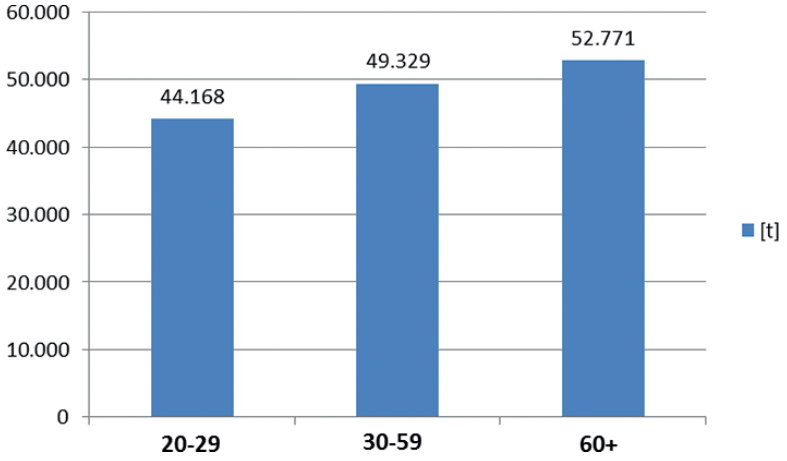


Figura 8 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [k] letto nella lista di parole e calcolato per genere che compone il corpus*

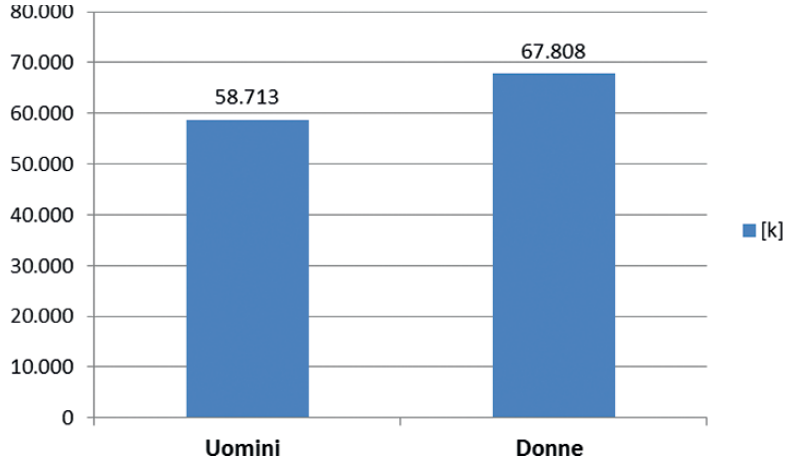
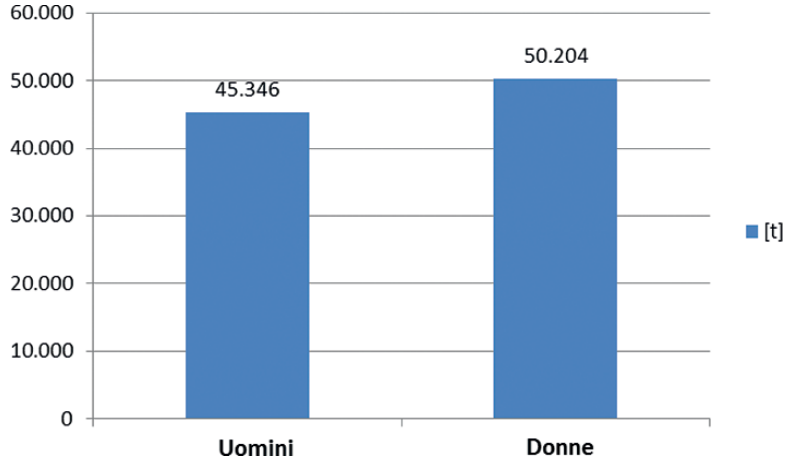


Figura 9 - *Valore medio di durata del VOT della consonante occlusiva sorda aspirata [t] letto nella lista di parole e calcolato per genere che compone il corpus*



Le analisi condotte sul corpus della ricerca confermano l'ipotesi di partenza (cfr. § 2). Nella lista delle frasi, come si può notare dalle figure 2, 3, 4 e 5, si attesta la presenza delle consonanti occlusive sorde aspirate nel dialetto sangiovannese[8] e, di contro, del fenomeno della sonorizzazione delle consonanti occlusive sorde del dialetto cosentino nello stesso contesto:

[8] Si considera come estremo destro del VOT il picco del primo ciclo periodico della vocale che segue l'occlusiva velare sorda considerata.

Figura 10 - *Aspirazione della consonante occlusiva velare sorda nella lista di frasi letta in dialetto sangiovannese da una parlante femminile*

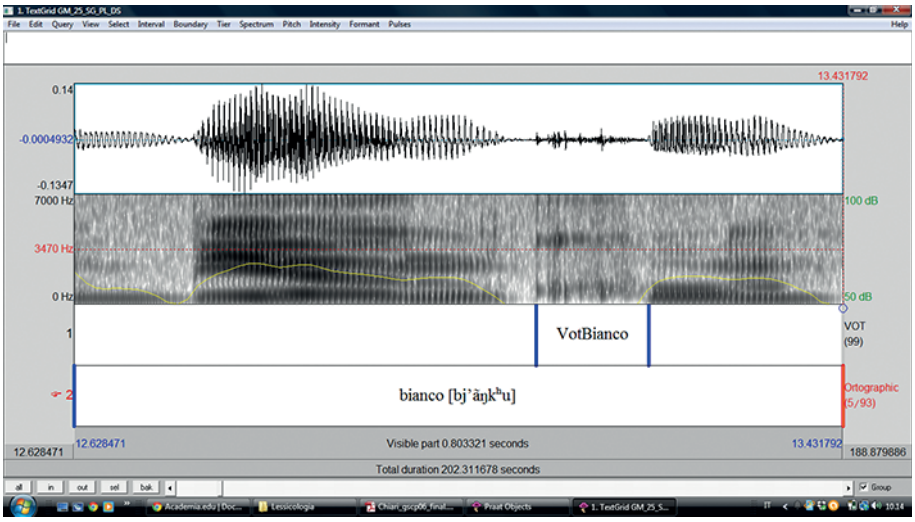


Figura 11 - *Aspirazione della consonante occlusiva dentale sorda nella lista di frasi letta in dialetto sangiovannese da una parlante femminile*

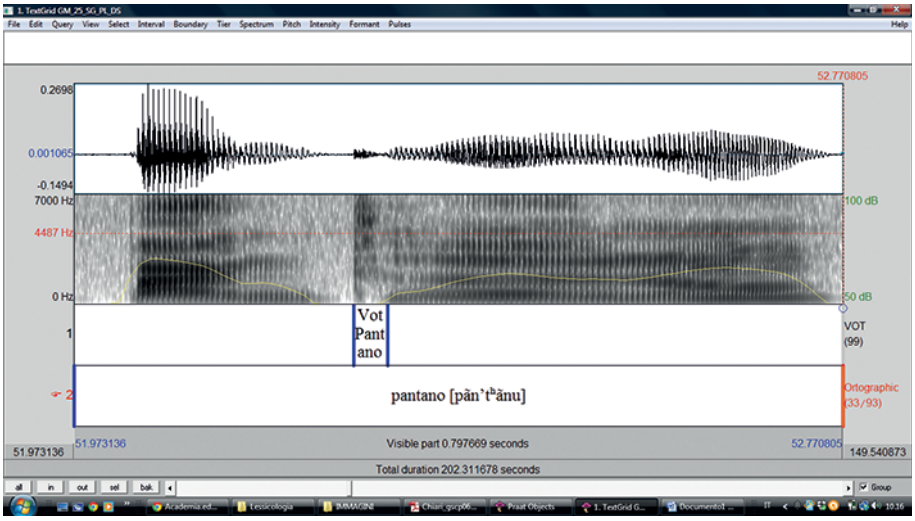


Figura 12 - *Sonorizzazione della consonante occlusiva velare sorda nella lista di frasi letta in dialetto cosentino da una parlante femminile*

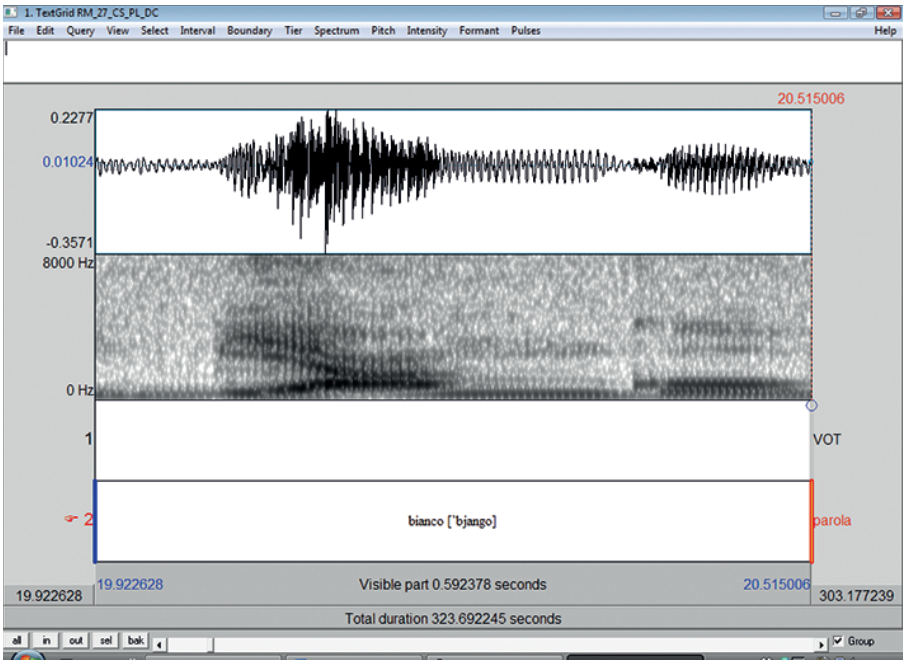
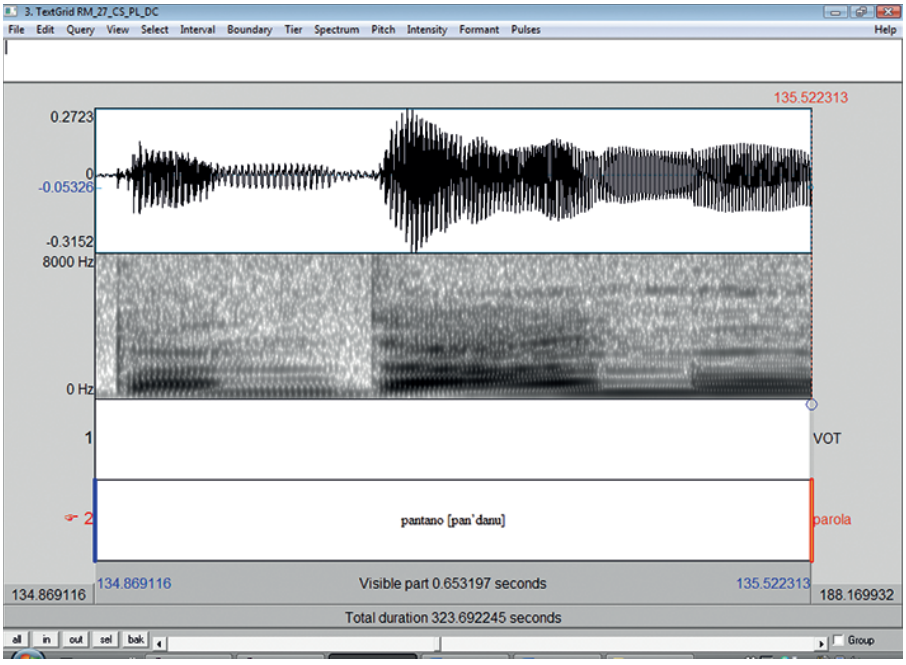


Figura 13 - *Sonorizzazione della consonante occlusiva dentale sorda nella lista di frasi letta in dialetto cosentino da una parlante femminile*



La stesso risultato acustico emerge dalla lettura della lista di parole, come si nota dalle figure 6, 7, 8 e 9:

Figura 14 - *Aspirazione della consonante occlusiva velare sorda nella lista di parole letta in dialetto sangiovannese da una parlante femminile*

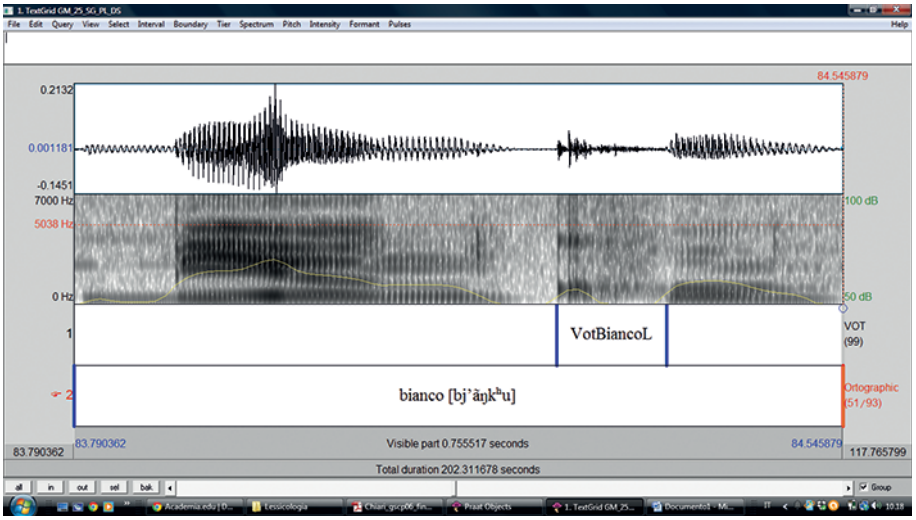


Figura 15 - *Aspirazione della consonante occlusiva dentale sorda nella lista di parole letta in dialetto sangiovannese da una parlante femminile*

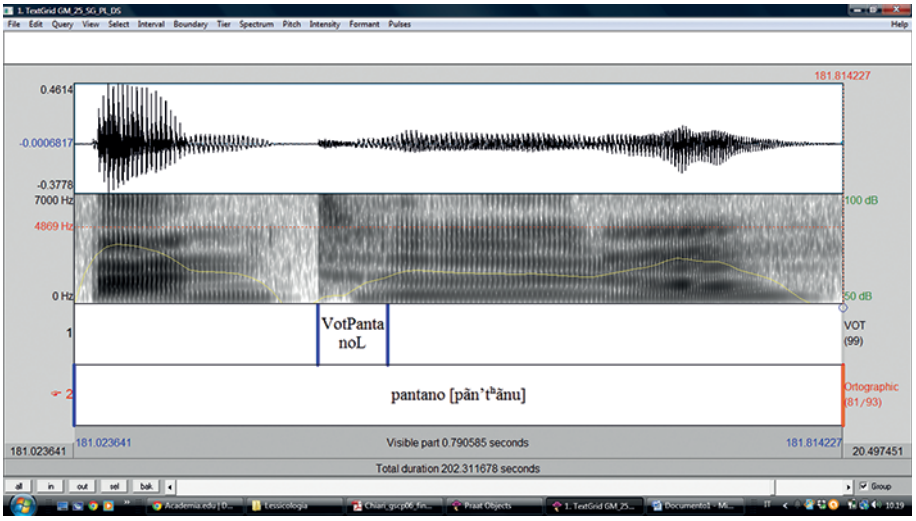


Figura 16 - *Sonorizzazione della consonante occlusiva dentale sorda nella lista di parole letta in dialetto cosentino da un parlante maschile*

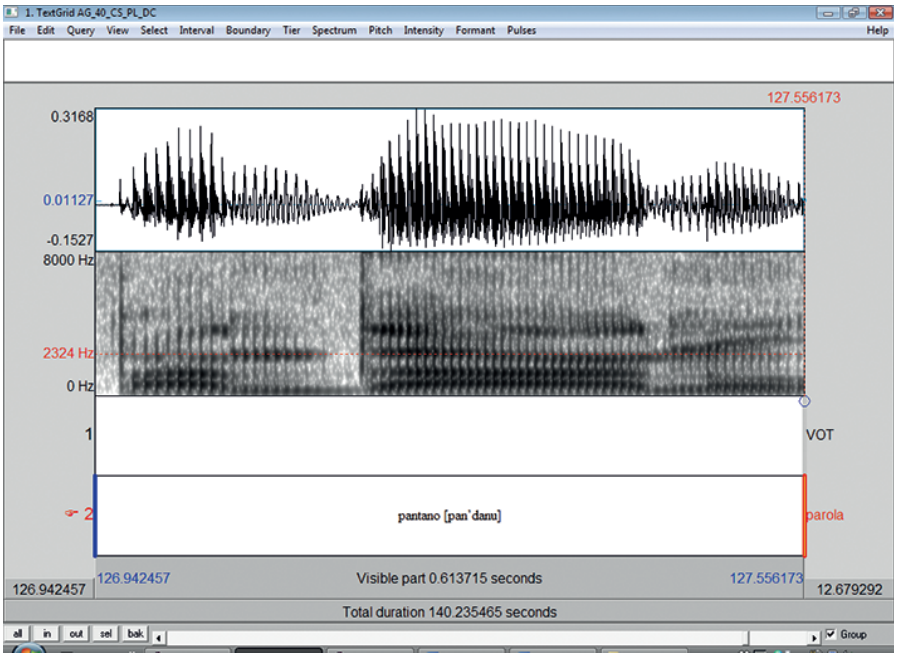
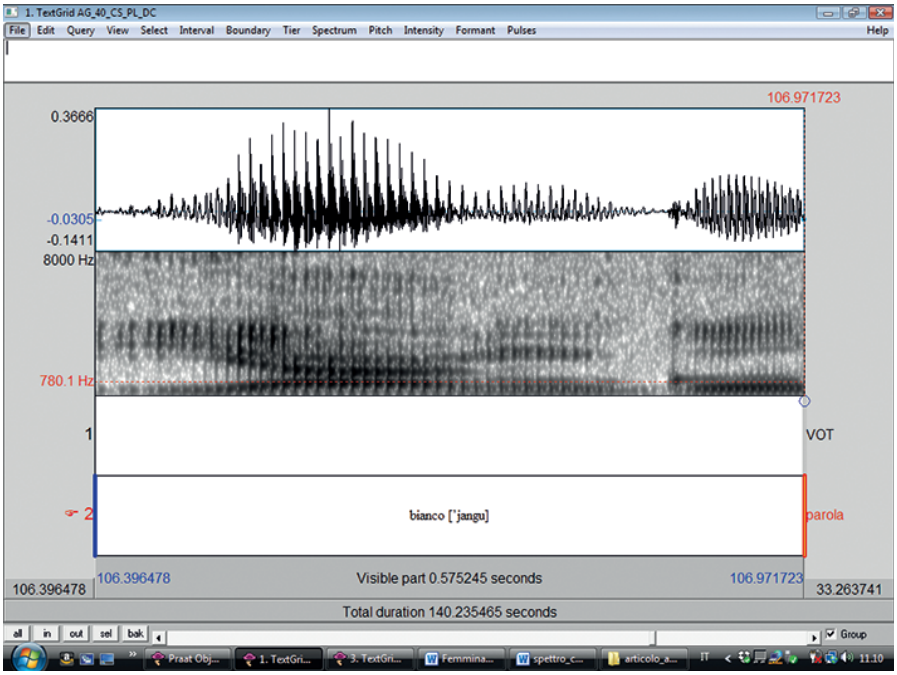


Figura 17 - *Sonorizzazione della consonante occlusiva velare sorda nella lista di parole letta in dialetto cosentino da un parlante maschile*



L'ispezione spettrografica dei dati mostra in maniera chiara che la totalità delle occlusive sorde prodotte dai sangiovannesi presenta un qualche grado di aspirazione, in ogni caso maggiore della soglia 30-35 ms (cfr. Cho, Ladefoged, 1965).

Dai risultati emerge con chiarezza anche quanto specificato in letteratura (cfr. § 2), cioè il VOT della consonante occlusiva velare sorda aspirata ha una maggiore durata rispetto alla dentale sorda aspirata, e questo vale per tutti i contesti accentuali considerati sia nelle frasi che nella lista di parole. A supporto di quanto appena detto, di seguito si riportano esclusivamente i valori medi del VOT per ciascun contesto accentuale esaminato:

Tabella 1 - *Contesti accentuali esaminati e rispettivi valori medi di durata del VOT*

Contesti nk- e -nk-	*Media*	*Contesti nt- e -nt-*	*Media*
parossitoni: nesso -nk- in posizione post tonica interna (frasi)	59.026	parossitoni: nesso -nt- in posizione post tonica interna (frasi)	48.333
parossitoni: nesso -nk- in posizione post tonica interna (lista parole)	75.222	parossitoni: nesso -nt- in posizione post tonica interna (lista parole)	62.689
parossitoni: nesso nk- in posizione pretonica iniziale (frasi)	48.249	parossitoni: nesso nt- in posizione pretonica iniziale (frasi)	32.284
parossitoni: nesso nk- in posizione pretonica iniziale (lista parole)	56.799	parossitoni: nesso nt- in posizione pretonica iniziale (lista parole)	45.847
parossitoni: nesso -nk- in posizione pretonica interna (frasi)	68.250	parossitoni: nesso -nt- in posizione pretonica interna (frasi)	45.622
parossitoni: nesso -nk- in posizione pretonica interna (lista parole)	74.688	parossitoni: nesso -nt- in posizione pretonica interna (lista parole)	48.908
proparossitoni: nesso nk- in posizione pretonica iniziale (frasi)	47.802	proparossitoni: nesso nt- in posizione pretonica iniziale (frasi)	35.841
proparossitoni: nesso nk- in posizione pretonica iniziale (lista parole)	56.425	proparossitoni: nesso nt- in posizione pretonica iniziale (lista parole)	38.120
proparossitoni: nesso -nk- in posizione post tonica interna (frasi)	62.147	proparossitoni: nesso -nt- in posizione post tonica interna (frasi)	60.995
proparossitoni: nesso -nk- in posizione post tonica interna (lista parole)	67.049	proparossitoni: nesso -nt- in posizione post tonica interna (lista parole)	51.209

Tale risultato conferma il dato articolatorio, che cioè il tratto [+/- posteriore] influenza la durata del VOT.

Un risultato molto interessante, che contrasta invece con i dati della letteratura (cfr. Lisker, Abramson, 1964), è la significativa variabilità della durata del VOT sia nella lista di frasi che di parole. Si presentano esclusivamente i valori medi, massimi e minimi in quanto, data la variabilità, risultano poco efficaci ulteriori indagini statistiche:

Figura 18 - *Valore medio, massimo e minimo di durata del VOT delle consonanti occlusive sorde aspirate [k] e [t] relativi alla lista di frasi e calcolati per l'intero corpus*

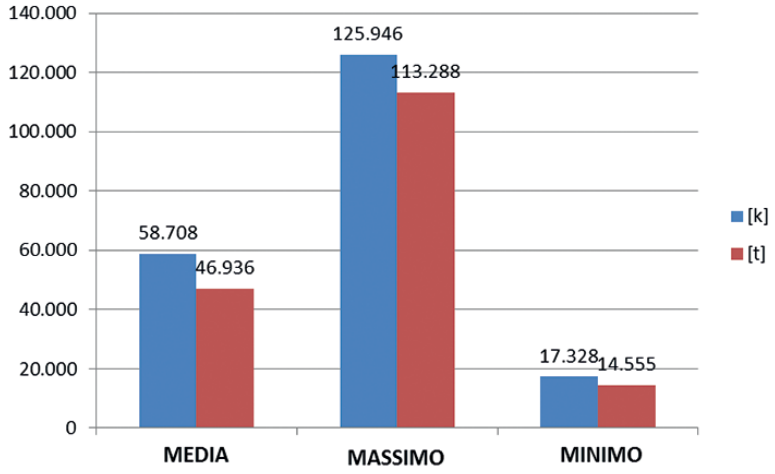
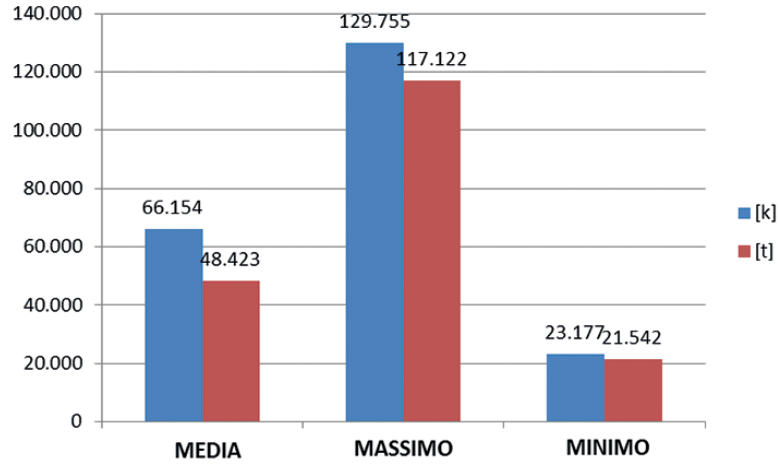


Figura 19 - *Valore medio, massimo e minimo di durata del VOT delle consonanti occlusive sorde aspirate [k] e [t] relativi alla lista di parole e calcolati per l'intero corpus*



Le figure 10 e 11 dimostrano come siano eterogenee le durate medie delle occlusive sorde aspirate, caratterizzate da valori massimi e minimi di VOT molto difformi. Queste oscillazioni di durata evidenziano la stretta connessione tra la durata del parametro acustico e il singolo locutore. Dunque appare evidente che la variabilità

intra-parlatore incida in maniera significativa sui risultati acustici, determinando notevole variabilità nei valori di durata del VOT (ricordiamo che non è presente una significativa differenza di durata del VOT all'interno delle fasce di età né tra i generi analizzati nelle parole in contesto frasale).

6. *Conclusioni*

Lo studio condotto dimostra in modo inequivocabile la presenza dell'aspirazione delle consonanti occlusive sorde in un centro collocato dalla letteratura nella seconda area dialettologica (cfr. § 1) caratterizzata dall'assenza di VOT e dalla presenza di sonorizzazione post-nasale.

Il VOT nel dialetto sangiovannese non è fonologicamente pertinente ed è presente in tutte le classi d'età, nei due generi e in tutti i contesti accentuali e diafasici.

Tale risultato richiede una profonda riflessione dialettologica:

1. la considerazione della distribuzione di questo fenomeno sul territorio calabrese postula una rivisitazione delle isoglosse finora riconosciute;
2. il dialetto di San Giovanni in Fiore deve essere considerato un'oasi linguistica nonostante lo stesso sia molto diverso rispetto ai dialetti appartenenti all'area 3 o 4 con sostrato greco, sia sotto il profilo sintattico (presenza/assenza dell'infinito, presenza/assenza del passato remoto) che lessicale (cfr. Trumper, 1997);
3. l'acquisizione di un processo fonologico da parte dei sangiovannesi, come l'aspirazione, favorito da frequenti contatti con popolozioni greche della non distante *Kroton* (l'attuale Crotone) per ragioni storico-commerciali.

Tra le spiegazioni presentate, riteniamo l'ultima quella più ragionevole. Infatti essa sarà oggetto di studio nel nostro prossimo lavoro.

Bibliografia

Abramson, A.S., Lisker, L. (1965). *Voice onset time in stop consonants: Acoustic analysis and synthesis*, Rapports du 5ᵉ Congrès International D'Acoustique, 1ª Conf. A51.

Cardona, G.R. (1988). *Dizionario di linguistica*. Roma: Armando.

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

Dubois, J. et al. (1979). *Dizionario di linguistica*. Bologna: Zanichelli [trad. it. di Dictionnaire de linguistique (1973), Paris].

Falcone, G. (1976). *Calabria*. In Profilo dei dialetti italiani 18. Pisa: Pacini.

Fuchs, S. (2005). Articulatory correlates of the voicing contrast in alveolar obstruent production in German. In *ZAS Papers in Linguistics*, 41. Berlin: ZAS.

Helgason, P. (2003). Faroese preaspiration. In *Proceedings of the XVth International Congress of the Phonetic Sciences*, Barcelona, 3-9 August 2003, 2517–2520.

Hrycyna, M. et al. (2011). VOT Drift in 3 Generations of Heritage Languages Speakers. In *Canadian Acoustics*, Toronto, 39 (3), 166-167.

Keating, A. (1984). Phonetic and phonological representation of stop consonant voicing. In *Language*, 60 (2), 286-318.

Kessinger, R.H., Blumstein, S.E. (1997). Effects of speaking rate on voice-onset time in Thai, French & English. In *Journal of Phonetics*, 25, 143-168.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lisker, L., Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. In *Word*, 20, 384-422.

Lisker, L., Abramson, A.S. (1971). Distinctive features and laryngeal control. In *Language*, 47, 776-785.

Lisker, L., Abramson, A.S. (1972). Voice-Timing Perception in Spanish Word-Initial Stops. In *83rd Meeting of the Acoustical Society of America*. Buffalo, N.Y., 15.

Lundeborg, I., Nordin, E., Zeipel-Stjerna, M. & Mcallister, A. (2014). Voice onset time in Swedish children with phonological impairment. In *Logopedics Phoniatrics Vocology*, Early Online: 1-7.

Marotta, G., Sorianello, P. (1992). Lenizione e sonorizzazione nell'italiano cosentino: un'analisi sperimentale. In *L'Italia Dialettale* LV. Pisa: Edizioni ETS, 62-122.

Martínez Celdrán, E., Fernández Planas, A.M. (2007). *Manual de fonética español*. Barcelona: Ariel.

Martinez Celdran, E., Fernandez Planas, A.M. & Carrera, J. (2003). Illustrations of the IPA: Spanish. In *Journal of the International Phonetic Association*, 33 (2), 255-260.

Mele, B., Schmid, S. (2009). Le occlusive palatali nel dialetto di San Giovanni in Fiore (CS). In Romito, L., Galatà, V. & Lio, R. (Eds.), *La fonetica sperimentale. Metodo e applicazioni*. Torriana: EDK editore, 349-371.

Mele, B. (2009). *Fonetica e fonologia del dialetto di San Giovanni in Fiore*. Tübingen-Basel: Francke Verlag.

Nance, C., Stuart-Smith, J. (2013). Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. In *Journal of the International Phonetic Association*, 43 (2), 129-153.

Pind, J. (1999). Speech segment durations and quantity in Icelandic. In *Journal of the Acoustical Society of America*, 106 (2), 1045-1053.

Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. In *Perception and Psychophysics*, 57, 291-304.

Rohlfs, G. (1970). *Grammatica storica della lingua italiana e dei suoi dialetti*, 1. Torino: Einaudi.

Romito, L., Scuticchio, P.A. (2008). La retroflessione in alcuni centri della provincia di Catanzaro: verso una mappa regionale. In Romito, L., Lio, R. & Galatà, V. (Eds.), *Proceedings del IV Convegno AISV*. Torriana: EDK Editore, 412-422.

Stevens, M., Hajek, J. (2007). Towards a phonetic conspectus of preaspiration: acoustic evidence from Sienese Italian. In *Proceedings of ICPhS XVI*, Saarbrücken, Germany, 429-432.

Thraisson, H. (1978). On the Phonology of Icelandic Preaspiration. In *Nordic Journal of Linguistics*, 1, 3-54.

Trudgill, P. (1983). *Sociolinguistics. An Introduction to Language and Society*. Harmondsworth: Penguin.

Trumper, J., Maddalon, M. (1988). Converging divergence and diverging convergence: the dialect-language conflict and contrasting evolutionary trends in moden Italy. In *Variation and Convergence*, Berlin - New York: W. de Gruyter, 217-259.

Trumper, J. (1997), Calabria and Southern Basilicata. In *The Dialects of Italy*. Routledge: London, 355-364.

Trumper, J., Chiodo, G. (1999). La pertinenza degli eventi catastrofici naturali per la dialettologia e la linguistica romanze. In *Rivista Italiana di Dialettologia. Lingue dialetti società*, a. XXII. Bologna: CLUEB.

Whiteside, S.P., Marshall, J. (2001). Developmental trends in Voice Onset Time: some evidence for sex difference. In *Phonetica*, 58, 196-210.

Whiteside, S.P., Dobbin, R. & Henry, L. (2003). Patterns of variability in voice onset time: a developmental study of motor speech skills in humans. In *Neuroscience Letters*, 347 (1), 29-32.

Zlatin, M.A. (1974). Voicing contrast: perceptual and productive voice onset time characteristics of adults. In *Journal of the Acoustical Society of America*, 56, 981-994.

PARTE II

ACQUISIZIONE, DESTRUTTURAZIONE E CODIFICA NEURALE DELLA STRUTTURA SONORA DEL LINGUAGGIO

Parte II – Sezione A

Acquisizione di L2

DOI: 10.17469/O2101AISV000012

ANNA DE MEO, MARILISA VITALE, YING XU

Lo sviluppo del sistema vocalico dell'italiano nell'interlingua di adulti sinofoni

The present study is a preliminary investigation on the ability of standard Chinese speakers to pronounce accurate vowel segments in L2 Italian. To this pourpose, six native Chinese learners of Italian, with different levels of L2 competence, were considered and the spectroacoustic analysis of Chinese and Italian vowel inventory, extracted from a corpus of read speech, was performed. Results have shown that the study participants, regardless of the L2 competence level, fail to differentiate the stressed and unstressed vowels, and to distinguish mid-open from mid-closed (front and back) vowel phonemes in Italian. Furthermore, it seems that only phoneme and not allophones are transfered in Italian from the Chinese L1.

1. *Introduzione*

L'accento straniero, che solitamente caratterizza l'interlingua di un parlante non nativo ed è riconosciuto dagli ascoltatori nativi come deviante rispetto alla produzione attesa di un locutore nativo, rappresenta la manifestazione delle difficoltà insite nella gestione del livello segmentale e soprasegmentale di una lingua seconda, sulla quale incidono numerosi fattori interni ed esterni (Piske, MacKay & Flege, 2001). Un parlato caratterizzato da divergenze fonetiche, anche di minima entità, rispetto alla produzione considerata nativa dalla comunità autoctona, può avere ripercussioni sulla comprensibilità degli enunciati e contestualmente indurre gli ascoltatori nativi ad attribuire al locutore un'identità sociale estranea, con possibili effetti sulla sfera professionale (Hosoda, Stone-Romero, 2010).

Lo sviluppo della competenza fonetico-fonologica nella L2 e le problematiche connesse con la pronuncia di una lingua non materna hanno visto negli ultimi anni un aumento di interesse nell'ambito della ricerca linguistica di ambito acquisizionale e didattico[1]. Le numerose indagini descrittive e sperimentali sono state affiancate anche da modelli teorici che hanno cercato di spiegare i meccanismi alla base della percezione, elaborazione e realizzazione dei suoni di una L2[2].

[1] Si possono citare, per il solo 2015, la pubblicazione del volume *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research* di Derwing e Munro (2015), la nascita della rivista *Journal of Second Language Pronunciation* e la settima conferenza annuale su *Pronunciation in Second Language Learning and Teaching*, PSLLT.

[2] Tra i più citati, si ricordano il *Native Language Magnet model* (Kuhl, Iverson, 1995), il *Perceptual Assimilation Model* (Best, 1995), lo *Speech Learning Model* (Flege, 1995), l'*Ontogeny and Philogeny Model* (Major, 2001).

Il presente studio, parte di una ricerca più ampia sulle caratteristiche fonetico-fonologiche dell'italiano L2 di apprendenti adulti, presenta i risultati di un'osservazione preliminare condotta sullo sviluppo delle vocali dell'italiano nell'interlingua di locutori cinesi. L'interlingua, un sistema linguistico in continua evoluzione, organizzato sulla base di un sistema di regole che l'apprendente elabora a partire dalle caratteristiche dell'input ricevuto (Giacalone Ramat, 2003: 18), rappresenta un laboratorio naturale per l'osservazione della nascita e la descrizione dello sviluppo di nuovi sistemi linguistici, e i dati di questa ricerca possono costituire una base di partenza per l'elaborazione di applicazioni didattiche e per la programmazione di interventi formativi mirati al miglioramento della competenza linguistica (Chini, 2011).

2. *Sistemi vocalici a confronto*

Per poter descrivere in maniera adeguata la resa dei segmenti vocalici nella lingua seconda, è opportuno identificare preliminarmente le caratteristiche dei sistemi vocalici di partenza, ovvero quello del cinese lingua materna e quello dell'italiano lingua target.

Per l'italiano si identificano sette fonemi vocalici, /i/ - /e/ - /ɛ/ - /a/ - /ɔ/ - /o/ - /u/, diversamente distribuiti all'interno di due sottoinsiemi, corrispondenti al sistema tonico e al sistema atono (per i valori formantici di riferimento si veda Ferrero, Magno Caldognetto, 1986). Le vocali atone non presentano i foni corrispondenti alle due vocali medio-basse anteriore /ɛ/ e posteriore /ɔ/.

Più complesso e meno definito appare il vocalismo della lingua cinese, per il quale non sembra esserci pieno accordo tra gli studiosi che si sono occupati di tale argomento. Le posizioni assunte riguardo al numero di fonemi vocalici presenti nel cinese standard risultano piuttosto diversificate e possono essere distribuite lungo un continuum che va dalla completa assenza di vocali fonologiche (Pulleyblank, 1984) a un'architettura vocalica articolata, costituita da dieci elementi (Zhu, 2011). Per il presente lavoro sarà adottato il modello presentato in Duanmu (2007), che identifica otto fonemi vocalici:

– cinque vocali di base, /i/ - /y/ - /u/ - /ə/ - /a/
– una vocale retroflessa, /ɚ/
– due vocali apicali, /ɿ/ (vocale apicodentale) e /ʅ/ (vocale apico-retroflessa)[3].

Per lo scopo della ricerca, si è ritenuto opportuno procedere a una verifica delle occorrenze di due varianti coarticolatorie del fonema centrale /ə/, che risultano comparabili a due fonemi vocalici dell'italiano: la variante medio-alta posteriore arrotondata [o], che si ritrova in sillaba aperta, dopo una consonante labiale, e la vocale

[3] Sebbene lo statuto di questi due suoni della lingua cinese sia ancora dibattuto, essi presentano nella loro struttura interna caratteristiche comuni alle vocali e rappresentano il nucleo delle sillabe in cui sono inseriti. I due foni, rappresentati talvolta dai simboli /z/, /ẓ/, sono denominati anche "consonanti sillabiche".

anteriore medio-alta non arrotondata [e], che compare prima della vocale anteriore alta /i/ e, in sillaba aperta, dopo le palatali[4].

3. *Lo studio*

Il presente studio ha come obiettivo l'analisi dello sviluppo della competenza fonologica dell'italiano L2, limitatamente alla componente vocalica, in apprendenti adulti, parlanti nativi di cinese standard.

3.1 I partecipanti

Ai fini della ricerca sono state selezionate sei parlanti cinesi madrelingua, studentesse universitarie con un'età compresa tra 21 e 23 anni, tutte residenti a Napoli al momento della ricerca e quindi esposte alla stessa varietà di italiano parlato. I sei soggetti hanno tutti effettuato un percorso di apprendimento formale dell'italiano durante la formazione universitaria in Cina e, con riferimento ai parametri del Quadro Comune Europeo di Riferimento (QCER) messo a punto dal Consiglio d'Europa (2002), sono differenziate per livello di competenza linguistica nella L2 come segue:

– 2 di livello elementare (A1);
– 2 di livello intermedio (B1);
– 2 di livello avanzato (C1).

La distribuzione delle partecipanti all'interno dei diversi livelli di competenza è stata effettuata mediante la somministrazione di un test di piazzamento, che ha previsto la verifica delle competenze sia attraverso la piattaforma Moodle del Centro Linguistico dell'Università degli Studi di Napoli "L'Orientale" sia mediante un colloquio con un docente di lingua italiana.

I soggetti coinvolti nella ricerca, parlanti della varietà standard di cinese mandarino, sono stati selezionati anche sulla base della propria regione di provenienza, al fine di ricreare una condizione linguistica che riproponesse in maniera più realistica possibile la composizione di una classe di italiano L2[5]. Come gruppo di controllo, sono state coinvolte nello studio anche due parlanti native di italiano, residenti in

[4] La rappresentazione degli allofoni della vocale centrale è ancora un problema aperto tra i fonetisti cinesi, in particolare per quanto riguarda la variante anteriore. Ad esempio, Chao (1968) e Cheng (1973) indicano la variante anteriore media con [e]; Xu (1980) rappresenta con /E/ il fono che occorre in sillaba aperta dopo le palatali e con [e] quello che appare prima della vocale anteriore alta /i/; Wang (1993) usa [e] per la variante che Xu trascrive con [E] e [ə] per quella indicata con [e]. Sulla base dei valori formantici rilevati nel corpus di riferimento, in questo lavoro si è scelto di utilizzare il simbolo [e].

[5] Per ciascun livello di competenza è stato selezionato un parlante cinese proveniente da una regione meridionale e uno proveniente da una regione settentrionale: Liaoning (nord) e Zhejiang (sud) per il livello A1; Municipalità Autonoma di Pechino (nord) e Anhui (centro-sud) per il livello B1; Shandong (nord) e Zhejiang (sud) per il livello C1.

Campania, di età media 29 anni, laureate in discipline umanistiche, docenti di italiano L2.

3.2 Il *corpus*

Per lo studio qui presentato è stato utilizzato un corpus costituito da parlato letto, al fine di consentire una maggiore comparabilità dei dati relativi ai diversi soggetti presi in esame. Nello specifico, a ciascun parlante è stato chiesto di leggere e registrare, dopo una breve fase di familiarizzazione, due brevi testi, che contenevano l'intero inventario fonetico-fonologico dell'italiano; ai parlanti cinesi è stato chiesto di leggere due testi anche nella lingua propria materna, i quali contenevano l'intero repertorio vocalico del cinese.

Sono quindi state complessivamente registrate 28 produzioni di parlato letto:
- 4 in italiano L1 (2 testi x 2 parlanti);
- 12 in italiano L2 (2 testi x 6 parlanti);
- 12 in cinese L1 (2 testi x 6 parlanti).

3.3 L'analisi

L'intero corpus raccolto è stato sottoposto ad analisi spettroacustica manuale, mediante il software Praat (Boersma, Weenink, 2015). Per ciascuna vocale prodotta dalle otto parlanti, native e non native, sono state rilevate le frequenze (in Hz) delle prime due formanti (F1, F2), in corrispondenza del punto centrale e dunque più stabile del segmento fonico, e le durate (in ms).

Prima di procedere all'analisi dei sistemi vocalici in L1 e L2, i valori in Hz di F1 e F2 sono stati normalizzati secondo la formula Nearey 2[6], allo scopo di annullare le variazioni dovute alle differenze anatomo-fisiologiche dei parlanti.

4. *I risultati*

Per verificare l'eventuale presenza di fenomeni di transfer nella realizzazione delle vocali dell'italiano L2, sono stati sottoposti ad analisi spettroacustica anche i segmenti vocalici prodotti dalle parlanti cinesi nella propria lingua materna.

I valori medi di F1 e F2 dei sei soggetti dello studio sono rappresentati graficamente nella Figura 1, che riporta anche i valori dei due allofoni della vocale centrale /ə/, rispettivamente [e] e [o]. Va osservata la posizione dell'allofono posteriore [o], che risulta collocato in stretta prossimità della vocale alta /u/.

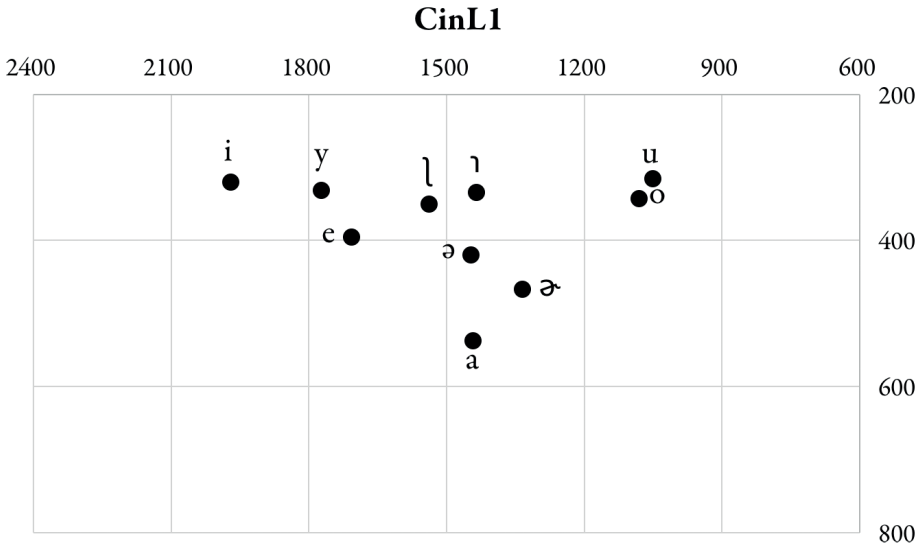
Per quanto riguarda le vocali dell'italiano L1, l'analisi del parlato letto delle due parlanti native del corpus mostra la diversa estensione del sistema tonico e atono.

[6] Metodo di normalizzazione di tipo estrinseco, che opera una trasformazione dei valori originali considerando le informazioni contenute nelle diverse formanti come parte di un sistema (Nearey, 1978). La procedura di normalizzazione Nearey 2 utilizza un fattore di scala identico per le diverse formanti analizzate e prende in considerazione la media dei valori logaritmici delle stesse. La formula applicata è: $F^*_{n[V]} = \text{anti-log}(\log(_{Fn[V]}) - \text{MEAN}_{log'})$.

Come evidente nella rappresentazione grafica di Figura 2, le vocali prodotte all'interno di sillabe atone risultano più centralizzate rispetto a quelle presenti nelle sillabe accentate.

Successivamente i dati relativi al cinese L1 saranno presentati suddivisi in tre sottogruppi (Figure 3-5), con riferimento al livello di competenza dell'italiano L2 delle parlanti (base A1, intermedio B1 e avanzato C1). Al fine di valutare in maniera più precisa gli esiti del processo di acquisizione tali dati saranno rappresentati graficamente insieme a quelli delle due locutrici native e dell'italiano L2 delle due parlanti cinesi inquadrate nello specifico livello di competenza.

Figura 1 - *Cinese L1: valori medi di F1 e F2*



La Figura 3 riporta i dati delle due parlanti di livello A1; la Figura 4 quelli delle parlanti di livello B1 e la Figura 5 i dati di quelle del livello C1. In ciascuna figura sono rappresentati graficamente i valori medi della F1 e F2 delle vocali prodotte in italiano L1 e L2 e del cinese L1. In tutte le figure, e dunque in tutti e tre i gruppi di partecipanti, si osserva la mancata riduzione dell'area vocalica atona tipica dell'italiano L1, fatta eccezione per la vocale centrale bassa /a/.

Figura 2 - *Italiano L1: valori medi di F1 e F2*

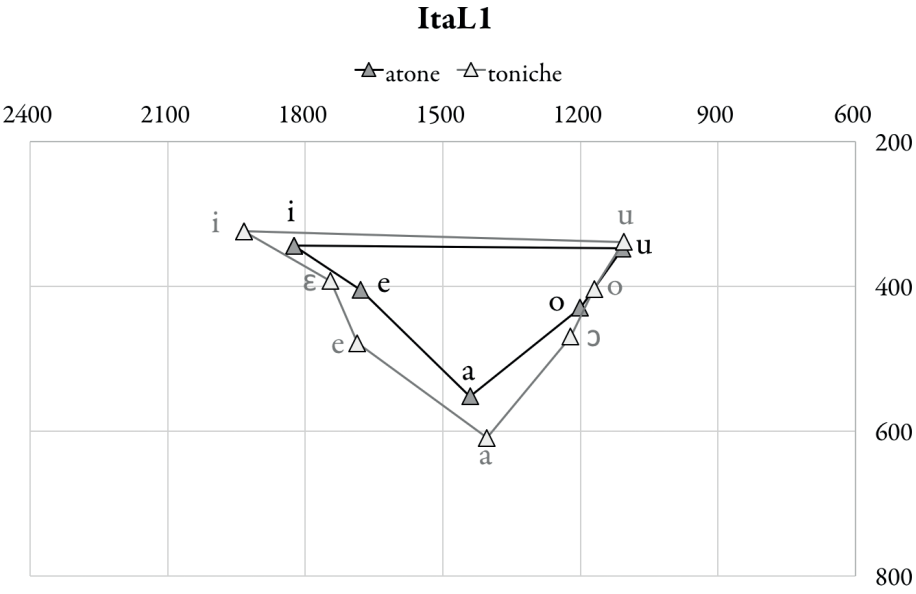
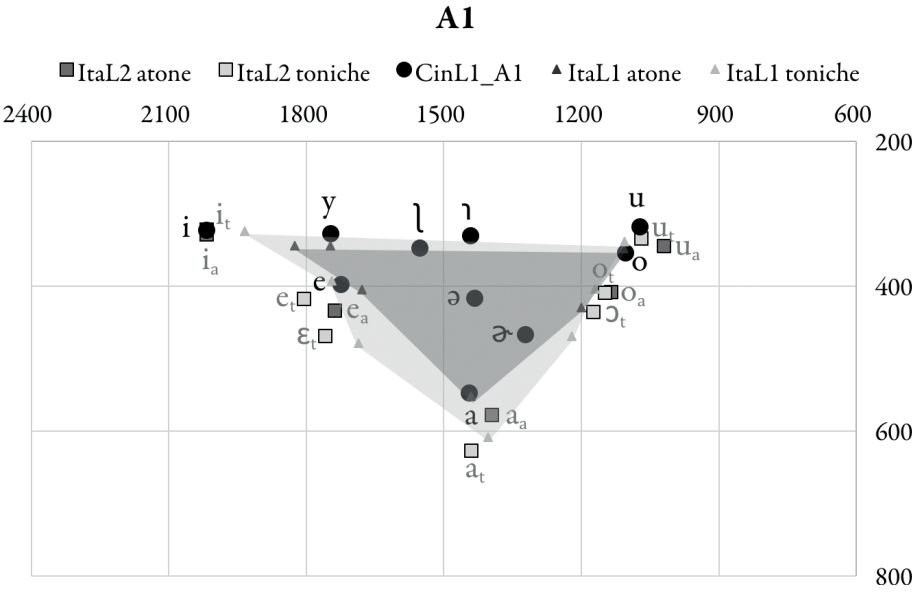


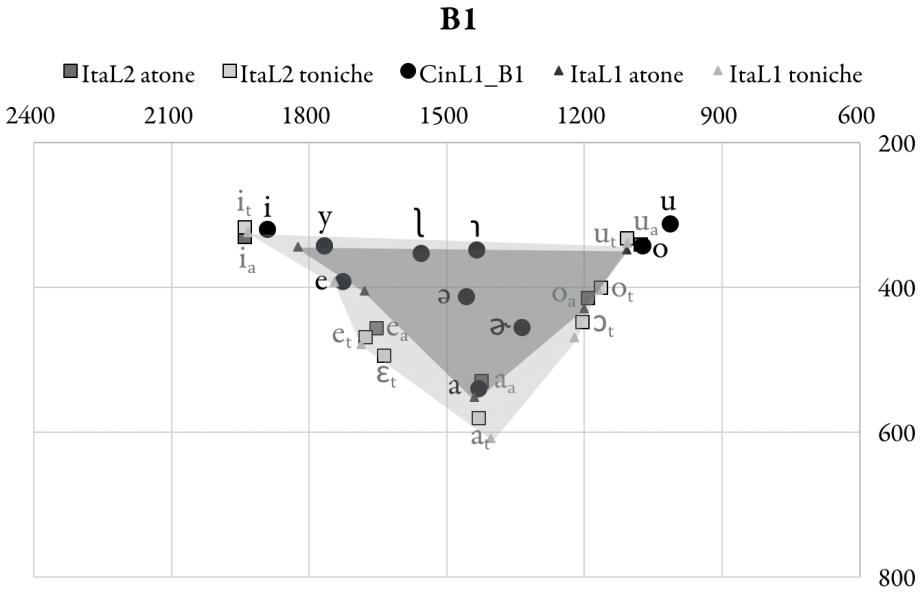
Figura 3 - *Livello A1, valori medi di F1 e F2: italiano L2 (□), italiano L1 (△), cinese L1 (○); v. toniche (t + grigio chiaro), v. atone (a + grigio scuro); sullo sfondo aree vocaliche tonica (chiara) e atona (scura) dell'italiano L1*



In italiano, in posizione accentata, i segmenti vocalici presentano una durata generalmente superiore rispetto a quella riscontrabile, per gli stessi segmenti, in posizione atona. Tale differenziazione temporale non ha valore fonologico, tuttavia

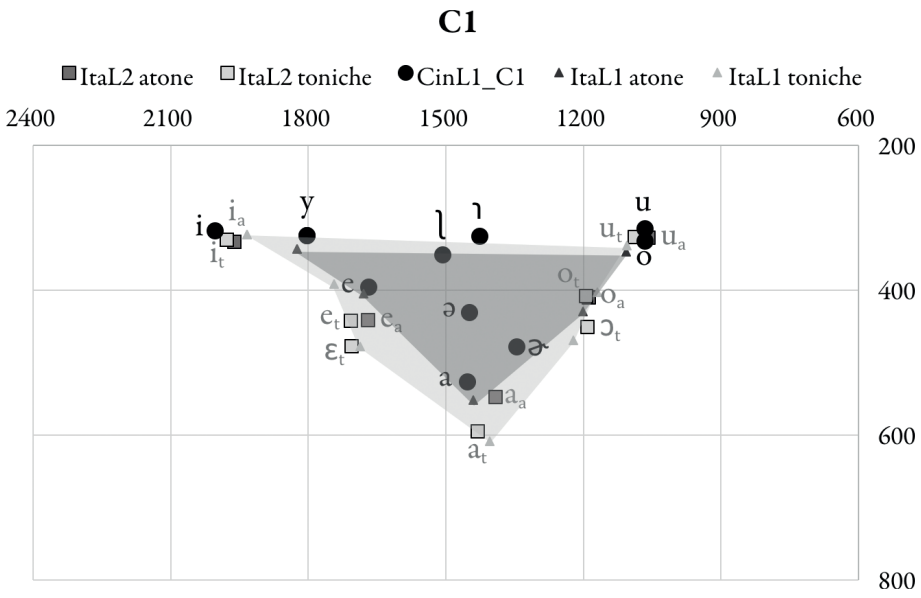
rappresenta il tratto prosodico che maggiormente partecipa alla caratterizzazione delle vocali toniche e atone (Bertinetto, 1981), con effetti evidenti sul piano dell'articolazione. Un intervallo temporale maggiore consente, infatti, al parlante di raggiungere in maniera più precisa il target articolatorio specifico delle diverse vocali, producendo come risultato un suono più accurato; nel caso delle vocali atone, invece, la durata più contenuta del segmento fonico comporta un maggior grado di approssimazione in termini di apertura e avanzamento degli organi articolatori, corrispondenti a un suono vocalico meno preciso e definito.

Figura 4 - *Livello B1, valori medi di F1 e F2: italiano L2 (□), italiano L1 (Δ), cinese L1 (○); v. toniche (t + grigio chiaro), v. atone (a + grigio scuro); sullo sfondo aree vocaliche tonica (chiara) e atona (scura) dell'italiano L1*



La distinzione operata dai parlanti nativi nell'articolazione dei segmenti vocalici tonici e atoni dell'italiano è certamente un fenomeno meno scontato per i non nativi e i risultati dell'analisi condotta nel presente studio lascia supporre che le apprendenti cinesi prese in esame, anche a un livello avanzato di competenza della lingua seconda, non differenzino le vocali in termini di durata.

Figura 5 - *Livello C1, valori medi di F1 e F2: italiano L2 (□), italiano L1 (△), cinese L1 (○); v. toniche (t + grigio chiaro), v. atone (a + grigio scuro); sullo sfondo aree vocaliche tonica (chiara) e atona (scura) dell'italiano L1*



Per poter valutare in maniera adeguata la durata dei segmenti vocalici del corpus, si è proceduto alla misurazione della velocità di articolazione rilevata nell'italiano L1 ed L2, ossia il rapporto tra il numero delle sillabe prodotte e la durata in secondi delle catene foniche (sill/s). La velocità di articolazione è un indice qualitativo, che esprime l'accuratezza del gesto articolatorio e la cui variazione è limitata dal grado di controllo del parlante rispetto ai contenuti espressi, oltre che dal tempo di fonazione e da costrizioni anatomo-fisiologiche. Nel nostro corpus questo indice appare positivamente correlato al livello di competenza linguistica dei locutori, come evidente dai valori riportati nella Tabella 1:

Tabella 1 - *Valori medi della velocità di articolazione in italiano L2 (livello A1, B1 e C1) e in italiano L1*

Italiano L2			*Italiano L1*
A1	B1	C1	
3,8 sill/s	4,1 sill/s	4,6 sill/s	5,4 sill/s

I valori medi in millisecondi (ms) delle durate relative alle vocali toniche e atone dell'italiano L2 delle parlanti coinvolte nello studio sono riportati nella Tabella 2, che presenta anche i valori rilevati nelle produzioni delle due parlanti native.

Tabella 2 - *Valori medi in ms delle durate vocaliche in italiano L2 (livello A1, B1 e C1) e in italiano L1*

		Italiano L2			*Italiano L1*
		A1	**B1**	**C1**	
Toniche	**/i/**	123	102	104	86
	/e/	140	112	98	72
	/ɛ/	137	109	102	110
	/a/	146	126	119	119
	/ɔ/	128	111	116	99
	/o/	139	122	117	112
	/u/	107	104	100	96
	media	131	112	108	99
Atone	**/i/**	107	76	72	55
	/e/	130	80	75	59
	/a/	117	88	81	70
	/o/	117	92	80	71
	/u/	97	83	74	52
	media	114	84	76	61

Dall'esame dei valori medi delle durate vocaliche e della velocità di articolazione si osserva un'accelerazione articolatoria, con conseguente riduzione della durata del segmento vocalico, associata al progredire della competenza linguistica. Tuttavia, i valori di tali durate restano mediamente più alti di quelli dell'italiano nativo, anche nel caso di soggetti con competenza avanzata C1.

Per verificare se all'interno del sistema dell'interlingua fosse preservata una differenza tra la durata delle vocali toniche e quella delle atone, è stato poi calcolato il differenziale Δd (valore percentuale della riduzione di durata tra vocali toniche e atone) per ciascun elemento dell'inventario fonetico-fonologico dell'italiano L1 e L2. I valori di Δd sono riportati nella Tabella 3. Sono state escluse dal calcolo le vocali medio-basse anteriori e posteriori, presenti solo nel sistema tonico dell'italiano.

Tabella 3 - *Valore percentuale della riduzione di durata tra vocali toniche e atone (Δd)*

	Italiano L2			*Italiano L1*
	A1	**B1**	**C1**	
/i/	13	25	31	36
/e/	7	29	23	18
/a/	20	30	32	41
/o/	16	25	32	37
/u/	11	13	11	46
% media	12	25	28	38

I valori di Δd mostrano, seppure con gradi diversi, una differenziazione temporale tra vocali toniche e atone all'interno di ciascun sistema di interlingua. Si nota un

progressivo aumento del Δd, a partire dal valore minimo rilevato nel livello base A1 (12% di riduzione della durata delle atone rispetto alle toniche) fino al valore massimo presente nel livello avanzato C1 (28% di riduzione), tuttavia ben distante dal valore medio delle parlanti native (38% di riduzione delle atone). Sebbene le parlanti non native realizzino le vocali atone con una durata inferiore rispetto a quella delle vocali toniche, la ridotta velocità di articolazione consente loro, comunque, una generale stabilità dei modelli formantici, con conseguente uniformità delle vocali toniche e atone.

Un confronto tra i dati dell'italiano L2 con quelli del cinese lingua materna ha permesso successivamente di verificare la presenza di eventuali fenomeni di transfer. Per tutti e tre i livelli di competenza appare una chiara sovrapposizione dei valori formantici della L1 e della L2 nella realizzazione delle due vocali alte, anteriore e posteriore, /i/ e /u/. La vocale centrale bassa /a/ del cinese mostra valori formantici che corrispondono a quelli della vocale atona dell'italiano L1; questo fono è l'unico a presentare, anche se in modo leggermente diverso nei tre livelli, una certa riduzione nella realizzazione atona. L'allofono [o] del cinese L1, la cui realizzazione appare molto vicina a quella della vocale posteriore alta /u/, non viene trasferito nell'italiano L2, dove compaiono due vocali medie posteriori scarsamente differenziate per grado di apertura ma con valori formantici vicini alla vocale medio-alta /o/ dell'italiano. Neppure l'allofono [e], i cui valori si collocano in corrispondenza di quelli della vocale atona italiana /e/, viene direttamente trasferito nel sistema dell'interlingua, dove appaiono due vocali medie anteriori, anch'esse scarsamente differenziate per apertura e variamente collocate tra la /e/ e la /ɛ/ dell'italiano.

5. *Conclusioni*

Il presente studio si configura come un'indagine preliminare, condotta attraverso un'analisi strumentale della produzione orale, sulla capacità di gestire la realizzazione dei segmenti vocalici dell'italiano L2 in apprendenti aventi il cinese standard come lingua materna. I soggetti selezionati per l'indagine risultano equamente distribuiti su tre diversi livelli di competenza della lingua obiettivo (A1, B1 e C1 del QCER).

I risultati ottenuti dall'analisi dei tracciati spettrografici delle vocali presenti nel corpus di italiano L2 e il confronto con le produzioni in cinese L1 e italiano L1, raccolte in funzione di campione di controllo, hanno permesso di rilevare diversi fenomeni del vocalismo L2 difformi rispetto al modello nativo italiano. Indipendentemente dal livello di competenza della lingua italiana, i parlanti non nativi non sono risultati in grado di differenziare il sistema tonico delle vocali da quello atono, in quanto per i segmenti vocalici non accentati non è stato osservato un chiaro fenomeno di centralizzazione, tipico dell'italiano L1. Questo dato è in parte dovuto alle caratteristiche del sistema vocalico della lingua materna degli apprendenti, in cui manca la distinzione tra sistema tonico e atono, in parte alla bassa velocità di articolazione.

Nelle produzioni non native, inoltre, è stata rilevata la mancata capacità di differenziare, all'interno del sistema tonico, le vocali medio-alte e medio-basse, collocate in posizione sia anteriore sia posteriore del trapezio vocalico, anche questa difficoltà connessa a una mancata presenza di tali suoni, con valore distintivo, nella lingua materna.

I nostri dati sembrano mostrare che allofoni della lingua materna, anche articolatoriamente vicini a fonemi della lingua obiettivo dell'apprendimento, non vengano trasferiti nell'interlingua. Il transfer riguarda piuttosto i fonemi della L1 che trovano una corrispondenza in fonemi della L2.

Un prossimo sviluppo della ricerca, oltre ad ampliare il numero dei soggetti coinvolti, per confermare statisticamente i dati rilevati in questa fase preliminare dello studio, intende verificare anche se la mancata differenziazione tra vocali toniche e atone nell'italiano di parlanti cinesi influisca percettivamente sull'identificazione di un parlato non nativo e contribuisca all'attribuzione di un determinato grado di accento straniero. I dati fin qui raccolti sembrano confermare la diretta relazione tra durata del segmento vocalico e accuratezza del gesto articolatorio, per cui il parlato italiano L2 di cinesi, caratterizzato da valori di velocità di articolazione ridotti, potrebbe apparire naturale e "nativo" nonostante la anomala staticità delle aree vocaliche.

Bibliografia

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

Best, C.T. (1995). A direct realist perspective on cross-language speech perception. In Strange, W. (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press, 167-200.

Boersma, P., Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.3.84. http://www.praat.org.

Chao, Y.R. (1968). *A grammar of spoken Chinese*. Berkeley, California: University of California Press.

Cheng, C.C. (1973). *A Synchronic Phonology of Mandarin Chinese*, Monographs on Linguistic Analysis n. 4. The Hague: Mouton.

Chini, M. (2011). Qualche riflessione sulla didattica di L2 ispirata alla recente ricerca acquisizionale. In *Italiano LinguaDue*, 2, 1-20.

Consiglio d'Europa (2002). *Quadro comune europeo di riferimento per le lingue. Apprendimento, insegnamento, valutazione.* Firenze: La Nuova Italia.

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

Duanmu, S. (2007). *The Phonology of Standard Chinese*, II edizione. Oxford: Oxford University Press.

Ferrero, F.E., Magno Caldognetto, E. (1986). Elementi di fonetica acustica. In Croatto, L. (Ed.), *Trattato di foniatria e logopedia*, vol. III. Padova: La Garangola, 155-196.

Flege, J.E. (1995). Second language speech learning: theory, findings, and problems. In Strange, W. (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press, 229-273.

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

Hosoda, M., Stone-Romero, E.F. (2010). The effects of foreign accents on employmentrelated decisions. In *Journal of Managerial Psychology*, 25, 113-132.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Kuhl, P.K., Iverson, P. (1995). Linguistic experience and the «perceptual magnet effect". In Strange, W. (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press, 121-154.

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

Nearey, T.M. (1978). *Phonetic Feature Systems for Vowels*. Indiana: Indiana University Linguistics Club.

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, 191-215.

Pulleyblank, E.G. (1984). *Middle Chinese: A Study in Historical Phonology*. Vancouver: University of British Columbia Press.

Strange, W. (1995). Phonetics of second-language acquisition: past, present, future. In *ICPhS*, 95 (4), 76-83.

Strange, W., Shafer, V.L. (2008). Speech perception in late second language learners: the re-education of selective perception. In Hansen Edwards, J.G., Zampini, M.L. (Eds.), *Phonology and Second Language Acquisition*. Amsterdam/Philadelphia: John Benjamins, 153-191.

Trouvain, J., Gut, U. (Eds.) (2007). *Non-Native Prosody. Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter.

Wang, Z. (1993). *The Geometry of Segmental Features in Beijing Mandarin*, Doctoral dissertation. The University of Delaware, Newark.

Xu, S. (1980). *Putonghua yuyin zhishi* [*Phonology of Standard Chinese*]. Beijing: Wenzi Gaige Chubanshe.

Zampini, M.L. (2008). L2 speech production research. In Hansen Edwards, J.G., Zampini, M.L. (Eds.), *Phonology and Second Language Acquisition*. Amsterdam/ Philadelphia: John Benjamins, 220-249.

Zhu, L. (2011). *Research on Chinese English Phonetics-The process from Chinese phonetics to Chinese English Phonetics*. Beijing: China Environmental Science Press.

DOI: 10.17469/O2101AISV000013

STEPHAN SCHMID, SARAH WACHTER

Le ostruenti sonore nella pronuncia dell'italiano di apprendenti svizzero-tedeschi

This study examines the pronunciation of voiced obstruents in Italian as a second language of Swiss German learners. On the basis of a contrastive analysis and the predictions of current models of L2 phonology, we claim that Swiss German learners 'assimilate' the voiced obstruents of Italian to the 'lax' (but unvoiced) sounds of their L1, and that devoicing is particularly frequent in the case of /dz/, /dʒ/ and /z/+C clusters.
These hypotheses are borne out by an acoustic analysis of a corpus of read speech gathered from 10 high school students in Zurich, from which we calculated the locally unvoiced frames of 644 tokens. We conclude that L2 pronunciation is determined both by the 'perceptual filter' of the L1 and by general patterns of markedness.

1. *Introduzione*

Il presente contributo esamina la pronuncia dell'italiano da parte di un gruppo di studenti liceali zurighesi, focalizzando la realizzazione del tratto [±sonoro] nelle ostruenti, il quale rappresenta una difficoltà notoria per apprendenti con un dialetto svizzero-tedesco come L1. Dopo brevi cenni ad alcuni modelli dell'acquisizione della fonologia di una L2 (1.1) si delinea uno schizzo di fonologia contrastiva (italiano *vs*. svizzero-tedesco) relativo ai sottosistemi delle ostruenti (1.2), il che permette di formulare delle ipotesi puntuali sulle difficoltà di pronuncia degli apprendenti svizzero-tedeschi (1.3). Dopo l'illustrazione del campione (2.1), del corpus (2.2) e della metodologia di indagine (2.3) si analizza il tasso di desonorizzazione delle ostruenti sonore secondo i fattori 'parlante' (3.2), 'consonante' (3.3) e 'contesto' (3.4). La discussione dei risultati (4) permette di confermare le ipotesi circa gli effetti congiunti della L1 (che agisce come filtro percettivo) e di principi generali di marcatezza (quali il contesto fonotattico e il luogo di articolazione) sulla pronuncia di una seconda lingua.

1.1 Pronunciare i suoni di un'altra lingua: alcuni modelli teorici

Negli ultimi decenni, l'acquisizione della pronuncia in una L2 è stata oggetto di un numero via via crescente di ricerche di fonetica sperimentale, alcune delle quali fanno riferimento a modelli teorici che sono stati elaborati per rendere conto dell'emergere della struttura sonora in una seconda lingua. Per molto tempo la ricerca internazionale ha preferito l'inglese come campo d'indagine del *second language speech*, ma più recentemente è stata studiata anche la pronuncia dell'italiano da par-

te di stranieri: v. per esempio i contributi riuniti nel volume curato da Costamagna e Marotta (2008) nonché vari lavori presentati ai convegni AISV degli ultimi anni.

Non è questa la sede per passare in rassegna la ricerca in questo campo né per discutere in modo approfondito le diverse teorie che si prefiggono di spiegare la genesi dell'accento straniero (per una discussione più approfondita rimandiamo a Schmid, 2012: 633-637), ma è doveroso menzionare innanzitutto lo *Speech Learning Model* (SLM, Flege, 1995) e il *Perceptual Assimilation Model* (nella sua versione PAM-L2, Best, Tyler, 2007). Secondo lo SLM, i suoni meglio distinguibili dal punto di vista percettivo vengono anche acquisiti più facilmente (Flege, 1995: 239; ma v. già Flege, 1987). Analogamente, il PAM-L2 – sviluppato sulla base di una teoria più generale della percezione di suoni sconosciuti che Catherine Best aveva elaborato negli anni Novanta (v., ad esempio, Best, 1995) – assume che "naïve listeners" [sic] tendano ad assimilare percettivamente i suoni della L2 ai fonemi articolatoriamente più simili della loro L1 (Best, Tyler, 2007: 22; cfr. anche Avesani et al., 2008). Tutto sommato, ambedue i modelli attribuiscono quindi alla lingua materna quella funzione di 'vaglio fonologico' che era già stata individuata da Trubetzkoy (1939 [1971: 126]).

Un altro modello dell'acquisizione della fonologia di una L2 al quale occorre accennare brevemente è l'*Ontogeny Philogeny Model* (OPM, Major, 2001), che mette invece in rilievo i processi evolutivi inerenti a principi universali del linguaggio verbale umano. A questo proposito non si può non ricordare il pensiero di Roman Jakobson (come Trubetzkoy membro del Circolo di Praga e in quanto tale uno dei fondatori dello strutturalismo europeo e della fonologia moderna), e in particolare il famoso testo del 1941 (la traduzione italiana esce nel 1971), che mette in relazione lo sviluppo del linguaggio nel bambino con la perdita della parola nell'afasia, ricollegando infine questi due processi speculari a 'leggi foniche generali' che si esplicitano anche nella tipologia fonologica delle lingue del mondo.

Nell'ambito della teoria dell'acquisizione di lingue seconde, la tipologia linguistica viene chiamata in causa anche dalla *Markedness Differential Hypothesis* (MDH, Eckman, 1977 e 2008: 97-100). Questa ipotesi postula che la difficoltà delle strutture sonore della L2 dipenda non soltanto dalle differenze tra L2 e L1, ma anche dal grado di marcatezza delle strutture nelle rispettive lingue. Ad esempio, è stato dimostrato che il tratto [±sonoro] delle ostruenti di una L2 viene acquisito difficilmente nei contesti fonotattici marcati – come la posizione finale di sillaba – se la L1 non possiede ostruenti in tale contesto.

1.2 Ostruenti sonore dell'italiano e ostruenti 'leni' dello svizzero-tedesco

Iniziamo lo schizzo di fonologia contrastiva con la descrizione dell'inventario delle ostruenti nella lingua di partenza dei nostri soggetti, ovvero nel dialetto di Zurigo (cfr. Fleischer, Schmid, 2006: 144):

Figura 1 - *Inventario delle ostruenti del dialetto di Zurigo*

	bilabiali	labiodentali	alveolari	palato-alveolari	velari
occlusive	p b̥		t d̥		k g̊
fricative		f v̥	s z̥	ʃ ʒ̊	x ɣ̊
affricate		p͡f	t͡s	t͡ʃ	k͡x

Tralasciando il fonema isolato /h/ (che non compare nella tabella di fig. 1), osserviamo che questo sistema distingue due serie di ostruenti omorganiche, ovvero tre coppie di occlusive e quattro coppie di fricative. Tuttavia, come si evince dalla trascrizione fonetica con l'aggiunta di un segno diacritico ai simboli delle consonanti sonore (ad esempio [b̥]), tale opposizione non si realizza tramite la correlazione di sonorità, bensì attraverso un contrasto di 'forza' che si manifesta acusticamente più che altro come una differenza di durata (nel senso che le 'forti' durano di più delle 'leni'). Difatti, per i dialetti svizzero-tedeschi si assume tradizionalmente un'opposizione tra ostruenti 'forti' e 'leni' piuttosto che un contrasto basato sul tratto [±sonoro] (v. Fleischer, Schmid, 2006: 244-246; Nocchi, Schmid, 2006: 519-522; cfr. anche Schmid, 1999: 62). In termini articolatori e fonatori ciò implica che sia le 'forti' che le 'leni' vengono realizzate senza vibrazione alcuna delle pliche vocali.

Dal punto di vista fonologico va inoltre notato che secondo l'inventario presentato in fig. 1 la correlazione di 'tensione' non è pertinente per le affricate, dove per ognuno dei quattro luoghi di articolazione troviamo soltanto l'elemento non marcato dell'opposizione, ovvero il fonema 'forte' o [-teso]. Infine occorre precisare che nello svizzero-tedesco opera un processo di fortizione che trasforma le ostruenti 'leni' in 'semi-forti' qualora queste si trovino a formare un nesso con un'altra ostruente (Fleischer, Schmid, 2006: 248).

Passando ora alla lingua bersaglio dei nostri apprendenti, diamo un rapido sguardo all'inventario delle ostruenti dell'italiano standard (v. Bertinetto, Loporcaro, 2005: 132; cfr. anche Schmid, 1999: 133).

Figura 2 - *Inventario delle ostruenti dell'italiano standard*

	bilabiali	labiodentali	alveolari	palato-alveolari	velari
occlusive	p b		t d		k g
fricative		f v	s z	ʃ (ʒ)	
affricate			t͡s d͡z	t͡ʃ d͡ʒ	

Ciò che salta subito all'occhio in questo inventario è che il contrasto fonologico fra ostruenti omorganiche è basato sulla correlazione di sonorità, il che implica che le pliche vocali vibrino durante l'intera fase di articolazione delle ostruenti sonore. La classe delle fricative è meno numerosa in italiano rispetto allo svizzero-tedesco a causa dello statuto marginale di /ʒ/ e dell'assenza di fricative velari; mancano altresì le affricate labiali e velari. L'inventario di fig. 2 mostra inoltre che il sistema

fonologico dell'italiano sfrutta la correlazione di sonorità anche nelle caselle delle affricate alveolari e palato-alveolari (v. la presenza dei fonemi sonori /dz/ e /dʒ/), mentre nell'inventario dello svizzero-tedesco tale posizione strutturale resta vuota (cfr. fig. 1).

1.3 Ipotesi e domande di ricerca

In base alle considerazioni teoriche esposte in 1.1 e all'analisi contrastiva fornita in 1.2 possiamo quindi formulare le seguenti ipotesi e domande di ricerca circa la difficoltà che gli apprendenti svizzero-tedeschi incontrano nella pronuncia delle ostruenti dell'italiano:

- Ipotesi 1: Essendo le ostruenti 'leni' dello svizzero-tedesco simili alle ostruenti sonore italiane (di cui condividono il modo e il luogo di articolazione nonché la posizione strutturale nell'inventario fonematico, ma non la sonorità), si può presumere che gli apprendenti difficilmente riescano a realizzare le ostruenti sonore dell'italiano come tali, sostituendole invece con i foni più simili della loro L1, cioè le 'leni' (cfr. Flege, 1995; Best, Tyler, 2007).
- Ipotesi 2: Tale tendenza alla desonorizzazione delle sonore sarà più forte nel caso delle due affricate /dz/ e /dʒ/ dell'italiano, viste le rispettive lacune nell'inventario della L1; va notato però che dal punto di vista della *equivalence classification* (Flege, 1987) si tratta sempre di suoni 'simili', dato che la stessa combinazione dei due tratti di modo e luogo di articolazione esiste anche nella L1.
- Ipotesi 3: La tendenza a desonorizzare le ostruenti sonore dell'italiano sarà inoltre più forte nei gruppi consonantici formati da due ostruenti (del tipo /zb/), a causa della regola di fortizione dei dialetti svizzero-tedeschi e del carattere tipologicamente marcato di questo contesto fonotattico.

In aggiunta a queste tre ipotesi specifiche possiamo porci anche delle domande di ricerca più generali:

- Domanda di ricerca 1: Esistono altre differenze nel tasso di desonorizzazione delle ostruenti sonore – ad esempio riguardo al modo e al luogo di articolazione – che sono in linea con tendenze tipologiche (ricavabili in base alla frequenza delle consonanti nelle lingue del mondo)?
- Domanda di ricerca 2: Esistono delle differenze riguardo al tasso di desonorizzazione tra i singoli parlanti?

2. *Dati e metodo*

2.1 Il campione

Al fine di verificare le tre ipotesi specifiche e le due domande di ricerca più generali formulate nel paragrafo precedente è stato allestito un *corpus* di parlato letto prodotto da dieci studenti di un liceo di Zurigo (6 ragazze e 4 ragazzi con un'età media

di 17 anni). Al momento della registrazione, i ragazzi stavano studiando l'italiano da due anni e mezzo con tre ore di lezione alla settimana.

Tutti gli allievi usano nella comunicazione quotidiana prevalentemente il dialetto zurighese. Tuttavia, essendo la Svizzera un paese con un forte tasso di immigrazione, in molte famiglie si parla anche un'altra lingua a causa della presenza di almeno un genitore alloglotto. Fra i dieci studenti del nostro campione, solo nel caso di Joe entrambi i genitori sono svizzero-tedeschi (indichiamo i parlanti con le prime tre lettere del loro nome). Negli altri nove casi le lingue presenti in famiglia sono il cinese (Yit), il giapponese (Mik, And), il farsi (Dar), lo svedese (Eli), il ceco (Kri), due diverse varietà di tedesco della Germania (Sab) e dell'Austria (Van), e infine anche l'italiano nel caso dell'allievo San, i cui genitori provengono dalla provincia di Lecce e dal Canton Ticino.

Dal punto di vista della competenza linguistica, lo svizzero-tedesco può comunque essere considerato la varietà dominante nel repertorio linguistico di tutti i parlanti, ed è ragionevole ipotizzare che il sistema fonologico del dialetto zurighese intervenga nella loro percezione delle ostruenti italiane. Invece, vi è un solo soggetto con genitori italofoni per il quale ci si può aspettare una realizzazione delle ostruenti sonore più simile a quella di un parlante nativo.

2.2 Il *corpus*

La procedura adottata per la raccolta e l'analisi dei dati ha seguito sostanzialmente il modello di uno studio precedente sulla pronuncia del francese come L2 da parte di studenti liceali zurighesi (Schmid, 2012).

I dieci allievi sono stati registrati durante la lezione di italiano in una stanza a parte messa a disposizione dalla scuola. La registrazione è stata effettuata con un registratore digitale Fostex FR-2FLE e un microfono a cravatta Sennheiser MKE 2 P-C (direttività omnidirezionale, gamma di frequenza di 20-20.000 Hz ±23dB e coefficiente di trasmissione a vuoto di 10 mV/Pa ±2,5 dB). I file sonori sono stati salvati con una frequenza di campionamento di 44.100 Hz e una quantizzazione di 16 bit.

Durante la registrazione, ognuno dei dieci parlanti ha letto ad alta voce 19 frasi contenenti le sette ostruenti sonore dell'italiano /b d ɡ v z dz dʒ/ in sei contesti fonotattici diversi, dove le ostruenti bersaglio sono sempre seguite da vocale, mentre varia il contesto precedente:

- ##_V in posizione iniziale di enunciato;
- V#_V dopo vocale all'inizio di parola;
- C#_V dopo sonorante all'inizio di parola;
- V_V dopo vocale all'interno di parola (consonante scempia);
- V_:V dopo vocale all'interno di parola (consonante geminata);
- z_V dopo sibilante sonora.

La tabella in fig. 3 elenca le parole inserite nelle 19 frasi da leggere; le cifre prima delle parole elicitate si riferiscono alla numerazione delle frasi che vengono riprodotte in Appendice. Il *corpus* complessivo ammonta a 670 realizzazioni possibili di ostruenti sonore (10 parlanti x 67 consonanti), visto che alcune combinazioni di consonante + contesto occorrono più di una volta. Delle 670 occorrenze possibili in base allo schema riportato in fig. 3 abbiamo però dovuto scartare 26 per motivi contingenti, legati all'esecuzione del compito di lettura, per cui sono state analizzate 644 realizzazioni.

Figura 3 - *Parole elicitate secondo la consonante e il contesto*

	b	d	g	v	z	dz	dʒ
##_V	***##bV*** 1) **b**uongiorno	***##dV*** 2) **d**i 10) **d**ove	***##gV*** 3) **g**uardate 6) **G**abriele	***##vV*** 1) vorrei 14) **v**ado		***##dzV*** 18) **z**ero	***##dʒV*** 4) **g**ioca
V#_V	***V#bV*** 7) una **b**irra	***V#dV*** 1) gusto **d**i 2) di **d**ove 6) nome **d**ella 13) collega **d**i 18) a **d**ue	***V#gV*** 13) di **G**abriele	***V#vV*** 1) di **v**aniglia 11) gente **v**a		***V#dzV*** 14) allo **z**oo	***V#dʒV*** 3) è **g**ià 11) la **g**ente
C#_V	***C#bV*** 7) al **b**ar 9) il **b**ambino 12) il **b**ambino	***C#dV*** 15) il **d**ialogo 16) un **d**ubbio 19) il **d**isgelo	***C#gV*** 1) al **g**usto	***C#vV*** 4) il **V**erona 17) in **V**altellina		***C#dzV*** 13) in **z**ona	***C#dʒV*** 1) un **g**elato
V_V	***VbV*** 10) a**b**itate 13) a**b**ita	***VdV*** 9) sgri**d**a 11) sta**d**io 14) va**d**o 16) enciclope**d**ia	***VgV*** 13) ccl**g**a 15) dia**g**ogo	***VvV*** 2) do**v**e 10) do**v**e 19) primave**r**a	***VzV*** 6) ro**s**a		***VdʒV*** 3) mezzo**g**iorno 14) cu**g**ino
V_:V	***Vb:V*** 16) a**bb**iamo 16) du**bb**io	***Vd:V*** 12) contra**dd**ice 17) A**dd**a	***Vg:V*** 15) le**gg**o	***Vv:V*** 5) a**vv**ersari		***V#d:zV*** 3) mezzogiomo 8) realizzare	***Vd:ʒV*** 4) o**gg**i 6) le**gg**e 16) le**gg**iamo
z_V	***zb*** 8) **s**bagliato	***zd*** 15) **s**drajo	***zg*** 9) **s**grida	***zv*** 2) **s**vizzero			***zdʒ*** 19) di**s**gelo

2.3 Procedura di analisi

L'analisi acustica delle registrazioni sonore è stata effettuata mediante il programma *Praat* (Boersma, Weenink, 2015). In un primo passo sono state segmentate e annotate su un *TextGrid* le realizzazioni delle ostruenti in questione, per calcolarne successivamente la percentuale di desonorizzazione (ovvero la percentuale di segnale non periodico in quella sezione temporale del segnale); in *Praat*, la funzione *fraction of locally unvoiced frames* viene fornita dal *voice report* che si trova nella finestra *edit* sotto il menu *pulses*.

L'analisi statistica dei dati è stata effettuata mediante i programmi *Microsoft Excel* e SPSS.

3. *Risultati*

3.1 Il tasso globale di desonorizzazione

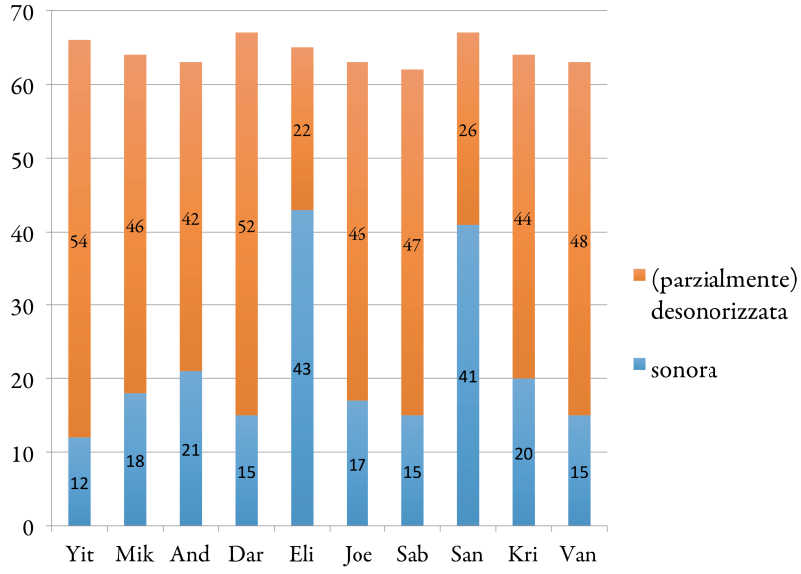
Delle 644 occorrenze effettivamente prese in considerazione, 217 sono state realizzate come pienamente sonore, mentre solo 12 sono state completamente desonorizzate. Invece, in ben 415 realizzazioni (ovvero il 64.4% dei casi) abbiamo incontrato delle soluzioni 'intermedie', con un tasso di desonorizzazione che va dall'11% fino al 97%. Considerando il numero esiguo delle ostruenti sonore completamente desonorizzate, d'ora in poi le considereremo assieme alle realizzazioni parzialmente desonorizzate (ottenendo quindi solo due categorie: 'sonore' e 'più o meno desonorizzate'). In effetti, se un allievo produce un fono [d] desonorizzato all'86%, possiamo dedurre che nella fonologia dell'interlingua tale fono corrisponda a un fonema /d̥/ e che la breve fase periodica all'inizio della fase di occlusione corrisponda – almeno nel caso che l'occlusiva venga preceduta da una vocale – a un fenomeno di coarticolazione, cioè alla cosiddetta *VoiceOFFsetTime*.

Come primo risultato globale possiamo comunque constatare che la realizzazione di ostruenti completamente sonore crea non pochi problemi agli apprendenti svizzero-tedeschi, dato che in quasi due terzi dei casi essi non riescono a produrle come tali. I dati mostrano però anche una notevole variazione secondo i fattori 'parlante', 'consonante' e 'contesto' che analizzeremo separatamente qui di seguito.

3.2 Il fattore 'parlante'

Cerchiamo innanzitutto di rispondere alla domanda di ricerca 2 formulata in 1.3: esistono delle differenze riguardo al tasso di desonorizzazione tra i singoli parlanti?

Figura 4 - *Numero delle realizzazioni di ostruenti sonore secondo i parlanti*



La fig. 4 mostra il grado di desonorizzazione delle ostruenti sonore per ognuno dei dieci parlanti, indicando il numero di realizzazioni nelle due categorie 'ostruente sonora' e 'ostruente (parzialmente o totalmente) desonorizzata'.

Tutto sommato possiamo dire che il quadro è piuttosto omogeneo, nel senso che in otto parlanti predominano le realizzazioni desonorizzate. Fanno però eccezione i due parlanti Eli e San, in cui le ostruenti completamente sonore (43 e 41 realizzazioni rispettivamente) superano quelle (parzialmente) desonorizzate (22 e 26 realizzazioni rispettivamente).

Per quanto riguarda la durata della fase aperiodica nelle singole realizzazioni desonorizzate (quelle all'interno della porzione arancione delle barre di fig. 4), la gamma delle percentuali varia notevolmente per tutti i parlanti; ad esempio, la parlante Van produce da un lato una [dz] desonorizzata al 92% e dall'altro una [b] desonorizzata al 14%. Considerando gli otto parlanti con un comportamento relativamente omogeneo (tutti meno Eli e San), la percentuale media di desonorizzazione calcolata per la durata delle ostruenti parzialmente sonore è del 47%, con una gamma che va dal 37% nel parlante And al 57% nel caso della parlante Yit e una deviazione standard del 7%.

Da questo quadro generale si distanziano però di nuovo i parlanti Eli e San che presentano un grado medio di desonorizzazione del solo 18%. Tenendo conto dei repertori linguistici parlati in famiglia (cfr. 2.1), possiamo forse spiegare il comportamento di San con l'italofonia dei suoi genitori (in un certo senso questo parlante può essere considerato bilingue). Nella famiglia di Eli si parla anche lo svedese, una lingua nella quale esistono ostruenti sonore a tutti gli effetti (cfr. Engstrand, 1999).

Considerando ora soltanto gli altri otto parlanti, troviamo comunque una certa correlazione tra il numero delle realizzazioni desonorizzate prodotte dai singoli parlanti e le durate della fase aperiodica all'interno di questi foni ($r = 0.76$, $R^2 = 0.57$): in altre parole, nei parlanti che pronunciano un numero più alto di ostruenti (parzialmente) desonorizzate, la fase aperiodica all'interno delle realizzazioni (parzialmente) desonorizzate è mediamente anche più lunga. Per l'analisi statistica delle realizzazioni desonorizzate abbiamo però dovuto rinunciare all'applicazione di un test parametrico (in questo caso di un'ANOVA univariata) a causa dei risultati di due test preliminari. Infatti, secondo il test di Shapiro-Wilk i valori sono distribuiti normalmente solo per 5 parlanti su dieci (Mik, And, Eli, Sab e Kri). Inoltre, in base al test di Levene non vi è uguaglianza della varianza tra i dieci parlanti ($p<.001$). Comunque, il test (non-parametrico) di Kruskal-Wallis indica che tra i dieci parlanti esistono delle differenze significative riguardo al grado di desonorizzazione delle ostruenti sonore ($\chi^2(9)=93{,}584$, $p<.001$).

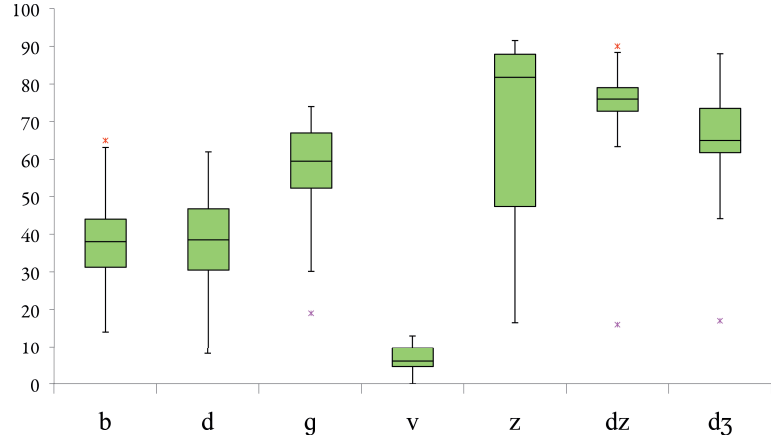
3.3 Il fattore 'consonante'

Nemmeno nel caso del fattore 'consonante' i test preliminari hanno permesso di applicare un test parametrico. Anche se consideriamo unicamente le realizzazioni (parzialmente) desonorizzate (cfr. fig. 4), il test di Shapiro-Wilk mostra una distribuzione normale soltanto per le tre ostruenti /v/, /ɡ/ e /dz/, ma non per /b/, /d/,

/z/ e /dʒ/. Inoltre, secondo il test di Levene le percentuali di desonorizzazioni non hanno la stessa varianza nelle sette ostruenti. Ricorrendo invece al test non parametrico di Kruskal-Wallis, si ottiene senz'altro un effetto significativo del fattore 'consonante' sulle percentuali di desonorizzazione ($\chi^2(6)=266{,}048$, $p<.001$).

Tale effetto è visibile anche a livello descrittivo nella fig. 5, i cui boxplot illustrano – per le sette ostruenti sonore prodotte da tutti e dieci i parlanti – la dispersione delle percentuali di desonorizzazione da 0% a 100% (comprese quindi le realizzazioni completamente sonore e sorde). I boxplot sono in effetti interpretabili in base alle ipotesi formulate nel paragrafo 1.3. Iniziando le nostre considerazioni con le tre occlusive /b d g/, notiamo ad esempio un tasso di desonorizzazione molto simile per le bilabiali e le alveolari (con una media del 37%), mentre le velari sono colpite più frequentemente dalla desonorizzazione (con una media del 56%).

Figura 5 - *Percentuale media della desonorizzazione secondo il fattore 'consonante'*

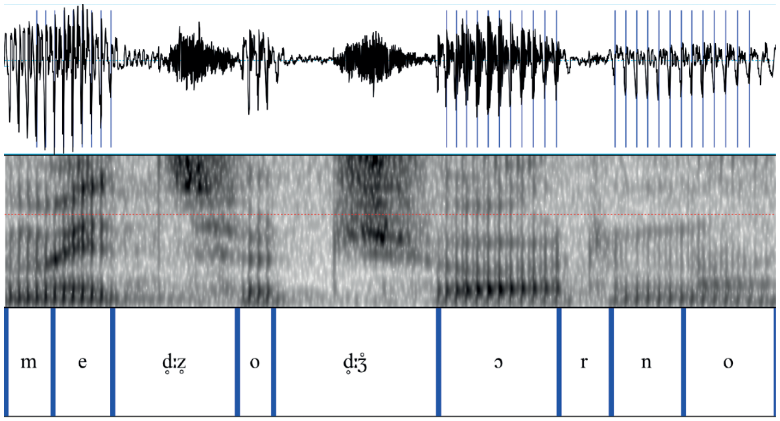


La maggiore desonorizzazione delle velari non può essere determinata dalla struttura della L1, per cui è più promettente cercarne la causa – sulla scia della più generale domanda di ricerca 2 – su eventuali tendenze universali nella distribuzione della consonante /g/. Ricordiamo, tra parentesi, che già Jakobson (1941 [1971]: 89-90) aveva postulato la precedenza delle consonanti labiali e dentali nell'acquisizione del linguaggio nei bambini. Ebbene, dagli studi di tipologia fonologica emerge che tra le occlusive sonore (di per sé marcate rispetto alle sorde) è più probabile che in un inventario fonematico manchi /g/ che non /b/ e /*d/ (Maddieson, 1984: 36). Questo risultato viene confermato da una consultazione della banca dati PHOIBLE allestita recentemente (Moran et al., 2014): le occlusive labiali (/b/) e coronali (/d/ e /d̪/) sono presenti rispettivamente nel 71% e 72% delle 2155 lingue documentate, laddove /g/ si trova solo nel 64% dei casi. Una possibile causa aerodinamica di questa differenza viene indicata da Ohala (1992: 8) nella difficoltà di mantenere la fonazione con un condotto vocale bloccato e piuttosto ristretto come nel caso del luogo di articolazione velare.

Tornando alla fig. 5 e volgendo lo sguardo verso il lato destro del grafico si nota un tasso di desonorizzazione molto elevato per i due fonemi /dz/ e /dʒ/. Per illu-

strare tale fenomeno riportiamo nella fig. 6 l'oscillogramma e lo spettrogramma della parola *mezzogiorno* pronunciata dal parlante Dar: la desonorizzazione delle due affricate viene evidenziata anche dall'assenza di strisce verticali blu che caratterizza le parti aperiodiche nel segnale acustico.

Figura 6 - *Oscillogramma e spettrogramma della parola* mezzogiorno *pronunciata dal parlante Dar*



Constatiamo a questo punto che la forte tendenza alla desonorizzazione delle due affricate sonore è conforme all'ipotesi 2 (v. 1.3). Oltre alla mancanza di questa posizione strutturale nell'inventario fonematico della L1 (ovvero all'assenza di affricate leni nello svizzero-tedesco), la difficoltà di pronunciare un'affricata sonora sembra però essere dovuta anche a una complessità articolatoria intrinseca di questo tipo di consonante. Sulla scia di Jakobson (1941 [1971]: 56, 52) possiamo infatti assumere che il modo di articolazione delle affricate sia marcato rispetto a quello delle fricative, le quali sarebbero a loro volta marcate rispetto alle occlusive.

Guardando infine la parte centrale della fig. 5, troviamo un comportamento molto diverso tra le due fricative, con soltanto 6% di desonorizzazioni della labiodentale rispetto a 68% nel caso della sibilante. Questo scarto non trova una spiegazione in considerazioni tipologiche (dato che /z/ e /v/ compaiono rispettivamente nel 31% e 29% delle lingue documentate nella banca dati PHOIBLE), ma sarà da ricondurre piuttosto all'azione della prima lingua degli apprendenti. Per quanto riguarda /v/, la facilità con cui essi riescono a produrre un suono labiodentale sonoro si spiega in effetti con la presenza dell'approssimante /ʋ/ nei dialetti svizzero-tedeschi (cfr. Fleischer, Schmid, 2006: 244; Nocchi, Schmid, 2006: 522-523). Invece, la difficoltà di /z/ non deriva tanto da proprietà 'paradigmatiche' di questa consonante quanto da un processo 'sintagmatico' che vedremo nel prossimo paragrafo.

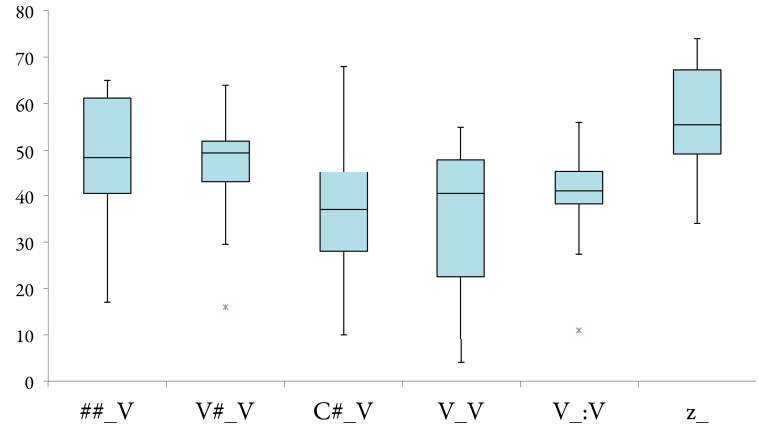
3.4 Il fattore 'contesto'

Anche per l'ultimo dei tre fattori presi in considerazione – il contesto fono(sin)tattico – abbiamo dovuto escludere l'applicazione di un test parametrico in base

ai risultati dei due test preliminari. Infatti, secondo il test di Shapiro-Wilk i valori sono distribuiti normalmente soltanto in due contesti, cioè all'inizio dell'enunciato (contesto ##_V) e all'inizio di parola dopo sonorante (contesto C#_V), ma non negli altri quattro contesti (V#_V, V_V, V_:V, z_); inoltre, il test di Levene non ha fornito una varianza uguale delle percentuali di desonorizzazione. Di conseguenza abbiamo optato anche in questo caso per un test non parametrico, il Kruskal-Wallis per l'appunto, che indica comunque un effetto significativo anche per il fattore 'contesto' ($\chi^2(5)=76{,}027$, $p<.001$).

Passando quindi alla statistica descrittiva, vediamo in fig. 7 dei boxplot che illustrano il diverso grado di desonorizzazione delle ostruenti sonore secondo la loro posizione nella parola e nell'enunciato, distinguendo i sei contesti fonotattici descritti in 2.2 per i dati dell'intero *corpus*:

Figura 7 - *Percentuale di desonorizzazione secondo il fattore 'contesto fonotattico'*



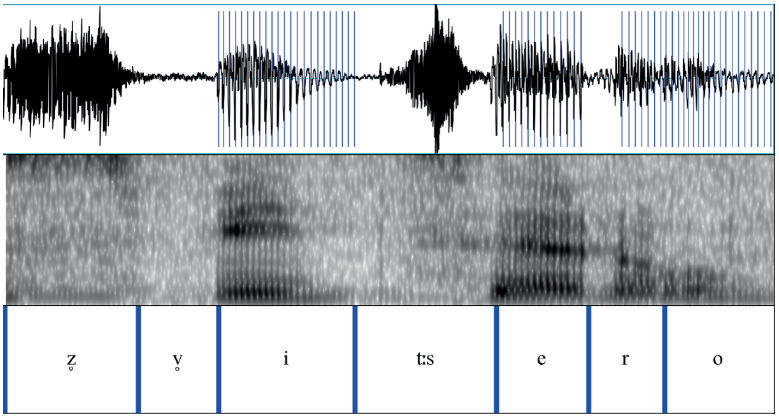
Dai boxplot si evince che la desonorizzazione è più contenuta nel contesto intervocalico all'interno di parola, il che viene confermato dalle medie che ammontano al 35% per le scempie (V_V) e al 38% per le geminate (V_:V). Questo risultato è in linea con le predizioni della *Markedness Differential Hypothesis* e può essere spiegato con la relativa facilità fisiologica di mantenere la fonazione quando l'ostruente è preceduta e seguita da una vocale. Un simile tasso di desonorizzazione, con una media del 38%, si trova al confine iniziale della parola quando precede una sonorante (C#_V), mentre sono più alti i valori medi dopo vocale (V#_V) e quando la consonante si trova in posizione iniziale di enunciato (##_V): le rispettive medie ammontano al 45% e al 48%. Il valore leggermente più alto risulta alquanto plausibile per la posizione iniziale di enunciato, ma non tanto per la posizione iniziale di parola.

Risulta invece chiaro che la desonorizzazione compare con maggiore incidenza quando l'ostruente sonora è preceduta da una sibilante (z_). In questo contesto troviamo una media di desonorizzazione del 57% che va ricollegata al 68% riscontrato per /z/ nel paragrafo precedente. È evidente come nel nesso /z/ + ostruente sonora

operi il processo di fortizione dello svizzero-tedesco descritto in 2.1, che favorisce notevolmente la pronuncia sorda di ambedue le consonanti nella L2.

Per illustrare questo fenomeno riportiamo infine l'oscillogramma e lo spettrogramma della parola *svizzero* pronunciata dalla parlante Mik:

Figura 8 - *Oscillogramma e spettrogramma della parola* svizzero *pronunciata dalla parlante Mik*



Anche in questo caso l'assenza di strisce verticali rivela la totale aperiodicità del segnale acustico durante l'articolazione dei primi due segmenti della parola, rivelando la completa desonorizzazione del nesso /zv/.

4. *Discussione e conclusioni*

Avviandoci a qualche riflessione interpretativa sui risultati appena esposti, ci pare opportuno addurre prima alcuni elementi di confronto con lo studio precedente sulla pronuncia delle ostruenti francesi da parte di un'altra classe di studenti liceali zurighesi (cfr. Schmid, 2012). Innanzitutto va rilevata una differenza sostanziale riguardo alla distribuzione delle tre categorie nel *corpus* di francese L2: su 340 occorrenze, 106 (31.2%) sono completamente sonore e 62 (18.2%) parzialmente desonorizzate, mentre ben 172 (50.6%, in pratica la metà) sono completamente desonorizzate. Invece, delle 644 occorrenze analizzate nel nostro corpus di italiano L2, 217 (33.7%) vengono realizzate come sonore e 415 (64.4%) come parzialmente desonorizzate, mentre vi sono soltanto 12 occorrenze (1.9%) completamente sorde (cfr. 3.1). Tale differenza è in parte dovuta al fatto che la lingua francese ammette ostruenti sonore anche in posizione finale di parola e quindi in un contesto particolarmente suscettibile di innescare il processo di desonorizzazione (cfr. Eckman, 2008). Tuttavia, nel *corpus* di francese L2 le realizzazioni desonorizzate occorrono anche in altri contesti fonotattici (cfr. Schmid, 2012: 645), per cui possiamo affermare che il campione analizzato nel presente studio riesce ad avvicinarsi di più alla struttura sonora della lingua bersaglio.

È difficile spiegare questa differenza nel comportamento delle due classi liceali (sul quale possono influire molteplici fattori quali la predisposizione all'acquisizione di una L2 dei singoli allievi, la qualità dell'insegnamento, ecc.), ma si può forse considerare il fatto che il francese rappresenta una materia scolastica obbligatoria (spesso non molto amata), laddove l'italiano è una lingua opzionale per la quale gli allievi potrebbero essere più motivati. È vero che nel nostro *corpus* quasi due terzi delle ostruenti sono almeno in parte desonorizzate, ma forse vale la pena insistere sia sul carattere parziale di questa desonorizzazione sia sulla presenza – in tutti i locutori – di ostruenti completamente sonore. È proprio questa forte variabilità riscontrata nei dati a portarci a una visione non tanto drammatica del livello raggiunto dagli allievi nella pronuncia dell'italiano. In base al carattere meno stabile delle interlingue in quanto sistemi in via di sviluppo, la produzione in una lingua seconda presenta spesso una maggiore variabilità rispetto a quella nella lingua nativa; in altre parole, si tratta di un tipo di variabilità che denota un potenziale di avvicinamento alla norma della lingua bersaglio.

Ciononostante non possiamo negare che la pronuncia delle ostruenti sonore costituisce una difficoltà notevole per gli apprendenti zurighesi; forse si tratta addirittura di una caratteristica saliente di ciò che potremmo definire come 'accento svizzero' nella pronuncia dell'italiano.

Tornando quindi al punto di partenza delle nostre riflessioni, i risultati di questo studio permettono di confermare in sostanza le tre ipotesi che avevamo formulato sulla base dell'analisi contrastiva tra L1 e L2 nonché sulla base dei postulati dei principali modelli teorici della fonologia delle interlingue:

- Ipotesi 1: In base alla somiglianza percettiva e all'equivalenza strutturale con le consonanti 'leni' della L1, le ostruenti sonore dell'italiano vengono in buona parte 'assimilate' alla rispettiva categoria dello svizzero-tedesco. Questi risultati sono in linea con le predizioni dell'SLM e del PAM-L2.
- Ipotesi 2: Tra le sette ostruenti sonore dell'italiano, le due affricate sono maggiormente soggette alla desonorizzazione, date l'assenza di un equivalente strutturale nella L1 e la complessità intrinseca di questi foni.
- Ipotesi 3: Un tasso di desonorizzazione più alto si trova anche nei nessi di sibilante + ostruente sonora come riflesso di un processo fonotattico della L1.

Per quanto riguarda la seconda domanda di ricerca più generale – le differenze tra i parlanti – abbiamo constatato un numero più alto di realizzazioni sonore presso due allievi, di cui uno è di origine italiana. Ai fini del nostro studio è però più rilevante la prima domanda di ricerca: l'eventuale coincidenza tra alcuni risultati nel *corpus* con tendenze note dalla tipologia fonologica. Possiamo interpretare in questo senso da un lato la minore desonorizzazione delle ostruenti intervocaliche e dall'altro la maggiore desonorizzazione delle occlusive velari rispetto agli altri luoghi di articolazione. Concludiamo quindi che il fenomeno della desonorizzazione nella pronuncia delle ostruenti sonore dell'italiano è dovuto sia all'influsso del sistema fonologico dello svizzero-tedesco sia a principi generali della struttura sonora del linguaggio.

Ringraziamenti

Siamo grati a Letizia Könz, insegnante di italiano nel *Realgymnasium Rämibühl* di Zurigo, nonché agli allievi della classe 5i per la loro partecipazione a questa ricerca. Grazie a Sandra Schwab per la preziosa consulenza statistica. Ringraziamo infine tre revisori anonimi per le osservazioni critiche molto dettagliate sulla prima versione del testo; ovviamente, i due autori rimangono gli unici responsabili di eventuali errori o imprecisioni.

Bibliografia

Avesani, C., Vayra, M., Best, C. & Bohn, O. (2008). Fonologia e acquisizione. In che modo l'esperienza della lingua materna plasma la percezione dei suoni del linguaggio. In Costamagna, L., Marotta, G. (Eds.), 15-41.

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. (1995). A direct realist view of cross-language speech perception. In Strange, W. (Ed.), 171-204.

Best, C., Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O., Munro, M. (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege.* Amsterdam: Benjamins, 13-33.

Boersma, P., Weenink, D. (2015). *Praat: doing phonetics by computer* [Computer program]. Version 5.4.09, retrieved 1 June 2015 from http://www.praat.org/.

Costamagna, L., Marotta, G. (Eds.) (2008). *Processi fonetici e categorie fonologiche nell'acquisizione dell'italiano*. Pisa: Pacini.

Eckman, F. (1977). Markedness and the Contrastive Analysis Hypothesis. In *Language Learning*, 27, 315-330.

Eckman, F. (2008). Typological markedness and second language phonology. In Hansen Edwards, J., Zampini, M. (Eds.), *Phonology and second language acquisition.* Amsterdam: Benjamins, 95-115.

Engstrand, O. (1999). Swedish. In *Handbook of the International Phonetic Association.* Cambridge: Cambridge University Press, 140-142.

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.), 233-277.

Fleischer, J., Schmid, S. (2006). Zurich German. In *Journal of the International Phonetic Association*, 36 (2), 243-255.

Jakobson, R. (1941). Kindersprache, Aphasie und allgemeine Lautgesetze. In *Uppsala Universitets Årsskrift*, 1-83 [trad. it. 1971, *Il farsi e il disfarsi del linguaggio. Linguaggio infantile e afasia.* Torino: Einaudi].

Maddieson, J. (1984). *Patterns of sound*. Cambridge: Cambridge University Press.

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

Moran, S., McCloy, D. & Wright, R. (2014). *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology (Available online at http://phoible.org, Accessed 4.06.15).

Nocchi, N., Schmid, S. (2006). Le consonanti labiodentali dello svizzero tedesco. In Savy, R., Crocco, C. (Eds.), *Analisi prosodica. Teorie, modelli e sistemi di annotazione* (Atti del 2° Convegno Nazionale AISV. Salerno, 30 Novembre-2 Dicembre 2005). Torriana: EDK Editore, 518-537.

Ohala, J. (1992). What's cognitive, what's not, in sound change. In Kellermann, G., Morrissey, M. (Eds.), *Diachrony within synchrony: language history and cognition*. Frankurt a.M.: Peter Lang, 309-355.

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 *Poznań Studies in Contemporary Linguistics*, 48 (4), 627-659.

Strange, W. (Ed.) (1995). *Speech perception and linguistic experience. Issues in cross-language research*. Timonium, MD: York Press.

Trubetzkoy, N. (1939). *Grundzüge der Phonologie.* Göttingen, Vandenhoeck & Ruprecht [trad. it. 1971, *Fondamenti di fonologia.* Torino: Einaudi].

Appendice

Le 19 frasi del questionario:

1. Buongiorno. Vorrei un gelato al gusto di vaniglia.
2. Di dove sei? – Sono svizzero.
3. Guardate, è già mezzogiorno.
4. Gioca il Verona oggi?
5. Gli avversari si abbracciano.
6. Gabriele legge il libro "*Il nome della rosa*".
7. Prendiamo una birra al bar.
8. È sbagliato non realizzare i suoi sogni.
9. La mamma sgrida il bambino.
10. Dove abitate?
11. La gente va allo stadio.
12. Il bambino contraddice il padre.
13. Il collega di Gabriele abita in zona.
14. Vado allo zoo con mio cugino.
15. Mi sdraio al sole e leggo il dialogo.
16. Se abbiamo un dubbio leggiamo l'enciclopedia.
17. L'Adda scorre in Valtellina.
18. Zero a due, questo è il risultato.
19. A primavera c'è il disgelo.

DOI: 10.17469/O2101AISV000014

YILUN TANG, BARBARA GILI FIVELA

Il cinese come L2: sulla percezione di alcune vocali da parte di studenti italiani poliglotti

The paper discusses whether L2 vowel-perception only depends on the L1 system and the exposure to the L2 system or it also correlates with the knowledge of L3 languages showing some phonetic-phonological characteristics that are similar to the L2 ones. Focus is on Italians learning Chinese, where phonemes such as /y/ and /ɤ/ are found, considering that some Italians we interviewed also know French or German, where /y/ is present and /ɤ/ is not. Results show that 1) Italians tend to identify Chinese stimuli as their L1 phonemes, with higher goodness-ratings for the most similar phones, 2) identification scores are correlated with the level of exposure to L2, though correlation/determination indexes are quite low, and 3) the influence of strong exposure to phonemes also found in L3 only partially emerges.

La letteratura scientifica relativa all'influenza esercitata dal sistema fonetico-fonologico della lingua materna (L1) sulla percezione dei suoni linguistici non nativi o dei suoni di una lingua di apprendimento (L2) è piuttosto ricca. Diversi lavori evidenziano come il sistema fonologico della lingua materna influenzi la percezione dei suoni linguistici già a partire da un anno di età (Kuhl, 1991) e rappresenti una sorta di filtro che condiziona l'identificazione e la discriminazione dei suoni non nativi. In particolare, i suoni non nativi, e in generale quelli di una L2, tendono ad essere assimilati a categorie native (e identificati come appartenenti ad esse) e, nel caso questo effettivamente avvenga, possono non essere percepiti come sufficientemente nuovi o diversi da richiedere la formazione di nuove categorie (Best, 1994; Flege, 1995). In diversi modelli, si ipotizza che questo sia il meccanismo alla base della possibilità di percepire differenze tra suoni nativi e non nativi, o L2 (ossia di discriminarli), e al quale si deve l'effettiva capacità di formare nuove categorie fonologiche nel corso dell'apprendimento di una seconda lingua (Best, Tayler, 2007). Anche l'influenza della lingua materna sulla produzione del parlato in altra lingua è stata oggetto d'indagine e discussione, ma i modelli proposti per rendere conto della produzione dei suoni linguistici non nativi, o dei suoni L2, sono meno numerosi e meno articolati. In questo senso, il modello attualmente più sviluppato è ancora quello di Flege, in base al quale difficilmente la produzione in L2 risulterà adeguata. Il modello infatti predice che l'assimilazione percettiva di un suono a una categoria nativa provochi un target undershoot in produzione, mentre la percezione delle differenze provochi addirittura un target overshoot (Flege, 1995).

Tuttavia, l'attenzione degli studi del settore è spesso incentrata sull'interferenza della lingua materna, mentre è noto che anche altre lingue conosciute dall'individuo

possono influenzare in qualche modo e in qualche misura le sue capacità di percezione e produzione. Non ci risulta, infatti, che la complessa influenza esercitata dalla varietà di codici linguistici noti al soggetto preso in esame sia spesso considerata esplicitamente negli studi sperimentali del settore (v. anche Cabrelli Amaro 2012; ma si veda lo *Special Issue on L3 Phonology*, pubblicato nel 2010 *sull'International Journal of Multilingualism*).

In questo lavoro, ci proponiamo quindi di mettere in relazione il livello di esposizione alla fonetica-fonologia della L2 (come indicativo della competenza) con le caratteristiche della lingua madre e con le caratteristiche di altre lingue conosciute dai soggetti sperimentali. In questo modo, intendiamo non solo verificare se l'influenza della L1 cambi a seconda del livello di esposizione alla L2, ma anche individuare interferenze molto specifiche dovute alla conoscenza di altre lingue il cui sistema fonetico-fonologico presenti elementi di somiglianza con quello della L2 oggetto di indagine (ad esempio, uno o più fonemi condivisi). In particolare, prenderemo in esame dati relativi ad italiani che stanno effettuando un percorso di apprendimento della lingua cinese e che in alcuni casi conoscano già altre lingue, come il francese o il tedesco, nelle quali sono usati fonemi e foni che esistono anche in cinese e che non si trovano in italiano. Nel nostro studio, invece di fare riferimento alla competenza dei soggetti sperimentali, abbiamo utilizzato un indice di esposizione alla lingua, individuato sulla base del numero di anni di studio e sul numero e la durata di soggiorni all'estero. Reputiamo, infatti, di non avere a disposizione un metodo ottimale per quantificare il livello di competenza in L2 dei nostri soggetti (e tantomeno quello di competenza fonetico-fonologica), per poi effettuare la correlazione con le loro capacità percettive (v. anche il paragrafo 2.4). Preferiamo quindi utilizzare un indice di esposizione alla/e lingua/e (e ai suoi/loro fenomeni fonetico-fonologici), che non rappresenta una quantificazione necessariamente significativa e che quindi non è ancora ottimale (perché anche a parità di anni di studio è possibile individuare livelli di esposizione diversi), ma che ci sembra adeguato in quanto misurabile in modo più preciso rispetto ad un generico livello di competenza in L2, che comunque difficilmente riguarderebbe solo le conoscenze fonetico-fonologiche.

1. *Il vocalismo del cinese (mandarino) e dell'italiano (standard e salentino)*

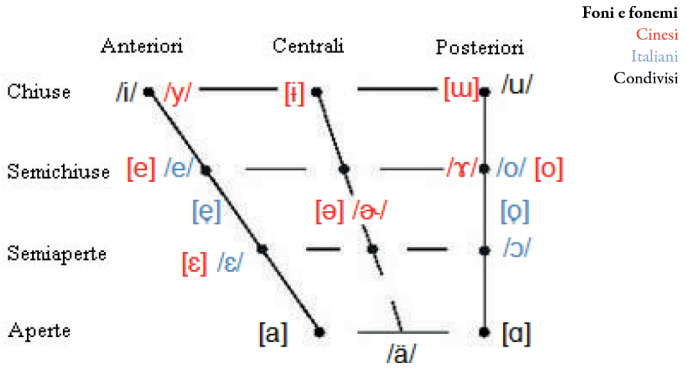
Il sistema fonologico del cinese include 6 fonemi vocalici, ossia /i y a ɚ ɤ u/ (Pan, 2010; Zhang, 2004), mentre quello dell'italiano include 7 fonemi vocalici /i e ɛ a ɔ o u/ – Figura 1, ridotti a 5 in molte varietà (Schmid, 1999). Ad esempio, il sistema fonologico della varietà parlata dai soggetti sperimentali di questo studio è composto da 5 vocali, senza distinzione tra le vocali medie semiaperte e semichiuse, sia sul piano anteriore che su quello posteriore (Grimaldi, 2003).

Apparentemente i fonemi vocalici cinesi non hanno molto in comune con quelli italiani, ma la situazione cambia se si considerano le varianti allofoniche. Ad esempio, i fonemi cinesi /ɚ/ e /ɤ/, che non esistono in italiano, hanno allofoni simili a quelli italiani. Il fonema /ɚ/ corrisponde agli allofoni [e, ɛ, ə], con [ɛ] che segue solo

le approssimanti [j] e [ɥ] (allofoni, rispettivamente, di /i/ e /y/) e forma un dittongo ascendente, [e] che precede la vocale /i/ formando un dittongo discendente, e [ə] che è seguito solo dalle consonanti nasali [n] e [ŋ] (con [ŋ] allofono di /n/) e forma una finale[1] nasalizzata; inoltre, /ə̹/ corrisponde al fono vocalico medio centrale [ə̹] che rappresenta l'intera finale. Esempio anche più rilevante ai fini di questo studio è /ɤ/, che ha solo un allofono, [o], il quale segue l'approssimante [w] o precede la vocale /u/ (di cui [w] è un allofono), facendo parte di un dittongo ascendente o discendente; in tutte le altre situazioni /ɤ/ corrisponde a [ɤ], che rappresenta la finale; alla stessa altezza di /ɤ/ si trova anche il fonema italiano /e/.

Altro fonema cinese degno di interesse è /y/. Si tratta di un fonema che non ha diversi allofoni e che si trova alla stessa altezza di /i/ e /u/, condivisi dalle due lingue e che si differenziano da /y/ solo per il tratto di arrontondamento o per quello di anteriorità. Peraltro, /y/ è uno dei fonemi cinesi che si trovano in lingue diverse dall'italiano, ma talvolta note agli italiani che apprendono il cinese. In particolare, /y/ compare sia in francese che in tedesco (in cui è una vocale lunga).

Figura 1 - *Sistema vocalico del cinese (mandarino, in rosso), dell'italiano (standard, in blu) e foni/fonemi condivisi tra i sistemi (in nero) – (Tang, 2014)*



I due sistemi che prendiamo in esame in questo studio, quindi, condividono alcuni fonemi vocalici e allofoni (oltre a quelli la cui descrizione esula dagli obiettivi di questo studio). In particolare, il sistema della varietà di italiano parlata dai soggetti qui esaminati è composto da cinque vocali, come già ricordato, per cui l'unica differenza di rilievo rispetto a quanto appena descritto e riportato in Figura 1 riguarda il fatto che manca l'opposizione tra le vocali medie, che sono solo /e/ e /o/ (Grimaldi, 2003).

Di particolare interesse per lo studio della percezione dei suoni della L2 da parte di italofoni, come evidenziato da Tang (2014) sulla base di osservazioni prevalentemente impressionistiche, sono i fonemi cinesi /y/ e /ɤ/, in quanto gli studenti

[1] La sillaba in cinese si analizza come composta di iniziale, finale e tono: l'iniziale è costituita dalla consonante che introduce la sillaba, mentre la finale è la parte della sillaba che segue l'iniziale, escludendo il tono (Masini et al., 2011: VIII).

italiani tendono a produrre il primo come [i], [u] oppure [ju], e il secondo come [i], [u], [o], [e]. In entrambi i casi, sembrano quindi essere le caratteristiche di labializzazione e antero-posteriorità a creare confusione negli apprendenti.

2. *Il test percettivo*

2.1 Obiettivi e ipotesi

L'obiettivo del presente contributo è studiare la classificazione di vocali cinesi da parte di studenti italofoni poliglotti che siano stati esposti in misura diversa alla lingua cinese, ipotizzando quindi che abbiano anche raggiunto livelli di competenza diversi. In particolare, si intende:

- verificare se l'influenza del sistema italiano cambi a seconda del livello di esposizione alla lingua e al sistema fonetico-fonologico del cinese;
- individuare interferenze dovute alla conoscenza di altre lingue il cui sistema fonetico-fonologico presenti elementi di somiglianza con quello del cinese (ad esempio, uno o più fonemi condivisi).

L'ipotesi di lavoro è che il livello di esposizione complessivo alla L2 (in particolare, alla sua fonetica-fonologia, e, di fatto, anche il livello di competenza che si assume sia stato raggiunto in L2) non sia necessariamente rappresentativo delle effettive capacità di classificazione/riproduzione di un fonema specifico. Ipotizziamo che all'interno del sistema, infatti, si possano individuare differenze dovute a conoscenze specifiche, come quelle derivate dall'esposizione a determinati foni, presenti in altre lingue conosciute. Come anticipato, più che a un ipotetico livello di competenza generale sviluppata dal soggetto, faremo riferimento all'entità dell'esposizione alla lingua, quantificata indicativamente sulla base del numero di anni di studio e della durata di eventuali soggiorni all'estero.

In particolare, prenderemo in esame la percezione di /y/ e /ɤ/, fonemi cinesi che gli italiani tendono a produrre come [i], [u] oppure [ju], e [i], [u], [o], [e] e a percepire, rispettivamente, come [i], [u] e [o], [e] (Tang, 2014). Di questi due fonemi, /y/ è presente anche in francese e in tedesco (benché in tedesco sia un fonema lungo, Maturi, 2009). Intendiamo quindi verificare se gli apprendenti italiani, che siano stati diversamente esposti alla lingua cinese e che in alcuni casi conoscano anche il francese e/o il tedesco, percepiscano diversamente /y/ e /ɤ/, manifestando una differente influenza delle categorie native dell'italiano sulla percezione degli esempi di queste due vocali.

2.2 Metodo

L'esperimento descritto in questo articolo è stato progettato facendo riferimento ai modelli della percezione di suoni non nativi o della L2 e ai numerosi lavori sperimentali volti a verificare l'assimilazione di questi suoni a categorie della L1 (e.g., Best 1994, Flege 1995 e i lavori sperimentali ispirati ad essi). In questi contributi, sono abitualmente utilizzate vocali in isolamento che si richiede vengano classificate rispetto ai fonemi vocalici della lingua materna degli ascoltatori. Si tratta ovviamen-

te di una semplificazione che, tuttavia, permette di controllare opportunamente l'esperimento percettivo al fine di poterne interpretare correttamente i risultati e che, proprio perché gli stimoli non rappresentano la complessità del parlato, consente di comprendere fino a che punto il sistema vocalico della L1 possa influenzare la percezione di suoni vocalici diversi.

All'esperimento percettivo hanno partecipato 24 soggetti italiani (7 maschi, 17 femmine), studenti di cinese presso l'Università del Salento – Lecce. Prima dell'esperimento ai soggetti è stato chiesto di rispondere ad un breve questionario, volto ad individuare le caratteristiche sociolinguistiche di riferimento. In particolare, oltre ai dati relativi al luogo di nascita, all'età, alle lingue parlate in famiglia e con amici (es. italiano e dialetto), alla lingua madre dei genitori, al livello di istruzione e alla professione, è stato chiesto di indicare in modo esplicito:

– il numero di anni di studio della lingua cinese e la durata di eventuali soggiorni in Cina;
– il numero di anni di studio di altre lingue e la durata dell'eventuale permanenza nei Paesi nei quali queste lingue rappresentano il codice linguistico più utilizzato.

Le informazioni relative al primo punto sono state usate come indice dell'esposizione alla lingua cinese, mentre le informazioni ricavate in relazione al secondo punto sono servite per individuare l'indice dell'esposizione a lingue come francese e tedesco e, quindi, l'ipotetico livello specifico di esposizione al fonema /y/. Come è già stato sottolineato, sia in francese che in tedesco esiste il fonema /y/ (lungo in tedesco; Maturi, 2009), e quindi studenti di queste lingue, almeno in teoria, sono stati maggiormente esposti ad esso, ascoltandone le realizzazioni e magari producendolo loro stessi.

Ai soggetti sono stati proposti esempi in ordine random delle vocali cinesi /i u y ɤ/, prodotte in isolamento da 3 parlanti cinesi di età inferiore ai 30 anni (2 femmine, 1 maschio); per controllo, sono stati proposti anche esempi delle vocali italiane /i u e o/, prodotte anch'esse in isolamento da parlanti salentini dell'area leccese di età inferiore ai 30 anni (2 femmine, 1 maschio). Gli stimoli vocalici sono stati normalizzati per durata, intensità e frequenza fondamentale e proposti in un test di identificazione e *goodness rating*[2]. Di fatto ai soggetti si chiedeva di classificare gli stimoli come vocali dell'italiano o suoni diversi (le opzioni fornite erano "i, e, a, o, u, nessuna"), valutando su una scala *Lickert* 1-7 la bontà di ogni stimolo rispetto alla classe scelta. Tutti gli stimoli sono stati proposti per tre volte in ordine random, all'interno dello stesso blocco. I soggetti, potevano effettuare una pausa ogni 25 stimoli ascoltati, ma, indipendentemente dal numero e dalla durata delle pause realizzate, il test è sempre stato svolto in meno di 20 minuti.

Nella sezione seguente, saranno presentati i risultati dell'identificazione e *goodness rating* delle vocali cinesi, mentre non saranno commentati i risultati dell'identificazione delle vocali italiane, usati solo per determinare l'affidabilità del sog-

[2] La normalizzazione è stata necessaria per controllare tutte le caratteristiche che non fossero relative al timbro della vocale e per evitare, quindi, che variazioni, come ad esempio quelle prosodiche, influenzassero la percezione da parte dei soggetti sperimentali.

getto nello svolgimento del compito sperimentale e valutare quindi una sua eventuale esclusione dall'analisi dei risultati. I risultati del test di identificazione sono stati analizzati prendendo in considerazione le percentuali di identificazione e i valori di correlazione tra le percentuali di identificazione e gli indici di esposizione individuale alle lingue straniere, stabiliti sulla base delle risposte al questionario. Per quanto riguarda il *goodness rating*, è stata presa in esame la valutazione media fornita per ogni item rispetto alla categoria scelta nel compito di identificazione, tranne quando è stata scelta l'opzione 'nessuna'. In questi casi, infatti, non è possibile avere la certezza che i soggetti pensino ad una categoria specifica e, quindi, non ha senso prendere in esame i valori di *rating*; inoltre, anche per il *goodness rating* sono stati considerati i valori di correlazione con gli indici di esposizione alle lingue straniere. In generale, per le correlazioni sono stati presi in esame i coefficienti di *Pearson* e di *Spearman* (rispettivamente parametrico e non parametrico)[3] e per ognuno di essi è stato calcolato anche l'indice di determinazione[4].

2.3 Risultati

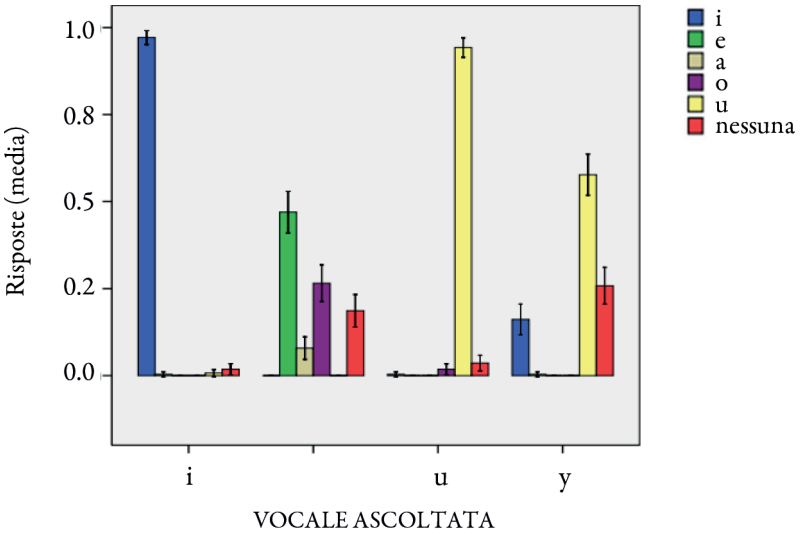
I risultati si riferiscono al test di identificazione svolto da tutti i soggetti reclutati per l'esperimento. Le percentuali di identificazione degli stimoli italiani, superiori all'85%, hanno infatti permesso di verificare l'accuratezza nello svolgimento del compito sperimentale e l'affidabilità dei risultati.

Nel grafico proposto in Figura 2 è riportata la media delle risposte fornite dai soggetti italiani in corrispondenza delle vocali prodotte dai parlanti cinesi. I dati, non ancora scorporati in base ai livelli agli indici di esposizione individuati, mostrano che i soggetti hanno un'ottima capacità di classificazione delle [i u] cinesi, che vengono identificate con /i u/ italiane in una percentuale molto elevata di casi (oltre il 95%). Al contrario, in corrispondenza delle vocali [ɤ] e [y] si osservano percentuali di identificazione che, secondo le attese, indicano problemi nella classificazione dei suoni percepiti.

[3] La scelta di presentare i risultati sia di un test parametrico che di un test non parametrico è dovuta al fatto che durante il controllo preliminare della matrice di dati, l'esclusione di alcuni soggetti comportava la perdita della condizione di normalità della distribuzione. Il prerequisito per la scelta del test ci è sembrato quindi molto precario, suggerendoci di effettuare entrambi i tipi di test (che peraltro non forniscono risultati molto diversi rispetto alla significatività dei dati).

[4] Come è noto, il valore numerico della correlazione varia tra 0 e 1 e più si avvicina a 1, più indica una relazione stretta tre le variabili considerate; il segno del coefficiente di correlazione corrisponde al modo in cui variano le variabili: se è positivo, indica che all'aumentare di una aumenta anche l'altra, mentre se è negativo indica che se una aumenta l'altra diminuisce. Il valore statistico della correlazione indica la probabilità di ottenere una certa correlazione per caso (se la significatività è minore di 0.05, le probabilità sono meno di 5 volte su 100 e quindi c'è una relazione tra le variabili). Infine, l'indice di determinazione o effetto della correlazione, ossia il valore di r2, può variare da 0 a 1 e indica in quale misura la proporzione di variazione può essere predetta dalla relazione tra le due variabili (ad esempio, un coefficiente di determinazione di 0.60 indica che il 60% della variazione può essere previsto sulla base della relazione tra le variabili, mentre il 40% non può essere spiegato dalla relazione considerata).

Figura 2 - *Media delle risposte fornite dai soggetti in corrispondenza delle vocali prodotte dai parlanti cinesi*



Osservando i dati, in particolare, sembra che venga identificato soprattutto il grado di altezza: in corrispondenza di [ʏ], la maggior parte delle risposte è a favore di /e/ ed alcune risposte sono a favore di /o/; per lo stimolo [y], la maggior parte delle risposte è a favore di /u/ e alcune risposte sono a favore di /i/. Inoltre, per entrambi i foni si nota un certo numero di risposte a favore di 'nessuna', ossia di risposte che potrebbero corrispondere alla reale corretta identificazione del fono come non appartenente al sistema della L1[5].

L'analisi delle valutazioni fornite durante il compito di *goodness rating* conferma che le [i u] cinesi sono abbastanza apprezzate come esempi delle vocali italiane alle quali sono associate (con valutazioni medie di 6.25 e 5.95 rispettivamente). Per quanto riguarda [ʏ] e [y], invece, i valori di gradimento sono inferiori e, soprattutto, non completamente in linea con i risultati di identificazione. Il fono [ʏ] è principalmente associato a /e/, ma ottiene una valutazione migliore come esempio di /o/ (4.12 vs. 3.8 per la scelta /e/)[6]; il fono [y] è principalmente associato a /u/, ma ottiene la stessa valutazione media anche quando viene identificato come /i/ (il *rating* è sempre 4).

I risultati sono poi stati scorporati in base all'indice di esposizione alla/e lingua/e straniera/e, ottenuto per ogni parlante prendendo in considerazione la quantificazione da noi operata in mancanza di una misurazione diretta del livello delle competenze fonetico-fonologiche. Nella tabella presentata in Figura 3 sono riportati i

[5] In ogni caso, si noti che da un punto di vista statistico le risposte la cui occorrenza è inferiore al 20% non sono comunque degne di nota in quanto al di sotto della soglia corrispondente alla scelta casuale (date le 5 opzioni possibili).

[6] La categoria /a/ viene scelta in un numero esiguo di casi e anche il *goodness rating* medio degli stimoli classificati come /a/ risulta abbastanza basso (2.9).

valori degli indici di correlazione e di determinazione relativi alla correlazione tra la percentuale di identificazione e, procedendo da sinistra verso destra, il numero degli anni di studio della lingua cinese, il numero degli anni di studio delle lingua tedesca e/o francese, il numero complessivo degli anni di studio di tutte le lingue considerate e, infine, la somma del numero degli anni di studio delle lingue e della durata dei soggiorni all'estero in Paesi in cui queste lingue sono parlate.

Come si può osservare dai valori di significatività della correlazione, riportati tra perentesi (es. per la prima cella in alto a sinistra, .000), l'indice di esposizione alla/e lingua/e dei soggetti è significativamente correlato con la percentuale di identificazione. Tuttavia i valori bassi di correlazione, ossia gli effettivi valori dell'indice di correlazione *r* o dell'indice di determinazione r^2 (es. per la prima cella in alto a sinistra, -0.260 e 6.7% oppure -0.278 e 7.7%, rispettivamente, a seconda che si consideri il risultato del test di *Pearson* o di *Spearman*) suggeriscono che non c'è una relazione stretta tra la percentuale di identificazione e l'indice di esposizione individuato. Infatti, i valori di *r* non superano mai 0.3 e quelli di r^2 non sono mai superiori a 9.6%. Inoltre, è evidente la mancanza di significatività del numero di risposte 'nessuna' ("non ita" in Figura 3) che, come già ricordato, dovrebbe indicare la percezione della presenza di foni non assimilabili a categorie native.

Figura 3 - *Correlazione tra la percentuale di identificazione e i diversi indici di esposizione da noi individuati (correlazione di Pearson e Spearman, rispettivamente in prima e seconda posizione nella cella)*

Pearson e Spearman's rho (n. cases 279)	**Studio di cinese**	**Studio di tedesco e/o francese**	**Studio di cinese e tedesco e/o francese**	**Studio di tre lingue e soggiorni all'estero**
y vs. i ita	-.260 (.000) -.278 (.000) r^2:6.7 e 7.7%	-.245 (.000) -.253 (.000) r^2:6 e 6.4%	-.269 (.000) -.288 (.000) r^2:7.2 e 8.2%	-.271 (.000) -.310 (.000) r^2:7.3 e 9.6%
y vs. u ita	n. s n. s.	.168 (.005) .159 (.008) r^2:2,8 e 2.5%	.164 (.006) .161 (.007) r^2:2.6 e 2.5%	.167 (.005) .157 (.009) r^2:2.7 e 2.4%
y vs. non ita	.184 (.002) .198 (.001) r^2:3.3 e 3.9%	n. s n. s.	n. s n. s.	n. s n. s.
ɤ vs. e ita	n. s n. s.	.183 (.002) .164 (.006) r^2:3.3 e 2.6%	.175 (.003) .172 (.004) r^2:3 e 2.9%	.174 (.003) .166 (.005) r^2:3 e 2.7%
ɤ vs. o ita	-.166 (.005) -.160 (.007) r^2:2.7 e 2.5%	-.286 (.000) -.265 (.000) r^2:8.1 e 7%	-.295 (.000) -.278 (.000) r^2:8.7 e 7.7%	-.297 (.000) -.278 (.000) r^2:8.8 e 7.7%
ɤ vs. non ita	n. s n. s.	n. s n. s.	n. s n. s.	n. s n. s.

Osservando i risultati ancor più nel dettaglio e prestando particolare attenzione al segno di *r* (v. nota 4), si può notare che sembra rilevante la sensibilità dei soggetti alle differenze acustiche tra gli stimoli, espressione delle diverse caratteristiche articolatorie. In particolare, a livelli bassi di esposizione alla/e lingua/e, i soggetti sembrano identificare gli stimoli a seconda della posizione antero-posteriore della lingua, mentre questo accade meno quanto più aumenta il livello di esposizione; inoltre, sembra aumentare progressivamente, e in modo direttamente correlato al livello di esposizione, l'attenzione al possibile arrotondamento delle labbra. Si consideri, infatti, la correlazione negativa che si riscontra nell'identificazione di [y] con /i/ e di [ɤ] con /o/ (correlazione che diventa più chiara, benché rimanga lieve, con l'aumentare del valore dell'indice di esposizione, v. celle da sinistra verso destra nella tabella). Si osservi invece la correlazione positiva che si riscontra nell'identificazione di [y] con /u/ e di [ɤ] con /e/. Quindi, all'aumentare del livello di esposizione alla/e lingua/e diminuisce l'identificazione di [y] con /i/ e di [ɤ] con /o/ e, in generale, aumenta quella di [y] con /u/ e di [ɤ] con /e/. In ogni caso, si tenga sempre presente che i valori delle correlazioni sono estremamente bassi.

Per quanto riguarda i valori di *goodness rating*, nessuna correlazione è significativa, ad indicazione del fatto che l'indice di esposizione alla/e lingua/e non è correlato alla valutazione della bontà del fono come esempio della categoria nativa scelta[7].

2.4 Discussione

I risultati presentanti nella sezione precedente mostrano una chiara propensione degli studenti italiani a identificare gli stimoli cinesi come fonemi della lingua materna, anche quelli completamente assenti dal sistema italiano, dato l'esiguo numero di risposte a favore di una categoria diversa da quelle corrispondenti alle vocali italiane (v. la risposta 'nessuna' nel paradigma sperimentale). Più dettagliatamente, in linea con le aspettative, le vocali [i u] cinesi vengono identificate con /i u/ italiane in una percentuale elevata di casi e, sulla base del goodness rating, risulta che sono anche piuttosto apprezzate come esempi di vocali italiane. Invece, per le vocali [ɤ] e [y] si registrano problemi di identificazione con le categorie date: [ɤ] è associato principalmente a /e/, oltre che ad /o/ e ad una possibile categoria non appartenente al sistema L1 (opzione 'nessuna'); la vocale [y] è associata principalmente a /u/, con alcune risposte a favore di /i/, ma soprattutto a una categoria non rappresentata tra quelle date e, di fatto, della L1 (opzione 'nessuna). Peraltro, i valori di gradimento delle vocali cinesi [ɤ] e [y] come esempi di vocali italiane sono inferiori a quelli appena discussi e, soprattutto, indicano che l'esempio non è comunque giudicato

[7] Si ricordi che non sono state considerate le valutazioni fornite in corrispondenza della scelta dell'opzione 'nessuna', che doveva corrispondere all'identificazione di una categoria 'altra', potenzialmente (ma ovviamente non solo e non necessariamente) quella corretta. Tuttavia, l'osservazione informale delle valutazioni fornite quando è stata scelta questa opzione conferma l'assenza di correlazione, ad indicazione dell'eccessiva vaghezza della definizione della categoria o, comunque, del fatto che gli esempi non sono visti come ottimi rappresentanti di un'ipotetica categoria 'non nativa', magari già esistente nel sistema dell'apprendente, benché non indicata in modo esplicito nel test.

come soddisfacente anche se si considera la categoria a cui il fono è principalmente assimilato. Infatti, il fono [ɤ] ottiene una valutazione migliore come esempio di /o/ piuttosto che di /e/ e il fono [y] ottiene la stessa valutazione media sia quando viene identificato come /u/ che quando viene identificato come /i/ (ovviamente non sono state considerate le valutazioni fornite per l'opzione 'nessuna').

I risultati, inoltre, indicano chiaramente che esiste una correlazione statisticamente significativa tra il livello di esposizione alla/e lingua/e considerato in questo contributo e le percentuali di identificazione dei foni; non esiste invece una correlazione significativa tra indice di esposizione alla/e lingua/e e *goodness rating*. Tuttavia, anche nel caso dei valori di identificazione, e quindi di significatività della correlazione, l'indice di correlazione e il coefficiente di determinazione sono sempre estremamente bassi, a dimostrare il fatto che le percentuali di identificazione sono spiegate dal livello di esposizione in proporzione decisamente limitata.

Fermo restando che la correlazione non sarebbe comunque da interpretare come causalità, è ovvio che i bassi valori ottenuti per essa possono essere dovuti a problemi metodologici, a cominciare dalla definizione del livello di esposizione alla/e lingua/e. Abbiamo, infatti, utilizzato l'indice di esposizione per avere un'indicazione che fosse il più possibile oggettiva sull'ipotetico livello di competenza, ma anche l'indice di esposizione è, di fatto, ricavato in modo indiretto, ipotizzando che possa essere correlato agli anni di studio o di soggiorni all'estero. Ovviamente una misura diretta e oggettiva della competenza fonetico-fonologica o dell'esposizione specifica ai fonemi indagati sarebbe stata migliore. Tuttavia, pensando, ad esempio, ai metodi usualmente utilizzati per verificare le competenze fonetico-fonologiche, riteniamo che ci sarebbe stato il rischio di circolarità, perché sarebbe stato necessario testare direttamente la capacità di identificazione e discriminazione dei foni per individuare il livello di competenza e poi si sarebbe cercata una correlazione tra le capacità percettive, ad esempio di identificazione, e il livello di competenza[8].

Peraltro, anche la procedura sperimentale adottata nel test percettivo potrebbe aver rappresentato un *bias*, visto che al soggetto italofono si fornivano come possibili opzioni le vocali che rappresentano le categorie fonologiche italiane e, in aggiunta, l'opzione 'nessuna', che doveva essere scelta in corrispondenza di un elemento percepito come non nativo. D'altra parte, come discusso all'inizio del paragrafo 2.2, la nostra metodologia è coerente con quella tipicamente usata nella verifica dei principali modelli relativi alla percezione di suoni non nativi o L2 e, anzi, abbiamo considerato utile introdurre una categoria non corrispondente ad una vocale nativa proprio per non escludere la possibilità che i soggetti fossero in grado di percepire

[8] Un revisore anonimo ha osservato che sarebbe stata utile un'autovalutazione del livello di competenza, effettuata dal soggetto con riferimento ai descrittori del Quadro di Riferimento Europeo. Tuttavia, per quanto questa osservazione sia valida in generale, non riteniamo che si adatti al nostro studio. In questo esperimento, infatti, eravamo interessati a correlare le competenze fonologiche relative a specifici fonemi con una misura della competenza o, meglio, dell'esposizione agli stimoli che fosse il più possibile oggettiva. La misura da noi adottata non è sicuramente ideale, ma riteniamo che un'autovalutazione non avrebbe fornito informazioni più precise o attendibili.

chiaramente differenze eclatanti rispetto alle categorie native (per rendere conto delle quali, quindi, non sarebbe stato considerato sufficiente scegliere una categoria nativa e assegnare un basso *goodness rating*); inoltre, pensiamo che non sia verosimile prevedere tutte le possibili risposte dei soggetti in relazione a foni non nativi e che, comunque, non sarebbe semplice trovare un metodo per riferirsi a tutte le diverse categorie senza rendere troppo articolato il compito sperimentale. Ad esempio, non si può assumere che i soggetti conoscano l'alfabeto fonetico a tal punto da poter svolgere il test facendo riferimento ad esso (senza contare il fatto che i risultati del test non misurerebbero solo le capacità percettive, ma anche quelle relative all'alfabeto fonetico). Per quanto i limiti dello studio messi in evidenza siano, secondo il nostro parere, difficilmente superabili, riteniamo si debba comunque fare riferimento ad essi nel momento in cui si osservi l'innegabile debolezza dei risultati ottenuti.

Tuttavia, l'analisi da noi condotta mette in evidenza alcuni aspetti piuttosto interessanti. In particolare, che:

– il livello di esposizione alla lingua cinese da noi considerato è significativamente correlato con i risultati di identificazione, benché la relazione tra i due aspetti non sia stretta (es. i coefficienti di determinazione, r2, non superano il 9.6%);
– l'importanza del considerare un livello di esposizione fonetico-fonologico specifico, relativo a foni presenti in altre lingue di studio, non emerge chiaramente, dato il lievissimo aumento di r e, soprattutto, l'osservazione della stessa tendenza per [y] e [ʏ] sulla base delle competenze in francese e/o tedesco (che dovrebbero aiutare solo per [y]);
– la sensibilità a caratteristiche specifiche (antero-posteriorità, labializzazione) sembra in qualche modo proporzionale alla quantità di esposizione alla/e lingua/e, ma presenta un quadro che suscita qualche perplessità. Infatti, nell'identificare gli stimoli, i soggetti sembrano farsi guidare maggiormente dalla posizione antero-posteriore della lingua a livelli bassi di esposizione al/ai codice/i; mentre, all'aumentare del livello di esposizione, sembra che aumenti progressivamente l'attenzione al, e l'influenza del, coefficiente di arrotondamento-procheilia. Tuttavia, questo non porta i soggetti a prendere in considerazione categorie "nuove", in cui, ad esempio, ad un certo tratto di anteriorità-posteriorità si associ un coefficiente di arrotondamento non atteso in base alle caratteristiche della L1. Il risultato suggerisce, quindi, che i soggetti non facciano riferimento a competenze approfondite sulle lingue di studio. Simili competenze dovrebbero portarli a considerare, ad esempio, che in cinese esistono delle vocali anteriori e arrotondate e delle vocali posteriori non arrotondate, che non fanno parte del sistema italiano. Se i soggetti facessero riferimento a competenze approfondite sulle lingue di studio, dovrebbero quindi considerare che un suono 'anomalo' che inizialmente può anche sembrare, ad esempio, una vocale anteriore media mal riuscita, ossia una possibile /e/, sia in realtà una vocale posteriore media, ma non arrotondata, ossia una [ɤ].

3. *Conclusioni*

In questo studio, abbiamo verificato se l'influenza della lingua materna cambi a seconda del livello di esposizione ad una o più lingue di studio, cercando in particolare di individuare specifiche interferenze dovute alla conoscenza di lingue il cui sistema fonetico-fonologico presenti elementi di somiglianza con quello della L2 oggetto di indagine. Per effettuare la nostra verifica, abbiamo preso in considerazione le capacità percettive di italiani che studino la lingua cinese - in cui ad esempio esistono i fonemi /y/ e /ʏ/, che mancano invece in italiano - e che in alcuni casi conoscano già altre lingue, come il francese o il tedesco, nelle quali è presente /y/ (o /y:/) e non /ʏ/.

Nonostante alcune necessarie limitazioni legate al disegno sperimentale (v. § 2.4), i risultati presentanti nella sezione precedente mostrano una chiara propensione degli studenti italiani a identificare gli stimoli cinesi, anche quelli completamente assenti dal sistema italiano, come fonemi della lingua materna, con indici di gradimento maggiori per le vocali che sono effettivamente più simili a quelle native (v. *goodness rating*). I risultati indicano anche piuttosto chiaramente che, mentre non esiste una correlazione significativa tra l'indice di esposizione alla/e lingua/e considerato nel nostro studio e *goodness rating*, esiste una correlazione significativa tra il livello di esposizione e le percentuali di identificazione dei foni. Tuttavia, anche nel caso di significatività della correlazione statistica, l'indice di correlazione e il coefficiente di determinazione non raggiungono mai valori elevati, suggerendo che il livello di esposizione influenzi e spieghi solo molto parzialmente le percentuali di identificazione.

L'aspetto forse più interessante che emerge dallo studio è che l'importanza dell'esposizione a fonemi e foni presenti in altre lingue di studio (in questo caso /y/) non emerge chiaramente, in quanto si osservano tendenze simili per questi fonemi e per quelli ai quali i soggetti sperimentali dovrebbero essere stati meno esposti (in questo caso /ʏ/). Peraltro, la sensibilità a caratteristiche specifiche (anteroposteriorità, labializzazione) sembra poter esser messa in relazione alla esposizione alla/e lingua/e, seppur in misura limitata. Tuttavia, su questo aspetto i risultati nel loro complesso suggeriscono che i soggetti non facciano riferimento a possibili conoscenze approfondite sulle lingue di studio, che dovrebbero portarli a individuare più facilmente stimoli non appartenenti alle categorie native (ad esempio, vocali anteriori arrotondate e vocali posteriori non arrotondate).

Lo studio ha quindi messo in evidenza solo parzialmente l'influenza esercitata dall'esposizione a fonemi e foni specifici. Tuttavia, secondo il nostro parere, questo risultato può essere dovuto, oltre che ai possibili limiti metodologici già discussi nell'articolo, anche al fatto che, dato l'apprendimento in contesto formale dei soggetti da noi intervistati, la quantità di esposizione durante lo studio può non essere stata sufficiente a determinare un reale apprendimento di tipo fonetico-fonologico, come dimostrato in ambito neurofisiologico da Grimaldi et al. (2014) per mezzo di altre procedure e tecniche sperimentali.

Bibliografia

Best, C.T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Goodman, J., Nusbaum, H. (Eds.), *The development of speech perception*. Cambidge: MIT Press, 167-224.

Best, C., Tayler, M. (2007). Nonnative and second-language speech perception. Commonalities and complementarities. In Munro, M.J., Bohn, O.-S. (Eds.), *Second language speech learning*. Amsterdam: John Benjamins.

Cabrelli Amaro, J. (2012). L3 phonology: An understudied domain. In Cabrelli Amaro, J., Flynn, S. & Rothman, J. (Eds.), *Third language acquisition in adulthood.* Philadelphia/Amsterdam: John Benjamins, 33-60.

Flege, J.E. (1995). Second-language speech learning: theory, findings, and problems. In Strange, W. (Ed.), *Speech Perception and Linguistic Experience, Issues in Crosslinguistic research*. Timonium, MD: York Press, 233-277.

Grimaldi, M. (2003). *Nuove ricerche sul vocalismo tonico del Salento meridionale*. Alessandria: dell'orso.

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P. & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood – A behavioral and ERP study. In *Frontiers in Human Neuroscience*, 8.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. In *Perception and Psychophysics*, 50, 93-107.

Maturi, P. (2009). *I suoni delle lingue, i suoni dell'italiano*. Bologna: il Mulino.

Pan, W. (2010). 汉英语言对比概论. Beijing: The Commercial Press.

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

Tang, Y. (2014). *L'analisi contrastiva dei sistemi fonologici del cinese mandarino standard e dell'italiano standard.* Tesi di laurea magistrale, Shanghai International Studies University.

Zhang Bin (2004). *A Concise Book of Modern Chinese*. Shanghai: Fudan University Press.

Parte II – Sezione B

Destrutturazione della forma sonora

DOI: 10.17469/O2101AISV000015

TYPHANIE PRINCE[1]

On the Jakobson's Postulate: the notion of markedness. A Case study of French Phonology of Aphasia and Acquisition

In this paper, we will suggest that the Jakobson postulate ([1941] 1968) concerning the mirror-image between language acquisition and speech language impairment is still relevant. This study presents the results of an investigation into the acquisition of phonology of 20 French children and the speech pathology of 20 French aphasic speakers. On the basis of French data, our claim is that markedness, which played a central role in Jakobson's work, is essential for the treatment of the production of children and aphasic speakers. We will show that similar strategies, such as deletions, epenthesis and substitutions are applied during acquisition but also during aphasia. These strategies correspond to the application of repair strategies made by aphasics and children to solve a conflict between their own constraints and the constraints applied in standard language.

1. *Introduction*

The starting point of this study is Jakobson's proposal ([1941] 1968) that a parallel can be drawn between the phonological system in acquisition and aphasia. In the famous *Child Language, Aphasia and Phonological Universals* (*Kindersprache, Aphasie und allgemeine Lautgesetz (1941)*), Jakobson suggests that universal laws govern the sound systems of all the language. These laws attest the strata development of the language system, they show that the system is not acquired randomly. Thus, "Jakobson put forward the thesis that the distinctive sounds of a language (i.e., phonemes) are acquired in an order that reflects their structural complexity, in terms of feature composition and basic syllabic structure" (Durand, Prince, 2015: 15). In the same way, this phonological system can be lost in aphasia, in the reverse order. Jakobson also claimed that we can observe similar strategies both in child speech and in the impairment of aphasic speakers. Finally, he proposes that these strategies respect a hierarchy of complexity.

> The dissolution of the linguistic sound system in aphasics provides an exact mirror-image of the phonological development in child language... The order in which

[1] I am very grateful to A. Tifrit, J. Durand, I. Kappa, M. Abuoudeh and the anonymous reviewer for their insightful comments and advices. I especially thank all of them for their constructive remarks. All errors are my own responsibility.

> speech sounds are restored in the aphasic during the process of recovery corresponds directly to the development of child language[2].

In order to validate or amend this postulate, the purpose of this study is to explore the different strategies used by normally-developed children during acquisition and by French aphasic speakers during the production of consonantal clusters / fricative+stop/, especially sC[3] and ʁC. The question we address in this paper is: do French aphasic speakers really have the same strategies as normally-developed children during acquisition?

Another claim in this article is to investigate to what extent syllabic complexity, segmental complexity or both of them play a role in phonological acquisition and phonological disorders. We will show that transformations applied to clusters also involve segments and we propose that the complexity observed results from the interface between constraints on segmental content and syllabic representation.

Most studies on phonological disorders in aphasia make a direct reference to acquisition data (Caramazza, 1994; Den Ouden, 2011; 2002; Den Ouden, Baastianse, 2003) but many studies focus on either the acquisition of phonology (Fikkert, 1994; Rose, Wauquier, 2007; Kirk, 2008; Rose, 2009) or the phonology of aphasic speakers (Nespoulous et al., 1987; Béland, Favreau, 1991; Den Ouden, 2002; Baqué et al., 2012). In our knowledge, there is no study that has explored the consequences of this phonological link between acquisition and aphasia, or has provided any information about the phonological complexity on the basis of empirical data.

Then, Den Ouden and Bastiaanse (2003: 83) also affirm that:

> [...] the logic of the general idea behind Jakobson's statement cannot be simply dismissed, especially if we are able to argue for an distinguish specific functional modules and can study patients with deficits in specific and well-defined language functions.

Furthermore, studies in the Theory of Constraints and Repair Strategies (TCRS), such as Blumstein (1973), Nespoulous et al. (1987), Béland, Favreau (1991), Valdois (1987), Moreau (1993) and Béland, Paradis (1997), have shown that phonological disorders correspond to a surface manifestation of the application of repair strategies made by aphasics to solve a conflict between their own constraints

[2] Jakobson (1968: 60-62).

[3] Phonology demonstrates that sC clusters have a special status for many reasons _ sC behave different from complex onsets in terms of sonority for example, data in typology and acquisition show that sC behave inconsistently, sometimes as a heterosyllabic sequence and other times as a tautosyllabic sequence (Goad, 2011). Many accounts have been proposed to address its particular behaviour (appendix, complex segment, extrasyllabicity, coda+onset, *inter alia*). However and because this is not the debate here, this point can not be treat here, but for a complete review, see Goad (2011) and Scheer (2015). In this paper, we will consider that sC is a kind of coda+onset sequence, as we will show in the analysis.

(they became too restrictive) and the constraints applied in standard language[4]. For example, in:

(1) *snail* 'lumaca' escargot /ɛskaʁgo/ → [tsɛtsaʁdo]

Generally, the aphasic speaker seems to have some difficulties to produce a complex word starting with a vowel and composed by two coda+onset clusters. Each complex element is simplified, i.e. the aphasic speakers use a repair strategy for the clusters and change them in a branching onset (or affricate). He also changes each velar into a coronal and reduplicates the branching onset in initial position. But, as suggested by Moreau (1993: 62) if an aphasic speaker applies a repair strategy, it may account for a less complex structure locally but it could generate a larger degree of complexity on the pattern (Béland, Paradis & Bois, 1993; Moreau, 1993). Besides, in acquisition, similar strategies are applied during the setting of parameters of the grammar. When the parameters are set, good productions occur. We prefer using the term of *transformations* rather than *errors*, because aphasic speakers or children apply different strategies to avoid something they cannot produce. These strategies lead to transformations in production.

On the basis on Jakobson's idea, we propose that markedness plays a role in acquisition and in aphasia. Children and aphasic speakers respect a hierarchy of complexity. This hierarchy is defined by markedness properties. Under this point, the goal is always to reduce the complexity adopting less marked structures. We also suggest that data from speech disorders and acquisition inform us about the complexity of clusters and segments. Complexity seems to appear on most phonological levels and much research has been devoted to understanding this notion. We propose that the complexity observed results from the interface between constraints on segmental content and syllabic representation.

First, we will start with a review of the literature on the phonology of aphasia and acquisition.

Secondly, we will present the different strategies applied in aphasia through the definition of phonological disorders called *phonemic paraphasias*. We will show that similar processes are applied in acquisition. Thirdly, we shall provide an overview of our experiment based on 20 aphasic French speakers and 20 French normally-developed children. Finally, we mainly focus on the process of reduction in clusters and provide a theoretical analysis of the data under discussion.

1.1 The Phonology of Aphasia and Acquisition

1.1.1 About Segments and Syllables in literature

If much research (Blumstein, 1978; Romani, Calabrese, 1998; Béland, Favreau, 1991) has demonstrated that deletions (consonant cluster reductions), segmental

[4] In speech language disorders, Calabrese (2005) suggested that brain damage involves a shift of parameters activation/deactivation. Repair strategies are issued to solve the violation resulting from the constraints (2005: 108).

substitutions and epenthesis (assimilation, consonantal harmony) are the most important strategies in aphasia, especially in English, French, Italian or Dutch, they are also frequent during the acquisition of phonology (Chin, Dinnsen, 1992; Fikkert, 2000; Kirk, 2008; Demuth, McCullough, 2009; Rose, 2009). The acquisition of consonant clusters is very problematic, and varies across the children and depends on the type of cluster, the position and the nature of the different consonants. Kappa (2002: 21) has shown that all the production of children tends to be unmarked in structure through different strategies such as deletion of the entire cluster, reduction and substitution (among others, Chin, Dinnsen, 1992 for English; Fikkert, 1994 for Dutch; Kappa, 2002 for Greek).

In aphasia, Blumstein, Baker & Goodglass (1977) suggest that the nature of the syllabic structure can generate transformations. For example, Blumstein (1973) has shown that deletion concerns mostly liquids in a cluster. Valdois (1987), Nespoulous and Moreau (1997) have shown that in a cluster pattern such as /obstruent+liquid/ or /fricative+liquid/, C^2 is more deleted than C^1 (Fikkert, 1994; 2000). But, in a cluster /fricative+obstruent/ as /sp/, C^1, the fricative will be deleted. In other words, deletion principally concerns the liquid or the fricative in a group in acquisition (Rose, Wauquier-Graveline, 2007) and also in aphasia. Freitas (1997) has shown that children acquired sC clusters before branching onset in Portuguese but Fikkert (1994) has shown, for Dutch, that children acquired first branching onset, as in French (Kehoe, Hilaire-Debove, Demuth & Lléo, 2008). In this way, sC clusters exhibits a different behaviour than branching onset. Dutch children also acquired onset clusters before coda clusters. Finally, in a sC clusters, children delete mostly the /s/ (Greenlee, 1974; Fikkert, 2000), like aphasics. Furthermore, the opposite strategy can be used. Valdois (1987) and Romani, Calabrese (1998) revealed that in French and Italian, the epenthesis is the most frequent repair strategy during aphasia to resolve a hiatus or if an item starts with a vowel. Another type of paraphasia in aphasia is substitution. Acquisition exposes similar transformations through consonantal harmony for example (Fikkert, 2000; Rose, 2009). At this point, Greenlee (1974) suggests three stages during acquisition of consonantal cluster; deletion of the entire cluster, deletion of one of the member of a cluster and substitution. Valdois (1987) highlights that substitution is the preferred transformation to solve a conflict between the internal parameters of the aphasic speaker and the constraints applied on language. Standard approaches to substitution in aphasia have relied on Interphonemic Distance (henceforth ID), which computes the number of feature changes. Blumstein (1973), Nespoulous et al. (1987) propose that most cases of substitutions show an ID of 1, i.e. most of the time, transformations imply only one feature value: place or manner of articulation (now PoAs and MoAs respectively). ID allows computing the degree of complexity of substitutions by comparing feature values of two matrices. The greater the distance, the more complex the substitution.

Furthermore, when segments contain the same number of features, ID is unable to reflect the complexity/the degree of markedness[5] between the different PoAs/MoAs. In this regard, we will suggest that element theory (e.g. Harris, 1994; Scheer, 1999; Cyran, 2010 or Backley, 2011) could explain the different patterns of substitution in aphasia and the asymmetrical behaviour of segmental classes, as we shall see later. We postulate that this particular scale is not universal but language-specific. According to Béland and Favreau (1991) and Béland, Paradis & Bois (1993) substitution involves particularly [coronal]. Moreau (1993) showed that [coronal] is also more deleted and used for epenthesis. [coronal], in speech disorders but also in acquisition and typology, seems to have a particular behaviour – coronals are less marked and the most frequent natural classes – as it is proposed by Jakobson, 1968; Paradis & Prunet, 1991; Scheer, 1999; Morissette & al., 2003; Kirk, 2008; Rice, 2009 for acquisition and typology.

1.1.2 Phonological Disorders in Aphasia vs Phonological Transformations in Acquisition

The term *phonological disorders* refers to a large scope of "phonological transformations" called *phonemic paraphasias*. Phonemic paraphasias occur in major kinds of aphasia: Broca, Wernicke and conduction. Roch-Lecours and Lhermitte (1979), Blumstein (1978) have proposed a typology of phonological disorders in aphasic speech involving substitution, addition, epenthesis, deletion or assimilation. Furthermore, similar phonological strategies are involved during the setting of parameters in acquisition or after a stroke, owing to the loss of the parameter settings.

1.1.3 Different Strategies

For the purpose of this study and on the basis of our data, we suggest to subsume phonemic transformations into 5 classes: deletion, epenthesis, metathesis, substitution and what we call total reduction (CCv → Cv):

Table 1 - *Kinds of strategies in cluster production*

Strategies		*Item*	*Ipa target*	*Productions*
Deletion C_1	$C_aC_b \rightarrow C_b$	*pen*	/stilo/	[tilo]
Deletion C_2	$C_aC_b \rightarrow C_a$	*garden*	/paʁk/	[paʁ]
Epenthesis	$C_aC_b \rightarrow C_wC_xC_b$	*snake*	/sɛʁpɑ̃/	[sɛʁəpɑ̃]
Metathesis	$C_aC_b \rightarrow C_bC_a$	*casque*	/kask/	[kaks]
Substitution C_1	$C_aC_b \rightarrow C_zC_b$	*snail*	/ɛskaʁgo/	[ɛʃkaʁdo]
Substitution C_2	$C_aC_b \rightarrow C_aC_X$	*snail*	/ɛskaʁgo/	[ɛstaʁdo]
CCv → Cv	$C_xC_b \rightarrow C_x$	*spatula*	/spatyl/	[ʃatyl]

[5] We suggest, as Den Ouden (2002), Romani, Galluzzi (2005) that complexity is equivalent to markedness.

In a consonantal sequence, the first member – C_1, the second member – C_2, or all the cluster, C_1C_2 can be deleted, like as in the example above. In this study, C_1 corresponds to a /s/ or a /ʁ/ while C_2 always corresponds to a stop.

(2) *card* 'carta' carte /kaʁt/ → [ka]

On the contrary, epenthesis corresponds to the insertion of a vowel, a glide or a consonant. Epenthesis appears frequently in hiatuses, especially in Italian (Calabrese, Romani, 1998).

(3) *snake* 'serpente' serpent /sɛʁpɑ̃/ → [sɛʁəpɑ̃]

Metathesis corresponds to the permutation of two segments in the string, we defined two kinds of metathesis:

– IntraClusterMetathesis – permutation of both members of the cluster $C_1C_2 \rightarrow C_2C_1$

(4) *helmet* 'casco' casque /kask/ → [kaks]

– InterClusterMetathesis – permutation with one of the member of the cluster and another consonant in the pattern. $C_1v[C_2C_3v]C_4vC_5 \rightarrow C_1v[C_2\mathbf{C_5}v]C_4v\mathbf{C_3}$

(5) *mosquito* 'zanzara' moustique /mustik/ → [muskit]

We can distinguish two different types of substitutions: one involves a gain or a loss of an element/feature and the other involves a complete change of segment. Finally, the last operation corresponds to a case of reduction, when a cluster is reduced to a C+V string: CCv → Cv like in the following example: *spatula:* /spatyl/ → [ʃatyl]. In this case, two different consonants: /s/ and /p/ became another consonant /ʃ/, as if it was a deletion and a substitution or a fusion of the features of the two segments.

1.2 The French Phonological System

1.2.1 Method

Population

The data is based on a cross-sectional study on a sample of 20 French monolingual children and 20 aphasic participants. 27 children were recorded but the data of 7 were discarded: 5 because the children were too shy or not interested in the task, 1 because he was bilingual, and another because he had suffered a stroke the year before. The child participants were normally developing children, 8 girls and 12 boys, aged between 2;1 and 3;8 years (M_{age} = 2;34 years). They were recorded at a kindergarten in France. All the children selected did not suffer from speech and hearing deficits, language delay, neurological or motor control disorders. All the children produced phonological transformations.

All the aphasic patients were selected at the hospital of Nantes, in the neurovascular unit. Each patients was diagnosed and classified on the basis of MT-86 a standardized battery test (Nespoulous, Joanette & Roch Lecours, 1986) by a speech therapist. The patients were selected based on the following criteria: native French speakers, right-handed, suffered from a right CVA (cerebral vascular accident), a lesion of the

left hemisphere and are selected in the acute and subacute phases of stroke. All of them suffering from phonetico-phonological disorders in production. In a set of 42 aphasic patients tested in hospital, 22 aphasic speakers were not included in the sample because their disorders did not correspond to our goal (for instance those who presented a major deficit of comprehension, an important jargonaphasia or more simply because they did not produce phonemic paraphasias). 18 of the aphasic speakers were recorded between Day+1 and Day+25 at the stroke (Neurovascular) unit of a hospital in Nantes, while the other 2 were recorded in a speech therapy practice. We have recorded 11 woman and 9 men, aged between 36 and 82 years (M_{age} = 63;10 years), 7 Broca's aphasics including 1 in therapy, 6 Wernicke's aphasics including 1 in therapy, 4 conduction aphasics and 3 Transcortical aphasics. All the aphasics suffer from a phonological disorder stemming from a lesion of the left hemisphere due to a stroke. All aphasic speakers produced phonemic paraphasias.

Phonemic paraphasias are the focus of this paper and involve patterns of error such as epenthesis, substitutions, deletions, metathesis and can be found across multiples tasks (spontaneous speech, repetition, picture naming, reading aloud). Furthermore, following Jakobson's postulate (1968) concerning the relation between acquisition and aphasia, we will compare all the productions to find out if children and aphasics use the same strategies.

Procedure

The experiment is composed of 40 items with clusters sC and ʁC in all positions and depending on the different classes for segments: [labial], [coronal] and [dorsal]. A naming and a repetition tasks were used to test the production of clusters. Words are monosyllabic, bisyllabic and trisyllabic (examples in table 2 below). Pictures were always black-and white drawings of common objects, vegetables and animals. Children were recorded at the kindergarten while the aphasic speakers were tested in a quiet room at the clinic. All aphasic speakers and children were recorded with a Tascam DR 100. Each session lasted approximately 20 minutes.

Table 2 - *Kinds of phonological structures*

Position	*Length*	*Syllabic structure*	*Patterns*		
# _.	1	CCVC	stage	*stade*	[stad]
	2	CCV.CV	pen	*stylo*	[stilo]
	3	CCV.CV.CV	beetle	*scarabé*	[skaʁabe]
. _ .	2	CVC.CVC	cap	*casquette*	[kaskɛk]
	3	VC.CVC.CV	snail	*escargot*	[ɛskaʁgo]
. _ #	1	CVCC	mask	*masque*	[mask]
	2	CV.CVC	dentist	*dentiste*	[dɑ̃tist]
	3	V.CVC.C	scarf	*écharpe*	[eʃaʁp]

All these data recorded were orthographically and phonetically transcribed by a French speaker using PHON software, developed by Rose, Y., MacWhinney, B.,

Byrne, R., Hedlund, G., Maddocks, K., O'brien, P. & Wareham, T. (2006). From the overall data, we excluded neologisms, misunderstanding productions, echolalic responses, perseverations and other paraphasias (for example lexical or semantic).

1.2.2 Data

A total of 906 phonemic transformations were explored depending on several types of transformations (table 3):

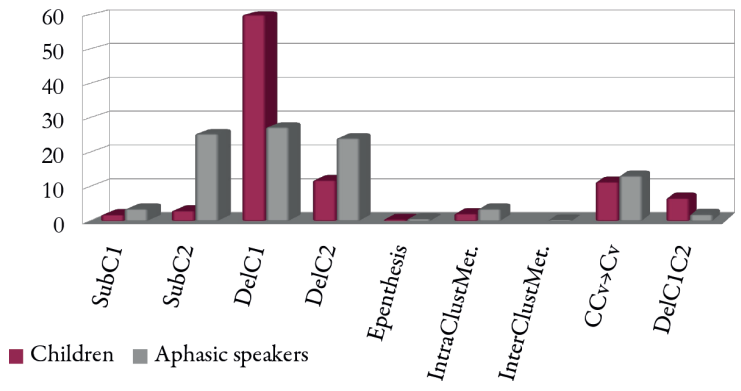
Table 3 - *Phonemic paraphasias*
Types of Transformations

	Children		*Aphasics*	
	N	*%*	*N*	*%*
Substitution C^1	11	1,85	11	3,54
Substitution C^2	18	3,03	**79**	**25,40**
Deletion C^1	**371**	**62,35**	**85**	**27,33**
Deletion C^2	**71**	**11,93**	**75**	**24,12**
Deletion C^1C^2	40	6,72	6	1,93
InterClusterMetathesis	0	0	1	0,32
IntraClusterMetathesis	13	2,18	11	3,54
Epenthesis	3	0,50	2	0,64
CCV → CV	**68**	**11,43**	41	13,18
TOTAL	595	100	311	100

The $\chi 2$ showed that there is a significant difference between the group (children *versus* aphasic speakers) and the type of transformations: $\chi 2_{[8]} = 179.6623$, p-value < 0.001.

We looked at these transformations in detail, depending on the type of the cluster, the position (initial, medial, final) and cluster's members C^1 or C^2. In figure (1) we can see the different transformations depending on the population:

Figure 1 - *Phonemic paraphasias*
Types of transformations .CC. (in %)



The main generalizations that can be formulated on the basis of our results are: (i) the most frequent transformation observed in both groups is the deletion of C^1 or C^2 (Barlow, 2001; Greenlee, 1974; Chin, Dinnsen, 1992) (ii) children delete the initial consonant more often than C_2 (iii) other important transformations observed are *substitution* of C^2 and the *reduction* – CCv → Cv (Kirk, Demuth, 2005; Barlow, 2004). Some examples are given in (6).

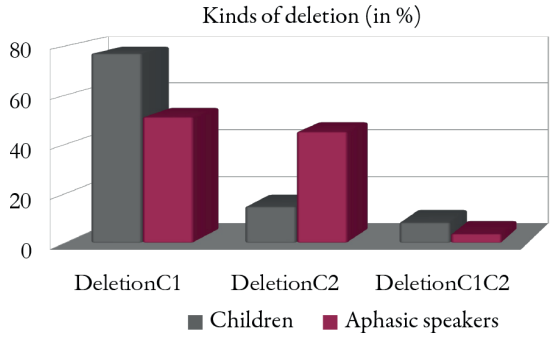
(6)	a. *snail*	'lumaca'	escargot	/ɛskaʁgo/	→	[ɛʃkaʁgo] [ɛkago] [esado]
	b. *mosquito*	'zanzara'	moustique	/mustik/	→	[muti] [musik] [muskit]
	c. *cap*	'beretto'	casquette	/kaskɛt/	→	[kakɛk] [kakɛt]
	d. *pen*	'penna'	stylo	/stilo/	→	[skilo] [tilo] [kijo]
	e. *satchel*	'cartella'	cartable	/kaʁtablə/	→	[kapat] [kaʁnat] [katab]

We propose to focus now on deletion, substitution and reduction. Despite the fact that different strategies are used, we suggest that all transformations applied by children and by aphasic speakers lead to a weakening process, a reduction of the complexity, through the syllabic structures or the internal representation of segments. To understand why children and aphasic speakers use these strategies, this study will be conducted depending on the position (initial, medial or final) and the nature of the consonants in the cluster [labial], [coronal] and [dorsal]. The listed transformations suggest that specific parameters are at play in aphasia and in acquisition. In fact, we will show that transformations involve the reduction of complexity in acquisition and also in aphasia.

1.3 Data: different paths to the weakening

1.3.1 Deletion

Figure 2 - *Kinds of deletions*



In acquisition, deletion concerns 81,00% of the total of transformations (482/595 cases). We have identified the cases of deletions depending on the position in the cluster: deletions of C^1 – 76,97%, (378/595 cases) deletions of C^2 – 14,73%, and C^1C^2 – 8,30% in the next table. Deletion of C^1 is clearly adopted. In aphasia, deletion concerns 53,38% of the total of transformations (166/311 cases). I have identi-

fied the cases of deletions depending on the position in the cluster: deletions of C^1 corresponds to 51,20%, (deletions of C^2 – 45,18%, and C^1C^2 – 3,62%).

Deletions are not random. There is a significant difference between the group and the kind of deletion: $\chi 2_{[2]} = 66{,}28$, p-value < 0.001. Children delete mainly the first consonant, which is always a fricative /s/ or /ʁ/. Two main reasons account for this deletion according to Jakobson, Demuth & Kehoe (2006), Wauquier (2008), Kirk (2008): first – fricatives are more complex than stops – and acquired later – that's why they are deleted – and second, we can also consider that the first member of the cluster is in coda position (Kaye, 1992; Goad, 2011; 2012). Furthermore the coda position is more marked than the onset position, and then, in a CVCV's model (Scheer, 2004) coda is weaker and prone to the deletion. In the literature, if ʁC is always an heterosyllabic sequence, a coda+onset, the status of sC remain to be determined. If, sC is prone, like ʁC, to the deletion of the first member, we can suggest that sC behaves similarly than ʁC. This fact constitutes an argument in favor of the coda+onset representation for sC. Thus, according to these data, we propose to treat sC as a coda+onset, like ʁC, in French.

At this point, if we observed specifically the deletion of C^2, *moustique*/mustik/ →[musik], it appears that this deletion involves specifically the final position in aphasia and in acquisition. We also observed that the rate of deletion of C^2 is important in aphasia. A way to justify why they prefer to delete the second member in final position (which is always a stop), instead of the first consonant is to appeal to the CVCV model: the final position is a special position because the presence of an empty nucleus leads to the variation. In this condition, the last consonant is always weak. Moreover, we have observed that more deletion concerns [dorsal] and [coronal], in acquisition but also in aphasia. The aphasic speaker preferred to use substitution of the second consonant before deleting it. In this way, we can also consider that the nature of the consonant plays also a role in deletion process. The observation of the nature of the segments through the substitution could predict which segments are more complex than others.

In this context, the syllabic position: coda *versus* onset, and the nature of consonant seem to play an important role during acquisition and speech disorders in aphasia.

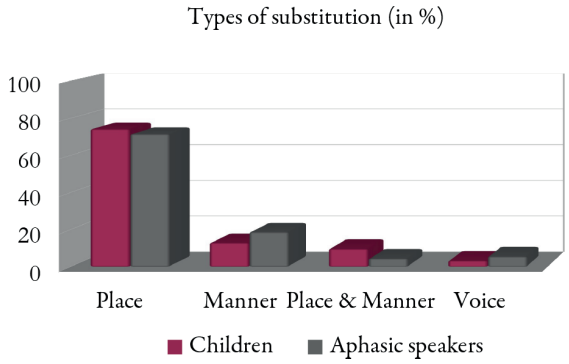
(7)	a. *mosquito*	'zanzara'	moustique	/mustik/	→	[musis]
	b. *beetle*	'scarabeo'	scarabée	/skaʁabe/	→	[saʁabe] [kaʁabe]
	c. *mask*	'maschera'	masque	/mask/	→	[mas]

If the syllabic structure is responsible for this type of transformation, we do not have an explanation for the targeting of some places of articulation. For example, when C^2 is deleted, it concerns principally [coronal] and [dorsal]. Fewer cases of [labial] were deleted. Beside deletions, another type of transformation plays an important role and can bring new elements about the complexity of place of articulations: the substitution's case.

1.3.2 Substitution

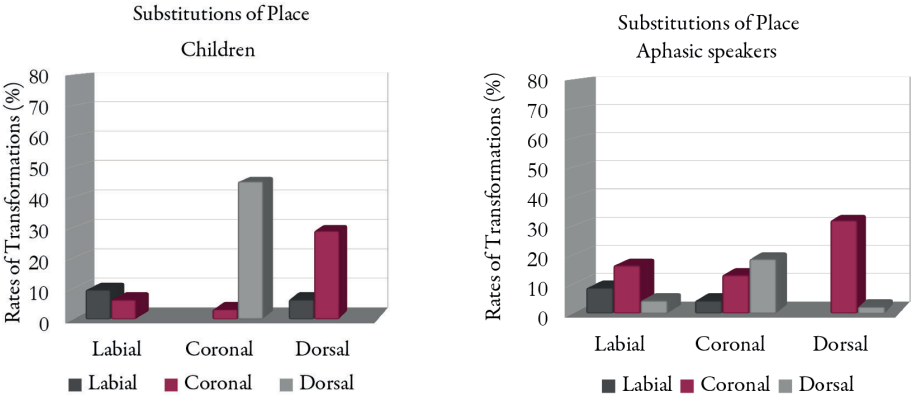
In the line of Blumstein (1973) and Nespoulous et al. (1987), we started by comparing all cases of substitutions depending on place of articulation, manner of articulation, both of them and voice.

Figure 3 - *Type of substitutions*



We can see, in the figure above that substitutions predominantly concern the place of articulations (72,13%) rather than manner or voice. Substitutions also involved mostly a single element (according to Blumstein, 1973). Data are as follows:

Figure 4 - *Substitutions' produced by children and aphasic speakers*



There is a significant difference between the kind of substitutions: for children: $\chi 2_{[4]}$ = 32,609, p-value < 0.001 and also for aphasic speakers: $\chi 2_{[4]}$ = 32,874, p-value < 0.001. We recall that C^1 is seldom substituted. In acquisition, substitution concerns 4,87% of the total of transformations (29/595 cases). Furthermore, an important point is that when C^2 is deleted, it concerns principally [coronal] and [dorsal]. Fewer cases of [labial] were substituted. [coronal] becomes [dorsal] and [dorsal] becomes [coronal]. Concerning this point, Jakobson (1969: 58) had shown that if a child produces /t/ earlier than the other consonants, at a particular moment, when a child

start to acquire /k/, there is a phase where /k/ and /t/ are used at the same stage. He also claimed that this change is not the result of a phonetic change. Unfortunately, he did not explain why /t/ and /k/ are employed similarly, contrary to the other classes.

During aphasia, substitution concerns 28,94% of the total of transformations (90/311 cases). We recall that C[1] is also rarely substituted. [dorsal] becomes [coronal] in 93,55 % of cases, [labial] is more steady but it can also be shifted into a [coronal]. Finally [coronal] stays a coronal or also becomes [dorsal]. We have noted that [labial] appears more stable and, for these reason, less complex than [coronal] and [dorsal]. Moreover, [coronal] mostly constitutes the target of transformations and are principally affected. Why? Kean (1975: 48), Paradis and Prunet (1991), Stemberger and Stoël-Gammon (1991) or Béland and Favreau (1991) have proposed that [coronal] have a special status, they are unspecified and less marked. For this reason, [coronal] can attract other elements/features[6] from other classes, like the dorsality and this is why [coronal] is mostly the target of substitution, deletion, epenthesis (in the French liaison) or assimilation, in sum, any phonological process. Following this, as many people argue, we suggest that [coronal] is a default value in the phonological system. In fact, why does [coronal] becomes mainly [dorsal] and not [labial]? The analysis shows that these cases are always cases of assimilation/harmony – due to a context – as Stemberger and Stoel-Gammon (1991), Béland and Favreau (1991), Rose (2009) and Rice (2009) have shown. For example, aphasic speakers replace /t/ with /k/, in (8), when another dorsal (or uvular) is present in the context:

(8)	*satchel*	'cartella'	cartable	/kaʁtablə/	→	[kaʁkwab]
	card	'carta'	carte	/kaʁt/	→	[kaʁk]
	watermelon	'anguria'	pastèque	/pastɛk/	→	[paskɛk] [pakɛk]

The important question here is, do [coronal] and [dorsal] have the same complexity? Like Rice (1996) argues, we will consider the default value of a segment or the special status attributed to a segment is a parameter, which depends of the kind of language. To represent [coronal] like a default value, we propose that [coronal] don't have an element for place (Paradis, Prunet, 1991; Scheer, 1999). In turn, [labial] involves a little number of transformation, it is more stable and robust, and acquired before the [dorsal] class in acquisition. [labial] includes only one element of PoA, which is {U}, for labiality/graveness, thus their relative stability in pathology. [dorsal] becomes [coronal] because it contains more structure, two elements of place of articulation: {I} and {U}, hence their capacity of rejection. Then, [dorsal] is more marked and more complex. [dorsal] into [coronal] corresponds to a lenition,

[6] More than *feature theory*, *element theory* (e.g. Harris, 1994; Scheer, 1999; Cyran, 2010; Backley, 2011) is a way to define the internal constitution of phonemes and can provide us with a direct measure of complexity and markedness. Elements are small units which characterize the properties of a segment, they are present (marked) or absent (unmarked) of a segment, like a tool to measure the complexity. For example the labial /p/ is {Uh?}: {U} for labiality, {h} for noise, contrary to fricative, because /p/ is a stop and {?} for occlusion, [+continuous]. /b/ share the same property plus {L} which mean is voiced. /b/ is more marked than /p/ because he received one element more.

in terms of structure. In this way, the following scale of element theory can explain most of these patterns:

[coronal]	<	[labial]	<	[dorsal]
{hʔ}		{UʔH}		{IUʔH}

In this line, we propose that most of the substitution are cases of segmental lenition: IU > Ø or U > Ø

(9)	a. *beetle*	'scarabeo'	scarabée	/skaʁabe/	→	[staʁabe]
				IU	>	Ø

Meanwhile, the observation of the last strategy: the reduction of CC cluster into another segment may allow us to make some progress.

1.3.3 Reduction of CC clusters CCv → Cv

CCv → Cv is the result of a substitution and a deletion, what we call *reduction*. In this context, cluster is becoming a consonant but neither C^1, nor C^2. They lead to a deletion of the CC structure for a Cv structure.

In acquisition, the reduction concerns 11,43% of the cases of transformation (68/595).

(10)	*beetle*	'scarabeo'	scarabée	/**sk**aʁabe/	→	[paʁape]
	skeleton	'scheletro'	squelette	/**sk**ølɛt/	→	[pəlɛt]
	snail	'lumaca'	escargot	/ɛ**sk**aʁgo/	→	[te**t**ago] [ge**g**ago]
	cap	'casco'	casquette	/ka**sk**ɛt/	→	[ta**t**ɛt]
	watermelon	'anguria'	pastèque	/pa**st**ɛk/	→	[pa**k**ɛk]
	porcupine	'porcospino'	porc-épic	/pɔ**ʁk**epik/	→	[potepik]

In aphasia speech, the reduction concerns 13,18% of the cases of transformation (41/311)

(11)	*spatula*	'spatola'	spatule	/**sp**atyl/	→	[**t**atyd]
	beetle	'scarabeo'	scarabée	/**sk**aʁabe/	→	[**t**aʁabe]
	snake	'serpente'	serpent	/sɛʁpɑ̃/	→	[tɛtɔ̃]
	computer	'computer'	ordinateur	/ɔʁdinatœʁ/	→	[ɔtinatœʁ]

Again, the position doesn't matter but the nature of the consonant does. We have noticed that total reduction is prominent in cluster with [dorsal] and not with [labial]. And, most of the time, a [coronal] segment emerges in C^3. [coronal], again, seems to have a special status and the strategies applied during acquisition and in pathology are not random, they follow a hierarchy of complexity as suggested by Jakobson.

To conclude, the following table recapitulates the principal transformations we have seen depending on the position:

Table 4 - *Transformations depending on the position*[7]

	Clusters	
	Children	**Aphasic speakers**
#sC[7]	deletion C^1	deletion C^1
		deletion C^2
$VC^1.C^2V$	deletion C^1	substitution C^2
	CCv → Cv	deletion C^1
		deletion C^2
VC^1C^2#	deletion C^2	substitution C^2
	deletion C^1	deletion C^2
		CCv → Cv

We have seen that different paths coexist to lead to the weakening. Children and aphasic speakers tend to reduce the most marked structures but in a different way. They can delete the CC, delete a member to adopt a CV structure or they can also reduce the segmental content through the substitution. To resume now, what does this reduction of complexity teach us?

Theoretical points: what does reduction teach us?
Concerning the empirical research findings for the phonological theory

About the consonantal clusters. Based on the transformations observed, children have shown a preference for the CV template. They produced more transformations on the syllabic structure than on the segmental content. Regardless of the position (initial, medial or final) they always reduced the consonantal clusters through the deletion of segment. Deletion of C^1 shows clearly that the fricative is in coda position and constitutes an evidence in favor of a representation in a government phonology. sC clusters, as ʁC, are heterosyllabic (Kaye, 1992; Goad, 2011). However, if the syllabic structure is affected in the same way during aphasia, the segmental structure posed more difficulties than in acquisition. If the transformations in initial position exhibit a preference for the CV structure, aphasic speakers adopted another kind of strategy for consonantal cluster in medial and final position: the substitution of the stop. This fact has reported that segmental content is more affected in adult speech disorders than in acquisition.

About the complexity between the nature of segments. If the [labial] class is more steady, the contrast between [coronal] and [dorsal] has not yet been set in acquisition. This is the reason why children used harmony/assimilation between [dorsal] and [coronal]. We can conclude that [labial] is less complex. Furthermore, if the data are similar in aphasia, they clearly show a preference for the [coronal]. [labial] and also [dorsal] turn into [coronal]. For these reasons, we have been proposed that [coronal] is the less marked structure, in terms of markedness, as argued by Paradis, Prunet (1991).

[7] We precise that ʁC doesn't exist in french in initial position.

The process of substitution results from adjustments and parameter setting. They correspond to a transformation, which disrupts the internal constitution of segments.

Concerning the Jakobson's postulate

Deletions of C^1, C^2 or substitutions showed a preference to reduce the syllabic structure or the segmental content. The collapsing of the whole cluster, which combines a substitution and a deletion, also leads to a reduction of complexity of the cluster into a Cv string. Consequently, if different paths are possible, the goal is still to reduce the complexity. We have also seen that aphasic speakers and children use the same processes – depending on the context – but statistics are useful to encourage further thinking.

As far as acquisition is concerned, deletion reflects the fact that the consonantal clusters and the coda position has not yet been apprehended by the phonological system. French children, during acquisition, show a preference for the weakening of the syllabic structure. Deletion of /s/ and /ʁ/ is the consequence of this weakening. Children have a clear preference for the CV template. The consonantal clusters are not yet acquired. After the setting of consonantal clusters parameter, another type of strategy, substitution, indicates that children try to test the possible values in different positions depending on the parameters of their language.

We notice that three main stages can be considered during the acquisition of a syllabic position and acquisition of the possible values in this position – depending on the parameters on the language.

The first one corresponding to the deletion C^1C^2, the second is the deletion of one of the member C^1 or C^2, then, when the CC group is acquired, substitution can appear (Greenlee, 1974). For these cases, these substitutions mainly concern the stop. According to Fikkert (2007):

> Moreover, development was formalized as the change of one or more parameter settings from the default to the marked value, thereby assuming continuity from children's early grammars to the final state grammar. Markedness played a key role in the principles and parameter framework: in the initial state all parameters have the default or unmarked value. In the course of development, parameters are changed to their marked setting if the input data provide positive evidence for that setting[8].

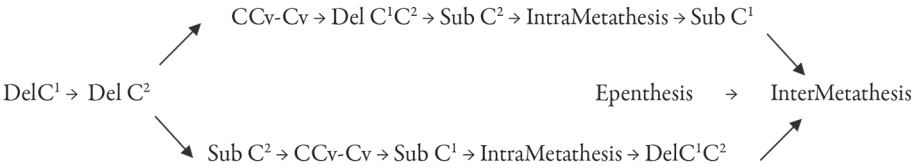
We have also seen that aphasic speakers tend to reduce the complexity but in a different way. If the access to the constraints on the syllabic structure is defective, the aphasic speaker will delete the coda position, like children. They also have a preference for a CV structure. In other cases, aphasic speakers have mostly deleted or substituted the second member, which is always a stop. We suggest that the nature of the consonant, particularly the place of articulation, is more problematic than the syllabic structure. Moreover, some other data, not presented here, have shown a big tendency wherein aphasics tend to substitute. To explain this, the proposal we are making is the following: during aphasia, access to the segmental constraints is

[8] Fikkert (2007: 548).

sometimes more problematic than the access to constraints of syllabic structure in acquisition.

To summarise the different strategies they made, the following scheme represents a scale where the deletion constitutes the most important transformation for both children and aphasic speakers and where epenthesis and metathesis are less employed.

Children



Aphasic speakers

To conclude for this point, we suggest that children favor segmental structure over syllabic structure while aphasic exhibit the opposite behaviour.

However, children seem to follow a pattern where deletion precedes the substitution and the reverse in aphasia. Therefore, regarding the strategies applied by children and aphasic speakers, Jakobson's postulate concerning a mirror image is still relevant but requires a more detailed analysis.

We have seen that when a subject suffers from aphasia, their phonological constraints undergo change. The value of certain parameters doesn't correspond to the value of the parameters of the language. The aphasic speaker uses different strategies to resolve these conflicts. This observation allows us to answer fundamental questions in phonology concerning the status of consonantal sequences and the internal structure of segments in particular. We saw in this study that constraints on segmental and syllabic content can generate different transformations; which allows us to account for a number of performance errors. Finally, studying data in acquisition and in speech disorders in adult helps our understanding of what is really the functioning of the phonological system, as Goldrick (2011: 19):

> Thus, although speech errors data are clearly influenced by grammatical principles, they present a mostly unexplored empirical domain for testing the predictions of grammatical theories[9].

References

Backley, P. (2011). *An introduction to Element Theory*. Edinburgh: Edinburgh university Press.

Baqué, L., Marczyk, A., Rosas, A., Estrada, M., Le Besnerais, M. & Nespoulous, JL. (2012). De la matière phonique à la structuration phonologique dans l'aphasie. In

[9] Goldrick (2011: 19).

Calvo, M.V., Murillo, J. (Eds.), *Perception phonique et parole.* Percepción fónica, habla y hablar. Mons (Belgique): Centre International de Phonétique Appliquée (CIPA), 75-98.

Barlow, J.A. (2004). Consonant clusters in phonological acquisition: Applications to assessment and treatment. In *CSHA Magazine*, 34, 10-13.

Barlow, J.A. (2001). The structure of /s/-sequences: evidence from a disordered system. In *Journal of Child Language.* Cambridge University Press, 291-324.

Barlow, J.A (2001a). A preliminary typology of initial clusters in acquisition. In *Clinical Linguistics & Phonetics*,15, 9-13.

Béland, R., Paradis, C. (1997). Principled syllabic dissolution in a primary progressive aphasia case. In *Aphasiology*, 12, 1171-1196.

Béland, R., Paradis, C. & Bois, M. (1993). Constraints and Repairs in Aphasic Speech: a group study. In *Revue canadienne de linguistique.* Toronto, 279-303.

Béland, R., Favreau, Y. (1991). On the special status of Coronals in Aphasia. In Paradis, C., Prunet, J.-F. (Eds.), *The special status of Coronals: Internal and External Evidence.* San Diego: Academic Press, 201-221.

Bernhardt, B., Stemberger, J. (2007). Phonological Impairment in children and adults. In De Lacy, P. (Ed.), *The Cambridge Handbook of Phonology.* Cambridge: Cambridge University Press, 575-595.

Blumstein, S.E. (1978). Segment structure and the syllable in aphasia. In Bell, A., Hooper, J.B. (Eds.), *Syllables and segments.* Amsterdam: North Holland Publishing Company, 189-200.

Blumstein, S.E., Backer, E. & Goodglass, H. (1977). Phonological factors in auditorycomprehension in aphasia. In *Neuropsychologia*, 15, 19-29.

Blumstein, S.E. (1973). A phonological investigation of aphasic speech. In *Janua Linguarum Serie Minor*, 153. The Hague: Mouton.

Calabrese, A. (2005). *Markedness and Economy in a Derivational Model of Phonology.* Mouton de Gruyter.

Caramazza, A. (1994). Parallels and Divergences in the Acquisition and Dissolution of Language. In *Philosophical Transactions; Biological Sciences.* 346 (1315), 121-127.

Chin, S.B., Dinnsen, D.A. (1992). Consonant clusters in disordered speech: constraints and correspondence patterns. In *Journal of Child Language*, 19, 259-85.

Cyran, E. (2010). *Complexity Scales an Licensing in Phonology*. Berlin-New-York: Mouton de Gruyter.

Demuth, K., Kehoe, M. (2006). The acquisition of Word-final Clusters in French. In *Catalan Journal of Linguistics*, 5, 59-81.

Demuth, K., McCullough, E. (2009). The longitudinal development of clusters in French. In *Journal of Child Language*, 36. Cambridge: Cambridge University Press, 425-448.

Durand, J., Prince, T. (2015). Phonological markedness, acquisition and language pathology: what is left of the Jakobsonian legacy? In Astesano, C., Jucla, M. (Eds.), *Neuropsycholinguistic Perspectives on Language Cognition.* Series: Explorations in Cognitive Psychology. London: Routledge, Taylor & Francis group.

Den Ouden, D.B., Bastiaanse, R. (2003). Syllable structure at different levels in the speech production process: Evidence from Aphasia. In Van de Weijer, J., Van Heuven, V.J. & Van der Hulst, H. (Eds.), *Current Issues in Linguistic Theory, The Phonological Spectrum, vol. II: Suprasegmental Structure*, 234. John Benjamins, 81-107.

Den Ouden, D.-B. (2002). *Phonology in Aphasia, syllables and segments in level-specific deficits. Doctoral dissertation. Grodil 39*: Rijkuniversiteit Groningen.

Fikkert, P. (2007). Acquiring Phonology. In De Lacy, P. (Ed.), *The Cambridge Handbook of Phonology.* Cambridge: Cambridge University Press, 537-555.

Fikkert, P., Freitas, M.-J., Grijzenhout, J., Levelt, C. & Wauquier-Gravelines, S. (2004). Syllabic markedness, segmental markedness, rhythm, and acquisition, *GLOW Conference*, Thessalonique.

Fikkert, P. (2000). Acquisition of Phonology. In Cheng, L., Sybesma, R. (Eds.), The first Glot International Test-of-the-Article Book. The latest in Linguistics. Studies in Generative Grammar 48. Berlin/New York: Mouton de Gruyter, 221-250.

Fikkert, P. (1994). On the acquisition of rhyme structure in Dutch. In Bok-Bennema, R., Cremers, C., *Linguistics in the Netherlands.* Amsterdam, Philadelphia: John Benjamins, 37-48.

Freitas, M.J. (1997). *Aquisição da estrutura silábica do Português Europeu*. Doctoral dissertation, University of Lisboa.

Goad, H. (2012). sC clusters are (almost always) coda-initial. In *The Linguistic Review*. Walter de Gruyter, 29, 335-373.

Goad, H. (2011). The representation of sC clusters. In van Oostendorp, M., Ewen, C., Hume, E. & Rice, K. (Eds.), *The Blackwell companion to phonology*. Oxford: Wiley-Blackwell, 898-923.

Goldrick, M. (2011). Linking Speech Errors and generative phonological theory. In *Language and Linguistics Compass*, 5, 397-412.

Greenlee, M. (1974). Interacting processes in the child's acquisition of stop-liquid clusters. In *Papers and Reports on Child Language Development*, 7, 85-100.

Harris, J. (1994). *English sound structure.* Oxford: Blackwell.

Jakobson, R. (1968). *Child Language: Aphasia and Phonological Universals.* The Hague: Walter de Gruyter (revised version of *Kindersprache, Aphasie und allgemeine Lautgesetze*, Uppsala, 1941). (1969 in French).

Kappa, I. (2002). On the acquisition of syllabic structure in Greek. In *Journal of Greek Linguistics*, 3, 1-52.

Kaye, J. (1992). Do you believe in magic? The story of s+C sequences. In SOAS *Working Papers in Linguistics and Phonetics*, 2, 293-313.

Kehoe, M., Hilaire Debove, G., Demuth, K. & Lleo, C. (2008). The structure of Branching Onsets and Rising Diphtongs: Evidence from the Acquisition of French and Spanish. In *Language Acquisition, A journal of Developmental Linguistics*, 15 (1), 5-57.

Kean, M.-L. (1975). *The theory of markedness in generative grammar.* Ph.D dissertation, MIT. Georgetown University.

Kirk, C. (2008). Substitution Errors in the Production of Word-initial and Word-final Consonant Clusters. In *Journal of Speech Language and Hearing Research*, 51, American Speech Language-Hearing Association, 35-48.

Levelt, C., Schiller, N. & Levelt, W. (1999/2000). The Acquisition of Syllables Types. In *Language Acquisition,* 8 (3), 237-264.

Moreau, N. (1993). *Apport de la théorie « des contraintes et des stratégies de réparation » à l'étude des paraphasies phonémiques dans l'aphasie*, Thèse de doctorat, Université de Toulouse-le-Mirail.

Morrisette, M.L., Dinnsen, D.A. & Gierut, J.A. (2003). Markedness and Context Effects in the Acquisition of Place Features. In *Canadian Journal of Linguistics*, 48 ¾, 329-355.

Nespoulous, J.-L. (1998). Segmental errors in aphasia as (Psycho)linguistic evidence of the nature of phonological representations. *Clinical phonetics and Linguistics.* In Ziegler, W., Deger, K. (Eds.). London: Whurr Publishers Ltd, 323-335.

Nespoulous, J.-L., Joanette, Y., Ska, B., Caplan, D. & Lecours, A.R. (1987). Production deficits in Broca's and Conduction aphasia: repetition vs reading. In Keller, E., Gopnik, M. (Eds.), *Motor and sensory processes of language.* Hillsdale: Lawrence Erlbaum Associate Inc., 53-81.

Pan, N., Snyder, W. (2004). Acquisition of /s/ initial Clusters: A Parametric Approach. Proceedings of the 28th Annual Boston University Conference on Language Development. In Brugos, A., Micciulla, L. & Smith, C.E. (Eds.). Somerville, MA: Cascadilla Press, 436-446.

Paradis, C., Prunet, J.-F. (1991). The Special Status of Coronals. Internal and External Evidence. In *Phonetics and phonology*, 2. San Diego: Academic Press.

Rice, K. (2009). On coronals: Are they special? In Arsenault, P., Jarmasz, LG., Kim, K. & Radišic, M. (Eds.), *Toronto Working Papers in Linguistics, Special Issue on Coronal Phonology*, 30, 1-15.

Roch Lecours, A., Lhermitte, F. (1979). *L'aphasie.* Flammarion médecine sciences, Université de Montréal.

Romani, C., Galluzzi, C., Bureca, I. & Olson, A. (2011). Effects of syllable structure in aphasic errors: implications for a new model of speech production. In *Cognitive Psychology*, 62, 151-192.

Romani, C., Galluzzi, C. (2005). Effects of syllabic complexity in predicting accuracy of repetition and direction of errors in patients with articulatory and phonological difficulties. In *Cognitive Neuropsychology*, 22 (7), 817-850.

Romani, C., Calabrese, A. (1998). Syllabic Constraints in the Phonological Errors of an Aphasic patient. In *Brain and language*, 64, 83-121.

Rose, Y. (2009). Internal and External Influences on Child Language Productions. In Marsico, E., Pellegrino, F. (Eds.), *Approaches to Phonological Complexity*. Berlin: Mouton de Gruyter, 329-353.

Rose, Y., Wauquier-Gravelines, S. (2007). French Speech Acquisition. In *The International Guide to Speech Acquisition*. In McLeod, S. (Ed.). Clifton Park, NY: Delmar Thomson Learning, 364-384.

Scheer, T. (2015). *Précis de structure syllabique, accompagné d'un apparat critique*. Paris: ENS édition.

Scheer, T. (2004). *A lateral Theory of Phonology: What is CVCV and what it should be?* Berlin: Mouton de Gruyter.

Scheer, T. (1999). A Theory of Consonantal Interaction. In *Folia Linguistica*, 32, 201-237.

Stemberger, J.P., Stoel-Gammon, C. (1991). The underspecification of coronals: evidence from language acquisition and performance errors. In Paradis, C., Prunet., J.-F. (Eds.), *The special status of coronals. Internal and external evidence.* San Diego: Academic Press, 181-199.

Valdois, S. (1987). *Les erreurs d'addition et d'omission dans l'aphasie: rôle du gouvernement phonologique*, Thèse de doctorat, University of Montréal.

DOI: 10.17469/O2101AISV000016

MANUELA BOLLANI

La percezione dello stimolo ritmico tra acquisizione, sintonizzazione e perdita: stato dell'arte e alcune riflessioni

The infant shows keen perceptive abilities and high responsiveness to languages in terms of both rhythmic and prosodic stimulations. However, during its first year, the scope of the infant's perceptive ability gradually narrows down, finally tuning into the typical acoustic traits of its native language and into the rhythmic and melodic features of the musical background provided by its cultural environment. This ability of accurately perceiving rhythm seems to be impaired in subjects with linguistic disorders: people affected by Dyslexia or Specific Language Impairment in fact show problems in rapid temporal processing. In this regard, a brief overview of the literature about rhythm perception will pinpoint significant analogies between acquisition, tuning and loss processes.

1. *Introduzione*

"Il ritmo è la denominazione dell'ordine del movimento". Così Platone (5°-4° secolo a.C.) nelle Leggi (II, 664E) definisce il ritmo unendo il concetto di movimento, quale fluire temporale del gesto, a quello di ordine, come periodicità e strutturazione. Il dibattito sulla relazione tra il concetto di periodicità e quello di ritmo è ancora aperto e, secondo Righini (1994), il ritmo sarebbe una categoria psicologica umana che solo per un atto della volontà percettiva andrebbe a creare un raggruppamento gerarchico laddove ci sarebbe solo pura ripetizione periodica, che di per sé non basterebbe a creare ritmo. D'altra parte, Patel (2008) sottolinea come la periodicità sia solo una delle possibili modalità di organizzazione ritmica, definendo il ritmo come "creazione di modelli sistematici di suoni in termini di tempo, accento e raggruppamento" ma non necessariamente regolari nel tempo.

La questione è particolarmente rilevante per comprendere il ruolo della componente ritmica del linguaggio, che ha avuto una lunga associazione con il concetto di periodicità (Allen, 1975; Keller, Port, 2007). In fonologia, il ritmo è stato studiato prevalentemente nell'ambito del modello della fonologia metrica (Hayes, 1995; Marotta, 2003), mentre in fonetica sperimentale lo studio del ritmo è associato al tema dell'isocronia e alla suddivisione in lingue a isocronismo accentuale, a isocronismo sillabico o a isocronismo moraico (Pike, 1945; Bertinetto, 1977; Bertinetto, 1981). Successivamente, la ricerca empirica si è mossa verso una concezione più ampia di ritmo del discorso sulla base di come le lingue si differenzino nella scansione temporale di vocali, consonanti e sillabe. In questa direzione, dai primi fondamen-

tali contributi di Dauer (1983) e Ramus e colleghi (1999), l'indagine fonetica sul ritmo linguistico si è arricchita di numerosi studi (Grabe, Low, 2002; Bertinetto, Bertini, 2008)[1].

Ad ogni modo, la percezione del ritmo svolge un ruolo centrale sia nell'acquisizione del linguaggio che, in senso più ampio, nello sviluppo psicologico e sensomotorio dell'individuo. I primissimi impulsi percepiti dal feto sono di natura ritmica e sarebbe proprio la biomeccanica del ritmo a regolare i processi di anticipazione percettiva del cervello che consentono di strutturare l'interazione con l'esterno (Benenzon, 1982; Trevarthen, 1999). In una prospettiva che integri diversi ambiti di ricerca legati allo sviluppo dell'individuo, il presente lavoro vuole fornire una breve panoramica degli studi riguardanti la percezione del ritmo nelle fasi di acquisizione e sintonizzazione linguistica e nei meccanismi connessi a deficit del linguaggio.

2. *Percezione ritmica e acquisizione linguistica*

2.1 La precoce discriminazione linguistica su base ritmica

L'udito è il primo tra i cinque sensi a svilupparsi nell'essere umano e, per il feto, il mondo esterno è originariamente un'esperienza uditiva, filtrata attraverso il liquido amniotico (Querleu et al., 1981). L'apparato uditivo comincia a formarsi dalla terza settimana di gestazione e completa la sua struttura intorno al quinto mese. Feti di 36-40 settimane reagiscono a cambiamenti delle caratteristiche fisiche dei suoni: alcuni esperimenti (Lecanuet, Granier-Deferre, 1993) hanno mostrato, attraverso la tecnica di decelerazione cardiaca[2], come i feti notassero il cambiamento tra due sequenze sillabiche simili. L'utero materno è quindi il primo ambiente sonoro e preparerebbe il futuro neonato a sintonizzarsi sulle caratteristiche acustiche dei suoni linguistici. Infatti, a pochi giorni di vita, il neonato dispone di abilità percettive molto raffinate e risulta in grado di discriminare tra diverse lingue: non solo tra la lingua materna ed una lingua straniera mai sentita prima d'allora (Mehler et al., 1988; Nazzi et al., 2000), ma anche tra due lingue straniere, entrambe mai sentite prima (Mehler et al., 1988; Nazzi et al., 1998; Nazzi et al., 2000).

Vari studi hanno proposto che i neonati riescano a discriminare tra lingue diverse sulla base del ritmo (Mehler et al., 1996; Ramus et al., 1999). Questa ipotesi presuppone che siano in grado di costruire rappresentazioni ritmiche in termini di sillabe, ovvero sequenze di vocali di varia durata ed intensità, interrotte da rumore consonantico. Vari studi hanno dimostrato, utilizzando per lo più la tecnica della suzione non nutritiva[3], che i neonati discriminino molto presto tra vocali diverse,

[1] Per una rassegna più approfondita sull'argomento si veda il numero tematico della rivista *Phonetica* (Kohler, 2009) e il contributo di Barry negli atti del V Convegno Nazionale AISV 2009 (Barry, 2010).

[2] Una decelerazione nel battito cardiaco dei feti in concomitanza con la presentazione di uno stimolo nuovo, dopo una fase di abituazione, indica l'avvenutadiscriminazione.

[3] Il neonato succhia una tettarella collegata ad un trasduttore, che misura la velocità di suzione del neonato. Dopo una fase di abituazione, i neonati succhieranno più velocemente quando percepiranno uno stimolo nuovo (Jusczyk, 2000).

anche se non presenti nella lingua di esposizione (Trehub, 1973; Kuhl et al., 1992) e che siano specificatamente sensibili al nucleo vocalico della sillaba (Bertoncini et al., 1988). In particolare, gli esperimenti di Bijeljac-Babic, Bertoncini e Mehler (1993) mostrano come i neonati siano in grado di discriminare tra sequenze con diverso numero di sillabe mentre non percepiscano alcun cambiamento nel numero dei segmenti fonici se il numero di sillabe rimane invariato. Questi risultati suggeriscono che i neonati siano in grado di crearsi una rappresentazione in termini di sillabe.

L'ipotesi secondo cui i neonati discriminino tra lingue diverse sulla base del ritmo si basa sull'idea che le lingue siano organizzate in classi ritmiche (Pike, 1945; Bertinetto, 1977; Bertinetto, 1981): lingue a isocronismo accentuale, come ad esempio inglese, olandese e russo; lingue a isocronismo sillabico, come ad esempio francese e italiano e lingue a isocronismo moraico, come ad esempio giapponese e turco. I risultati dei maggiori studi riguardo la precoce discriminazione linguistica (Mehler et al., 1988; Christophe, Morton, 1998; Nazzi et al., 1998; Nazzi, Jusczyk, 1999)[4] andrebbero a supporto dell'ipotesi che i neonati siano in grado di discriminare tra classi ritmiche di lingue. Dai dati riportati nella tabella 1, che riassume i risultati di alcuni dei maggiori studi in questione, emerge che i neonati, a pochi giorni di vita, riescono a discriminare tra lingue appartenenti a classi ritmiche differenti, sia nel caso in cui il contrasto includa la loro lingua nativa (Mehler et al., 1988) sia nel caso in cui il contrasto riguardi due lingue entrambe non familiari (Nazzi et al., 1998). D'altraparte, i neonati non risultano in grado di discriminare tra coppie di lingue appartenenti alla stessa classe ritmica (ad esempio inglese ed olandese) se non dopo i 5 mesi di vita (Christophe, Morton, 1998; Nazzi et al., 1998; Nazzi, Jusczyk, 1999). È possibile che dopo i 5 mesi, la rappresentazione che gli infanti si costruiscono della lingua sia più raffinata oppure si basi su informazioni diverse dal ritmo, come i vincoli fonotattici (Jusczyk et al., 1993).

Tabella 1 - *Risultati di esperimenti di discriminazione linguistica (adattata da Guasti, 2007)*

Lingue Discriminate	*L1*	*Età*	*Riferimento*
Francese-Russo	Francese	4 giorni	Mehler et al., 1988
Inglese-Olandese*	Francese	4 giorni	Nazzi et al., 1998
Inglese-Giapponese	Francese	5 giorni	Nazzi et al., 1998
Inglese-Italiano	Inglese	2 mesi	Mehler et al., 1988
Inglese-Olandese*	Inglese	2 mesi	Christophe, Morton, 1998
Inglese-Olandese	Inglese	5 mesi	Nazzi, Jusczyk, 1999

* Non si ha discriminazione

[4] Gli studi in questione utilizzano per lo più il metodo della suzione non nutritiva e tecniche di manipolazione e filtraggio dello stimolo linguistico.

2.2 L'utilizzo degli indizi ritmici: la segmentazione del flusso sonoro continuo e l'ipotesi del Bootstrapping fonologico della sintassi

La precoce capacità percettiva dell'infante risulta fondamentale per il processo di acquisizione linguistica e, in particolare, la sensibilità agli indizi ritmici e prosodici del linguaggio renderebbe possibile la segmentazione del flusso sonoro continuo del parlato percepito. L'*Infant Directed Speech* è solitamente molto ricco da un punto di vista prosodico e i confini di frase e dei costituenti prosodici sono solitamente segnalati dai fenomeni di declinazione naturale: pause, F0 discendente e sillaba finale allungata (Fernald et al., 1989). Ciononostante, pause non affidabili, fenomeni di co-articolazione, assenza di riferimenti lessicali di partenza e variabili sonore legate alla vocalità del parlante rendono la segmentazione in unità discrete e l'individuazione dei confini di parola nel parlato spontaneo un'attività particolarmente complicata per il bambino.

Grazie all'integrazione tra un meccanismo generale di elaborazione statistica (*Statistical Learning Mechanisms*, Saffran et al., 1996) e l'utilizzo di indizi prosodico-ritmici, specificatamente linguistici, i neonati già verso i 7-8 mesi sarebbero in grado di segmentare il parlato (Jusczyk, Aslin, 1995). Attraverso il meccanismo di *Statistical Learning* che calcola le probabilità di transizione tra eventi, i bambini risultano in grado di elaborare regolarità distribuzionali, ovvero informazioni statistiche circa la probabilità di adiacenza tra due eventi linguistici. In un esperimento di Saffran e colleghi (1996), neonati di 8 mesi sono risultati in grado di elaborare le differenti probabilità di transizione tra sillabe adiacenti e, di conseguenza, individuare le sequenze che avrebbero potuto formare delle parole. I neonati sarebbero quindi in grado di calcolare la frequenza tra sillabe adiacenti ed usare questa informazione per segmentare il parlato.

Ad ogni modo, la segmentazione del flusso continuo del parlato è un processo che richiede l'integrazione tra i meccanismi dello *Statistical Learning* e l'elaborazione di molteplici indizi linguistici, quali il numero di sillabe, la posizione dell'accento e i vincoli fonotattici. Si è visto come i neonati di pochi giorni siano in grado di crearsi una rappresentazione in termini di sillabe e notino se il numero di sillabe cambi o meno (Bijeljac-Babic et al., 1993). Di conseguenza, è possibile che i bambini facciano congetture sulla lunghezza delle parole e possano attendersi dei confini di parola dopo un certo numero di sillabe a partire dalla struttura ritmica di una lingua e quindi dal repertorio sillabico (Mehler, Nespor, 2004).

Inoltre, in base alla posizione dell'accento, i bambini possono prevedere dove collocare un confine di parola. In un esperimento di Christophe e colleghi (1994) con l'utilizzo della tecnica della suzione non nutritiva (vedi nota 3), dei neonati francesi di solo 3 giorni di vita sono riusciti a discriminare tra due diverse tipologie della stessa sequenza bisillabica: in un caso appartenente ad un'unica parola, in un altro caso utilizzata come ultima sillaba e prima sillaba di due parole differenti. Le due sequenze differivano nella configurazione accentuale, che nel francese prevede che la sillaba tonica sia l'ultima della parola. La posizione dell'accento sembra essere un indizio rilevante in particolare per le lingue tradizionalmente considera-

te come *stress-timed* (Lee, 2010). Per quanto riguarda l'inglese, Jusczyk e colleghi (1993) ipotizzano che i neonati futuri parlanti inglesi utilizzino una Strategia di Segmentazione Metrica (SSM) che consisterebbe nel supporre che ogni volta che viene percepita una sillaba forte inizi una nuova parola. Questa strategia si basa sull'analisi di un corpus linguistico inglese che mostrerebbe la maggiore frequenza nella lingua inglese di strutture bisillabiche trocaiche rispetto a quelle giambiche (Cutler, Carter, 1987). In accordo con la SSM, i neonati esposti all'inglese a 7 mesi riconoscono le parole trocaiche ma non quelle giambiche. Il fatto che a partire dai 10-11 mesi i bambini risultino in grado di riconoscere anche le parole giambiche suggerisce che inizino ad utilizzare altri indizi, oltre a quelli legati alla posizione accentuale, per segmentare il parlato (Jusczyk, Houston, 1999).

La sensibilità per i diversi indizi emerge in differenti momenti dello sviluppo: tra i 7 mesi e mezzo e gli 8 emerge la sensibilità per gli indizi riguardanti la forma tipica delle parole e le regolarità distribuzionali (Saffran et al., 1996; Jusczyk, Houston, 1999), e verso i 9 mesi la sensibilità per i vincoli fonotattici (Jusczyk et al., 1993), che regolano quali sequenze di suoni siano permesse all'interno, all'inizio o alla fine di una parola in una data lingua e quali non lo siano. In un esperimento di Jusczyk e colleghi (1993) i bambini americani a 9 mesi, ma non a 6, sono risultati in grado di discriminare tra liste di parole olandesi e liste di parole inglesi formate dallo stesso repertorio di suoni ma differenti per quanto riguarda i vincoli fonotattici. Nel caso di filtraggio che rimuovesse l'informazione fonemica, i bambini non erano in grado di discriminare tra parole olandesi ed inglesi in quanto, come si è già visto (Christophe, Morton, 1998; Nazzi et al., 1998; Nazzi, Jusczyk, 1999), l'informazione prosodica può non essere sufficiente per discriminare tra lingue appartenenti alla stessa classe ritmica.

Infine, secondo l'ipotesi del *bootstrapping* fonologico della sintassi il bambino potrebbe utilizzare gli indizi ritmico-prosodici per elaborare aspetti del sistema sintattico di una lingua (Morgan, Demuth, 1996; Fikkert, 2007; Demuth, 2011). Secondo il principio di attivazione ritmica, il neonato sarebbe in grado di percepire la prominenza associata alla vocale accentata all'interno del sintagma ed identificare il parametro d'ordine dei costituenti di una data lingua, ovvero se testa-complemento o complemento-testa: se all'interno del sintagma fonologico il bambino sente la prominenza a destra allora l'ordine dei costituenti è testa-complemento, come nel caso di lingue come l'italiano e il francese, se all'interno del sintagma fonologico il bambino sente la prominenza a sinistra allora l'ordine dei costituenti è complemento-testa, come nel caso di lingue come il giapponese e il turco (Nespor et al., 1996). I risultati di un esperimento di Christophe, Nespor, Guasti e Van Ooyen (2003) mostrano come i neonati francesi a 2 mesi siano in grado di discriminare tra due lingue con diversa distribuzione della prominenza, come il francese, con prominenza a destra, ed il turco, con prominenza a sinistra.

3. *Verso la sintonizzazione percettiva*

Durante il primo anno di vita la capacità percettiva del bambino va sintonizzandosi sui parametri sonori della lingua nativa. Fino ai 6-8 mesi i neonati possono ancora discriminare non solo tra suoni appartenenti alla lingua a cui sono esposti, ma anche tra suoni appartenenti a lingue mai sentite. Quest'abilità tende a scomparire verso gli 8-10 mesi ed è del tutto scomparsa a 11-12 mesi, quando l'infante inizia a produrre le prime parole nella lingua madre (Werker, Tess, 1984).

Questa sintonizzazione avviene più precocemente per i suoni vocalici rispetto a quelli consonantici: neonati di 4-6 mesi sono in grado di discriminare tra contrasti vocalici non-nativi, mentre neonati di 6-8 mesi non risultano più in grado (Kuhl et al., 1992; Polka, Werker, 1994) mentre, per quanto riguarda i contrasti non-nativi tra consonanti, i bambini risultano in grado di discriminarli a 6-8 mesi ma non dopo i 10-12 mesi (Werker, Tees, 1984). Inoltre, attorno ai 9 mesi, i bambini diventano particolarmente sensibili al contorno accentuale della loro lingua nativa (Jusczyk, 2000). Ad ogni modo, il processo di sintonizzazione percettiva rimane un fenomeno indagato soprattutto a livello segmentale e molto meno a livello soprasegmentale.

La precoce e ampia capacità discriminatoria del neonato ed il fenomeno di sintonizzazione percettiva fanno pensare, come già osservato da Jakobson (1941), che il processo di acquisizione del sistema fonologico di una lingua possa consistere in una sorta di selezione sonora da un bagaglio più ampio. Secondo il modello *Maintenance/Loss* (Aslin, 1981; Werker, Tees, 1984), la capacità percettiva del bambino sarebbe già al suo apice prima dell'acquisizione linguistica vera e propria e, successivamente, verrebbero mantenute quelle abilità che trovano riscontro nello stimolo percepito nella lingua nativa mentre verrebbero perse quelle abilità relative a parametri e categorie non rilevanti per la lingua nativa.

Il meccanismo di sintonizzazione percettiva ha un corrispondente analogo in ambito musicale (Scott et al., 2007) e ciò fa pensare che possa essere un fenomeno più generale che coinvolge il trattamento dello stimolo acustico, non specificatamente linguistico. Alcuni esperimenti mostrano come i bambini occidentali a 6 mesi siano in grado di processare ritmi in metri complessi, mentre a 12 mesi mostrino una capacità percettiva nettamente specializzata nel trattamento di modelli ritmici occidentali (Hannon, Trehub, 2005)[5].

In conclusione, il processo di sintonizzazione percettiva che avviene durante il primo anno di vita viene regolato dalla specifica esperienza culturale del bambino e sembra essere un fenomeno di ordine generale che si esprime attraverso differenti ambiti sonori, quali la musica e il linguaggio (Scott et al., 2007).

[5] Gli esperimenti di Hannon e Trehub (2005) prevedevano una fase di abituazione in cui veniva proposta una struttura ritmica e una fase successiva in cui la struttura veniva disturbata con alcune variazioni. Veniva misurata la frequenza di sguardi dei bambini su due monitor con stimoli visivi associati agli stimoli sonori.

4. *Correlati neurali e cognitivi della percezione ritmica tra musica e linguaggio*

L'abilità musicale e quella linguistica seguono percorsi paralleli ed integrati e, anche a livello cognitivo, i confini di una sfumano nell'altra rendendo talvolta difficoltosi i tentativi di circoscrivere precisi ambiti di indagine. Le risposte neurali a stimoli linguistici e stimoli musicali mostrano attivazioni ampiamente sovrapposte nei bambini (Kotilahti et al., 2010), mentre nei soggetti adulti il trattamento di stimoli linguistici e stimoli musicali mostra configurazioni di dominanza emisferica, con gli stimoli linguistici processati prevalentemente dall'emisfero sinistro e quelli musicali dall'emisfero destro (Callan et al., 2006). Questi risultati suggeriscono che la specializzazione emisferica nell'adulto emergerebbe gradualmente durante lo sviluppo (Brandt et al., 2012) e che i processi cognitivi potrebbero condividere meccanismi cerebrali mentre i domini potrebbero avere rappresentazioni separate nella memoria a lungo termine (Patel, 2012).

In linea con Patel (2012), Zatorre e Gandour (2008) propongono che la specializzazione emisferica possa non riguardare una differenza astratta tra il dominio musicale e il dominio linguistico ma una differenza di trattamento dello stimolo acustico. Secondo questa ipotesi, già presente nello studio di Zatorre e colleghi (2002), l'emisfero sinistro sarebbe specializzato nel trattamento di cambiamenti acustici rapidi, importanti per la percezione linguistica, mentre l'emisfero destro sarebbe specializzato nel trattamento di minime modulazioni di frequenza dello stimolo acustico, rilevanti per la percezione musicale[6]. Musica e linguaggio si mostrano particolarmente connessi anche da questo punto di vista, non permettendo una vera e propria specificità di trattamento. Il trattamento di cambiamenti acustici rapidi risulterebbe cruciale per il riconoscimento del timbro sia linguistico che musicale (Robinson, Patterson, 1995) in quanto la percezione di cambiamenti temporali dell'ordine di 25-50 millisecondi risulta cruciale per l'estrazione di informazioni fonemiche dal segnale linguistico (Tallal, Piercy, 1973) e questa piccola finestra temporale risulta essenziale anche per il riconoscimento dei timbri degli strumenti musicali, in quanto nessuno strumento attacca il suono con una frequenza immediatamente stabile (Hall, 1991). Di conseguenza, l'elaborazione dello stimolo ritmico risulta essere il dominio di maggiore connessione tra attività musicali ed attività linguistiche (Hausen et al., 2013).

Inoltre, il trattamento dello stimolo ritmico, sia linguistico che musicale, attiva le aree della corteccia motoria e l'area pre-motoria (Grahn, Brett, 2007; Chen et al., 2008), dando supporto alla *Motor Theory of Rhythm Perception* (Stewart et al., 2006), parallela alla *Motor Theory of Speech Perception* (Liberman, Mattingly, 1985), per la quale la percezione del ritmo si baserebbe sui meccanismi motori necessari per la sua produzione. Nelle scienze motorie, infatti, la capacità di ritmo è una delle capacità coordinative specifiche ed è intesa come la abilità che permette di organiz-

[6] Per una discussione più approfondita di questa ipotesi si vedano i lavori di Sophie K. Scott, in particolare Scott, Mcgettigan (2013) e la rassegna di Ott e Meyer negli atti del V Convegno Nazionale AISV 2009 (Ott, Meyer, 2010).

zare i movimenti in maniera che l'azione risulti il più fluida ed armoniosa possibile attraverso un corretto dosaggio di tempi ed intensità (Blume, 1982)[7].

5. *Percezione del ritmo e disturbi linguistici*

In linea con la *Motor Theory of Rhythm Perception* (Stewart et al., 2006), Fujii e Wan (2014) propongono l'ipotesi SEP (acronimo di *Sound Envelope Processing* e *Synchronization and Entrainment to a Pulse*) come modello di interazione tra abilità linguistiche e musicali, che assegna al ritmo un ruolo fondamentale nella riabilitazione di deficit in entrambi gli ambiti. L'ipotesi SEP si pone come completamento di un modello precedente, l'OPERA di Patel (2011; 2014), un impianto concettuale che descrive come le attività musicali possano agevolare e affinare il *processing* linguistico. OPERA è un acronimo che riassume le cinque condizioni in cui le attività musicali possono essere di maggior beneficio a quelle linguistiche: (1) *Overlap* – indica la sovrapposizione nell'attivazione neurale durante il trattamento di stimoli musicali e di stimoli linguistici, (2) *Precision* – su questa rete neurale condivisa gli stimoli musicali richiedono un livello di precisione maggiore nel trattamento rispetto agli stimoli linguistici, (3) *Emotion* – le attività musicali che coinvolgono la rete neurale provocano forti emozioni positive, (4) *Repetition* – le attività musicali che impegnano questa rete sono caratterizzate da frequenti ripetizioni, (5) *Attention* – le attività musicali che attivano questa rete neurale richiedono un'attenzione molto focalizzata. In altre parole, la prima condizione (*Overlap*) descrive il contesto neurale condiviso da musica e linguaggio mentre le condizioni seguenti (*Precision, Emotion, Repetition, Attention*) indicano le proprietà delle attività musicali che azionano la plasticità neurale intensificando il *processing* del linguaggio.

Con l'ipotesi SEP, Fujii e Wan (2014) aggiungono due componenti ad integrazione del modello OPERA di Patel, per sottolineare come sia lo stimolo ritmico, nello specifico, ad essere altamente benefico per la riabilitazione dei disturbi linguistici. Le due componenti sono (1) il trattamento dell'inviluppo sonoro (*Sound Envelope Processing*), che postula che la percezione del ritmo sia la fonte di maggiore sovrapposizione nelle reti neurali di musica e linguaggio (Goswami, 2011), e (2) la sincronizzazione e accordo ad un impulso (*Synchronization and Entrainment to a Pulse*), che postula che la fonte di maggiore sovrapposizione nelle reti neurali di musica e linguaggio sia la percezione ritmica in particolare se accompagnata da produzione ritmica abbinata ad un'attivazione senso-motoria (Tourville, Guenther, 2011).

L'impianto prettamente teorico della SEP trova applicazione pratica in interventi terapeutici come la *Melodic Intonation Therapy* (MIT). Ispirata dall'osservazione clinica che i pazienti con afasia non-fluente potessero cantare parole pur essendo

[7] L'impianto teorico alla base della *Motor Theory of Rhythm Perception* (Stewart et al., 2006) si avvale anche degli studi neuro-fisiologici sui neuroni specchio (Rizzolatti, Craighero, 2004) e delle indagini neuro-cognitive di Hickok e Poeppel (2000), che hanno contribuito ad approfondire il rapporto tra percezione e produzione.

incapaci di parlare (Yamadori et al., 1977), la MIT utilizza le funzioni preservate come la melodia e il ritmo per migliorare il linguaggio espressivo. Le principali componenti di questa tecnica di riabilitazione del linguaggio sono (1) l'intonazione melodica, (2) l'utilizzo di frasi e ripetizioni formulaiche e (3) lenta e periodica verbalizzazione accompagnata dal battere della mano sinistra (Sparks et al., 1974; Schlaug et al., 2010). Le componenti ritmiche della MIT, ovvero la ritmica del cantato e il battere della mano sinistra, possono aiutare a facilitare il trattamento dell'inviluppo sonoro e la sincronizzazione e accordo ad un impulso motorio, così come postulato dalla SEP (Fujii, Wan, 2014).

La centralità del ritmo nella riabilitazione del deficit linguistico emerge anche in considerazione delle difficoltà di elaborazione dello stimolo ritmico in soggetti affetti da disturbi linguistici. Già nel 1973, Tallal e Piercy ipotizzavano che la dislessia fosse il risultato di un deficit nel trattamento temporale rapido, in particolare nella veloce transizione formantica che permette di distinguere una consonante da un'altra. Studi più recenti (Goswami et al., 2002; Surányi et al., 2009) dimostrano che, rispetto ai bambini con sviluppo nella norma, i bambini dislessici sono meno sensibili alle modulazioni di intensità di uno stimolo sonoro e, nello specifico, presentano grosse difficoltà nella percezione del tempo di salita (*Rise Time*) e del centro percettivo (*P-Centre*)[8] di un suono. Tra l'altro, la sensibilità ai tempi di salita risulta strettamente connessa alle abilità di lettura e i lettori precoci mostrano migliori capacità percettive verso i tempi di salita rispetto ai soggetti di controllo.

Un altro disturbo linguistico che potrebbe essere correlato con difficoltà nella percezione ritmica e nell'ascolto musicale è il Disturbo Specifico del Linguaggio (DSL - *Specific Language Impairment*). Il DSL è un fallimento del normale sviluppo linguistico nonostante un'intelligenza normale, un contesto di crescita e acquisizione linguistica normale e l'assenza di disturbi all'udito o problemi emotivi. Come nel caso della dislessia, i bambini che sviluppano il DSL mostrano una ridotta sensibilità alla durata dei suoni fin dai primissimi mesi di vita e hanno difficoltà nel trattamento degli stimoli temporali rapidi e nella discriminazione di contrasti relativi ai tempi di salita (Benasich, Tallal, 2002; Corriveau et al., 2007).

La sensibilità al tempo di salita e al *P-Centre* è essenzialmente sensibilità al ritmo e, infatti, i bambini con disturbi linguistici hanno grosse difficoltà in compiti di coordinamento ritmico, come parlare o fare un semplice movimento sincronizzandosi ad un metronomo, e in compiti di percezione ritmica, quali discriminare tra due ritmi differenti (Goswami et al., 2002; Huss et al., 2011). In questi soggetti, attività che collegano esplicitamente le strutture ritmiche musicali a quelle linguistiche possono aiutare a migliorare il coordinamento e la percezione ritmica (Bhide et al., 2013).

[8] Il *Rise Time* è il tempo in cui uno stimolo sonoro raggiunge il massimo dell'ampiezza e il *Perceptual Centre* di un evento sonoro indica il momento in cui ne viene percepita l'occorrenza (Goswami et al., 2002).

6. *Conclusioni e prospettive di ricerca*

Fin dai primi mesi, il neonato mostra una forte sensibilità per le caratteristiche prosodiche del linguaggio ed un'elevata capacità percettiva che, nel corso del primo anno di vita, va gradualmente riducendosi, sintonizzandosi sulle proprietà acustiche della lingua nativa, nonché sulle caratteristiche melodico-ritmiche della musica della cultura dominante (Werker, Tees, 1984; Hannon, Trehub, 2005). Nonostante la modalità di trattamento diventi via via più modulare durante lo sviluppo (Patel, 2012), musica e linguaggio sono facoltà cognitive strettamente correlate e, a livello neurale, è il trattamento dello stimolo ritmico ad attivare la maggior parte delle sovrapposizioni, in particolare nelle aree motorie (Stewart et al., 2006; Fujii, Wan, 2014). Le abilità connesse al trattamento temporale rapido sono proprio quelle che risultano danneggiate nei dislessici e nei soggetti affetti da Disturbo Specifico del Linguaggio, che mostrano grosse difficoltà nella percezione ed elaborazione del *Rise Time* e del *P-Centre* (Goswami et al., 2002). È possibile che molti disturbi dell'apprendimento linguistico possano essere meglio compresi se considerati come deficit nel trattamento di input acustici complessi (Brandt et al., 2012).

Questa breve rassegna sulla percezione del ritmo apre le porte a stimoli, prospettive e ipotesi. Un quesito su tutti potrebbe essere il seguente: se i deficit nell'apprendimento linguistico possono essere considerati come deficit nel trattamento di input acustici complessi, il meccanismo di sintonizzazione percettiva che avviene a circa un anno di vita può essere considerato come un restringimento delle abilità di trattamento acustico? Un approfondimento dei meccanismi di sintonizzazione percettiva, in particolare a livello soprasegmentale, e l'integrazione di differenti ambiti di ricerca, dall'acquisizionale al terapeutico, possono fornire importanti indizi sul reale meccanismo di percezione acustica in generale e ritmico-linguistica in particolare e, in linea con le intuizioni jakobsoniane, sulle possibili analogie tra acquisizione, sintonizzazione e perdita.

Bibliografia

Allen, G.D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. In *Journal of Phonetics*, 3, 75-86.

Aslin, R.N. (1981). Experimental influences and sensitive periods in perceptual development: a unified model. In Aslin, R.N., Alberts, J.R. & Petersen, M.R. (Eds.), *Development of Perception*, vol. 2. Academic Press, 45-94.

Barry, W.J. (2010). Rhythm measures in retrospect. Reflections on the nature of spoken-language rhythm. In Schmid, S., Schwarzenbach, M. & Studer, D. (Eds.), *Atti del V Convegno Nazionale AISV* (Università di Zurigo, 4-6 febbraio 2009). Torriana (RN): EDK Editore, 3-12.

Benasich, A.A., Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. In *Behavioural Brain Research*, 136, 31-49.

Benenzon, R. (1982). *Manuel de musicothérapie*. Toulouse: Privat.

Bertinetto, P.M. (1977). Syllabic blood, ovvero l'italiano come lingua ad isocronismo sillabico. In *Studi di Grammatica Italiana*, 6, 69-96.

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

Bertinetto, P.M., Bertini, C. (2008). On modelling the rhythm of natural languages. In Barbosa, P., Arantes, P. (Eds.), *Speech prosody, Proceedings of the 4th conference* (Campinas, Brazil, 6-9 May 2008). Capes, 427-430.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P.W., Kennedy, L.J. & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. In *Journal of Experimental Psychology: General*, 117, 21-33.

Bhide, A., Power, A. & Goswami, U. (2013). A rhythmic musical intervention for poor readers: a comparison of efficacy with a letter-based intervention. In *Mind, Brain and Education*, 7, 2, 113-123.

Bijeljac-Babic, R., Bertoncini, J. & Mehler, J. (1993). How do four-day-old infants categorize multisyllabic utterances. In *Developmental Psychology*, 29, 711-721.

Blume, D.D. (1982). *Posizioni tecniche essenziali per lo sviluppo delle capacità coordinative.* Roma: CONI SdS.

Brandt, A., Gerbian, M. & Slevc, L.R. (2012). Music and early language acquisition. In *Frontiers in Psychology*, 3, 327, 1-17.

Callan, D.E., Tsytsarev V., Hanakawa, T., Callan, A.M., Katsuhara, M., Fukuyama, H. & Turner R. (2006). Song and speech: brain regions involved with perception and covert production. In *Neuroimage*, 31, 1327-1342.

Chen, J.L., Penhune, V.B. & Zatorre, R.J. (2008). Listening to musical rhythms recruits motor regions of the brain. In *Cerebral Cortex*, 18, 2844-2854.

Christophe, A., Dupoux, E., Bertoncini, J. & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. In *Journal of the Acoustical Society of America*, 95, 1570-1580.

Christophe, A., Morton, J. (1998). Is Dutch native English? Linguistic analysis by 2-month-olds. In *Developmental Science*, 1, 215-219.

Christophe, A., Nespor, M., Guasti, M.T. & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. In *Developmental Science*, 6, 2, 211-220.

Corriveau, K., Pasquini, E. & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: a new look at an old hypothesis. In *Journal of Speech Language and Hearing Research*, 50, 647-666.

Cutler, A., Carter, D.M. (1987). The predominance of strong initial syllable in the English vocabulary. In *Computer Speech and Language*, 2, 133-142.

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. In *Journal of Phonetics*, 11, 51-62.

Demuth, K. (2011). The acquisition of phonology. In Goldsmith, J., Riggle, J. & Yu, A. (Eds.), *The Handbook of Phonological Theory*. Malden, MA: Blackwell, 571-595.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. In *Journal of Child Language*, 16, 477-501.

Fikkert, P. (2007). Acquiring phonology. In De Lacy, P. (Ed.), *Handbook of phonological theory*. Cambridge, MA: Cambridge University Press, 537-554.

Fujii, S., Wan, C.Y. (2014). The role of rhythm in speech and language rehabilitation: the SEP hypothesis. In *Frontiers in Human Neuroscience*, 8, 777, 1-15.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. In *Trends in Cognitive Sciences*, 15, 3-10.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S. & Scott, S.K. (2002). Amplitude envelope onsets and developmental dyslexia: a new hypothesis. In *Proceedings of the National Academy of Sciences*, U.S.A. 99, 10911-10916.

Grabe, E., Low, E.L. (2002). Durational variability in speech and the rhythm class hypothesis. In *Papers in Laboratory Phonology*, 7, 515-546.

Grahn, J.A., Brett M. (2007). Rhythm and beat perception in motor areas of the brain. In *Journal of Cognitive Neuroscience*, 19, 893-906.

Guasti, M.T. (2007). *L'acquisizione del linguaggio - Un'introduzione*. Milano: Raffaello Cortina Editore.

Hall, D. (1991). *Musical Acoustics*, 2nd Edn. Pacific Grove: Brooks/Cole Publishers.

Hannon, E.E., Trehub S.E. (2005). Tuning in to musical rhythms: infants learn more readily than adults. In *Proceedings of the National Academy of Sciences*, USA, 102, 35, 12639-12643.

Hausen, M., Torppa, R., Salmela, V.R., Vainio, M. & Särkämö, T. (2013). Music and speech prosody: a common rhythm. In *Frontiers in Psychology*, 4, 566, 1-16.

Hayes, B. (1995). *Metrical stress theory. Principles and case studies*. Chicago: University of Chicago Press.

Hickok, G., Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. In *Trends in Cognitive Sciences*, 4, 131-138.

Huss, M., Verney, J.P., Fosker, T., Mead, N. & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. In *Cortex*, 47, 674-689.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Jusczyk, P. (2000). *The discovery of spoken language*. Cambridge: MIT Press.

Jusczyk, P.W., Aslin, N. (1995). Infants' detection of sound patterns of words in fluent speech. In *Cognitive Psychology*, 29, 1-23.

Jusczyk, P.W., Cutler, A. & Redanz, N. (1993). Preference for the predominant stress patterns of English words. In *Child Development*, 64, 675-687.

Jusczyk, P.W., Houston, D.M. (1999), The beginning of word segmentation in English-learning infants. In *Cognitive Psychology*, 39, 159-207.

Keller, E., Port, R. (2007). Speech timing: Approaches to speech rhythm. In *Proceedings of the 16th International Congress of Phonetic Sciences* (Saarbrücken, Germany, August 6-10 2007), 327-329.

Kohler, K.J. (Ed.) (2009). *Rhythm in speech and language*. In *Phonetica*, 66, 1-2.

Kotilahti, K., Nissilä, I., Näsi, T., Lipiäinen, L., Noponen, T., Meriläinen, P., Huotilainen, M. & Fellman, V. (2010). Hemodynamic responses to speech and music in newborn infants. In *Human Brain Mapping*, 31, 595-603.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. In *Science*, 255, 606-608.

Ott, C., Meyer, M. (2010). Functions of the left and right posterior temporal lobe during segmental and suprasegmental speech perception. In Schmid, S., Schwarzenbach, M. & Studer, D. (Ed.), *Atti del V Convegno Nazionale AISV* (Università di Zurigo, 4-6 febbraio 2009). Torriana (RN): EDK Editore, 393-424.

Lecanuet, J.P., Granier-Deferre, C. (1993). Speech stimuli in the fetal environment. In De Boysson-Bardies, B., Halle', P., Sagart, L. & Durand, C. (Eds.), *Developmental Neurocognition: Speech and Voice Processing in the First Year of Life*. Dordrecht: Kluwer.

Lee, C.S. (2010). Speech rhythm and word segmentation: a prominence-based account of some crosslinguistic differences. In Schmid, S., Schwarzenbach, M. & Studer, D. (Eds.), *Atti del V Convegno Nazionale AISV* (Università di Zurigo, 4-6 febbraio 2009). Torriana (RN): EDK Editore, 33-43.

Liberman, A.M., Mattingly, I. G. (1985). The motor theory of speech perception revised. In *Cognition*, 21, 1-36.

Marotta, G. (2003). What does phonology tell us about stress and rhythm? Some reflections on the phonology of stress. In Solé, M.J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the XVth International Congress of Phonetic Sciences*, 3 (1). Barcelona: Universitat Autònoma de Barcelona, 333-336.

Mehler, J., Dupoux, E., Nazzi, T. & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infant's viewpoint. In Morgan, J., Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*. Mahwah: Erlbaum, 101-116.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. In *Cognition*, 29, 143-178.

Mehler, J., Nespor, M. (2004), Linguistic Rhythm and the development of language. In Belletti, A. (Ed.), *Structures and Beyond. The Carthography of Syntactic Strucutures.* Oxford: Oxford University Press, 3, 213-222.

Morgan, J., Demuth, K. (1996) (Eds.). *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Mahwah: Lawrence Erlbaum.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998), Language discrimination by newborns: Towards an understanding of the role of rhythm. In *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766.

Nazzi, T., Jusczyk, P.W. (1999). Five-month-old infants' discrimination of languages, *Proceedings of the 23rd Annual Boston Conference on Language Developmental*. Sommerville: Cascadilla Press, 519-528.

Nazzi, T., Jusczyk, P.W. & Johnson, E.K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. In *Journal of Memory and Language*, 43, 1-19.

Nespor, M., Guasti, M.T. & Christophe, A. (1996). Selecting word order: the rhythmic activation principle. In Kleinhenz, U. (Ed.), *Interfaces in Phonology*. Berlin: Akademic Verlag, 1-26.

Patel, A.D. (2008). *Music, language and the brain*. Oxford: Oxford University Press.

Patel, A.D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. In *Frontiers in Psychology*, 2, 142.

Patel, A.D. (2012). Language, music, and the brain: a resource-sharing framework. In Rebuschat, P., Rohrmeier, M., Hawkins, J. & Cross, I. (Eds.), *Language and Music as Cognitive Systems*. Oxford: Oxford University Press, 204-223.

Patel, A.D. (2014). Can non-linguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. In *Hearing Research*, 308, 98-108.

Pike, K.L. (1945). *The intonation of American English*. Ann Arbor. Michigan University Press.

Polka, L., Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. In *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.

Querleu, D., Renard, X. & Versyp, F. (1981). Les perceptions auditives du foetus humain. In *Medicine et Hygiéne*, 39, 2101-2110.

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

Righini, P. (1994). *L'acustica per il musicista. Fondamenti fisici della musica*. Milano: Ricordi.

Rizzolatti, G., Craighero, L. (2004). The mirror-neuron system. In *Annual Review of Neuroscience*, 27, 169-192.

Robinson, K., Patterson, R.D. (1995). The duration required to identify the instrument, the octave, of the pitch chroma of a musical note. In *Music Perception*, 13, 1-14.

Saffran, J.R., Aslin, R.N. & Newport, E.L. (1996). Statistical learning by 8-month-old infants. In *Science*, 274, 1926-1928.

Schlaug, G., Norton, A., Marchina, S., Zipse, L. & Wan, C.Y. (2010). From singing to speaking: facilitating recovery from non-fluent aphasia. In *Future Neurology*, 5, 657-665.

Scott, L.S., Pascalis, O. & Nelson, C.A. (2007). A domain general theory of the development of perceptual discrimination. In *Current Directions in Psychological Science*, 16, 197-201.

Scott, S.K., Mcgettigan, C. (2013). Do temporal processes underlie left hemisphere dominance in speech perception? In *Brain and Language*, 127, 36-45.

Sparks, R., Helm, N. & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. In *Cortex*, 10, 303-316.

Stewart, L., Vonkriegstein, K., Warren, J.D. & Griffiths, T.D. (2006). Music and the brain: disorders of musical listening. In *Brain*, 129, 2533-2553.

Surányi, Z., Csépe, V., Richardson, U., Thomson, J.M., Honbolygó, F. & Goswami, U. (2009). Sensitivity to rhythmic parameters in dyslexic children: a comparison of Hungarian and English. In *Reading and Writing*, 22, 41-56.

Tallal, P., Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. In *Nature*, 241, 468-469.

Tourville, J.A., Guenther, F.H. (2011). The DIVA model: a neural theory of speech acquisition and production. In *Language and Cognitive Processes*. 26, 952–981.

Trehub, S.E. (1973). Infants' sensitivity to vowel and tonal contrasts. In *Developmental Psychology*, 9, 91-124.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. In *Musicae Scientiae*, Special Issue "Rhythms, musical narrative and the origins of human communication", 157-213.

Werker, J.F., Tees, R.C. (1984). Cross language speech perception: evidence for perceptual reorganization during the first year of life. In *Infant Behavior and Development*, 7, 49-63.

Yamadori, A., Osumi, Y., Masuhara, S. & Okubo M. (1977). Preservation of singing in Broca's aphasia. In *Journal of Neurology, Neurosurgery, and Psychiatry*, 40, 221-224.

Zatorre, R.J., Belin, P. & Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. In *Trends in Cognitive Sciences*, 6,1.

Zatorre, R.J., Gandour, J.T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. In *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 363, 1087-110.

DOI: 10.17469/O2101AISV000017

ELISA PELLEGRINO, VALERIA CARUSO, ANNA DE MEO

Discriminating between pathological and foreign-accented speech: an acoustic-perceptual study on Italian

This study investigates native Italian listeners'ability to discriminate pathological from foreign-accented speech, comparing the read performance in Italian of Italian and Ukrainian deaf talkers with those produced by hearing Japanese and Ukrainian L2 speakers. Seventeen native Italian listeners rated the speech samples for 'accentedness', 'comprehensibility' and 'speech type'. The comparison between perceptive and acoustic data shows the difficulties of listeners to distinguish between native and non-native deaf speech. The mixed prosodic profile of the Japanese speaker and the great number of segmental misarticulations lead listeners to rate her speech similarly to those of deaf speakers, namely as 'pathological' and 'scarcely comprehensible'. Conversely, the Vowel Space Area and the consonantal misarticulations type differentiate her from the deaf speakers, and are probably responsible for the judgement concerning her 'non-nativity' to the Italian language.

1. *Introduction*

This study focuses on the oral productions of prelingual deaf speakers, whose perceptive deficit has been recognized by many researchers as a type of impairment that prevents the development of a native-like competence in the vocal languages:

> [...] one can report that all studies in the literature agree on the fact that young and adult deaf quite never reach a linguistic competence similar to that of a native speaker, neither in the oral nor in the written language" (Caselli, Maragna, Volterra, 2006: 235).

Starting from these assumptions, an analysis of deaf speech seems to be valuable for the double aim of enriching the existing literature on deaf people's skills, and of researching the features responsible for the perceptual discrimination of different kinds of oral productions. More precisely, this study will give insights into the ability of native Italians to distinguish non-native speech from the pathological oral productions of deaf people. As it will be clear in a moment, in fact, it is rather difficult for native listeners to discriminate between these two types of voices. In particular, the oral productions of deaf people prove to be undistinguishable with reference to the additional native/non-native parameter. Italian listeners are in fact unable to recognize if a deaf speaker has learned Italian as a first or second vocal language.

Moreover, the listeners' inability to detect any foreign accent in deaf speech is somehow paralleled by the analogous misinterpretation of some elementary learners' speech productions, which are mistaken for a kind of pathological speech.

Similar misinterpretations will be in focus in the following pages while discussing the results of a perceptive test along with data from the spectro-acoustical analysis of different types of voices.

Before going into the details of the experimental study, a brief description of deaf people's voices will be given by surveying the existing literature.

1.1 Features of deaf people's voice reported by the literature

The bulk of literature on deaf people's speech is made up of studies carried out by medical researchers interested in the quality estimation of their vocal productions. As a consequence, the analyses carried out so far have mainly evaluated the mispronunciations and the general comprehensibility of these atypical speakers on the segmental level.

Many of these studies, however, have also described the quality of deaf people's voices which, since the 1960s (Calvert, 1962), have been described as 'tense', 'hollow', 'breathy', 'flat', but also as 'harsh' and 'throaty', two characteristics that Shriberg, Kwiatkowski & Rasmussen (1990) refer to the 'laryngeal quality' of this pathologic speech. Deaf voices tend to be 'tense' or even strident since they receive most of the vocal linguistic input from tactile sensations (Lenden, Flipsen, 2007), while the reverberation of sounds in the pharynx is responsible for their hollow character (Boone, 1971; Finkelstein et al., 1993). On the other side, the excessive force put on plosives before a vowel (Lenden, Flipsen, 2007) determines the breathy quality of their speech productions and is also responsible for improper pauses and fragmented segmentation (Pellegrino, Caruso, 2012).

Even more distinguishable are, however, the prosodic traits of this atypical speech. The irregular phrasing is indeed the most perceptible prosodic feature, since deaf people pronounce utterances word by word, lacking co-articulations and sound blends. There is also a general tendency to prolong vowels (Boone, 1971) and to slow speech rate, two factors that weaken fluency. Additionally Nickerson (1975) pointed out how deaf speakers, being unable to vary pitch contours properly, produce a monotonous speech and use pauses to signal word and sentence boundaries (Stathopoulos et al., 1986). On the other hand, some other researchers have underlined how pitch, among other traits, is influenced by the degree of hearing loss in these subjects. Bonetti, Utović & Dulčić (2008) found that prelingual mild deaf adolescents, for example, were able to control specific vocal features, such as intonation, accent, pauses and speech rate. This proved that prosody was not uniformly affected in the category of deafness because some prosodic features remain preserved with lower degrees of deafness.

1.1 Other relevant features of deaf speech

The current study is more focused on deaf speakers' linguistic skills and competences than it is in medical research.

In a previous study, Pellegrino and Caruso (2012) have highlighted that even the most disadvantaged deaf speakers, such as foreigners with a profound hearing

impairment, have communicative competences in the oral L2 that are independent of their segmental accuracy and the level of their hearing loss. Put to test in a semi-spontaneous communicative task, profound deaf foreigners proved in fact their ability to pronounce speech acts as single utterances, performing requests at a speech and articulation rate that are comparable to those of hearing native speakers. This suggests that the chief concern of whatever kind of speaker is that of preserving communicative units, even if the uttered sentences prove to be scarcely intelligible.

In the present study, however, the viewpoint adopted is perceptive, and the focus is on the ability of the native speakers to recognize pathological and non-native traits in the voices they are listening to. Considering, in fact, the results of a preliminary investigation (Pellegrino, Caruso & De Meo, 2012) showing that some L2 spoken productions can be mistaken for 'impaired' or even 'synthesized' speech, it seems interesting to investigate which acoustic features are responsible for mistaking the voices of non-native hearing speakers for those of impaired subjects.

In order to achieve this aim, samples of different kinds of voices were collected: the oral productions of deaf speakers, both Italian and foreign citizens, as well as those produced by hearing non-native speakers. In the first research step, the speech samples were administered to a group of native Italian listeners, who rated for different parameters the voices they had been listening to. Secondly, the perceptive judgments were compared with the spectro-acoustical analysis of the oral corpus in order to identify:

– the existence of any transfer phenomena from the language of the first logopaedic treatment to the L2 vocal Italian, and, additionally,
– the presence of any acoustic features that allow native listeners to recognize this type of pathological speech.

2. *The study*

The study thus consists of two parts: a perceptive test and a spectro-acoustical analysis. The perceptive test needed the recruitment of two groups of participants: one of readers, providing the necessary speech samples; one of native listeners, expressing judgments about the collected voices.

2.1 Participants

As for the readers, the group is composed as follows:

– 2 Italian sisters (henceforth D_L1_Fr; D_L1_Mo), with severe hearing impairment, aged 34 and 36 respectively, grown up in a hearing family, educated with an oral method, and having the vocal Italian as their L1 (see Pellegrino, De Meo & Caruso, 2014) and a low proficiency in the Italian Sign Language (LIS).
– 1 Ukrainian deaf woman (henceforth D_L2_Al), with severe hearing impairment, aged 42, living in Italy for 8 years, having a hearing son and employed as healthcare worker. She is proficient both in LIS and both in signed and vocal Ukrainian.

– 1 Japanese female learner of L2 Italian (henceforth H_L2_Ay), who had been recognized as 'impaired' or 'synthesized' in a previous, preliminary investigation (Pellegrino, Caruso & De Meo, 2012).

Additionally, two voices make up the control group in order to:
a. identify the presence of any transfer phenomena in the vocal L2 productions from the language of the first logopaedic treatment of the deaf foreigner;
b. assess the difference between the speech of a native Italian and that of a severe deaf.

The control group is thus composed by:
– 1 Ukrainian female, who is an L2 Italian learner (henceforth H_L2_Ha). The spectro-acoustic analysis of her voice has been compared to that of the Ukrainian deaf in order to verify point a) of the previous list.
– 1 native Italian speaker (henceforth H_L1_El), offering the necessary comparisons in order to estimate point b).

All foreign readers involved in the test have an elementary level of linguistic competence in Italian.

As for the listeners, the group is composed of 17 native Italians, aged between 20 and 28. Six of them declared to be unexperienced with any kind of impaired speech, six instead said to be familiar with deaf people's speech, and six others with the speech disorders caused by Down's Syndrome, Parkinson's Disease, stuttering, and schizophrenia.

2.2 Materials and Methods

In order to prevent different factors that might have biased the results, participants were involved in a read speech task. This made comparisons possible, since many voices were collected for the oral corpus and they were extremely different among each other. Moreover, the complexity of the text used as a stimulus was carefully controlled in terms of both word length, syntactic and informative structure. Participants were thus given as a reading material a simplified short excerpt from an article of fashion (see below):

> G-Star è uno stilista famoso e fa jeans molto belli.
> Oggi a Milano G-Star vende i suoi jeans a solo 20 euro, un prezzo molto basso.
> L'attrice Laura Chiatti fa pubblicità ai jeans di G-Star nel negozio di Milano.

The speakers' voices recordings were taken in single sessions in an anechoic chamber, at 44.100 Hz sampling rate, at the Language Center of 'L'Orientale' University of Naples.

After that, the 17 native subjects listened, in a randomized order, to the single excerpts of the read speech uttered by the seven readers. Each speech sample was then rated for 'accentedness', 'comprehensibility' and 'speech'. Before taking the per-

ception test, the listeners were explained that 'accentedness' serves to assess whether the speakers' accent is judged as different from that of the L1 community (Derwing, Munro, 2005). 'Comprehensibility', on the other hand, expresses the effort required from the listener in order to understand the non-native speaker utterances (Warren, Elgort & Crabbe, 2009).

For the ratings regarding accentdness, listeners had to judge if the speakers sounded native or non-native like; whereas 'comprehensibility' had to be rated on a five-point scale (0 = 'poor'; 5 = 'very good'). Although in studies on foreign accentedness, the term 'comprehensibility' often overlaps with that of 'intelligibility' and the two measures seem to be very well correlated (Munro, Derwing, 1999), in this study the main focus was on 'comprehensibility' rather than 'intelligibility'. This choice was made in accordance to the study aims and materials: 'comprehensibility' is usually assessed with human judgments and relies on prosodic cues, while 'intelligibility' is assessed with transcription tasks and mainly depends on segments (Munro, Derwing, 1995; 1999; 2001).

With respect to the evaluation of the 'kind of speech' parameter, instead, listeners could choose among four options: 'normal, 'pathological', 'synthesized' or 'other'.

2.3 Results

2.3.1 The perception test

Results of the perception test show for the 'accentedness' parameter that Italian listeners, regardless of their experience with impaired speech, succeed in distinguishing between native and non-native speech, when the utterances are produced by 'non-pathological', and 'non-synthesized' speakers (table 1), who in the following pages will be referred to as 'normal'.

Table 1 - *Rates for accentedness*

	Native	*Non Native*
Subjects	Nr. listeners (%)	Nr. listeners (%)
H_L1_El	17 (100%)	
H_L2_Ay		17 (100%)
H_L2_Ha		17 (100%)
D_L1_Fr	16 (94%)	1 (6%)
D_L1_Mo	15 (88%)	2 (12%)
D_L2_Al	15 (88%)	2 (12%)

In fact, all listeners recognize which one is the Italian native speaker's voice, but only the hearing foreigners are identified as non-native. On the contrary, all deaf participants are considered as 'native', and no exception is made for the Ukrainian deaf women (D_L2_Al).

However, the fact that the speakers' provenience is unrecognizable when they are affected by a severe hearing loss is not surprising. This result is in fact compatible with the outcomes of another study by Pellegrino, Caruso, De Meo (in press) that demonstrates

how difficult it is to find strict correlations between prosodic profiles and deaf people's origin, degree of hearing loss and quantity of exposure to the vocal language. More precisely, the research has highlighted that fluency is the most unpredictable feature of deaf people's speech, since every deaf subject develops a personal pausing strategy in order to overcome articulatory difficulties and lack of self-control on his/her own utterances.

Additionally, the complexity of the 'non-native speech' category emerges also with respect to the assessment of another parameter, such as 'nature of speech'. When, in fact, the Italian listeners were asked to judge if the speaker they had been listening to was 'normal', 'pathological' or 'synthesized', they were able to recognize the hearing native as a 'normal' speaker and the three deaf women as 'pathological', while the hearing L2 speakers received contradictory rates (table 2). The Ukrainian woman (H_L2_Ha) was judged in fact as 'normal' by 11 listeners, and 'pathological' by six listeners. Even more ambiguous is the case of the hearing Japanese (H_L2_Ay), whose voice was considered 'pathological' by almost half of the listeners (8), 'normal' and 'synthesized' by four listeners each; while one listener considered her as both synthesized and pathological.

Table 2 - *Rates for kind of speech*

Subjects	*Normal*	*Pathological*	*Synthesized*	*Synthesized/Pathological*
H_L1_El	17 (100%)			
H_L2_Ay	4 (23.5%)	8 (47%)	4 (23.5%)	1 (6%)
H_L2_Ha	11 (64%)	6 (36%)		
D_L1_Fr	17 (100%)			
D_L1_Mo	17 (100%)			
D_L2_Al	17 (100%)			

Lastly, speech samples were rated on 'comprehensibility' (table 3).

Table 3 - *Rates for comprehensibility*

Subjects	*Average rate (St. Dev.)*
H_L1_El	5 (± 0)
H_L2_Ha	4.3 (± 0.46)
H_L2_Ay	2.5 (± 0.9)
D_L2_Al	2.3 (± 0.8)
D_L1_Fr	2.3 (± 0.7)
D_L1_Mo	1.4 (± 0.5)

As expected, the hearing native speaker (H_L1_El) was found very comprehensible (5), on the contrary, the three deaf subject were the least intelligible among the participants, with grades ranging from 2.3 of D_L2_Al to 1.4 of D_L2_Mo. Again, the more surprising results are those concerning the hearing L2 speakers. Although both of them were considered as non-native speakers, the Ukrainian hearing woman (H_L2_Ha) achieves a high level of comprehensibility (4.3), whereas the Japanese speaker (H_L2_Ay) gets scores (2.5) that are in line with those received by the deaf

subjects. One can thus infer that the scarce comprehensibility of this speaker is responsible for the fact that many native listeners (9 out of 17) have considered her as an impaired speaker.

2.4 Acoustic analysis

The next step of the research was to spectro-acoustically analyze the spoken performance of the six subjects in order to verify whether the listeners' rates on 'nature of speech', 'accentedness' and 'comprehensibility' could correlate with some shared prosodic profile.

To the purpose, using the software Praat (Boersma, 2001), the speech corpus was segmented into 2 different tiers: CV segments, syllables. CV boundaries were identified by visual inspection of speech spectrograms and waveforms. The segmentation of glides followed the rules adopted by Ramus et al. (1999): [w] and [j] were treated as consonants and the boundary was placed between the approximant and the vowel. Diphthongs were segmented in one or two vowels depending on the spectro-acoustic characteristics of the tract. If both vowels presented a specific steady-state formant pattern, the diphthong was divided into two vowels; otherwise, it was treated as a single interval. Voice Onset Time in stop consonants was included within the consonantal interval. In vowel+nasal consonant sequence, the nasalized portion of the vowel was assigned to the vowel. In case of initial voiced stop consonants, the first boundary corresponded to the onset of glottal pulses. Post-pausal voiceless plosives were assigned a duration equal to the mean value of single plosives in the same utterance. Respiratory and syntactic pauses were labeled as silent pauses (SP). Different kinds of disfluencies (laryngealizations, nasalizations, vocalizations, etc.) were treated as filled pauses (FP). The duration of all segments was extracted using a Praat script. Additionally, for every speech chain, defined as the portion of an utterance within two silences (Pettorino, Giannini, 2010), the lowest and highest f_0 values were manually measured.

On the basis of these measures, and following the Prosody–Voice Screening Profile for pathological speech by Schriberg et al. (1992), the following rhythmic-prosodic parameters were calculated:

- Articulation rate (AR), i.e. the ratio between the number of syllables really uttered and phonation time (syll/s).
- Speech rate (SR), i.e. the ratio between the number of syllables really uttered and the total time of the utterance time (syll/s).
- Fluency (F), i.e. the ratio between the number of syllables and the number of speech chains (syll/SC).
- Tonal range (TR), i.e. the interval between the f_0 minum and maximum, measured in semitones.

At segmental level, contrastive analyses on vowel formant patterns were carried out. From the CV tier, the steady state formant pattern of stressed and unstressed vocalic segments, without transitions to and from adjacent sounds, was identified. Then,

the frequency of F1 and F2 were measured manually. The raw vowel measurements were normalized through the web-based software package NORM, using Neary 2 (formant extrinsic) as normalization method.

Consonantal pronunciation mistakes were also taken into consideration. Given deaf speakers' inability to rely on acoustic cues in the process of speech perception, the first objective was to verify whether deaf speakers were able to reproduce correctly the information mainly provided by the visual system (place of articulation), and the cues derived primarily from the acoustic signal (manner of articulation and voicing) (McGurk, MacDonald, 1976). A further goal was to investigate whether deaf talkers would produce errors that are of a different type from those made by the hearing speakers.

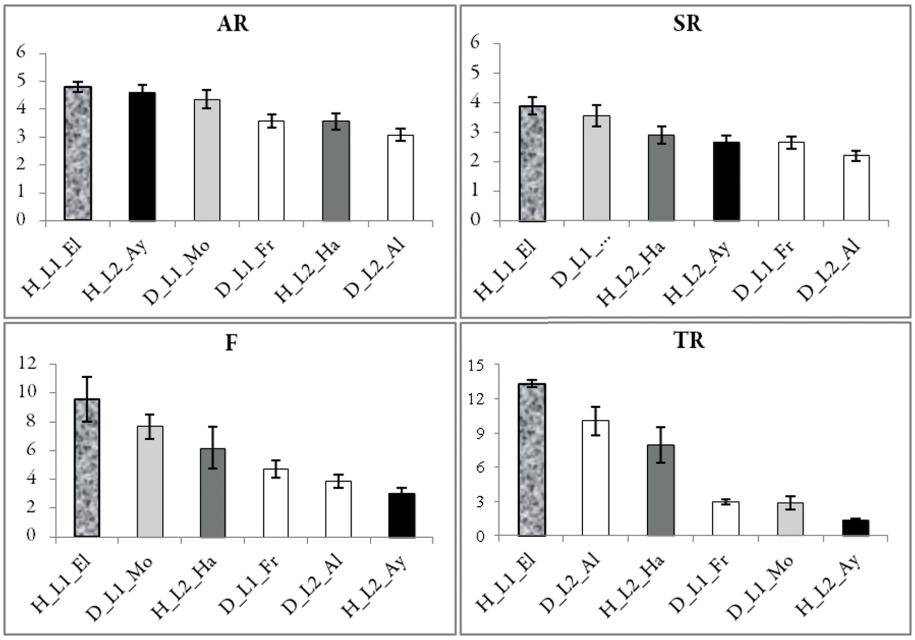
2.5 Acoustic correlates of accentedness, perceived nature of speech, and comprehensibility

2.5.1 Suprasegmental level

The results of spectro-acoustic analysis carried out on the corpus shows no strict relationship between the native Italian listeners' judgements and the prosodic profiles of the different speakers' groups. This means that the suprasegmental clues alone are not robust enough to explain the assessment results of the perceptive test. However, this is still in line with the results of the study by Pellegrino, Caruso, De Meo (in press) showing that the spoken productions of deaf people can have the same prosodic features, regardless of the speakers' geographical origin, level of hearing impairment and the amount of exposure to the vocal Italian. Moreover, the present study provides evidence of an imperfect correspondence between prosodic and perceptual data that seems to indicate the stronger relevance of some other segmental clues for specific categories of judgments, like 'native/non-native' and 'pathological' (figs. 1-4).

Among the clearest mismatches between acoustic and perceptual data one can find the case of the Italian deaf D_L1_Mo (light grey column in graphs above). The Italian listeners judged her spoken productions as an example of scarcely comprehensible, pathological speech even though her prosodic profile is similar to the one of the native hearing speaker (H_L1_El, granite column) in terms of Articulation rate, Speech Rate and Fluency. In addition, the Tonal range is the only parameter that distinguishes this same Italian deaf (D_L1_Mo) from the native speaker and equates her to the other speakers who receive the same grades for comprehensibility and speech type (D_L1_Fr and H_L2_Ay).

Figures 1-4 - *Cross-comparison of the oral corpus per AR, SR, F and TR*



The hearing Ukrainian speaker represents, instead, a different case (H_L2_Ha, dark grey column). Her lower values of Articulation Rate, Speech Rate, Fluency and Tonal Range than the hearing native Italian (H_L1_El) seem not to increase the listeners' effort to understand her speech, but they are probably responsible for the judgement of 'non-nativeness'. The majority of the listeners, in fact, considers her speech as 'highly comprehensible' and 'non-native'. On the other hand, her Fluency values are similar to those of one deaf Italian, D_L1_Mo, and her Tonal range is in line with the Ukrainian deaf woman (D_L2_Al), two clues that have probably biased the correct recognition of her voice by some listeners. Six out of 17 of them have in fact considered her voice as 'pathological'.

Completely different is the case of the hearing Japanese (H_L2_Ay, black column). Her mixed prosodic profile might have confused the listeners, who assigned very low comprehensibility ratings and chose extreme options when asked to rate the kind of speech they had been listening to. She has in fact some vocal traits in common with all the groups of participants. For the Articulation Rate, she is similar to the native speaker and to one Italian deaf, D_L1_Mo. In terms of Speech Rate, instead, she resembles the Ukrainian speaker and the other Italian deaf (D_L1_Fr). In terms of Fluency, her speech flow is as fragmented as the one produced by the Ukrainian deaf (D_L2_Al). What on the contrary differentiates her from the rest of the group is the narrow interval between f_0 maximum and minimum. This type of non-native intonation, therefore, might have been perceived as unnatural as the flat and artificial sounding intonation produced by text-to-speech systems (Clark,

1999). This can explain the 'synthesized speech' judgments expressed by four listeners.

2.5.2 The segmental level

At segmental level, contrastive analyses on vowel formant patterns were conducted. Before starting to discuss the results, it is important to underline that the text used for the read speech task was not originally created to allow for a punctual analysis of the Italian vowel system. The objective, instead, was to look at the effect of the hearing impairment and non-nativeness on the size of vowel areas used by the six speakers reading the same text. The passage includes all seven phonetic vowels, though they occur with a different frequency (table 4).

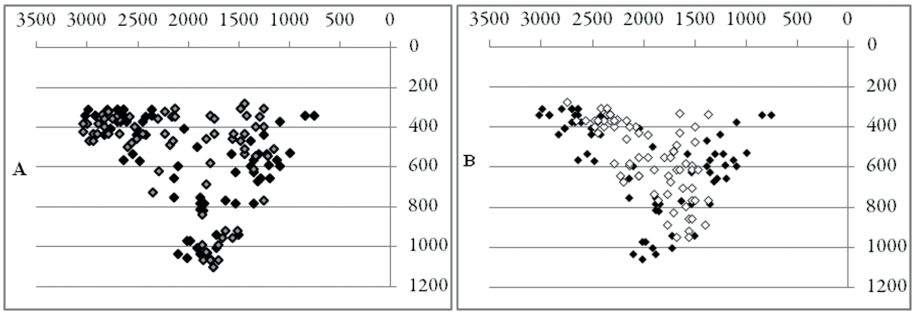
Table 4 - *Vowel frequency*

[i]	[ɛ]	[e]	[a]	[ɔ]	[o]	[u]
29%	8%	11%	26%	3%	20%	3%

The diagrams in figs 5A-7 plot the normalized F1/F2 values of the vowels produced by the Italian hearing speaker in comparison to the ones of:
1. Italian deaf readers (figs. 5A-5B).
2. Ukrainian deaf and hearing readers (figs. 6A-6B).
3. Japanese and Ukrainian hearing readers (fig. 7).

As regards the first comparison (figs. 5A-5B), it is interesting to point out that even in the case of vowel articulation, the Italian deaf speaker D_L1_Mo is very close to the native model, especially for the area of front-close and open vowels (fig. 56A). The same does not apply to her sister.

Figures 5 A-B - *Vowel areas A of the hearing native speaker (black rhombi) in comparison with D_L1_Mo (grey rhombi in fig. A, on the left) and D_L2_Fr (white rhombi in fig. B, on the right)*



In line with the results of acoustic and kinematic studies on vowel articulation by deaf and normal speakers (see Pratt, Tye-Murray, 2011 for a review), D_L1_Fr presents vowel space reduction with a general centralization pattern in the degrees of

fronting and opening. This means that D_L1_Fr differs from the native speaker as well as from her sister not only on the suprasegmental level, but also on the degree of variability in tongue and shape position across vowels. Tye-Murray (1991) had already shown that deaf subjects are not capable of producing vowel distinct movement trajectories, unlike normal subjects whose tongue movements vary in view of the vowel to be uttered.

Figures 6 A-B - *Vowel areas of the hearing native speaker (black rhombi) in comparison with D_L2_Al (white squares, in fig. A, on the left) and H_L2_HA (grey x marks, in fig. B, on the right)*

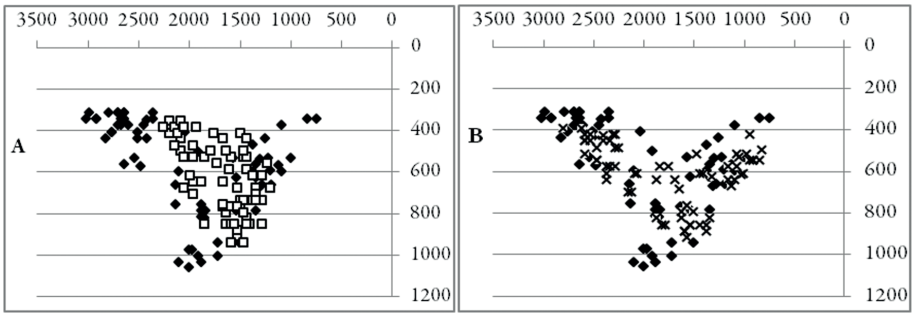
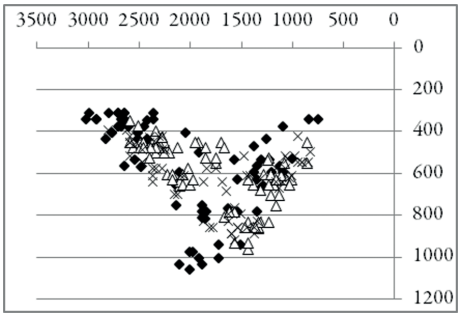


Figure 7 - *Vowel areas of the hearing native speaker (black rhombi) in comparison with H_L2_HA (grey x marks) and H_L2_Ay (white triangles)*



The centralization pattern is even more evident in the case of the Ukrainian deaf speaker (D_L2_Al) (fig. 6A). Probably due to reduced tongue mobility, she succeeds in uttering only close mid and open mid vowels, but is not able to distinguish between anterior and posterior vowels. The drastic centralization trend observed in D_L2_Al cannot be explained only as the effect of the language of the logopedic treatment she received, namely the vocal Ukrainian. The VSA of the hearing Ukrainian reader (H_L2_Ha) (fig. 6B), similarly to the one of the Japanese (H_L2_Ay) (fig. 7), is more reduced than those of the native speakers, but she was at least able to produce front and back vowels. Behavioral data similar to those of the Ukrainian deaf speaker were already found by Stein (1980), who demonstrated that hearing impaired children were able to produce tongue height distinctions between

high and mid vowels, but they were unable to differentiate front and back vowels. All this things considered, one can affirm that the trend to vowel centralization of the deaf speakers' voices analyzed so far is partially imputable to the sensorial impairment, and partially to the fact that they are L2 speakers.

2.5.3 The segmental level: consonants

In addition to the vowel areas, attention was also paid on consonantal articulation, since a sizeable body of research has demonstrated that deaf subjects produce a "myriad of consonantal errors" affecting speech intelligibility (Pratt, Tye-Murray, 2011: 208). This further level of analysis had a twofold objective: on the one hand, it was intended to verify whether Italian and foreign deaf speakers commit the same kind of mistakes while reading aloud in Italian. On the other hand, it was specifically aimed at ascertaining further reasons for the judgment expressed by the group of listeners for the speech produced by the Japanese, who was considered as an 'impaired' speaker.

The pronunciation mistakes were analyzed both quantitatively and qualitatively. For each speaker, the number of misarticulations, with respect to the native model, was calculated. We also considered the kind of segmental errors committed and the sound category mostly affected by mispronunciations.

The results of quantitative analysis have shown that the speakers that score limited comprehensibility (the 3 deaf participants and the Japanese learner) were also the ones producing consonantal errors (table 5). Out of the 83 consonantal sounds uttered by the native speaker, the deaf Ukrainian (D_L2_Al) gets 10 sounds wrong, D_L1_Mo and the Japanese speaker err 15 consonants, D_L1_Fr 23. No segmental errors were instead produced by the hearing Ukrainian speaker, whose spoken productions were judged as highly comprehensible.

Table 5 - *Number and Percentage of consonantal mistakes per speaker and comprehensibility level*

Subjects	*Target number of consonants*	*N. of mispronunciations per speaker (%)*	*Comprehensibility level*
H_L2_Ha	83	–	4.3 (± 0.46)
D_L2_Al	83	10 (12%)	2.3 (± 0.8)
D_L1_Mo	83	15 (18%)	1.4 (± 0.5)
H_L2_Ay	83	15 (18%)	2.5 (± 0.9)
D_L1_Fr	83	23 (28%)	2.3 (± 0.7)

The errors produced by the least comprehensible speakers were then classified according to the commonest taxonomy available in the literature for the errors produced by deaf and non-native speakers. Therefore, accent shift, insertions, omissions, substitutions, errors in consonantal clusters, voiced-voiceless confusion were considered (Major, 2001; Pratt, Tye-Murray, 2011; Tye-Murray et al., 1996). As table 6 shows, the most recurrent type of mistake is phone substitution rather than

omission, as Pratt and Tye-Murray (2011) on the contrary report. In descending order, there are voiced-voiceless confusions, omissions and lastly insertions, which have considerably lower errors percentage.

Table 6 - *Percentage of consonantal mistake type*

insertions	*omissions*	*substitutions*	*voiced-voiceles confusion*
11%	14%	51%	24%

With respect to the distribution of the error type across speakers, fig. 8 shows that all the speakers produce substitution errors. To this typology, the Japanese speaker adds also insertions and deletions; the Italian deaf subjects add only omissions, while the Ukrainian only insertions.

Figure 8 - *Number of errors for speech type and speaker*

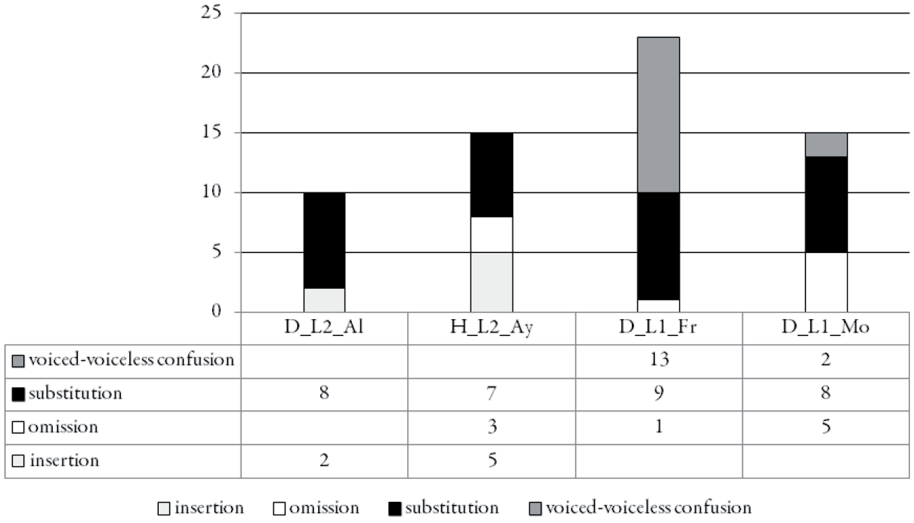
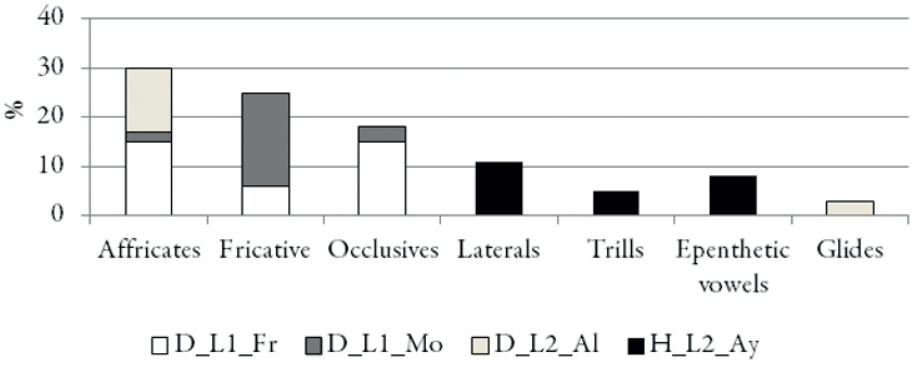


Figure 9 - *Mistaken sound category*



It is interesting to underline that, despite the across-speakers similarity in the type of mistakes made, what differentiates the hearing Japanese from all deaf speakers is the mistaken sound category (fig. 9)[1].

The Japanese, in fact, errs particularly with liquids. For example she tends to delete the alveolar trill in coda position (i.e. 'G-star' uttered like [dZ"ist"a]; |dZ| i| s| t |a|- r|); she substitutes the long alveolar lateral in intervocalic position with the voiced alveolar plosive (i. e. 'belli' uttered like [b"eddi]; |b |e |dd |i |) and single alveolar lateral with alveolar trill (i.e. 'solo' uttered like [s "oro] |s |o |r|o |). She usually adds epenthetic central vowels in loan words ending with consonant (i.e. 'jeans' uttered like [dZ"ins@]; |dZ| i| n|s|+@|) and in sequences involving complex consonantal clusters (i.e. 'pubblicità' uttered like [pub@ditSit"a]; |p| u |b |+@|d|i |tS|i|t|a |) in order to simplify the syllable structure.

As regards deaf speakers, instead, regardless of their provenience, the sounds mostly affected by errors are the plosives, the fricatives and the affricates both in initial and in intervocalic positions. In the examined corpus, there is evidence of:

- shift in manner of articulation ts > s (es. 'prettso' uttered like [pr"esso]; |p |r |e |ss |o); tʃ > ʃ ('attrice' uttered like [attr"iSe];|a |tt |r |i |S |e |);
- shift in place of articulation: st>sk ('stilista' uttered like [stil "iska]; |s |t | i | l |i |s |t |a |);
- devoicing: b > p ('basso' uttered like [b_0"asso]; |b_0| a |s |s|o|); dZ > tS ('oggi' uttered like ["oddZ_0i]; |o|ddZ_0|i|); v > f ('venti' uttered like [v_0"enti], [|v_0| e|n| t| i);
- omission of fricative sounds ('stilista' uttered like [til "itta]; | – s | t | i| l| i |tt|a|).

Considering the latter category of error, this was produced only by one Italian deaf speaker, D_L1_Mo, whose hypernasal speech reflects the inability to close the soft palate completely, and thus to produce fricatives. A more detailed analysis of the kinds of errors listed above seems to suggest that deaf people do not find it hard to produce complete closure. The plosives indeed are primarily object of a shift in the place of articulation and devoicing. On the contrary, all the errors related to a shift in the manner of articulation involve sounds with a mixed articulatory mechanics. The affricates were substituted by the corresponding fricatives, probably because the subjects find it difficult to coordinate two consecutive movements in uttering a single sound. Rather than fully stopping the air from leaving the vocal tract, then releasing it through a constricted opening, they produce only the second part of the movement, resulting in a fricative.

3. *Conclusions*

Summing up the results achieved so far, it is important to underline, in the first place, the difficulties encountered by the native listeners to distinguish the spoken productions of deaf speakers with reference to the native/non-native parameter.

[1] The consonantal mistakes were transcribed and annotated according to the rules adopted for the corpus CLIPS (see Savy, 2005).

Despite the spectro-acoustical evidence of some phonetic transfers from the language of the logopaedic treatment, these features are overheard by native listeners, who fail to distinguish between native and non-native Italian deaf speakers.

Similarly, the mixed prosodic profile of the Japanese speaker and the high number of segmental mistakes she makes might have confused the native listeners, who judged her oral productions similarly to those of deaf speakers, namely as 'pathological' and 'scarcely comprehensible'. On the contrary, her VSA and consonantal misarticulations differentiate her from the deaf speakers and thus may have caused the different judgment concerning her 'non-nativity' to the Italian language. In fact, when the Japanese speaker reads in Italian, she errs mainly with liquids and with complex consonantal clusters, while the sounds that are more affected by errors in deaf speech are the plosives, fricatives and affricates, both in initial and in intervocalic positions. These features derive from the difficulties to control voicing and the degrees of constriction that deaf speakers commonly have.

All this things considered, one can underline the relevance of the data collected so far for better outlining the types of deaf speakers' linguistic inabilities and for offering further evidence of the wide-ranging variations admitted by the category of 'non-native Italian speech'.

Bibliografia

Boersma, P. (2001). Praat, a system for doing phonetics by computer. In *Glot International*, 5, 341-345.

Bonetti, L., Utović, V. & Dulčić, A. (2008). Prosody in Speech Produced by Deaf Persons. In *Croatian Review of Rehabilitation Research*, 44, 1-13.

Boone, D.R. (1971). *The voice and voice therapy*. Englewood Cliffs, NJ: Prentice-Hall.

Calvert, D.R. (1962). Deaf voice quality: A preliminary investigation. In *The Volta Review*, 64, 401-403.

Clark, R.A.J. (1999). Using Prosodic Structure to Improve Pitch Range Variation in Text to Speech Synthesis. In Ohala, J.J. (Ed.), *Proceedings of the XIVth International Congress of Phonetic Sciences*, vol. 1. San Francisco: University of California Press, 69-72.

Derwing, T.M., Munro, M.J. (2005). Second language accent and pronunciation teaching: A research-based approach. In *TESOL Quarterly*, 39, 379-397.

Finkelstein, Y., Bar-Ziv, J., Nachmani, A., Berger, G. & Ophir, D. (1993), Peritonsillar abscess as a cause of transient velopharyngeal insufficiency. In *Cleft Palate-Craniofacial Journal*, 30, 421-428.

Lenden, J.M., Flipsen, P.J. (2007). Prosody and voice characteristics of children with cochlear implants. In *Journal of Communication Disorders*, 40, 66-81.

Major, R.C. (2001). *Foreign Accent: The Ontogeny and Philogeny of Second Language Phonology*. Mahwah, NJ: Erlbaum.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. In *Nature*, 264, 746-748.

Munro, M.J., Derwing, T.M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. In *Language and Speech*, 38, 289-306.

Munro, M.J., Derwing, T.M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. In *Language Learning*, 49 (supp. 1), 285-310.

Munro, M.J., Derwing, T.M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech. In *Studies in Second Language Acquisition*, 23, 451-568.

Nickerson, R. (1975). Characteristics of the speech of deaf persons. In *The Volta Review*, 77, 342-362.

Pellegrino, E., Caruso, V. (2012). "And Nevertheless they do speak Italian". Prosodic explorations in speech of deaf immigrants. In De Meo, A. & Pettorino, M. (Eds.), *Prosodic and Rhythmic Aspects of L2 Acquisition*. *The case of Italian*. New Castle upon Tyne: Cambridge Scholar Publishing, 83-102.

Pellegrino, E., Caruso, V. & De Meo, A (2012). Other voices: prosodic discriminative features of foreign and impaired speech, oral presentation to *Prosody in Typical and Atypical Populations Congress*, University of Reading, UK, 3-4 September.

Pellegrino, E., Caruso, V. & De Meo, A (in press). La competenza prosodica nell'italiano parlato di sordi nativi e stranieri. In *Il linguaggio disturbato. Modelli, strumenti, dati empirici*, Bari, Università degli studi di Bari.

Pellegrino, E., De Meo, A. & Caruso, V. (2015). "Chi compie l'azione?" L'applicazione del Competition Model su sordi italiani. In Favilla, M.E., Nuzzo, E. (Eds.), *Grammatica applicata: apprendimento, insegnamento, patologie*. Milano: studi AItLA, 165-178.

Pettorino, M., Giannini, A. (2010). Il parlato dei mass media. In Pettorino, M., Giannini, A. & Dovetto, F. (Eds.), *La Comunicazione Parlata* 3, Napoli, OPAR, Università degli studi di Napoli 'L'Orientale', 71-83.

Pratt, S.R., Tye-Murray, N. (2011). Speech Impairment Secondary to Hearing Loss. In Mc Neil, M.R. (Ed.), *Clinical Management of Sensorimotor Speech Disorders*. New York: Thieme, 204-234.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. In *Cognition*, 72, 265-292.

Savy, R. (2005). Specifiche per l'etichettatura dei livelli segmentali. In Albano Leoni, F., Giordano, R. (Eds.), *Italiano Parlato. Analisi di un dialogo*. Napoli: Liguori.

Shriberg, L.D., Kwiatkowski, J., Rasmussen, C., Gregory, L.L. & Miller, J.F. (1992). The prosody-voice screening profile (pvsp): psychometric data and reference information for children. Phonology Project Technical Report n. 1, Waisman Center on Mental Retardation and Human Development, University of Wisconsin-Madison, 1-54.

Stathopoulos, E., Duchan, J., Sonnenmeier, R.M. & Bruce, N.V. (1986). Intonation and pausing in deaf speech. In *Folia Phoniatrica*, 38, 1-12.

Stein, D. (1980). A study of articulatory characteristics of deaf talkers, Doctoral dissertation, University of Iowa.

Tye-Murray, N., Spencer, L., Bedia, E.G., Woodworth, G. (1996). Differences in children's sound production when speaking with a cochlear implant turned on and turned off. In *Journal of speech and hearing research*, 39, 604-10.

Tye-Murray, N. (1991). The establishment of open articulatory postures by deaf and hearing talkers. In *Journal of Speech and Hearing Research*, 34, 453-459.

Warren, P., Elgort, I. & Crabbe, D. (2009). Comprehensibility and prosody ratings for pronunciation software development. In *Language Learning & Technology*, 13, 87-102.

DOI: 10.17469/O2101AISV000018

BARBARA GILI FIVELA, MASSIMILIANO M. IRACI,
MIRKO GRIMALDI, CLAUDIO ZMARICH

Consonanti scempie e geminate nel morbo di Parkinson: la produzione di bilabiali

This study aims to investigate speech articulation by Parkinson Disease's speakers and, in particular, to check whether their speech gestures show temporal and spatial specificities in comparison to matching healthy controls in the realization of 1) singleton vs. geminate and 2) voicing and manner of articulation distinction; moreover we want to check if 3) the age of subject may have an impact too on speech gestures characteristics.
Results on the singleton vs. geminate distinction are in line with the literature on both healthy and Parkinson Disease's subjects and show that the latter often correspond to reduced gestures, e.g. in amplitude, with the exception of antero-posterior tongue dorsum movement. Voicing and manner of articulation show complex changes in speech production and age seems to affect speech in terms of greater values for elderly people.

1. *Introduzione*

I soggetti affetti da Morbo di Parkinson spesso sviluppano la forma ipocinetica di disartria (Duffy, 2005), solitamente associata alla realizzazione di movimenti di ampiezza ridotta. In effetti, studi acustici hanno confermato la tendenza in soggetti affetti da morbo di Parkinson a produrre gesti di ampiezza ridotta rispetto ai soggetti di controllo (Skodda et al., 2011; 2012), in linea con la realizzazione di *target undershooting* (Ackermann, Ziegler, 1991). Tuttavia, studi articolatori effettuati con articulografia elettromagnetica (EMA) hanno permesso di riscontrare ampiezze maggiori nelle produzioni dei soggetti disartrici (Wong et al., 2010; 2011), talvolta anche solo lungo un asse di riferimento (Gili Fivela et al., 2014; Iraci et al., in stampa). In uno studio preliminare su due soggetti italiani affetti da Morbo di Parkinson, ad esempio, Gili Fivela et al. (2014) hanno mostrato che l'ampiezza dei gesti articolatori può risultare minore o maggiore di quella dei soggetti di controllo se un dato gesto, in particolare quello linguale, viene osservato sull'asse del movimento verticale o su quello antero-posteriore. Peraltro l'ampiezza del gesto è un parametro rilevante nel quadro della Articulatory Phonology (Browman, Goldstein, 1990) che considera il gesto articolatorio come unità fonologica di riferimento, definita anche in base all'ampiezza.

Nella lingua italiana, ad esempio, è stato dimostrato sulla base di dati relativi ad una varietà di italiano del nord-est, che le consonanti bilabiali geminate sono caratterizzate da maggiore durata e ampiezza del gesto di apertura labiale rispetto alle corrispettive scempie (Gili Fivela, Zmarich, 2005; Zmarich, Gili Fivela, 2005). Il primo confronto effettuato con il parlato di soggetti colpiti dal Morbo di Parkison ha permesso di evi-

denziare che il contrasto linguistico corrispondente alla presenza di scempie e geminate viene mantenuto anche nel parlato patologico, nonostante le summenzionate differenze in ampiezza, differenze nella durata dei gesti consonantici e anche nel *phasing* dei gesti. Quest'ultimo, in particolare, è stato misurato come l'intervallo tra il bersaglio articolatorio raggiunto dal dorso della lingua per la realizzazione della vocale tonica e il bersaglio raggiunto dal labbro inferiore per la consonante target e risulta ridotto nel caso di un soggetto patologico (Gili Fivela et al., 2014). Le considerazioni sul *phasing* nella produzione di geminate e scempie sono peraltro molto importanti, perché sono tradizionalmente messe in relazione con i modelli della coordinazione inter-gestuale (Öhman 1967; Browman, Goldstein, 1990). Per l'italiano, ad esempio, è stato proposto da Smith (1995) il modello Vowel-to-Vowel di Öhman, che prevede un intervallo costante tra le vocali e una relazione costante tra le vocali e la costrizione consonantica e quindi, di fatto, una sorta di imposizione del gesto consonantico su quello vocalico: quest'ultimo resta invariato, mentre il primo inizia prima e finisce dopo, in modo che rimanga inalterata la relazione tra la massima costrizione consonantica e le vocali adiacenti. Tuttavia, dati più recenti sull'italiano (Gili Fivela, Zmarich, 2005; Zmarich, Gili Fivela, 2005; Gili Fivela et al., 2007; Zmarich et al., 2007; 2009, 2011; Iraci et al., in stampa) hanno mostrato che la situazione è più complessa e che di fatto l'intervallo tra le vocali non è costante, così come non è costante per tutti i soggetti la relazione tra la costrizione consonantica e le vocali adiacenti. Inoltre, per quanto riguarda la coordinazione inter-gestuale, mentre per le geminate si osserva un anticipo del gesto consonantico nella vocale precedente, non si osserva un suo prolungamento in quella seguente. La coordinazione inter-gestuale osservata, quindi, è in linea con il modello di Öhman solo per quanto riguarda l'inizio della costrizione consonantica, mentre le altre caratteristiche sono coerenti con il modello di Browman e Goldstein (1990), piuttosto che con quello di Öhman. Browman e Goldstein, infatti, predicono che sia la distanza tra i bersagli vocalici che quella tra bersaglio vocalico e massima costrizione consonantica varino in base al numero di consonanti che sono presenti, quindi in base alla realizzazione di una geminata piuttosto che di una scempia. Sulla base di queste considerazioni è stato osservato, quindi, che i dati sono coerenti con un modello ibrido tra quello di Öhman e quello di Browman e Goldstein.

Lo studio delle consonanti scempie e geminate rappresenta quindi un'ottima opportunità per confrontare il comportamento di soggetti colpiti da Morbo di Parkinson e soggetti di controllo rispetto ad un compito linguistico che può fornire dati interessanti di per sé, circa la resistenza del contrasto linguistico nel parlato patologico, e anche in relazione ai modelli teorici per la coordinazione inter-gestuale.

2. *Indagine sperimentale*

2.1 Obiettivi e ipotesi

L'obiettivo principale di questo lavoro è verificare se i gesti articolatori di soggetti affetti da Morbo di Parkinson mostrino alterazioni, sia spaziali che temporali, rispetto ai gesti di soggetti non patologici. In particolare siamo interessati a gesti

articolatori linguisticamente significativi, come quelli coinvolti nella realizzazione di consonanti scempie e geminate, differenziate per il tratto di sonorità e/o di modo di articolazione. Rispetto ai contesti di indagine, intendiamo quindi verificare se:

1. la realizzazione del contrasto tra consonanti scempie e geminate venga realizzata coerentemente con quanto descritto in letteratura anche dal punto di vista articolatorio e se i parlanti patologici preservino la differenza tra scempie e geminate (ad esempio grazie a strategie compensatorie); particolare attenzione sarà prestata ai parametri relativi alla realizzazione delle consonanti bersaglio e delle vocali che rappresentano il contesto immediatamente adiacente, ma anche alla temporizzazione inter-gestuale;
2. sonorità e modo di articolazione abbiano un effetto significativo sulle misure acustiche e articolatorie relative alla distinzione tra scempie e geminate.

Infine, poiché i nostri soggetti patologici hanno un'età che varia dai 64 agli 81 anni, avremo anche modo di verificare se:

3. l'età dei soggetti possa influire sia sulla durata che sull'ampiezza dei gesti (con conseguenti ricadute sul piano acustico, cfr. F1 ed F2).

Sulla base di quanto descritto in letteratura ipotizziamo che la realizzazione del contrasto tra consonanti scempie e geminate venga realizzata anche dal punto di vista articolatorio e che, coerentemente con i primi risultati relativi all'italiano nel Morbo di Parkinson (Gili Fivela et al. 2014, Iraci et al. in stampa), i parlanti patologici preservino la differenza tra scempie e geminate nonostante una riduzione media dei valori dei parametri considerati; ci aspettiamo anche che alcune differenze riguardino la coordinazione inter-gestuale, sempre in linea con i lavori descritti in letteratura. Circa l'influenza di sonorità e modo di articolazione, le attese sono che la prima, come riportato in Iraci et al. (in stampa), svolga un ruolo significativo, anche solo per via delle caratteristiche tipiche delle varietà regionali del sud nella produzione delle occlusive bilabiali sonore, tipicamente più lunghe (Romano, 2003; Gaillard-Corvaglia, Kamiyama, 2006); ci aspettiamo inoltre che il modo di articolazione influenzi la produzione di scempie e geminate occlusive orali e nasali almeno dal punto di vista articolatorio, data la diversa coordinazione del gesto del velo palatino per la produzione di consonanti orali e nasali. Infine, l'attesa è che l'età dei soggetti possa influire sia sulle caratteristiche acustiche che su quelle articolatorie, date le modificazioni del tratto orale dovute all'invecchiamento, descritte da Xue e Hao (2003) e in parte ritrovate in Gili Fivela et al. (2014), per cui i parlanti più anziani tenderebbero, ad esempio, a produrre vocali più lunghe e aperte (Xue, Hao, 2003).

2.1 Corpus e soggetti

Il corpus è composto da pseudoparole del tipo 'CVC(C)V, in cui le consonanti sono nasali o occlusive bilabiali, sorde o sonore, da realizzare come scempie o geminate e in cui la vocale tonica è /i/ e la post-tonica /a/ (es. /'pipa/, /'pippa/, /'biba/, /'bibba/, /'mima/, /'mimma/). Le pseudoparole sono state inserite nella frase cornice

"La 'CVC(C)V blu" e ai soggetti è stato chiesto di produrre almeno sette ripetizioni di ogni frase, presentata in ordine random.

Sono state acquisite le produzioni di dieci soggetti di cui cinque patologici e cinque non patologici, di controllo. Tutti i soggetti provengono dall'area di Lecce e hanno un'età compresa tra i 64 e gli 81 anni (patologici: media 71,8, d.s ±7,19; controllo: media 70,6, d.s. ±7,02). Ai fini dell'analisi, i soggetti sono stati raggruppati in tre insiemi sulla base della loro età anagrafica: nel primo confluiscono i soggetti di età inferiore ai 69 anni (Giovani: PD-1, PD-5, CTR-1, CTR-5), nel secondo quelli di età compresa tra i 69 e i 75 (Intermedi: PD-3, PD-4, CTR-3 e CTR-4), e nell'ultimo i soggetti più anziani (Anziani: PD-2 e CTR-2).

I soggetti patologici sono affetti da Morbo di Parkinson e hanno sviluppato una disartria di tipo ipocinetico, e in base alla valutazione dell'intellegibilità del loro parlato compiuta dagli autori hanno raggiunto un livello di compromissione che varia da lieve-medio a medio-alto. I soggetti di controllo, invece, hanno dichiarato di non aver mai riportato alcun tipo di compromissione a livello neuro-cognitivo e di non essere mai stati affetti da disturbi del linguaggio o della parola. Tutti i soggetti hanno letto il foglio informativo e firmato il modulo predisposto per il consenso informato.

2.2 Strumenti e registrazione

Le acquisizioni sono state effettuate all'interno di una stanza silenziosa presso il CRIL di Lecce, per mezzo di un Articulografo Elettromagnetico, AG501 (Cartsens GmbH) e, quindi, della simultanea acquisizione del dato cinematico e del corrispondente segnale acustico. Sono stati applicati in totale sette sensori su ogni soggetto: 2 sulla lingua (di cui 1 sulla punta e 1 sul dorso), 2 sulle labbra (1 sul labbro superiore e 1 su quello inferiore), 1 sul naso e 2 dietro alle orecchie per la normalizzazione dei dati. I sensori presi in considerazione nel corso di questo studio sono LL (*lower lip*) e TD (*tongue dorsum*): LL corrisponde al sensore applicato sul labbro inferiore, rappresentativo dei gesti consonantici; TD è il sensore applicato sul dorso della lingua ed è rappresentativo dei gesti vocalici.

I soggetti sedevano comodamente e familiarizzavano con il corpus presentato su copia cartacea; dopo il termine del set-up potevano continuare a familiarizzare con il corpus, questa volta presentato sullo schermo. Durante la registrazione, i soggetti potevano decidere di effettuare interruzioni in caso di necessità (es. bere, riposarsi). La durata complessiva delle registrazione variava da un minimo di 45 minuti ad un massimo di un'ora e 15 minuti, soprattutto in relazione al numero e durata delle interruzioni.

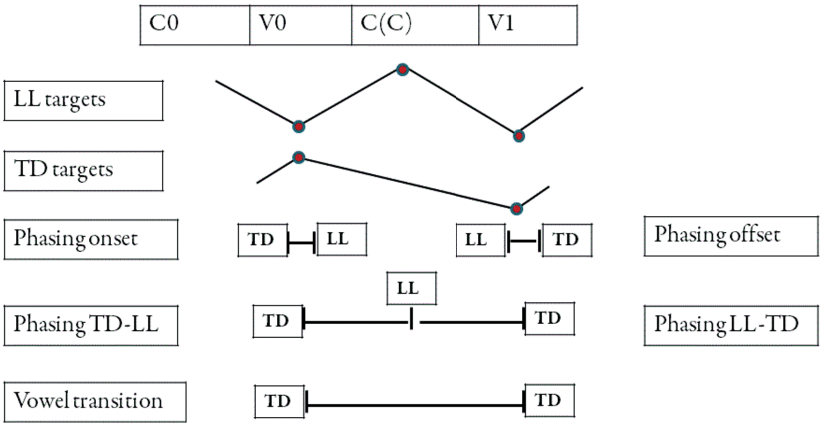
2.3 Etichettatura e misure

Il segnale acustico è stato etichettato manualmente grazie a Praat (Boersma, Weenink, 2009), in modo da indicare i confini segmentali e tre punti della parte stabile della vocale (uno al centro e gli altri due a ±20 msec; v. Fig. 1). Sulla base delle etichette, sono state misurate le durate acustiche della vocale tonica (V0), della con-

sonante target (C) e della vocale post-tonica (V1) e la prima e la seconda formante di entrambe le vocali, ricavate grazie alla media dei tre punti etichettati sulla parte stabile della vocale.

Il segnale cinematico è stato etichettato in modo semiautomatico per mezzo di MAYDAY (Sigona et al., 2015). Sul tracciato di velocità di ogni sensore, è stato etichettato ogni picco di velocità associato al gesto ed il valore minimo successivo, corrispondente al target sul tracciato di posizione (v. Fig. 1). Per ogni articolatore preso in considerazione (il labbro inferiore, LL, e il dorso della lingua, TD) sono state misurate le durate e le ampiezze dei gesti articolatori sia di approssimazione al target (*closing gesture*) che di allontanamento da esso (*opening gesture*); le misure hanno riguardato sia l'asse verticale, sia quello antero-posteriore. Ad esempio (v. Figura 1) la durata e l'ampiezza del gesto di chiusura di LL corrispondono rispettivamente al tempo e allo spazio tra il target di V0 e quello di C (da /i/ a /m/ nella parola /'mima/); la durata e l'ampiezza del gesto di chiusura di TD, corrispondono rispettivamente al tempo e allo spazio tra il target della vocale pre-tonica contenuta nella frase cornice e quello di V0 (da /a/ ad /i/ in /la'mima/); il gesto di apertura di TD corrisponde all'intervallo tra V0 e V1 (da /i/ ad /a/ in /la'mima/). Infine, la misura tradizionale del *phasing* tra due gesti corrisponde all'intervallo tra il target di posizione dell'uno e quello dell'altro (es., il *phasing* TD-LL è l'intervallo tra il target di posizione su TD per /i/ e il target di posizione su LL per C(C), mentre il *phasing* LL-TD è l'intervallo tra il target di posizione su LL per C(C) e il target di posizione su TD per /a/); il *phasing* onset/offset corrisponde all'intervallo tra l'inizio/fine del gesto consonantico LL per C(C) e il target di posizione su TD per /i/ o /a/.

Figura 1 - *Etichette utilizzate e alcune misure ricavate*



2.4 Statistica

L'analisi statistica è stata effettuata per mezzo del software *R* (R Core Team, 2015) e grazie a modelli ad effetti misti (Bates et al., 2014: pacchetto *lme4*) che prevedevano sei fattori fissi e un termine di errore (*random slope*). I fattori fissi riguardavano la presenza/assenza della patologia (Tipo, 2 livelli: soggetti patologici e di controllo), l'e-

tà dei soggetti (Età, 3 livelli: giovani, intermedi e anziani), lo status della consonante (Condizione, 2 livelli: scempia e geminata), la sonorità della consonante (Sonorità, 2 livelli: sorda e sonora), il modo di articolazione che la caratterizza (Modo, 2 livelli: occlusiva e nasale) e il ricorso a ripetizione nell'elicitazione dei materiali (Ripetizione, 7 livelli: numero ripetizioni analizzate per ogni item); il termine d'errore è stato aggiunto al modello per tenere conto della variabilità inter-soggetto nella produzione degli item[1].

Per ogni misura da noi considerata, ossia per ogni variabile dipendente, è stata verificata l'eventuale violazione di omoschedasticità e normalità attraverso un'analisi dei residui. La mancanza di violazioni ci ha garantito la possibilità di applicare i modelli misti e di verificare l'effetto di cinque dei sei fattori, ad esclusione del fattore Ripetizione, grazie alla costruzione del modello generale appena descritto e di un modello "ridotto", nel quale era escluso il fattore da testare: la significatività di ogni fattore è stata verificata attraverso il confronto tra il modello generale e il modello "ridotto", effettuato per mezzo del *Likelihood Ratio Test* (procedura che prevede il test chi quadro; Winter, 2013). Al fine di verificare la significatività delle interazioni, è stata adottata una procedura analoga, ma piuttosto che un modello "ridotto" è stato creato un modello "ampliato", che sommasse una possibile interazione al modello generale. Per effettuare i test post-hoc abbiamo usato il test di Tukey (Hothorn *et al.* 2008: pacchetto *multcomp*). I test sono stati considerati significativi con $p \leq 0.05$.

3. *Risultati*

3.1 Dati acustici

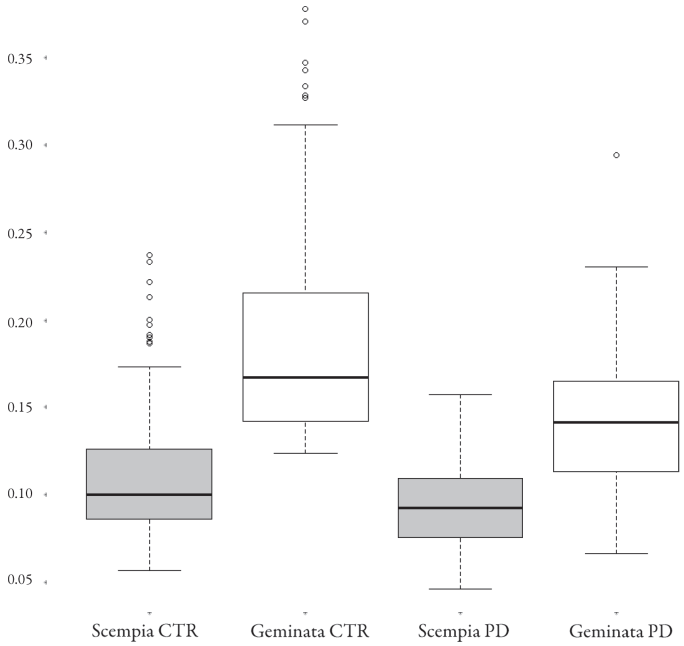
La durata della C target varia in base ai fattori Condizione ($\chi2(1)=18.449$, $p<0.001$) e Modo ($\chi2(2)=6.071$, $p=0.048$) (Figura 2). In particolare, la durata delle geminate è maggiore di quella delle scempie e la durata delle occlusive è maggiore di quella delle nasali. Inoltre risultano significative le interazioni tra Tipo e Condizione, tra Condizione e Sonorità e tra Condizione e Modo. La prima interazione mostra che le geminate hanno durata maggiore delle scempie per entrambe le popolazioni di soggetti, ma il post-hoc rivela che le geminate dei soggetti di controllo sono significativamente diverse solo dalle scempie dei patologici, mentre le geminate dei patologici non sono significativamente più lunghe né delle scempie dello stesso gruppo, né di quelle prodotte dai soggetti di controllo (sebbene tendenzialmente la loro durata sia sempre maggiore rispetto ad entrambi i casi); l'interazione tra Condizione e Sonorità sottolinea che le geminate sono più lunghe sia delle scempie sorde che delle sonore, mentre non c'è differenza dovuta alla sonorità tra le scempie e tra le geminate; infine, l'interazione tra Condizione e Modo mostra che i valori sono inferiori per le nasali tranne che nel caso delle geminate, per le quali non si registra una differenza significativa.

Per quanto riguarda le vocali, la durata della tonica, /i/, varia in base a tutti i fattori considerati, ossia Tipo, Età, Condizione, Sonorità e Modo ($\chi2(1)=15.38$, $p<0.001$;

[1] Modello = variabile dipendente ~ tipo + età + condizione + sonorità + modo + repetition + (1 + condizione + sonorità + modo | speaker).

$\chi2(1)=6.372$, p=0.041; $\chi2(1)=6.1219$, p=0.013; $\chi2(1)=10.14$, p=0.001; $\chi2(2)=10.69$, p=0.004). In particolare, la durata della vocale è maggiore nei patologici che nei soggetti di controllo, mentre è minore prima di geminata che di scempia ed è minore prima delle sorde e delle occlusive. Per quanto riguarda il fattore Età, le vocali dei giovani risultano significativamente più brevi dei quelle degli altri parlanti, mentre le interazioni tra i fattori non sono significative. La durata acustica della vocale postonica /a/ varia significativamente in base ai fattori Sonorità e Modo ($\chi2(1)=8.35$,p=0.003; $\chi2(2)=8.381$,p=0.015) e per l'interazione tra Tipo e Età. In particolare, la durata della vocale è maggiore dopo consonante sonora che dopo sorda e dopo occlusiva piuttosto che nasale; per quanto riguarda l'interazione, le vocali dei patologici di età intermedia sono più lunghe di quelle dei giovani (sia soggetti di controllo che patologici), del gruppo di controllo intermedio e dei patologici anziani e sono quindi simili solo a quelle del gruppo di controllo deglianziani.

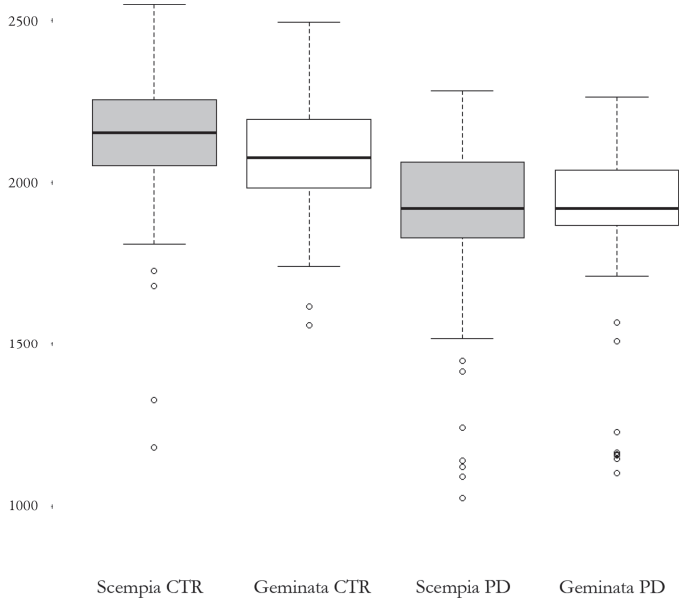
Figura 2 - *Durata acustica della consonante*



Per quanto riguarda i valori formantici della vocale tonica /i/, F1 varia in base al Modo ($\chi2(1)=15.53$, p=0.0004), con valori minori per le occlusive che per le nasali, mentre non si riscontrano interazioni significative. Non si riscontrano invece variazioni significative di F2 in base ai fattori considerati o a loro interazioni. Tuttavia i valori di F2 sono leggermente inferiori nei patologici (Figura 3).

Nel caso della vocale postonica /a/, non si riscontrano fattori significativi per quanto riguarda i valori di F1, mentre la F2 è influenzata dal fattore Sonorità ($\chi2(1)=3.870$, p=0.049), con interazione tra Età e Condizione e tra Età e Sonorità. Infatti per i giovani la geminata corrisponde a valori minori di quelli individuati per la scempia, e la sonora ha valori maggiori della sorda solo per gli anziani.

Figura 3 - *Valori di F2 della tonica /i/*



3.2 Dati articolatori

3.2.1 Gesto labiale

I test effettuati sulla durata del gesto consonantico di chiusura (Figura 4a) hanno dato risultati significativi per il fattore Condizione, Sonorità e Modo (rispettivamente, $\chi2(1)=5.718$, $p=0.016$; $\chi2(1)=8.783$, $p=0.003$; $\chi2(2)=8.960$, $p=0.011$) e per l'interazione tra Tipo e Età e tra Condizione e Modo. In particolare, i gesti sono più lunghi per le geminate, le sonore e le occlusive orali, e, per quanto riguarda le interazioni, la durata del gesto dei soggetti di controllo anziani è significativamente maggiore di quella dei patologici, sia giovani che anziani, e le geminate nasali risultano corrispondere a valori maggiori di tutte le scempie, mentre le geminate occlusive risultano maggiori rispetto alle sole scempie nasali.

La durata del gesto consonantico di apertura (Figura 4b) è significativamente influenzata dal fattore Tipo e dal fattore Condizione (rispettivamente $\chi2(1)=3.933$, $p=0.047$ e $\chi2(1)=6.198$, $p=0.012$) con una forte interazione tra Condizione e Modo. I gesti durano di più nella realizzazione di geminate e nelle produzioni dei soggetti di controllo e, per quanto riguarda l'interazione, la scempia nasale mostra valori inferiori rispetto a tutte le altre, inclusa la scempia occlusiva, ma le scempie corrispondono comunque a valori inferiori rispetto alle geminate.

Figura 4a - *Gesto consonantico di chiusura: durata*

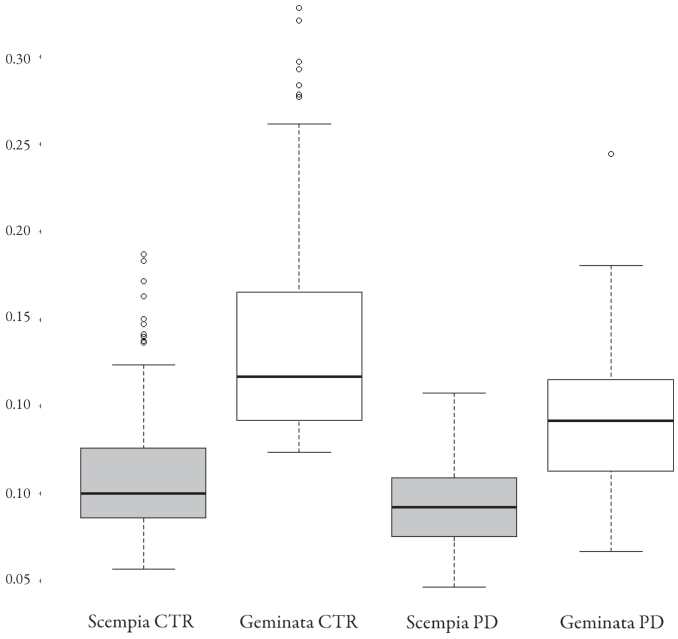
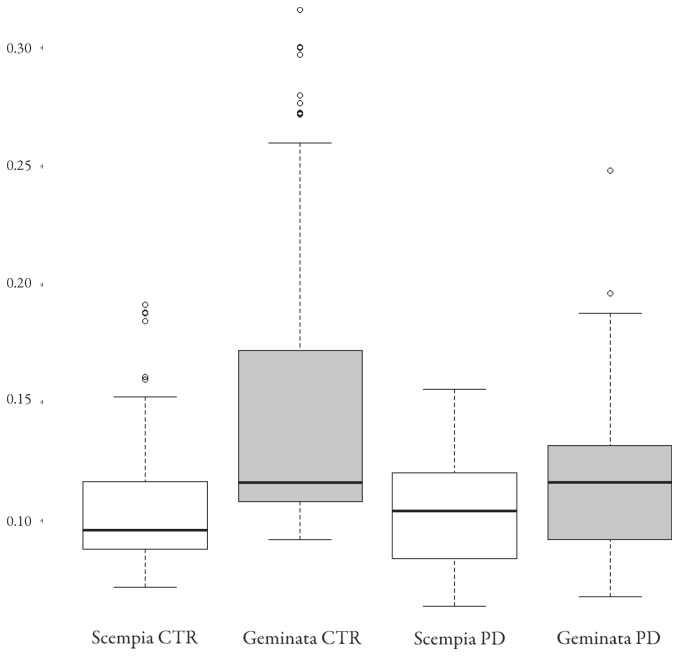


Figura 4b - *Gesto consonantico di apertura: durata*



La durata totale del gesto labiale (chiusura + apertura) varia significativamente in base ai fattori Condizione e Età (rispettivamente, $\chi 2(1)=17.852$, $p<0.0002$

e $\chi2(2)=7.376$, $p=0.025$), con una leggera interazione tra Tipo e Condizione (0.051) e tra Condizione e Modo. La durata del gesto è sempre maggiore per le geminate e per il fattore Età il post-hoc di Tukey rivela che il gesto dei giovani risulta significativamente più breve di quello dei soggetti intermedi e degli anziani. Per quanto riguarda le interazioni, quella tra Condizione e Modo corrisponde al fatto che si registra una differenza statisticamente significativa tra tutte le sottocondizioni, tranne che tra geminate nasali e occlusive; l'interazione tra Tipo e Condizione è dovuta al fatto che la differenza tra scempie e geminate è significativa solo all'interno delle due popolazioni, ma non attraverso esse (ossia, ad es., la scempia dei soggetti di controllo non è diversa significativamente dalla geminata dei patologici).

Per quanto attiene all'ampiezza del gesto labiale di chiusura lungo l'asse verticale (Figura 5a), risultano significativi tutti i fattori (Tipo: $\chi2(1)=14.673$,$p=0.0001$; Età: $\chi2(2)=20.159$,$p<0.0001$; Condizione: $\chi2(1)= 11.672$, $p=0.0006$; Sonorità: $\chi2(1)=8.736$,$p=0.0031$; Modo: $\chi2(2)=8.948$, $p=0.0114$), con un'interazione tra Condizione e Modo. L'ampiezza è maggiore nelle produzioni dei soggetti di controllo, delle geminate, delle sonore e delle occlusive orali piuttosto che nelle condizioni opposte. Il post-hoc su Età, mostra che l'ampiezza è maggiore per i parlanti di età intermedia. Infine, l'interazione ci mostra però che non c'è differenza tra la geminata nasale e la scempia occlusiva, oltre al fatto che non ci sono differenze di Modo tra scempie o tra geminate.

Figura 5a - *Gesto di chiusura: ampiezza*

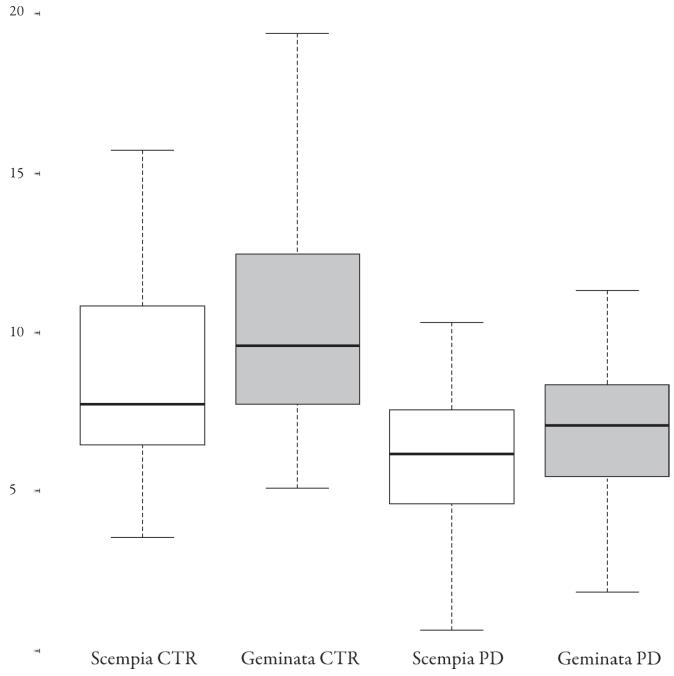
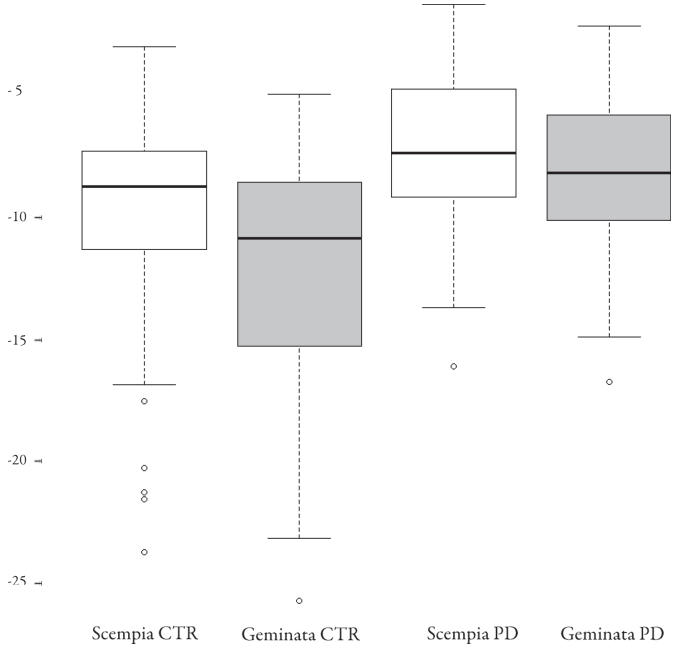


Figura 5b - *Gesto di apertura: ampiezza*



Anche per l'ampiezza del gesto labiale di apertura lungo l'asse verticale (Figura 5b) risultano significativi tutti i fattori (Tipo: $\chi2(1)=5.365$, $p=0.0205$; Età: $\chi2(2)=5.7102$, $p=0.057$; Condizione: $\chi2(1)=15.729$, $p<0.0001$; Sonorità: $\chi2(1)=8.4783$, $p=0.0035$; Modo: $\chi2(2)=10.382$, $p=0.0055$), con un'interazione leggera tra Tipo e Condizione (0.057) e tra Età e Condizione, Età e Sonorità, Età e Modo. In particolare, come per la chiusura, l'ampiezza è maggiore nelle produzioni dei soggetti di controllo, delle geminate, delle sonore e delle occlusive orali piuttosto che nelle altre condizioni. Il post-hoc rivela che per quanto riguarda Età sono i valori dei soggetti di età intermedia ad essere superiori a quelli degli altri. L'analisi dell'interazione tra Tipo e Condizione mostra che c'è una differenza significativa dell'ampiezza del gesto tra i valori di tutte le combinazioni considerate, tranne che tra quelli delle geminate prodotte dei patologici e delle scempie prodotte dai soggetti di controllo, in linea con un aumento dell'ampiezza dalle scempie dei patologici alle geminate dei soggetti di controllo, con le scempie dei soggetti di controllo e le geminate dei patologici in posizione intermedia (Figura 5b). Per quanto riguarda l'interazione tra Condizione e Età, risulta che è il gruppo di età intermedia a differire dagli altri per i valori di ampiezza, tranne che per le scempie rispetto alle geminate degli anziani: di fatto né giovani né anziani sembrano differenziare scempie e geminate (ma si consideri che questo riguarda dati in cui i soggetti di controllo non sono scorporati dai patologici). Circa l'interazione tra Sonorità e Età, risulta che solo i soggetti di età intermedia differenzino i valori per sorde e sonore e infine, per l'interazione tra Età e Modo, solo i soggetti di età intermedia sembrano differen-

ziare nasali e occlusive, in quanto le loro occlusive sono statisticamente diverse da occlusive e nasali di anziani e di giovani.

3.2.2 Gesto linguale

Il gesto linguale di chiusura sull'asse verticale risulta influenzato da tutti i fattori tranne Condizione (Tipo: $\chi2(1)=7.196$; $p=0.007$; Età: $\chi2(2)=11.726$, $p=0.002$; Sonorità: $\chi2(1)=9.841$, $p=0.001$; Modo: $\chi2(2)=9.856$, $p=0.007$), con interazione tra il fattore Età e Modo e tra Condizione e Modo. Il gesto dura maggiormente nel caso della produzione dei soggetti di controllo, in presenza di consonanti sonore e occlusive orali e, in base al post-hoc di Tukey, per i soggetti di età intermedia rispetto che per tutti gli altri. Per quanto riguarda l'interazione tra Modo e Condizione, i valori cambiano per tutte le combinazioni tranne che per le scempie vs. geminate all'interno dei due diversi modi di articolazione; circa l'interazione tra Età e Modo, all'interno di ogni gruppo di età le nasali corrispondono a valori minori rispetto alle occlusive, ma quelle dei soggetti di età intermedia corrispondono a valori maggiori di quelle degli altri gruppi, e le occlusive del gruppo di età intermedia hanno durata maggiore di quelle dei giovani.

Sulla durata del gesto di apertura si riscontra invece un effetto significativo dei fattori Condizione e Età (rispettivamente $\chi2(1)=13.591$, $p=0.0002$ e $\chi2(2)=8.547$, $p=0.013$) e un'interazione significativa dei fattori Condizione e Modo. Infatti, i gesti durano maggiormente in presenza di consonanti geminate e, come indicato dal post-hoc di Tukey, per soggetti anziani e di età intermedia rispetto a soggetti più giovani. Per quanto riguarda le interazioni, i valori sono diversi per tutte le combinazioni di condizioni, tranne che per scempie nasali e occlusive (peraltro le nasali geminate risultano corrispondere ai valori maggiori).

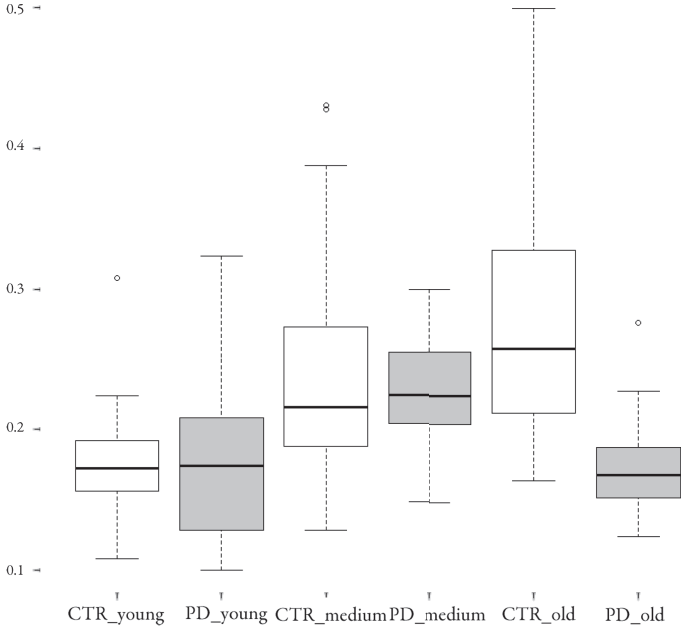
L'analisi dell'ampiezza della chiusura del gesto vocalico sull'asse verticale mette in evidenza un effetto significativo del fattore Condizione ($\chi2(1)=5.855$, $p=0.0155$), ma con interazioni significative tra Tipo e Età e tra Condizione e Età e tra Sonorità e Modo. Il gesto risulta di ampiezza minore nel contesto di geminate che in quello di scempie e, circa le interazioni, risulta infatti che esiste una differenza significativa solo tra i soggetti di controllo che appartengono ai gruppi di età intermedia e di giovani, così come, per l'interazione tra Età e Condizione, tra le geminate e le scempie degli anziani; per l'interazione tra Condizione e Modo, la differenza risulta solo tra occlusive scempie e geminate, e come mostrato dall'interazione tra Condizione e Sonorità, solo tra le sorde scempie e geminate e tra le sorde scempie e le sonore geminate. Infine, per quanto riguarda l'apertura, l'ampiezza è significativamente influenzata dai fattori Condizione, Sonorità e Modo (rispettivamente $\chi2(1)=3.579$, $p=0.058$, $\chi2(1)=13.085$, $p=0.0002$, e $\chi2(2)=13.234$, $p=0.001$), senza alcuna interazione. L'ampiezza dell'apertura è infatti maggiore per le geminate, le consonanti sonore e le occlusive orali.

Per quanto riguarda l'asse antero-posteriore, la durata del gesto di chiusura (Figura 6) varia in relazione ai fattori Sonorità e Modo ($\chi2(1)=8.942$, $p=0.002$, $\chi2(1)=10.168$, $p=0.006$) con l'interazione significativa di Tipo e Età, di Condizione

e Sonorità e di Condizione e Modo. In particolare, i gesti durano maggiormente per le consonanti sonore e le occlusive orali. Circa le interazioni, i più giovani (sia soggetti di controllo che patologici), i soggetti di controllo di età intermedia e i patologici anziani presentano valori minori dei soggetti di controllo anziani e dei patologici di età intermedia.

Invece il gesto di apertura è influenzato dai fattori Età e Condizione (rispettivamente $\chi2(1)=8.258$, $p=0.016$ e $\chi2(1)=13.527$, $p=0.0002$) con un'interazione tra Condizione e Modo. I gesti hanno infatti maggiore durata in presenza di consonanti geminate e il post-hoc chiarisce che la durata è significativamente inferiore nelle produzioni dei più giovani piuttosto che in quelle di tutti gli altri. Per quanto riguarda l'interazione tra Condizione e Modo, esiste una differenza significativa tra tutte le condizioni tranne che tra le scempie occlusive e nasali.

Figura 6 - *Gesto di chiusura di TD su X: durata*



Per quanto riguarda l'ampiezza, il gesto di chiusura (Figura 7a) è significativamente influenzato dai fattori Tipo e Modo (rispettivamente $\chi2(1)=8.237$, $p=0.004$ e $\chi2(1)=7.795$, $p=0.0209$), con una forte interazione tra Tipo e Età (v. Fig. 6). In particolare, il gesto è più ampio nelle produzioni dei parlanti patologici e nelle nasali piuttosto che nelle occlusive orali. Circa l'interazione, risulta chiaramente che i valori dei patologici di età intermedia sono superiori a quelli di tutti gli altri, mentre quelli dei patologici anziani sono superiori solo rispetto ai quelli dei soggetti di controllo anziani e di età intermedia; tra i valori più bassi si registrano quelli dei soggetti di controllo anziani e di età intermedia, diversi da quelli dei soggetti di controllo più giovani.

Anche il gesto di apertura (Figura 7b) varia in ampiezza ma solo a seconda del valore del fattore Tipo ($\chi 2(1)=6.074$, $p=0.013$), e con un'interazione tra Tipo e Età, tra Tipo e Sonorità e tra Condizione e Modo. In particolare, il gesto è più ampio nelle produzioni dei parlanti patologici e dall'analisi delle interazioni si comprende che non c'è differenza statistica tra patologici e soggetti di controllo giovani, i patologici anziani non differiscono dai soggetti di controllo anziani (che a loro volta non differiscono dai soggetti di controllo di età intermedia) e i patologici giovani non differiscono dagli altri patologici. L'interazione tra Tipo e Sonorità risulta dovuta al fatto che i patologici non fanno differenza tra sorda e sonora mentre i soggetti di controllo sì, e che le produzioni sonore dei soggetti di controllo non sono diverse da quelle dei patologici (e quindi la differenza è essenzialmente dovuta alle sorde). Per quanto riguarda l'interazione tra Condizione e Modo, risulta che siano le geminate occlusive ad essere più ampie dalle nasali, sia scempie che geminate.

Figura 7a - *Gesto di chiusura di TD su X: ampiezza*

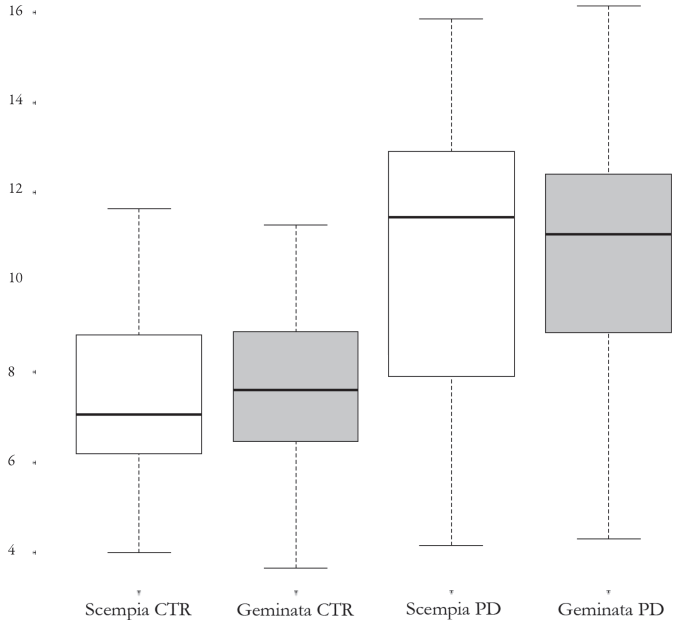
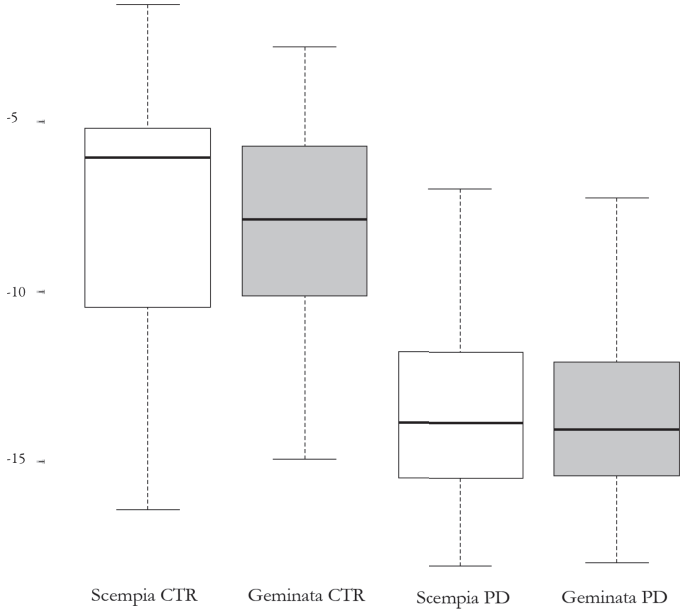


Figura 7b - *Gesto di apertura di TD su X: ampiezza*



3.2.3 Il *phasing* e la coordinazione inter-gestuale

Il controllo effettuato solo sulle nasali e sui parlanti non patologici offre risultati in linea con i precedenti (Gili Fivela et al., 2007; Zmarich et al., 2007; 2009; 2011). In particolare, i test relativi al *phasing onset* e *offset* (Figure 8a e 8b) mostrano che il fattore Condizione ha un effetto significativo o tendente alla significatività (per onset: $\chi2(1)=5.675$, $p=0.0172$; per offset: $\chi2(1)=3.511$, $p=0.06$, senza interazioni significative) e, in particolare, le geminate corrispondono a un intervallo significativamente più ampio delle scempie per l'onset e a uno meno ampio per l'offset. I dati indicano quindi che il gesto della geminata inizia prima di quello della scempia ed è quindi anticipato rispetto al target della vocale tonica; peraltro finisce dopo, quasi in concomitanza con il target per la vocale seguente.

I risultati del *phasing* tradizionale sono invece meno interessanti, perché solo il *phasing* tra il labbro inferiore e la lingua (LL-TD) risulta influenzato dal fattore Condizione ($\chi2(1)=5.978$, $p=0.0144$) con le geminate che presentano valori maggiori delle scempie. Il *phasing* tra la lingua e il labbro inferiore (TD-LL) risulta significativamente influenzato solo dal fattore Età ($\chi2(2)=8.029$, $p=0.0180$), senza interazioni, con il test post-hoc di Tukey che mostra che i valori sono significativamente maggiori per i soggetti anziani rispetto agli altri soggetti.

Figura 8a - *Phasing onset*

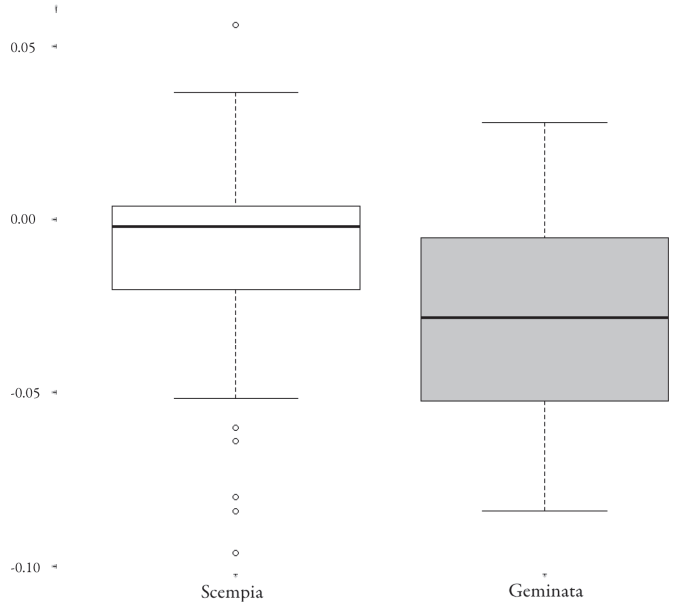
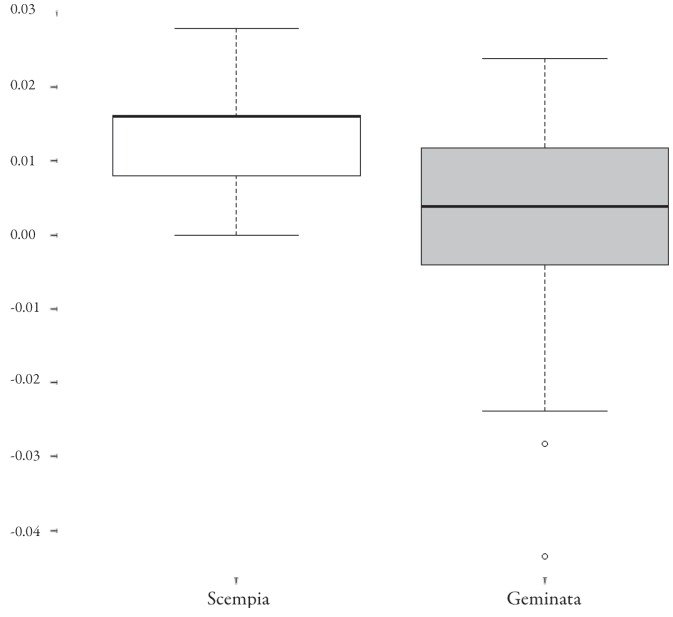


Figura 8b - *Phasing offset*

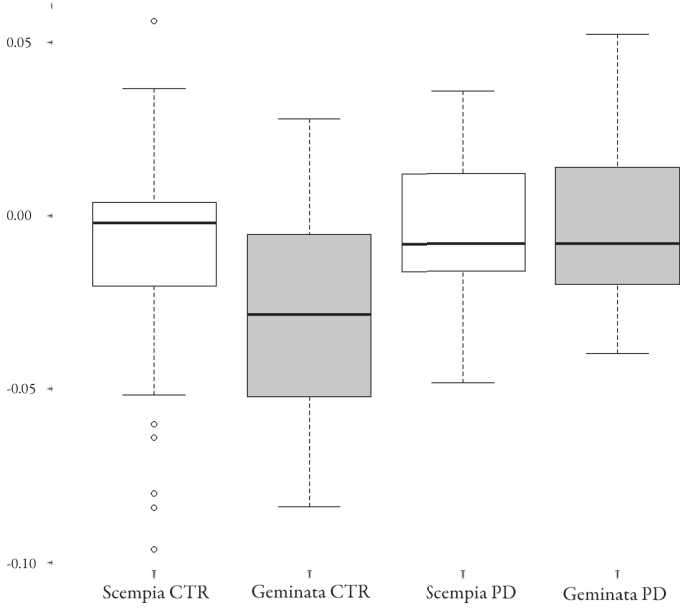


Per quanto riguarda l'effetto della patologia sul *phasing*, a giudicare dai risultati dei test statistici con i quali abbiamo verificato l'esistenza di una differenza tra i materiali prodotti dai parlanti patologici e dai soggetti di controllo, sempre relativamente alle nasali, emerge che per il *phasing* onset non ci sono effetti significativi dei fattori

fissi da noi considerati, tanto meno di Tipo; risultano invece significative le interazioni tra Tipo e Età (per il fatto che si registrano valori maggiori per i soggetti di controllo di età intermedia rispetto a quelli osservati per i patologici anziani e di età intermedia) e tra Tipo e Condizione (perché in realtà solo per i soggetti di controllo esiste la differenza tra scempie e geminate). L'analisi delle altre misure non permette di evidenziare differenze tra soggetti patologici e di controllo. L'insieme dei dati delle due popolazioni, infatti, mostra che per il *phasing* offset non ci sono risultati significativi dei fattori considerati, né interazioni.

Invece le misure relative al *phasing* tradizionale mostrano che, considerando i materiali dei patologici oltre che dei soggetti di controllo, TD-LL varia significativamente in base al fattore Condizione [$\chi 2(1)=5.922$, $p=0.0149$], con le geminate che corrispondono a valori maggiori rispetto alle scempie, senza interazioni di sorta. Il *phasing* tra LL-TD, infine, corrisponde a quello osservato per i soli soggetti di controllo, ossia i valori delle geminate sono significativamente maggiori rispetto a quelli delle scempie [$\chi 2(1)=7.994$, $p=0.0046$]; senza interazioni significative.

Figura 9 - *Phasing onset: soggetti di controllo vs. patologici*



4. *Discussione*

Il modello statistico usato per analizzare i risultati ha permesso di far emergere l'influenza di diversi fattori e anche interazioni interessanti tra essi. Nella discussione ci concentreremo soprattutto sui fattori fissi, evidenziando la presenza di interazioni nei casi più interessanti.

I risultati delle misure acustiche sono coerenti con quelli attesi in base alla letteratura per quanto riguarda l'opposizione di scempie e geminate, in quanto queste ultime risultano essere significativamente più lunghe delle scempie e precedute da vocali più brevi. Anche dal punto di vista articolatorio risultano confermate molte osservazioni riportate nella letteratura precedente. In particolare, i gesti consonantici (labiali) durano di più e sono più ampi per le geminate che per le scempie. Per quanto riguarda i gesti linguali, quello di apertura sull'asse verticale ha durata e ampiezza maggiore per le geminate che per le scempie. Sull'asse orizzontale, antero-posteriore è nuovamente il gesto linguale di apertura ad avere durata maggiore per le geminate, mentre la sua ampiezza e i valori del gesto di chiusura non sembrano cambiare in relazione al contrasto linguistico. I risultati sono in linea con le attese dovute alle indagini precedenti (Gili Fivela, Zmarich, 2005; Zmarich, Gili Fivela, 2005; Zmarich et al., 2007) e aggiungono informazioni circa i gesti linguali (si veda anche Iraci et al., in stampa).

Sempre in linea con la letteratura precedente, si riscontra che i parlanti patologici distinguono scempie e geminate in termini di misure acustiche, perché le geminate hanno durata maggiore delle scempie. Tuttavia in media i patologici presentano valori minori dei soggetti di controllo e infatti le loro geminate non sono significativamente più lunghe delle scempie dei soggetti di controllo e tanto meno delle loro geminate; la vocale tonica, invece, nei patologici ha durata media maggiore e, anche se la differenza non è statisticamente significativa, presenta valori di F2 leggermente minori che nei soggetti di controllo. Per quanto riguarda l'articolazione, la patologia influenza l'ampiezza dei gesti consonantici che risultano meno ampi sia in chiusura che in apertura, mentre incide sulla durata soprattutto per il gesto di apertura, che risulta significativamente più breve nei patologici che nei soggetti di controllo; peraltro l'influenza della patologia si osserva anche sul gesto consonantico nel suo complesso (chiusura+apertura) che non risulta sempre diverso tra scempie e geminate attraverso le popolazioni di soggetti (in quanto, ad esempio, la scempia dei soggetti di controllo non è significativamente diversa dalla geminata dei patologici, risultato che emerge anche nell'analisi dell'ampiezza del gesto di apertura). Circa i gesti linguali, è solo il gesto di chiusura a mostrare valori minori per i patologici, benché solo l'ampiezza, ma non la durata, cambi in base al contrasto linguistico (con ampiezza sorprendentemente minore per le geminate). Sull'asse antero-posteriore, la differenza tra soggetti di controllo e patologici si riscontra solo nei valori di ampiezza, che sono maggiori per i patologici sia per la chiusura che per l'apertura. Si noti però che l'ampiezza del gesto di apertura per i patologici non varia tra sorde e sonore e la differenza tra soggetti di controllo e patologici sembra riguardare più le sorde. In realtà, l'analisi delle interazioni per il gesto di apertura chiarisce che non c'è differenza statistica tra patologici e soggetti di controllo giovani, così come i patologici anziani non differiscono dai soggetti di controllo anziani (che a loro volta non differiscono dai soggetti di controllo di età intermedia), facendo intuire che le differenze riguardino soprattutto i parlanti di età intermedia. I risultati descritti in questa sede, quindi, sono coerenti con le nostre aspettative e i risultati precedenti

(Gili Fivela et al. 2014; Iraci et al. in stampa), in quanto i soggetti patologici presentano spesso valori ridotti per le misure considerate, fatto salvo ciò che riguarda l'articolatore linguale e lo spostamento sull'asse antero-posteriore. La riduzione dei valori di ampiezza e di durata è coerente con i tratti tipici della disartria ipocinetica, mentre la ragione dell'incremento dei valori relativi al gesto linguale sull'asse antero-posteriore non ci è chiara, dato che le vocali prodotte dai soggetti patologici nel set di dati qui considerato non risultano diverse da quelle prodotte dai soggetti di controllo (spostate posteriormente nel caso di /a/ e anteriormente nel caso di /i/), a giudicare dai valori formantici (si veda ad esempio la tendenza osservata circa i valori formantici della /i/ tonica). L'ipotesi che avanziamo è che differenze in questo senso esistano e siano rilevanti nel complesso (differenza tra /i/ ed /a/), benché non emergano statisticamente nell'analisi della singola vocale. Ci riproponiamo di verificare questa ipotesi nei prossimi contributi.

Circa l'influenza della sonorità e del modo, effetti significativi della sonorità si sono registrati sulla durata acustica delle vocali, che sono più brevi in contesto di consonante sorda, e sulla F2 della postonica, più alta in contesto di consonante sonora. Per il modo, le consonanti occlusive hanno durata maggiore di quella delle nasali, corrispondono a vocali precedenti più brevi, con F1 minore, e sono seguite da vocali più lunghe rispetto alle nasali. Per quanto riguarda l'articolazione e i fattori sonorità e modo, l'influenza è chiara sull'ampiezza dei gesti labiali di chiusura e apertura, ed è maggiore nel contesto di consonante sonora e di occlusive orali; per quanto riguarda la durata, invece, sono soprattutto i gesti consonantici di chiusura ad essere più influenzati, gesti che sono più lunghi in contesto di consonanti sonore (in corrispondenza dei quali la differenza di durata si registra anche per il gesto consonantico nel suo complesso) e di occlusive orali. Il contesto di consonante sonora sembra quindi corrispondere a misure di ordine maggiore, in linea con le aspettative relative alle occlusive sonore, anche perché è il modo occlusivo, piuttosto che quello nasale, a corrispondere a valori maggiori. Per i gesti linguali, è la durata del gesto di chiusura ad avere valori maggiori in contesto di sonore e occlusive orali, coerentemente con i gesti labiali; per il gesto di apertura è l'ampiezza ad essere maggiore nel contesto di occlusive e sonore. Sull'asse antero-posteriore, risulta che, in modo analogo a quanto osservato per gli altri gesti, la chiusura abbia valori di durata maggiori in contesto di consonanti sonore e occlusive orali; tuttavia l'ampiezza del gesto di chiusura risulta stranamente maggiore per le nasali che per le occlusive orali. Ci sembra quindi che questi risultati si possano mettere in relazione con le caratteristiche tipiche delle varietà regionali del Sud nella produzione delle occlusive bilabiali sonore, normalmente rafforzate (Romano, 2003, Gaillard-Corvaglia & Kamiyama 2006; v. anche Iraci et al., in stampa). Se il modo esercita poi un'influenza legata alla necessaria diversa coordinazione del gesto del velo palatino, oltre ai gesti necessari per la produzione di consonanti orali, per ora possiamo solo dire che questa non si esprime per mezzo di incremento generalizzato dei valori delle misure relative alle nasali o almeno che l'eventuale incremento è minore di quello osservato per le occlusive nel loro complesso (quindi incluse le sonore). Un elemento degno di

approfondimento futuro riguarda però l'ampiezza del gesto di chiusura, maggiore per le nasali che per occlusive.

L'età dei parlanti sembra giocare un ruolo importante solo in relazione alla durata della tonica e del gesto consonantico nel suo complesso, che risultano più brevi per i parlanti più giovani, e in relazione all'ampiezza dei gesti, che risulta minore anche per gli anziani rispetto ai parlanti di età intermedia. Anche per i gesti linguali, l'età sembra influire. In particolare sulla durata della chiusura, con valori maggiori per i soggetti di età intermedia, e su quella dell'apertura, per cui i valori sono maggiori sia per soggetti di età intermedia che per anziani. Sull'asse antero-posteriore, l'età gioca un ruolo significativo solo per il gesto di apertura, che dura meno nelle produzioni dei più giovani piuttosto che in quelle di tutti gli altri. Nel complesso, quindi, i nostri risultati sono coerenti con quelli descritti da Xue e Hao (2003) e in parte ritrovati in Gili Fivela et al. (2014), per cui i parlanti più anziani tendono a differire dagli altri, ad esempio per vocali più lunghe e aperte (Xue, Hao, 2003). Infatti, benché nel nostro caso anche i parlanti più giovani siano in realtà dei sessantacinquenni, le modificazioni osservate sono comunque coerenti con valori maggiori per le produzioni di coloro che hanno un'età più avanzata, ossia anche per i parlanti di età intermedia che di fatto hanno 74-75 anni.

Infine, l'analisi del *phasing* effettuata per le produzioni delle nasali dei soggetti di controllo mostra che il gesto della geminata inizia prima di quello della scempia, ed è quindi anticipato rispetto al target della vocale tonica, in linea con i risultati di Smith (1995), Gili Fivela et al. (2007) e Zmarich et al. (2007; 2009; 2011). Inoltre, in linea con Smith (1995), ma non con gli altri lavori citati, il gesto finisce dopo nella geminate che nelle scempie (di fatto quasi in concomitanza con il raggiungimento del target per la vocale seguente). Tuttavia, *contra* Smith (1995) e in linea con gli altri risultati sull'italiano, la costrizione non sembra stabile rispetto al target della vocale: infatti le geminate presentano valori maggiori delle scempie per il *phasing* LL-TD. Inoltre, come è stato già puntualizzato, il gesto di apertura linguale varia in termini di durata (e di ampiezza) a seconda che siano presenti consonanti geminate o scempie, di nuovo *contra* Smith (1995). Questi risultati, a partire dall'ultimo messo in evidenza, non sono quindi coerenti con la presenza di un gesto vocalico invariabile sul quale si innesti quello consonantico, come predetto dal modello di Öhman; da questo punto di vista, sono perfettamente coerenti con il modello di Browman e Goldstein. Anche l'anticipo del gesto della geminata nella vocale tonica è coerente con questo modello. Infatti, coerentemente con un'analisi bifonematica delle geminate (per una discussione, si veda Gili Fivela, Zmarich, 2005; Zmarich, Gili Fivela, 2005), l'anticipo del gesto consonantico può essere motivato dall'affiliazione consonantica del primo fonema che compone la geminata, ossia la coda delle sillaba tonica. L'aspetto che invece non sembra coerente con il modello di Browman e Goldstein riguarda invece la posticipazione, osservata nella geminata rispetto alla scempia, della fine del gesto consonantico nella vocale della sillaba postonica. Infatti, l'affiliazione del secondo elemento della geminata non dovrebbe differire da quello della scempia e in entrambi i casi la posizione occupata dovrebbe

essere quella di attacco della sillaba. In questo senso, quindi, continua ad essere vero quanto osservato da Gili Fivela et al., 2007 e Zmarich et al., 2007; 2009; 2011, ossia che sarebbe necessario un modello ibrido tra quello di Öhman e quello di Browman e Goldstein. Ci riproponiamo comunque di verificare ulteriormente questi risultati, ad esempio prendendo in considerazione le misure relative al *plateau* di *attainment* della costrizione consonantica.

Infine, per quanto riguarda l'effetto della patologia sul *phasing*, i dati analizzati in questo studio mostrano che non si riscontrano differenze tra soggetti di controllo e patologici nel *phasing*. Tuttavia, i dati relativi alle interazioni confermano che sono solo i soggetti di controllo a differenziare scempie e geminare in relazione alla misura del *phasing* onset. Peraltro considerare anche i patologici nella popolazione di riferimento, permette di ottenere risultati significativi sia sul *phasing* TD-LL che su quello LL-TD, mentre sui soli soggetti di controllo era solo quest'ultimo a variare in modo significativo in base alla presenza di scempie e geminate. La patologia sembra quindi modificare in positivo o in negativo le significatività osservate nella popolazione di soggetti di controllo, ad indicare il fatto che sono necessarie altre analisi per comprendere appieno quali siano le differenze cruciali che interessano il parlato nel Morbo di Parkinson rispetto al parlato non patologico.

5. *Conclusioni*

L'obiettivo principale di questo lavoro era verificare se i gesti articolatori di soggetti affetti da Morbo di Parkinson mostrassero alterazioni rispetto ai gesti di soggetti non affetti dal Morbo, sia in termini spaziali che temporali, prendendo in considerazione in modo particolare:

1. la realizzazione del contrasto tra consonanti scempie e geminate;
2. la sonorità e il modo di articolazione delle consonanti;
3. il fatto che l'età dei soggetti vari e possa influire sia sulla durata che sull'ampiezza dei gesti (con ovvie ricadute sul piano acustico, cfr. F1 ed F2).

In generale, i risultati discussi sulle consonati scempie e geminate sono in linea con i precedenti (Gili Fivela, Zmarich, 2005; Zmarich, Gili Fivela, 2005) e aggiungono informazioni circa i gesti linguali. Inoltre, circa l'effetto della patologia, è stato confermato che i soggetti patologici presentano spesso valori ridotti per le misure considerate, fatto salvo ciò che riguarda l'articolatore linguale e lo spostamento sull'asse antero-posteriore. Anche in questo caso, i risultati sono quindi in linea con la letteratura precedente (per l'italiano, Gili Fivela et al. 2014, Iraci et al. in stampa). L'influenza della sonorità e del modo di articolazione, risulta abbastanza chiara e in linea con le attese almeno per quanto riguarda il primo fattore, in relazione alla presenza di consonanti bilabiali sonore rafforzate (Romano, 2003; Gaillard-Corvaglia, Kamiyama, 2006). Per quanto riguarda il modo di articolazione, per ora possiamo solo affermare che la maggior complessità articolatoria prevista per le nasali non si esprime per mezzo di incremento generalizzato dei valori delle misure considerate o

almeno che l'eventuale incremento è minore di quello osservato per le occlusive nel loro complesso (quindi incluse le sonore).

Circa l'effetto del fattore età, nel complesso abbiamo riscontrato dati coerenti con quelli descritti da Xue e Hao (2003) e in parte ritrovati in Gili Fivela et al. (2014), per cui i parlanti più anziani tendono a differire dagli altri (Xue, Hao, 2003). Benché nel nostro caso anche i parlanti più giovani siano in realtà dei sessantacinquenni, le modificazioni osservate sono comunque coerenti con l'incremento dei valori nelle produzioni di coloro che hanno un'età più avanzata, ossia anche per i parlanti di età intermedia che di fatto hanno 74-75 anni. Tuttavia, dal momento che il raggruppamento per età accorpa soggetti patologici e di controllo nei tre livelli di età considerati, è ipotizzabile un'influenza della patologia sull'andamento di questi risultati (per una discussione, cfr. Iraci et al., in stampa).

Infine, per quanto riguarda il *phasing*, i risultati ottenuti confermano quanto osservato da Gili Fivela et al., 2007 e Zmarich et al., 2007; 2009; 2011, ossia che la coordinazione inter-gestuale per scempie e geminate è coerente con un modello ibrido tra quello di Öhman e quello di Browman e Goldstein. A questo riguardo, la patologia sembra modificare in vario modo le significatività osservate nella popolazione di soggetti di controllo, ad indicare il fatto che sono necessarie altre analisi per comprendere appieno quali siano le differenze cruciali che interessano il parlato nel Morbo di Parkinson rispetto al parlato non patologico. In ogni caso, l'indagine descritta è preliminare e i risultati dovranno quindi essere verificati tramite ulteriori analisi.

Ringraziamenti

Un sentito ringraziamento va all'ing. Sigona (CRIL) per il cruciale supporto tecnico, a Bodo Winter (University of California, Merced) per i consigli relativi all'analisi statistica mediante modelli misti, ai dott. Danilo Patrocinio e Vincenzo Sallustio (Centro di Foniatria e disturbi della comunicazione, Dip. Riabilitazione, ASL Lecce) per il supporto medico durante il corso delle acquisizioni e per il reperimento dei soggetti patologici, per il quale ci preme anche ringraziare i dott. Filomena My e Giorgio Trianni (Unità Neurologia, Osp.V. Fazzi, Lecce).

Bibliografia

Ackerman, H., Ziegler, W. (1991). Articulatory deficits in Parkinsonian dysarthria: an acoustic analysis. In *J. of Neurology, Neurosurgery, and Psych.*, 54, 1093-8.

Bertinetto, P.M. (1981). Strutture Prosodiche dell'Italiano, *Studi di grammatica italiana, Accademia della Crusca*. Firenze.

Browman, C., Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. In *Journal of Phonetics* (18), 299-320.

Douglas, B., 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.

Boerma, P., Weeknink, D. (2009). *Praat: Doing Phonetics by Computer* (Version 5.1.20). http://www.praat.orghttp://www.praat.org.

Duffy, J.R. (2005). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 2° ed. Elsevier Mosby.

Forrest, K., Weismer, G. & Turner, G.S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. In *Journal of the Acoustical Society of America*, 85, 2608-22.

Gaillard-Corvaglia, A., Kamiyama, T. (2006). La /b/ "forte" in salentino (Puglia): uno studio acustico, percettivo e fisiologico. In *Convegno internazionale "La Comunicazione Parlata 2006"*, Febbraio 2006, Napoli, 87-99.

Iraci, M.M., Zmarich, C., Grimaldi, M. & Gili Fivela, B. (in stampa). Il parlato nel morbo di Parkinson: ampiezza dei gesti articolatori e distintività dei suoni linguistici. In *Giornate di Studio – Il linguaggio disturbato*, 28-30/11/14, Bari.

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 *10th International Seminar of Speech Production (ISSP)*, 5-8 Maggio 2014, Colonia, Germania, 146-9.

Gili Fivela, B., Zmarich, C. (2005). Italian Geminates under Speech Rate and Focalization Changes: Kinematic, Acoustic, and Perception Data. In *InterSpeech 2005*, Lisbona, Portogallo, 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 agosto 2007, 469-472.

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. In *Biometrical Journal*, 50 (3), 346-363.

Maiden, M., Parry, M. (1997), *The Dialects of Italy*. New York: Routledge.

Ohamn, S.E.G. (1967). Numerical Model of Coarticulation. In *Journal of the Acoustical Society of America*, 41, 310-20.

R Core Team (2015). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. http://www.R-project.org.

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.

Sigona, F., Stella, A., Grimaldi, M. & Gili Fivela, B. (2015). MAYDAY: A Multimodal Articulatory Data Analyzer. In *Atti del X Convegno AISV*, 22-24 Gennaio 2014, Torino, 173-184.

Smith, C.L. (1995). Prosodic patterns in the coordination of vowel and consonant gestures. In Connell, B., Arvaniti, A. (Eds.), *Phonology and phonetic evidence. Papers in Laboratory Phonology IV*. Cambridge University Press, 205-222.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications, *arXiv*:1308.5499. http://arxiv.org/pdf/1308.5499.pdf.

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 J. 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.

Xue, S.A., Hao, G.J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. In *Journal of Speech Language and Hearing Research*, 46, 3, 689-701.

Zmarich, C., Gili Fivela, B. (2005). "Consonanti scempie e geminate in italiano: studio cinematico e percettivo dell'articolazione bilabiale e lebiodentale". In *Atti del Convegno Nazionale AISV "Misura dei parametri"*, Padova, 2-4 dicembre 2004. 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 "Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche"*, Povo di Trento, 29-30 novembre-1 dicembre 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., *La Fonetica Sperimentale: Metodo e Applicazioni. Atti del 4° Convegno AISV*. Torriana: 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, 401-404.

DOI: 10.17469/O2101AISV000019

GIOVANNA LENOCI, CATERINA PISCIOTTA, CLAUDIO ZMARICH

Profili evolutivi della coarticolazione per una diagnosi precoce di balbuzie cronica in età prescolare

The purpose of the study is to examine the relationship between abnormal coarticulation profiles in two stuttering children and the risk to develop persistent stuttering. According to Subramanian et al. (2003), in fact, shortly after the onset of the disorder, children whose stuttering will eventually persist show a higher degree of coarticulation than children who will eventually recover. Comparisons were made among one child known to be a persistent stutterer, one child known to be a recovered stutterer and 26 normally fluent controls. Anticipatory coarticulation was analyzed with the *Locus Equation* metric. Contrary to expectations, data show that the child who eventually persist presents a lower degree of coarticulation than the control group while the other one, who eventually recovered from stuttering, shows a degree of coarticulation very close to the control group *range*.

1. *Introduzione*

La balbuzie è uno dei disordini motori della comunicazione verbale più studiati sebbene molti aspetti del disturbo rimangano ancora sconosciuti, soprattutto a causa del suo carattere multidimensionale. Essa implica infatti, in varia misura, fattori linguistici, psicologici e ambientali che interagiscono continuamente e rendono lo studio e la definizione del disturbo alquanto articolati. Van Riper (1990) la definisce come "a neuro-muscolar disorder whose core consists of tiny lags and disruptions in the timing of the complicated movements required for speech" (317).

Nonostante la complessa interazione dei suddetti fattori possa contribuire all'emergere e al persistere della balbuzie, attualmente l'opinione prevalente fra gli studiosi è che la caratteristica saliente del disturbo è l'*output* anomalo del sistema motorio del parlato.

Nella produzione delle disfluenze[1], infatti, il sistema motorio non riesce a generare e/o inviare i comandi motori appropriati ai muscoli per far sì che il parlato continui ad essere fluente[2] (Olander et al., 2010). A partire dalla definizione di balbuzie data da

[1] Le disfluenze sono i sintomi primari della balbuzie e interrompono bruscamente e frequentemente il flusso continuo del parlato. Le disfluenze tipiche della balbuzie sono definite *Stuttering-Like Disfluencies* e sono le ripetizioni di fono o di sillaba, le ripetizioni di parole monosillabiche e i prolungamenti silenti o udibili di fono (Yairi, Ambrose, 2005).

[2] I movimenti articolatori dei balbuzienti sono risultati essere diversi da quelli dei normofluenti in termini di variabilità, velocità e coordinazione anche durante il parlato percettivamente fluente (es. Kleinow, Smith, 2000).

Van Riper (1990), Van Lieshout et al. (2004) hanno sviluppato un approccio teorico (lo *Speech Motor Skill*) nel quale si sostiene che le disfluenze da balbuzie sono il frutto di 'errori' nei meccanismi del sistema di controllo motorio del parlato[3] e che quest'ultimo rappresenta l'anello debole nella catena di eventi che portano alla produzione linguistica. Un'attività motoria come la produzione verbale (cfr. *speech*), per potersi definire specializzata richiede un certo periodo di pratica, necessario per acquisire una precisa organizzazione spaziale e temporale e una minimizzazione dei costi energetici e temporali dell'attività stessa (Van Lieshout et al., 2004). Nell'approccio proposto da Van Lieshout et al. (2004) i balbuzienti si posizionano all'estremità inferiore del *continuum* non patologico che caratterizza l'abilità motoria a fini linguistici dell'intera popolazione, perché hanno un sistema di controllo motorio del parlato 'limitato' e immaturo. Le ricerche interessate all'analisi della coordinazione temporale (coproduzione) tra vari sottosistemi (di solito sono analizzati quello laringale e articolatorio), e tra i gesti diversi dello stesso sottosistema, hanno tutte, in vario modo, confermato l'esistenza di deviazioni rispetto alla norma (Kleinow, Smith, 2000; Namasivayan, van Lishout, 2011). Questi risultati sono coerenti anche con quelli della letteratura sulla coarticolazione nei balbuzienti. A livello articolatorio, la coarticolazione si manifesta attraverso l'anticipazione o la posticipazione dei movimenti degli articolatori impegnati nella realizzazione di un dato fono, in intervalli di tempo in cui altri articolatori sono impegnati nella realizzazione dei foni vicini. Pertanto, nel parlato adulto normofluente, la coarticolazione è sintomo di un alto grado di abilità motoria in quanto permette l'organizzazione spazio-temporale dei gesti articolatori all'interno del tratto vocale e una coordinazione stabile ed efficiente raggiunta attraverso l'interazione sinergica fra muscoli e articolatori.

Lo stabilizzarsi di queste sinergie funzionali è un processo che, nel bambino con normale sviluppo linguistico, si affina nel corso degli anni: uno dei risultati più frequentemente riportati in letteratura nella ricerca sullo sviluppo motorio del parlato è che le produzioni articolatorie dei bambini sono più variabili rispetto a quelle degli adulti (Smith, Goffman, 1998; Green et al., 2000) e che alti livelli di variabilità persistono fino all'età adolescenziale. Superata questa fase, i loro *pattern* articolatori diventano più stabili e più vicini al target adulto (Smith, Zelaznik, 2004).

Gli studi più significativi sulla coarticolazione nei balbuzienti sono quelli relativi alle transizioni della seconda formante (F2), il correlato acustico del luogo di occlusione nel cavo orale, lungo la direzione antero-posteriore (Fant, 1970).

Van Riper (1982) è stato uno dei primi a sostenere che la balbuzie, come disordine motorio del parlato, è il risultato di una 'rottura' nella coarticolazione o nella transizione tra un fono e quello successivo. Secondo Kent (1994) la balbuzie si manifesta proprio nella difficoltà ad eseguire compiti articolatori complessi o nel mantenimento dell'organizzazione spaziale degli organi articolatori, come dimostrato dallo studio di Robb e Blomgren (1997) i quali hanno analizzato la coarticolazione anticipatoria in sillabe CV prodotte fluentemente da soggetti balbuzienti adulti. I risultati delle analisi, condotte misurando i cambiamenti frequenziali delle transizioni di F2, hanno evidenziato che,

[3] Per controllo motorio s'intende la capacità del sistema nervoso di regolare e/o dirigere i movimenti necessari alla produzione verbale.

rispetto al gruppo di controllo, i balbuzienti mostravano pendenze maggiori, ad indicare che la lingua, nel passare dall'articolazione del *target* consonantico a quello vocalico, si muove velocemente nel cavo orale. I due gesti articolatori vengono realizzati, dunque, con una coordinazione spazio-temporale ridotta (vale a dire con basso grado di coarticolazione intrasillabica), ma con una velocità maggiore e questa potrebbe essere la strategia articolatoria adottata per riuscire a mantenere l'eloquio fluente.

Il tentativo di riuscire a interpretare le anomalie riscontrate nel parlato dei balbuzienti adulti come limitazione/restrizione del sistema articolatorio dei balbuzienti o come strategia di adattamento al disturbo incentiva la ricerca a potenziare gli studi sui soggetti prescolari: solo in questo modo è infatti possibile intercettare la balbuzie sul nascere, riducendo al minimo le possibilità che siano intervenute strategie di reazione e compensazione al disturbo.

La balbuzie, infatti, è un disordine della prima infanzia che compare all'età media dei 33 mesi, periodo particolarmente delicato per il bambino, caratterizzato dallo sviluppo anatomofisiologico e delle capacità cognitive e motorie. Oltre all'interesse scientifico di studiare il disturbo nelle sue manifestazioni dirette, ne esiste anche uno etico e propriamente clinico che riguarda la necessità di iniziare precocemente il trattamento terapeutico per quei bambini candidati alla cronicizzazione. Secondo le più recenti stime epidemiologiche, infatti, la balbuzie ha un alto tasso di remissione spontanea (Yairi, Ambrose, 2013): circa il 90% dei bambini che inizia a balbettare sembra debellare completamente il disturbo entro i sei anni d'età, in maniera spontanea e senza intervento clinico. Secondo gli esperti, però, il restante 10% che svilupperà balbuzie cronica deve essere identificato e sottoposto a trattamento terapeutico il prima possibile, in quanto maggiore è l'intervallo temporale che intercorre dalla comparsa del disturbo, minori sono le probabilità di eliminarlo e, una volta che la balbuzie è diventata persistente (sembra che lo diventi dopo 4 anni dall'insorgenza, Yairi, Ambrose, 2005), il trattamento diventa lungo, difficile e quasi mai risolutivo (Starkweather, 1993). Ad oggi, infatti, non esistono cure conosciute, universalmente efficaci, per la guarigione completa da una condizione di cronicità.

Per tutte queste ragioni una delle principali linee di ricerca, intraprese soprattutto negli ultimi anni, è quella di identificare e confrontare i sottogruppi del campione balbuziente: il fenomeno del recupero spontaneo, infatti, solleva parecchi interrogativi riguardo alla natura delle differenze fra balbuzie cronica e remissiva e ai differenti processi evolutivi che caratterizzano le due forme del disturbo.

Inoltre, lo studio dei *pattern* coarticolatori di bambini che hanno da poco iniziato a balbettare permette di osservare il disturbo nelle sue effettive manifestazioni dato che questi, a differenza degli adulti, potrebbero non aver ancora sviluppato strategie di adattamento articolatorio[4].

Lo studio della coarticolazione nei bambini vicini all'insorgenza della balbuzie ha quindi un altro notevole vantaggio: offre la possibilità di identificare precocemente

[4] "coarticulatory behaviors of adults who do and do not stutter are not readily generalizable to the coarticulatory behavior of children who stutter and children who do not stutter, due to many developmental differences in speech/language production between adults and children" (Chang et al., 2002: 677).

eventuali anomalie nell'organizzazione dei gesti articolatori e di facilitarne la riabilitazione nel trattamento terapeutico (Hardcastle, Tjaden, 2008).

Gli studi finora condotti sui bambini balbuzienti sono però scarsi e fra questi uno dei più recenti è quello di Subramanian et al. (2003) i quali hanno analizzato la coarticolazione intrasillabica del parlato percettivamente fluente di 20 bambini balbuzienti, registrati fra i 7 e i 9 mesi successivi alla comparsa del disturbo. Nello studio è stato incluso anche un gruppo di controllo, formato da bambini con normale sviluppo linguistico. L'obiettivo del lavoro era duplice: da un lato confrontare i risultati acustici del campione balbuziente con quello di controllo, dall'altro identificare *pattern* coarticolatori differenti fra quei bambini balbuzienti che alla fine del periodo di osservazione previsto dal progetto avrebbero recuperato spontaneamente, rispetto a quelli che invece avrebbero cronicizzato il disturbo. I parametri acustici analizzati erano la durata e l'estensione della transizione della seconda formante, misurando i valori frequenziali di F2 all'inizio e alla fine della transizione nelle sillabe CV. Dal confronto tra i gruppi, è emerso che un deficit della transizione formantica è significativo e forse centrale in presenza di balbuzie cronica. Nello specifico, i risultati relativi alle variazioni della transizione di F2 hanno mostrato che i soggetti che in seguito avrebbero sviluppato balbuzie cronica mostravano cambiamenti frequenziali minori, sia rispetto al gruppo di controllo sia rispetto all'altro sottogruppo di bambini balbuzienti che successivamente sarebbe guarito spontaneamente dal disturbo. Nessuna differenza è stata registrata, invece, per l'altro parametro acustico analizzato, la durata delle transizioni.

Gli autori dello studio hanno sostenuto, pertanto, che i bambini cronici, per riuscire ad essere fluenti, adottano la strategia articolatoria di un'alta coarticolazione intrasillabica, eseguendo i movimenti articolatori in un range spaziale ridotto rispetto agli altri due gruppi.

Sulla base di tutte le osservazioni teoriche e sperimentali fin qui menzionate, nel presente contributo verrà presentata una parte dei dati relativi all'analisi acustica della coarticolazione anticipatoria effettuata sul parlato di bambini balbuzienti italiani di età prescolare, reclutati nel progetto longitudinale intitolato "Indici fonetici predittivi di balbuzie cronica in età prescolare", coordinato dal Dott. Zmarich dell'ISTC-CNR, sede di Padova. L'obiettivo del progetto è quello di individuare e convalidare il valore prognostico di alcuni indici proposti in letteratura, di tipo percettivo[5], acustico e psicologico[6], ritenuti buoni marcatori di balbuzie cronica in bambini di età prescolare. La convalida del potere predittivo di tali indici permetterebbe di indirizzare, il prima possibile, al trattamento terapeutico solo quei bambini predisposti a sviluppare la forma cronica del disturbo.

[5] Uno degli indici sperimentali del progetto è il *Profilo delle Disfluenze*, che per ragioni di spazio non verrà presentato in questa sede ma per il quale si rimanda a Zmarich (in stampa).

[6] È stato dimostrato che a partire dai tre anni i bambini balbuzienti mostrano consapevolezza delle difficoltà di fluenza e attitudini più negative rispetto ai coetanei normofluenti. Un test di facile somministrazione, per valutare l'attitudine comunicativa dei soggetti prescolari è il *KiddyCAT* (Vanryckegem, Brutten, 2006; Bernardini et al., submitted). Il test prevede che venga chiesto al bambino di rispondere a 12 domande, dirette e poste verbalmente, riguardanti la percezione che il soggetto ha del proprio modo di parlare e dei sentimenti che lo accompagnano ogni volta che parla.

2. *Il Metodo*

I bambini reclutati nel progetto del CNR appartengono a famiglie con genitori o fratelli balbuzienti conclamati, residenti nel Veneto: un risultato frequentemente riportato in letteratura è che la balbuzie ha delle basi ereditarie (Kraft, Yairi, 2012) e che la probabilità di iniziare a balbettare per i bambini con storia familiare di balbuzie è 5-6 volte maggiore rispetto a bambini nati in famiglie di non balbuzienti (Kloth et al., 1999).

Il progetto ha l'obiettivo di seguire longitudinalmente i bambini, dai 24 mesi d'età, prima dunque che insorga la balbuzie, fino ad almeno 15 mesi successivi. In questo modo è possibile monitorare l'evoluzione naturale del disturbo e seguire i soggetti nelle diverse fasi. Passati i 15 mesi successivi all'insorgenza, se i soggetti continuano a balbettare e se i genitori lo richiedono, si interrompe il periodo di osservazione sperimentale e inizia il trattamento terapeutico.

Secondo il protocollo del progetto al raggiungimento del ventiquattresimo mese d'età il bambino veniva audio e video registrato, durante una sessione di interazione semistrutturata con la madre.

Alla fine della sessione di registrazione ai genitori veniva lasciato un pro-memoria per riconoscere formalmente i primi sintomi della balbuzie (le disfluenze) e dei questionari anamnestici. Inoltre veniva loro raccomandato di contattare i referenti del progetto non appena il bambino avesse iniziato a balbettare. Se per i genitori erano comparsi i primi segni della balbuzie, e dopo la verifica della corrispondenza dei sintomi da parte degli operatori del progetto, il bambino veniva accompagnato al Centro Medico di Foniatria per una diagnosi formale del disturbo che comprendeva una serie di accertamenti clinici, tra cui quello otorinolaringoiatrico, psicologico, cognitivo e logopedico che sarebbe stata ripetuta alla fine del periodo di osservazione sperimentale. Parallelamente alla diagnosi clinica formale, iniziava per il bambino una serie di video-audio registrazioni domiciliari, con scadenza trimestrale, effettuate dagli operatori inviati dall'ISTC-CNR, per raccogliere dati sullo sviluppo fonetico (Zmarich et al., 2012; Bortolini, 1995); sullo sviluppo lessicale (Caselli et al., 2007); sull'attitudine comunicativa del bambino (Vanrychegem, Brutten, 2006); sull'evoluzione della gravità della balbuzie (Riley, 1994).

Sempre durante la sessione di registrazione veniva chiesto ai genitori di creare le condizioni per un'interazione spontanea col bambino necessaria allo studio delle variabili sperimentali (cfr. 'Profilo delle disfluenze', analisi acustica della coarticolazione intrasillabica).

I bambini venivano comunque audio-registrati per altre due tappe successive ai 15 mesi dall'esordio, a 18 e 24 mesi, e l'avvenuta risoluzione o meno del disturbo veniva certificata attraverso un'intervista telefonica fatta ai genitori a distanza di 36 e 48 mesi dalle prime manifestazioni del disturbo.

Sono entrate a far parte del progetto 50 famiglie con storia familiare di balbuzie e 44 bambini sono stati registrati a 24 mesi. Tre famiglie si sono ritirate dal progetto subito dopo la prima registrazione e dei 41 bambini rimasti, 17 hanno sviluppato la balbuzie (41,46%). Altre tre famiglie di bambini balbuzienti hanno successivamente rinunciato a partecipare al progetto e il numero complessivo di soggetti si è così ridotto a 14.

2.1 I soggetti dello studio

I soggetti del lavoro sperimentale che qui presentiamo sono due, un maschio e una femmina: Alessandro e Anna. Tutti e due i soggetti hanno storia familiare di balbuzie: Anna è figlia di padre balbuziente e ha iniziato a balbettare all'età di 30 mesi mentre Alessandro ha padre e fratello maggiore balbuzienti, e i primi sintomi della balbuzie sono comparsi ai 31 mesi d'età.

Alessandro ha intrapreso più cicli di trattamento terapeutico alla fine del periodo di osservazione e alla valutazione finale, prevista dal protocollo del progetto per accertare lo *status* di balbuziente/non balbuziente, risulta che la balbuzie ancora persiste. Anna, invece, sembra essersi rimessa spontaneamente dal disturbo.

Per entrambi i soggetti sono state analizzate una o più registrazioni relative alle tappe significative del disturbo (cfr. Tab. 1): Tappa -1, corrispondente ai mesi precedenti la comparsa del disturbo; Tappa 0, relativa al momento della comparsa (o ai primi mesi successivi ad essa); Tappa ≤12 corrispondente ad un intervallo temporale di massimo un anno dall'insorgenza e la Tappa finale, relativa alla registrazione effettuata attorno ai 15 mesi successivi alla comparsa del disturbo, che coincide con la fine dell'osservazione sperimentale. Nel protocollo del progetto la tappa ≤12 è di fondamentale importanza per la predizione di cronicità. Secondo lo studio di Subramanian et al. (2003), riportato nell'introduzione, i bambini destinati a sviluppare balbuzie cronica mostravano, nel secondo semestre dalla comparsa del disturbo, un alto grado di coarticolazione nelle sillabe percettivamente fluenti, sia rispetto al gruppo di controllo sia rispetto ai bambini balbuzienti che in seguito avrebbero debellato il disturbo. Se questi risultati venissero convalidati sarebbe possibile ritenere l'andamento della seconda formante un indice precoce di cronicità in bambini di età prescolare.

Tabella 1 - *Nella tabella vengono riportate le tappe significative dell'evoluzione del disturbo dei due soggetti sperimentali e l'età (in mesi) dei soggetti al momento delle audio-registrazioni*

Soggetti	*Tappa -1*	*Tappa 0*	*Tappa ≤12*	*Tappa finale*
Anna	24 mesi; 29 mesi	30 mesi; 31 mesi	43 mesi;	
Alessandro	24 mesi	34 mesi	42 mesi;	45-48 mesi;

Le audio registrazioni, tutte della durata di circa 75 minuti, sono state effettuate al domicilio dei soggetti, avendo cura di ridurre al minimo i rumori che potevano alterare la qualità del segnale acustico registrato. Per Anna, che faceva parte di uno studio pilota precedente al progetto longitudinale, l'audio-registrazione è stata realizzata tramite un registratore digitale MiniDisc Sony MZ-R70, collocato in un marsupio indossato dalla bambina e con un microfono esterno Sony ECM DS70P tipo Lepelier, fissato sul colletto della maglietta. Per Alessandro, invece, è stato utilizzato un registratore digitale Edirol R-09 con microfono incorporato, posizionato vicino al bambino. Nel progetto è stato incluso anche un gruppo di bambini con normale sviluppo linguistico, attualmente composto di 26 soggetti, d'età compresa fra i 28 e i 54 mesi d'età. I bambini del gruppo normofluente sono stati audio-registrati presso alcuni istituti d'infanzia di Padova e la selezione dei soggetti è avvenuta in seguito al vaglio di questionari sottoposti ai genitori, per escludere la

presenza di anomalie nello sviluppo psicofisiologico dei soggetti. L'analisi acustica della coarticolazione intrasillabica, per tutti i bimbi, ha permesso il confronto fra i due gruppi.

2.2 Trascrizione fonetica

La prima fase dell'analisi è consistita nella trascrizione fonetica manuale (in IPA) del segnale acustico registrato a 44.1 kHz e 16 bit (eccetto i suoni di tipo vegetativo o di difficile identificazione) e acquisito sul PC con una frequenza di campionamento di 22.05 kHz, tramite il software PRAAT[7]. Successivamente, sulla base della trascrizione fonetica, sono stati individuati ed etichettati tutti gli enunciati contenenti sillabe CV o CVC, trascritte con consonante occlusiva[8] bilabiale (p/b), dentale (t/d) e velare (k/g) seguite da vocali, necessarie per l'analisi della coarticolazione. Per ogni file audio è stato creato un *Textgrid* in PRAAT, composto da quattro *Tiers* in cui sono state annotate ortograficamente le sillabe potenzialmente utili all'analisi, i foni effettivamente prodotti dal bambino (in alfabeto SAMPA, perché utilizza i soli caratteri ASCII presenti in una normale tastiera del PC), le sillabe aspirate[9] e le produzioni disfluenti. A causa della relativa rumorosità del segnale la segmentazione tra foni e tra sillabe è stata svolta manualmente, apponendo dei confini all'interno dei *Text-grid*, sulla base dell'andamento delle formanti, in particolare di F2 (Salza, 1990).

2.3 Analisi Acustica

L'analisi acustica della coarticolazione intrasillabica è stata effettuata, inizialmente, soltanto sulle sillabe prodotte in modo percettivamente fluente[10], usando come indice l'andamento della transizione della seconda formante (F2). La metrica utilizzata per quantificare il grado di coarticolazione anticipatoria è quella delle *Locus Equations* (cfr. LE; Sussman et al., 1999), calcolate sui valori frequenziali della transizione della seconda formante, misurati subito dopo il rilascio dell'occlusione consonantica (sul primo ciclo glottico, $F2_{onset}$; cfr. Fig. 1 a)) e a metà della vocale successiva ($F2_{vowel}$; cfr. Fig. 1 a)). Questi valori frequenziali sono stati misurati per ogni singola categoria di luogo (bilabiali, dentali e velari) prodotta con un'ampia gamma di vocali e rappresentati in un piano cartesiano, con le coordinate di $F2_{onset}$ riportate sull'asse delle ordinate e quelle di $F2_{vowel}$ sull'asse delle ascisse (cfr. Fig. 1 b)).

[7] Il programma è scaricabile dal sito www.PRAAT.org.

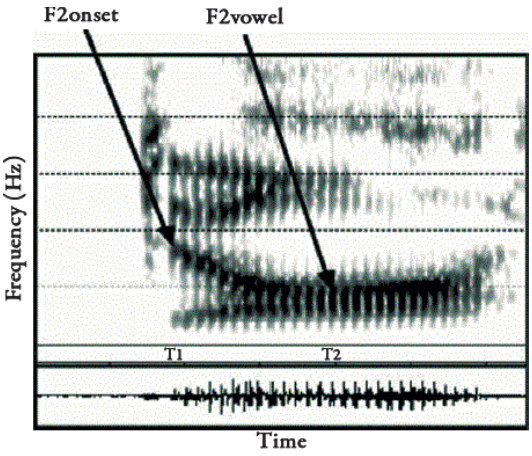
[8] La scelta di selezionare esclusivamente le sequenze con consonanti occlusive è dovuta alla metrica utilizzata, le *Locus Equations* (cfr. LE). Lindblom (1963) derivò inizialmente le LE dalle produzioni di parole contenenti sillabe CV con consonanti occlusive e osservò che le pendenze delle LE variavano in modo sistematico in funzione del luogo di articolazione dell'occlusiva.

[9] Le sillabe aspirate sono da considerare separatamente perché introducono un artefatto nella misurazione rispetto alle non aspirate.

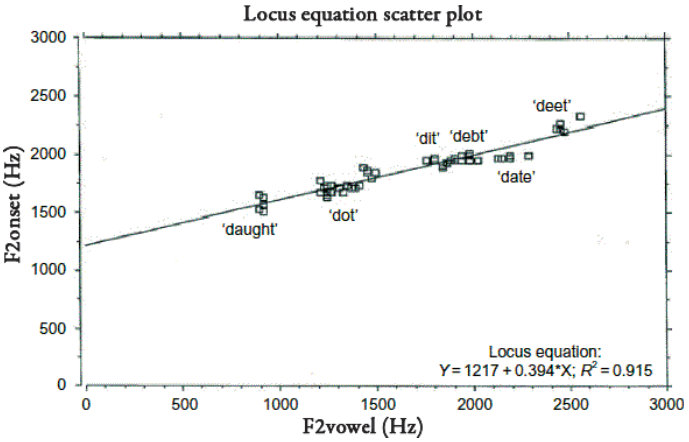
[10] Il numero di sillabe analizzate attraverso le varie tappe sono state, rispettivamente per Anna e Alessandro, le seguenti: bilabiali (358; 136), dentali (512; 282), velari (240; 95).

Figura 1 - a) *Spettrogramma con i punti di misurazione delle coordinate delle LE: punto frequenziale prelevato sul primo ciclo glottidale della vocale (F2onset) e a metà del nucleo vocalico (F2vowel).* b) *Grafico rappresentativo delle LE e della retta di regressione lineare per le sequenze [dVt] (tratto da Sussman, Guitar, 2011: 171)*

a)



b)



Disponendo le coordinate frequenziali, che codificano le transizioni di F2 di varianti allofoniche di una data categoria di luogo, nel piano cartesiano, emerge un evidente ordine lineare (cfr. Fig. 1 b)) in quanto i punti si addensano in maniera compatta lungo una retta di regressione descritta dalla seguente equazione: $F2_{onset}$ (V) = k × $F2_{vowel}$ (V) + c (Lindblom, 1963), con k e c costanti reali che rappresentano rispettivamente la pendenza della retta di regressione (cfr. *slope*)[11] e l'intercetta[12] con l'asse delle ordinate (Sussmann et al., 1999). I valori della slo-

[11] "the slope indicates the amount of change of F_2 at the beginning of the vowel transition for a unit of change of F_2 at the vowel midpoint" (Iskarous et al., 2010: 2021).

[12] "the intercept indicates F_2 at the beginning of the vowel transition for a zero F_2 at vowel midpoint, that is, the intercept is at the intersection of the regression line with the dependent variable axis"

pe[13] e dell'intercetta, per i tre luoghi articolatori, sono distinti nello spazio acustico e rappresentano "partly distinctive invariant properties" (Nearey, Shamass, 1987: 17), come dimostrato da diversi studi condotti sul parlato sia di adulti che di bambini in età prescolare (Sussman et al., 1992; Sussman et al., 1999).

Il valore della *slope* di queste rette di regressione lineare, serve a quantificare la misura in cui, in una sequenza CV, la vocale influenza la frequenza dell'*onset* della transizione di F2, e quindi la coarticolazione fra il gesto consonantico e quello vocalico (Krull, 1988). L'analisi acustica della coarticolazione intrasillabica è stata automatizzata, grazie alla creazione di appositi *script* di PRAAT[14]: il primo è servito al calcolo automatico dei valori di F1 e F2 tramite l'algoritmo LPC del menu *Formant* sulla porzione stabile della vocale e al calcolo del valore centrale e dei valori medi delle stesse formanti (al 30%, 50% e 70% della sua durata), oltre che delle durate di tutti i foni e di tutte le sillabe indicati nei *Tier*. Col secondo *script*, invece, è stato calcolato il V.O.T.[15] delle sillabe iniziali di enunciato e la F2 della consonante. Una volta calcolati[16] i valori di F2 relativi alla consonante e alla vocale di ciascuna sillaba, essi sono stati messi in relazione per calcolare, attraverso le *Locus Equations*, il grado di sovrapposizione spazio-temporale della vocale sulla consonante, per ogni luogo articolatorio.

3. *Risultati*

3.1 Profili evolutivi della coarticolazione

Nei grafici 1 e 2 sono rappresentati, per ogni categoria di luogo articolatorio (bilabiali, dentali e velari), i valori delle *slopes* calcolati rispettivamente per Alessandro e Anna, nelle diverse tappe di evoluzione del loro disturbo. Questi grafici permettono di visualizzare le variazioni nel grado di coarticolazione anticipatoria nel tempo (sull'asse delle ascisse è riportata l'età, in mesi, dei due bambini balbuzienti), dal periodo precedente la comparsa del disturbo (Tappa -1) fino ai mesi vicini alla fine del periodo di osservazione.

3.1.1 Alessandro

Nel grafico 1 sono illustrati i profili evolutivi della coarticolazione di Alessandro attraverso i tre luoghi articolatori:

(Iskarous et al., 2010: 2021).

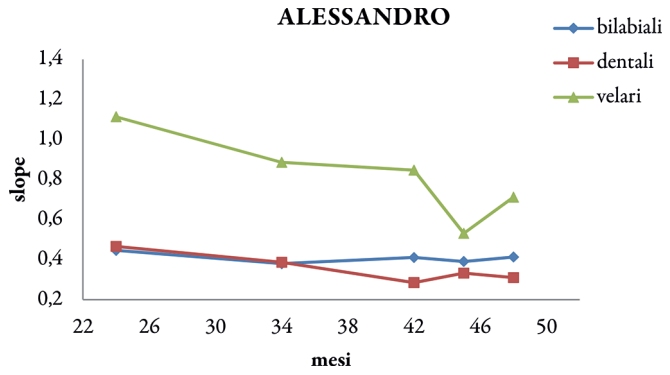
[13] Il lavoro di Lindblom (1963) offre il primo chiaro riconoscimento che le *slope* delle LE variano in modo sistematico in funzione del luogo di articolazione dell'occlusiva.

[14] Gli script sono stati creati da Vincenzo Galatà (ISTC-CNR).

[15] Il *Voice Onset Time* misura l'intervallo di tempo che intercorre tra il rilascio di una consonante occlusiva e l'inizio della vibrazione delle corde vocali per la vocale successiva. Il V.O.T. è considerato uno dei correlati acustici più affidabili per distinguere tra consonanti sorde e sonore.

[16] Per un numero consistente di sillabe è stata effettuata un'analisi di tipo manuale, che ha prodotto un sostanziale accordo con le misure automatiche: sullo spettrogramma sono stati calcolati, manualmente, con la tecnica della FFT (*Fast Fourier Transform*) gli effettivi valori frequenziali della F2 misurati all'inizio della transizione e a metà della vocale, negli stessi istanti temporali calcolati dagli *script*.

Grafico 1 - *In ascissa è espressa l'età (in mesi) di Alessandro e in ordinata il valore della slope per il luogo di articolazione bilabiale (in blu), dentale (in rosso) e velare (in verde). Nell'ellisse rossa sono raccolti i valori delle slope nella tappa si insorgenza della balbuzie*



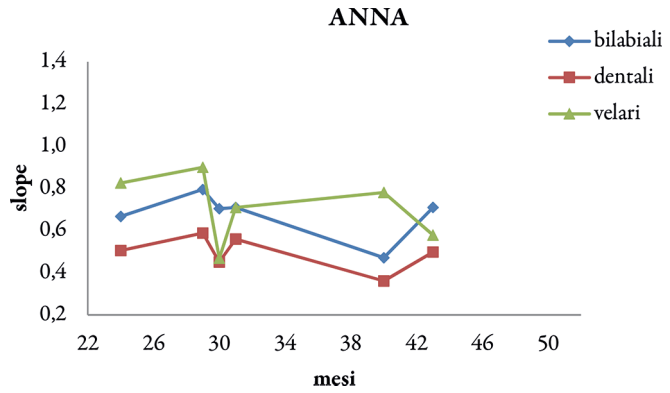
Il primo risultato che merita di essere osservato è che già all'età di 24 mesi (tappa -1), quando Alessandro ancora non balbetta, i valori delle *slopes* registrati per il luogo di articolazione bilabiale e dentale sono sovrapposti nello spazio acustico. Questo rappresenta un fenomeno nient'affatto comune nella letteratura sulle *Locus Equations* (studiate perlopiù sulla popolazione non patologica), le quali si caratterizzano per essere un buon marcatore del luogo di articolazione delle consonanti occlusive, presentando valori di pendenza che variano in funzione del luogo articolatorio.

Il valore della *slope* per il luogo velare si posiziona, invece, su livelli alti, conformemente ai risultati riportati in letteratura. La distinzione coarticolatoria fra il luogo bilabiale e dentale non è realizzata anche nella tappa successiva, quando la balbuzie compare, e i rispettivi valori delle *slopes* si attestano su livelli piuttosto bassi. Rispetto alla tappa (-1), quando il disturbo insorge, il grado di coarticolazione per il luogo velare si abbassa leggermente mentre si mantiene stabile nella tappa successiva, a distanza di 8 mesi dall'insorgenza. Esso si presenta, invece, meno stabile nelle due tappe successive: il valore della *slope* si abbassa notevolmente a distanza di 11 mesi dalla comparsa del disturbo e aumenta nella tappa finale. Il grado di coarticolazione per gli altri due luoghi articolatori, bilabiale e dentale, si attesta invece sempre su livelli ancora molto bassi nelle tappe successive.

3.1.2 Anna

Nel grafico 2 sono rappresentati, nello spazio acustico, i profili evolutivi della coarticolazione misurati per Anna:

Grafico 2 - *In ascissa è espressa l'età (in mesi) di Anna e in ordinata il valore della slope per il luogo di articolazione bilabiale (in blu), dentale (in rosso) e velare (in verde). Nell'ellisse rossa sono raccolti i valori delle slope nella tappa si insorgenza della balbuzie*



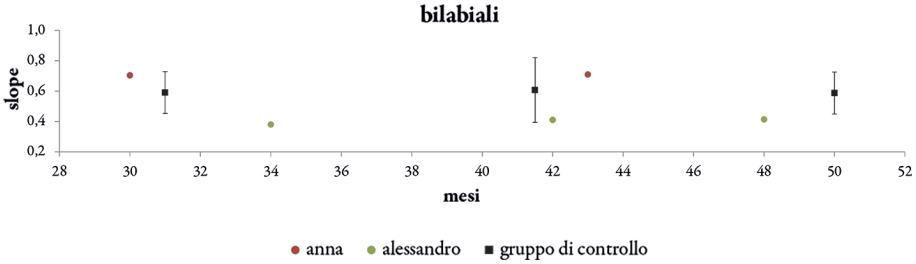
I risultati emersi dall'analisi acustica della coarticolazione intrasillabica effettuata sul parlato fluente di Anna mostrano delle differenze rispetto a quelli osservati per Alessandro. Innanzitutto nelle fasi precedenti la comparsa del disturbo (quando la bambina ha 24 e 29 mesi d'età) i valori delle *slopes* registrati per i tre luoghi articolatori sono ben distinti nello spazio acustico. La distinzione coarticolatoria fra i tre luoghi si perde, invece, quando il disturbo compare (età di 30 mesi): un brusco calo del grado di coarticolazione per il luogo velare, fa sì che il valore della *slope* si sovrapponga a quello del luogo dentale, e con quello del luogo bilabiale nel mese successivo l'insorgenza. Altro aspetto da sottolineare è la coerenza dei *pattern* coarticolatori attraverso i tre luoghi: i valori delle *slope* subiscono una diminuzione (maggiore per il luogo velare) quando il disturbo insorge e un aumento nel mese successivo. Nelle tappe seguenti, i profili evolutivi della coarticolazione per i luoghi dentale e bilabiale sono simili (discendente a distanza di 10 mesi dall'insorgenza e ascendente nei 3 mesi successivi) mentre quello velare segue un pattern inverso.

3.2 Confronto col gruppo di controllo

Nei grafici 3, 4 e 5 vengono visualizzati, per i tre luoghi di articolazione, i valori delle *slopes* calcolati per Alessandro e Anna in due tappe specifiche del disturbo: l'insorgenza e la tappa (≤ 12 mesi). Nei grafici vengono riportati anche i valori medi e le deviazioni standard (per delimitare il *range* normativo) delle *slopes* calcolate per i 26 bambini con normale sviluppo linguistico, divisi in due fasce d'età, corrispondenti all'età dei due bambini balbuzienti: la prima è la fascia 28-34 mesi, la seconda è la 39-47. Siccome per Alessandro si dispone anche di un'altra registrazione, effettuata quando il bambino aveva 48 mesi d'età, è stato calcolato il grado di coarticolazione ed è stato confrontato con quello del gruppo di controllo (fascia d'età 46-54 mesi).

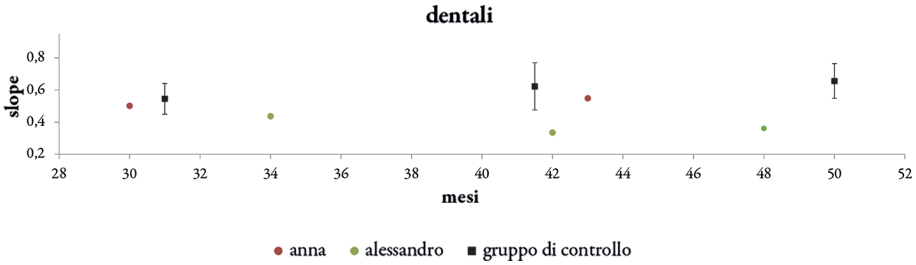
Questi grafici permettono, pertanto, di confrontare, da una parte, i dati della coarticolazione dei due bambini balbuzienti col gruppo di controllo di pari età e di convalidare, dall'altra, il valore prognostico della transizione di F2 nella predizione precoce di balbuzie cronica, proposto da Subramanian et al. (2003).

Figura 2 - *In ordinata vengono riportati i valori delle slope delle LE per il luogo bilabiale e in ascissa l'età (in mesi) dei bambini analizzati. In nero sono riportati i valori medi e le deviazioni standard delle slope registrate per il gruppo di controllo e nei colori descritti nella legenda quelli dei 2 bambini balbuzienti*



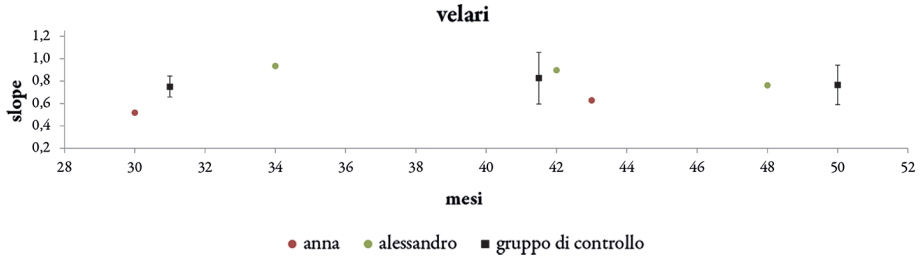
Dal confronto dei dati per il luogo bilabiale, emerge chiaramente che Anna e Alessandro presentano valori di *slope* molto diversi, anche rispetto al gruppo di controllo. Anna infatti si posizione nel limite superiore alla norma mentre Alessandro presenta un valore di *slope* molto più basso, che si posiziona al di fuori del *range* normativo. Nel secondo semestre dalla comparsa, invece, il valore della *slope* di Anna si avvicina al valore medio registrato per il gruppo di controllo mentre quello di Alessandro si posiziona ancora su valori bassi, posizionandosi nel limite inferiore del gruppo di controllo, anche nella tappa finale.

Figura 3 - *In ordinata vengono riportati i valori delle slope delle LE per il luogo dentale e in ascissa l'età (in mesi) dei bambini analizzati. In nero sono riportati i valori medi e le deviazioni standard delle slope registrate per il gruppo di controllo e nei colori descritti nella legenda quelli dei 2 bambini balbuzienti*



Per quanto riguarda il luogo di articolazione dentale è possibile osservare come, nella tappa (0), il grado di coarticolazione di Anna rientra nei valori normativi mentre Alessandro si posiziona, come per le bilabiali, nel limite inferiore alla norma. Anche nel secondo semestre dalla comparsa, mentre Anna presenta un grado di coarticolazione anticipatoria simile ai coetanei con normale sviluppo linguistico, Alessandro presenta un basso grado di coarticolazione. Anche a distanza di 14 mesi dalla comparsa del disturbo, il grado di coarticolazione eccede il *range* normativo), posizionandosi sempre nel limite inferiore.

Figura 4 - *In ordinata vengono riportati i valori delle slope delle LE per il luogo velare e in ascissa l'età (in mesi) dei bambini analizzati. In nero sono riportati i valori medi e le deviazioni standard delle slope registrate per il gruppo di controllo e nei colori descritti nella legenda quelli dei 2 bambini balbuzienti*



I dati ottenuti per il luogo di articolazione velare mostrano invece una situazione diversa rispetto agli altri due luoghi: al momento della comparsa del disturbo, infatti, il valore della *slope* registrato per Anna eccede il valore medio del gruppo normofluente, posizionandosi nel limite inferiore mentre quello registrato per Alessandro si pone oltre il limite superiore. Nel secondo semestre dalla comparsa, il valore della *slope* di Alessandro si avvicina al valore medio del gruppo normofluente mentre quello registrato per Anna rimane su valori bassi rispetto al gruppo di controllo.

In generale è possibile osservare come le strategie articolatorie adottate dai bambini balbuzienti per ottenere la fluenza variano a seconda dei soggetti, dei luoghi consonantici esaminati e della gravità del disturbo.

Alessandro, ad esempio, che ha cronicizzato il disturbo nella sua forma severa, presenta rispetto ad Anna, una coarticolazione anormalmente bassa rispetto al gruppo normofluente, soprattutto per i luoghi dentale e bilabiale. In questo caso la bassa coarticolazione è dovuta ad una ridotta capacità di coordinazione di due gesti dello stesso articolatore (apice e dorso della lingua) o di due articolatori differenti (labbra e dorso della lingua).

Nel caso delle velari, invece, la coarticolazione fra i due gesti consonantico e vocalico potrebbe essere facilitata dall'adattamento di vincoli biomeccanici già esistenti: la sovrapposizione anatomica dei due gesti articolatori è totale, essendo l'articolatore lo stesso (il dorso della lingua).

4. *Conclusioni*

Scopo di questo contributo è quello di presentare una parte dei risultati sperimentali del progetto longitudinale 'Indici fonetici predittivi di balbuzie cronica in età prescolare'. Sebbene il numero dei soggetti non sia sufficiente per stimare il valore prognostico della transizione della seconda formante come indice di cronicità, i risultati delle analisi permettono delle considerazioni preliminari, seppur di tipo ancora qualitativo, sulle fasi evolutive del disturbo, sia nella sua forma cronica che in quella remissiva. Ricordiamo che, dalla valutazione finale dei genitori dei due bambini balbuzienti, risulta che mentre Anna si è rimessa con un solo ciclo di tratta-

mento, Alessandro balbetta ancora nonostante si sia sottoposto a più cicli di trattamento. Quest'ultimo dunque sembra aver sviluppato la forma cronica del disturbo con un livello di gravità di tipo severo.

Dai dati acustici sulla coarticolazione anticipatoria emerge che le strategie articolatorie adottate per riuscire ad essere fluenti, sono diverse nei due bambini balbuzienti. Alessandro, infatti, presenta una sovrapposizione spazio-temporale minima fra il gesto consonantico (bilabiale e dentale) e quello vocalico, sia rispetto ad Anna che rispetto al gruppo di controllo.

La bassa coarticolazione riscontrata nell'eloquio di Alessandro potrebbe essere indice del fatto che gli articolatori, nel passare dalla posizione assunta per l'articolazione consonantica a quella vocalica, si muovono in un *range* spazialmente ampio, come dimostrato anche dallo studio di Robb e Blomgren (1997) realizzato con soggetti balbuzienti adulti. Le ipotesi di un'alta coarticolazione nell'eloquio dei bambini destinati a cronicizzare, proposta da Subramanian et al. (2003), non viene confermata dai risultati preliminari di questo studio.

Bibliografia

Bernardini, S., Cocco, L., Zmarich, C., Vanryckeghem, M., Di Pietro, M. & Natarelli, G. (submitted to *The Journal of Fluency Disorders*). The KiddyCAT©: A Normative and Comparative study of Italian CWS and CWNS.

Chang, S.E., Ohde, R.N. & Conture, E.G. (2002). Coarticulation and formant transition rate in young children who stutter. In *Journal of Speech, Language and Hearing Research,* 45, 676-688.

Fant, G. (1970). *Acoustic theory of speech production*. Berlin: Mouton De Gruyte.

Green, J.R., Moore, C.A., Higashikawa, M. & Steeve, R.W. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. In *Journal of Speech, Language, and Hearing Research*, 43, 239-255.

Iskarous, K., Fowler, C. & Whalen, D.H. (2010). Locus equations are an acoustic expression of articulator synergy. In *Acoustical Society of America,* 128 (4), 2021-2032.

Kleinow, J., Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. In *Journal of Speech, language and Hearing Research*, 43, 548-559.

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. (2012). Genetic bases of stuttering: the state of the art. In *Folia Phoniatrica et Logopedica*, 64, 34-47.

Krull, D. (1988). Acoustic properties as predictors of perceptual responses: a study of Swedish voiced stops. Phonet. Exp. Res. Inst. Ling., Univ. Stockh. (PERILUS) 7, 66-70.

Lindblom, B. (1963). *On vowel reduction*. *The Royal Institute of Technology Speech Transmission Laboratory.* Stockholm: Report N. 29.

Namasivayam, A.K., Van Lieshout, P.H.H.M. (2011). Speech Motor Skill and Stuttering. In *Journal of Motor Behavior*, 43 (6), 477-489.

Nearey, T., Shammass, S. (1987). Formant transitions are partly distinctive invariant properties in the identification of voiced stops. In *Canadian*.

Olander, L., Smith, A. & Zelaznik, H.N. (2010). Evidence That a Motor Timing Deficit Is a Factor in the Development of Stuttering. In *Journal of Speech, Language, and Hearing Research*, 53, 876-886.

Robb, M., Blomgren, M. (1997). Analysis of F2 transitions in the speech of stutteres and nonstutterers. In *Journal of fluency Disorders*, 22, 1-16.

Salza, P.L. (1990). La problematica della segmentazione del segnale vocale. In *Atti della 1° Giornata di Studio di G.F.S*, Padova, 3-6 novembre.

Smith, A., Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. In *Journal of Speech, Language, and Hearing Science*, 41, 18-30.

Smith, A., Zelaznik, H.N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. In *Developmental Psychobiology*, 45, 22-33.

Subramanian, A., Yairi, E. & Amir, O. (2003). Second formant transitions in fluent speech of persistent and recovered preschool children who stutter. In *Journal of Communication Disorders*, 36, 59-75.

Sussmann, H.M., Hoemeke, K.A. & Mccaffrey, H.A. (1992). Locus equations as an index of coarticulation for place of articulation distinction in children. In *Journal of Speech and Hearing Research*, 35, 769-781.

Sussmann, H.M., Duder, C., Dalston, E. & Cacciatore, A. (1999). An acoustic analysis of the developmental of CV coarticulation: A case study. In *Journal of Speech, Language and Hearing Research*, 42, 1080-1096.

Van Lieshout, P.H.H.M., Hulstijn, W. & Peters, H.F.M. (2004). Searching for the weak link in the speech production chain of people who stutters: A motor skill approach. In Maassen, B., Kent, R.D., Peters, H.F.M., Van Lieshout, P.H.H.M. & Hulstijn, W., *Speech motor control in normal and disordered speech*. Oxford, Vanryckegem & Brutten, 2006, 313-355.

Van Riper, C. (1982). *The nature of stuttering*. Englewood Cliffs, NJ.

Van Riper, C. (1990). Final thoughts about stuttering. In *Journal of Fluency Disorders*, 15, 317-318.

Yairi, E., Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. In *Journal of Fluency Disorders*, 38 (2), 66-87.

Yairi, E., Ambrose, N. (2005). *Early Childhood Stuttering: for clinicians by clinicians.* Austin (TX): Pro-Ed.

Zmarich, C. (in stampa). 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.), *Miscellanea in onore di Alberto Mioni.* Edizioni CLEUP: Padova.

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 del VIII convegno nazionale AISV. Roma: Bulzoni Editore, 51-66.

DOI: 10.17469/O2101AISV000020

DANIELA BELTRAMI, LAURA CALZÀ, GLORIA GAGLIARDI, ENRICO GHIDONI, NORINA MARCELLO, REMA ROSSINI FAVRETTI, FABIO TAMBURINI

Strumenti di *screening* linguistico per l'identificazione precoce della fragilità cognitiva

This paper presents the preliminary results of the OPLON project, co-funded by the Ministry of Education as part of the contract "Smart Cities and Communities and Social Innovation". It aimed at identifying early linguistic symptoms of cognitive decline in the elderly. This pilot study was conducted on a corpus composed of spontaneous speech sample collected from 20 subjects, who underwent a neuropsychological screening for visuo-spatial abilities, memory, language, executive functions and attention. A rich set of linguistic features was extracted from the digitalized utterances (at phonetic, suprasegmental, lexical, morphological and syntactic levels) and the statistical significance in pinpointing the pathological process was measured. Our results, far from being complete, show remarkable trends for what concerns both the linguistic traits selection and the automatic classifiers building.

1. *Introduzione*

Il progetto OPLON ["OPportunities for active and healthy LONgevity", Smart Cities and Communities – DD 391/RIC] si pone l'obiettivo di sviluppare strumenti di diagnosi precoce della fragilità finalizzati alla prevenzione del declino e alla promozione della salute dei soggetti anziani.

Uno dei fenomeni epidemiologici più rilevanti dal punto di vista sanitario e socio-economico che ha interessato la società occidentale negli ultimi decenni è l'aumento significativo della popolazione anziana: in Italia vi sono oltre un milione di pazienti fragili e tale numero è destinato a raddoppiare entro i prossimi vent'anni (ISTAT, 2006).

Tra gli effetti importanti della maggior aspettativa e qualità della vita vi è l'aumento delle patologie neurodegenerative a carattere progressivo, il cui sintomo più rilevante è il decadimento cognitivo. È oramai noto che il processo che conduce alla demenza inizia molti anni prima dell'esordio dei sintomi, quindi molto prima dell'inizio dell'effettiva fase "clinica".

La possibilità di identificare precocemente la malattia si pone come un imprescindibile obiettivo scientifico. Una diagnosi precoce, infatti, non solo permetterebbe di ottimizzare gli interventi assistenziali nei confronti dei pazienti e dei loro familiari (miglior definizione del quadro clinico; accesso precoce ad alcuni trattamenti; programmi di prevenzione sempre più efficaci; maggior *compliance* al percorso terapeutico; miglior risposta ai cambiamenti dello stile di vita; riduzione dei rischi secondari della patologia), ma consentirebbe anche al paziente di pianificare

la propria vita in una fase in cui le capacità sono ancora sostanzialmente mantenute (presa di decisioni riguardanti il futuro; riduzione dello stigma sociale; assunzione di consapevolezza graduale) e agevolerebbe la ricerca e la sperimentazione di terapie in grado di modificare il decorso clinico, fornendo ai pazienti la possibilità di accedere precocemente ai trattamenti farmacologici (i farmaci oggi disponibili agiscono efficacemente soltanto se somministrati in fase altamente precoce).

Vari studi (Kempler, 1987; Caramelli, 1998) mostrano che oltre alla memoria (specialmente quella episodica), anche il linguaggio (soprattutto quello spontaneo) può subire cambiamenti significativi già diversi anni prima dell'esordio clinico della malattia. Tali cambiamenti, tuttavia, non sono identificati facilmente dai test di *screening* tradizionali che, purtroppo, prestano insufficiente attenzione all'eloquio spontaneo (Lezak, 1995) e dovrebbero essere valutati con strumenti alternativi costruiti *ad hoc*.

2. *Prospettive linguistico-computazionali*

Nella letteratura specifica degli ultimi anni si rilevano lavori che si occupano dell'identificazione automatica dei vari stadi e tipi di demenze a partire dall'analisi statistica della produzione linguistica spontanea dei pazienti. La maggior parte di questi studi, quasi totalmente rivolti all'analisi della lingua inglese, assume come base d'indagine la rilevazione di opportuni tratti linguistici negli enunciati dei pazienti, identificati di solito con strumenti automatici, tentando di correlare tali tratti con i risultati dei test cognitivi utilizzati in ambito neuropsicologico per l'identificazione precoce di disturbi cognitivi (Lehr, 2012; Roark et al., 2011; Jarrold et al., 2010; Satt et al., 2013; Chapman et al., 2002).

Si è giunti all'elaborazione di metodologie automatiche e di procedure di analisi che aprono nuove prospettive di progettazione metodologica e nuovi obiettivi. In certa misura, tuttavia, può restare aperta una domanda sulla legittimità della correlazione fra piano linguistico e piano cognitivo. Questa trova ragione, a nostro avviso, in una pluralità di ambiti della linguistica, della psicologia, della neurologia e, in particolare, della neurolinguistica. Ambiti e ricerche che affondano le loro radici in tempi che appaiono lontani. Doveroso, in questa sede, appare far precedere alla presentazione del contributo un riferimento a Jakobson e agli studi da lui condotti sul linguaggio patologico, in cui l'afasia è indagata non di per sé, ma in relazione al farsi e al disfarsi del linguaggio, come campo produttivo per lo studio linguistico nella sua globalità (Jakobson, 2006; Jakobson, Halle, 1956).

L'opera di Jakobson, letta alla luce delle procedure sviluppate nell'ambito delle tecnologie informatiche si pone, oggi, come fondante nelle intuizioni e nei contributi. Che la patologia del linguaggio obbedisca a un insieme di regole è asserito fin dalle prime fasi della sua ricerca, così come che nessuna regola sottesa alla regressione del linguaggio possa essere individuata prescindendo dai prin-

cipi, dalle metodologie e dalle procedure proprie della linguistica. Come ben si chiarisce nelle seguenti parole:

> Se l'afasia è un disturbo del linguaggio, come dice il termine stesso, ne consegue che ogni descrizione e classificazione dei disturbi afasici deve preliminarmente chiarire quali aspetti del linguaggio sono alterati nei diversi tipi di questo disturbo. (Jakobson, 1956: 22)

Costante, nella ricerca jakobsoniana, è l'asserzione dell'importanza di un'analisi dei deficit linguistici che sorgono in concomitanza con danni cerebrali e della correlazione che fra questi si instaura. Correlazione evidenziata già negli studi ancora precedenti di Hughlings Jackson, ricordato da Jakobson come "primo vero interprete dell'afasia", ed oggetto di analisi nelle ricerche di Kurt Goldstein, André Ombredane e Alexander R. Luria, indicati come autori dei migliori trattati apparsi negli ultimi anni '40 sull'afasia, nei cui libri è possibile trovare i primi sforzi dei neurologi verso l'utilizzazione sistematica dei principi linguistici moderni per l'analisi delle sindromi afasiche. Così come è ripetutamente asserita l'opportunità, se non la necessità, di un approccio interdisciplinare in cui in cui sia presente l'apporto della teoria matematica della comunicazione e della teoria dell'informazione, e in cui possa costantemente accrescersi la cooperazione di linguisti, psicologi, psichiatri, neurologi. Nella nuova dimensione di analisi un ruolo primario spetterebbe alla linguistica quale punto di convergenza di ambiti disciplinari diversi[1].

Con le opportune modifiche queste affermazioni, formulate intorno alla metà del secolo scorso, potrebbero essere accettate ancora oggi. Certo, gli ambiti di indagine nei gruppi di lavoro si sono modificati. Oggi facciamo riferimento al Natural Language Processing, alle biotecnologie, alle neuroscienze, alle scienze cognitive. Al generico concetto di afasia si sostituisce in letteratura una classificazione più specifica dei disturbi del linguaggio. In particolare appaiono superati i limiti del modello afasiologico: già da alcuni decenni, con gli sviluppi delle neuroscienze e con il diffondersi delle tecniche di neuroimmagine, si è rilevata la necessità di operare modifiche alla rigidità posta dal modello classico nella localizzazione e nell'organizzazione cerebrale delle funzioni linguistiche (Bambini, 2009). Nuovi aspetti sono presi in esame, come la modalità o i livelli di strutturazione; sono indicati nuovi oggetti di ricerca, ma non è confutata la validità che i risultati ottenuti nello studio delle afasie mantengono sia a livello linguistico sia, in una dimensione multidisciplinare, clinico. Non è negata la rilevanza che gli studi linguistici possono assumere nell'analisi del linguaggio patologico ma appare dimenticata una lunga tradizione risalente al secolo scorso, che può essere

[1] "Mi sembra che siano ormai pochi i ricercatori nel campo dei disturbi del linguaggio che credono ancora alla non rilevanza del ruolo della linguistica nello studio dell'afasia. [...] Vi sono parecchi centri in cui neurologi, psicologi, linguisti, e altri specialisti, lavorano insieme per descrivere, esaminare, analizzare l'afasia, e ottenere le diagnosi e prognosi più esatte." (Jakobson, 2006: 171).

riduttivo fare risalire a Jakobson ma che in Jakobson certamente trova uno studioso particolarmente attento agli aspetti teorici e metodologici.

Negli ultimi decenni si è sviluppato un nuovo interesse verso la possibilità di giungere all'individuazione delle caratteristiche distintive del linguaggio dei soggetti con diverso grado di fragilità cognitiva utilizzando anche approcci di analisi delle produzioni verbali spontanee (March et al., 2003). Vari ed importanti sono stati gli studi sull'analisi del discorso e del testo[2]. Sono state oggetto di indagine le abilità conversazionali e comunicative e aspetti più specifici come la ricchezza lessicale, la complessità sintattica, il livello di competenza pragmatica. Si è configurato un panorama alquanto articolato che mostra, ad esempio, come la struttura sintattica sia meglio preservata della lessico-semantica (Kempler et al., 1987) e come alcune alterazioni linguistiche siano apprezzabili a livello di discorso (Caramelli et al., 1998) nei soggetti affetti da demenza di Alzheimer. L'eloquio di tali pazienti, descritto in letteratura come tangenziale e incoerente, è caratterizzato in particolare dall'abuso di termini deittici e pronomi senza antecedente e da un aumento di revisioni e ripetizioni, a cui si accompagnano difficoltà nel *retrieval* lessicale e deficit delle abilità pragmatiche (Taler, Phillips, 2008).

Tuttavia, occorre ammettere che, pur nella varietà e molteplicità delle ricerche, i contributi che sono stati dati nell'ambito della linguistica appaiono insufficienti nella definizione di un quadro concettuale di riferimento.

Nuove prospettive sembrano oggi configurarsi nell'attuale alto interesse verso tecniche di analisi computerizzate: questo ha portato alla creazione di "corpora linguistici elettronici", intendendo con ciò collezioni di testi in forma elettronica scelti in virtù della loro rappresentatività nel caratterizzare una particolare lingua o varietà linguistica, e allo sviluppo di numerosi algoritmi per l'analisi e la classificazione del "parlato".

Applicato al linguaggio patologico, questo approccio e le tecnologie relate hanno il notevole vantaggio di rappresentare una registrazione del linguaggio naturale e spontaneo, al di fuori quindi degli schemi dei test neuropsicologici del linguaggio, potenzialmente applicabili ad ampie fasce di popolazione con strumenti a basso costo (Lehr, 2012).

L'analisi computazionale di dati linguistici e paralinguistici può dare un contributo rilevante allo *screening* delle patologie neurodegenerative di tipo dementigeno: tecniche proprie del Natural Language Processing sono state recentemente utilizzate con successo nell'analisi automatica di testi scritti, testi orali elicitati in contesto clinico controllato e produzioni spontanee di soggetti affetti da demenza, in particolare Alzheimer e Demenza Fronto-Temporale (Afasia Primaria progressiva e Demenza Semantica) (Peintner et al., 2008; Jarrold et al., 2010; Roark et al., 2011; Fraser et al., 2014).

[2] "In clinical groups, discourse research can provide important insights into the pragmatic aspects of language changes and the role of non-linguistic factors in communication (e.g., communicative context, conversational partner) as associated to brain damage." (March et al., 2006: 311-312).

Le prospettive di analisi computazionali coprono vari ambiti linguistici, spaziando dall'esame approfondito della produzione orale spontanea alla ricerca di disfluenze, incertezze e/o modificazioni nella sequenza fonica e nei parametri soprasegmentali fino a un'approfondita analisi lessicale, morfosintattica e sintattica con l'obiettivo di individuare regolarità che evidenzino una maggiore semplificazione dell'eloquio nei pazienti rispetto ai controlli. La letteratura ha prodotto, negli ultimi anni, una moltitudine di indicatori, misurabili spesso con strumenti automatici, che, da studio a studio, si sono rivelati significativi nel discriminare tra popolazioni diverse di soggetti. Tali metodologie si sono quindi dimostrate in grado di rilevare *pattern* latenti e regolarità nel linguaggio di queste popolazioni di pazienti, potenzialmente utili all'identificazione precoce, classificazione e descrizione del decadimento cognitivo (Elvevåg, Garrard, 2014).

Date queste premesse, l'obiettivo a lungo termine del progetto OPLON riguarda lo sviluppo di tecniche facilmente somministrabili di raccolta e analisi di campioni di linguaggio spontaneo per lo screening sistematico della popolazione potenzialmente a rischio. Tali strumenti devono essere economici, semplici e utilizzabili anche in ambiti non specializzati (Medicina di Base, ad esempio).

A medio termine, questo studio si propone di dimostrarne la fattibilità in ambiente controllato, visto che in letteratura non risultano al momento studi simili sulla lingua italiana.

3. *Costruzione, annotazione e analisi del corpus*

3.1 Raccolta dati

La ricerca prevede il reclutamento di novanta soggetti bilanciati per sesso, età (*range* 50-75) e scolarità (licenza media o licenza elementare con buona stimolazione intellettiva nel corso della vita), di cui quarantacinque soggetti di controllo (neurologicamente indenni) e altrettanti affetti da declino cognitivo. Tale declino fa riferimento a due diverse categorie:

1. Mild Cognitive Impairment (MCI): deterioramento cognitivo lieve non in grado di interferire in modo significativo con le attività della vita quotidiana, riguardante una singola funzione cognitiva (ad es. memoria) o più funzioni (ad es. memoria, attenzione e linguaggio). I soggetti con MCI sono stati divisi in due sottogruppi:
   a. MCI con isolato deficit di memoria (a-MCI single domain; n = 15);
   b. MCI con deficit di funzioni cognitive ma non di memoria (non a-MCI *single* or *multiple domain*; n = 15).

   In entrambi i casi la compromissione cognitiva non interferisce con il funzionamento quotidiano. Tale differenziazione (a-MCI vs. non-a-MCI) è stata fatta poiché sembra che diversi tipi di MCI possano essere prodromici di differenti tipi di demenza (*Alzheimer Disease* – AD, demenza vascolare, demenza fronto-temporale, ecc.). La variante amnesica, ad esempio, sembrerebbe

evolvere più frequentemente verso una demenza (AD in particolare) rispetto ad altri tipi di MCI.
2. Demenza in fase iniziale (*early*-D; n = 15): i soggetti appartenenti a questo gruppo hanno ricevuto la diagnosi al termine di un articolato iter diagnostico. In questo caso i deficit cognitivi influenzano (sebbene ancora solo parzialmente) le autonomie nelle attività della vita quotidiana (ad es. assunzione autonoma della terapia farmacologica, cura personale, cura dell'ambiente domestico, ecc.). Un ulteriore requisito per fare parte di questo gruppo sperimentale era quello di ottenere un punteggio maggiore o uguale a 18 al MMSE (Mini Mental State Examination).

Ai soggetti sono stati somministrati i test neuropsicologici più efficaci nella discriminazione tra soggetti normali, con MCI e con Demenza. Tali prove, infatti, valutano le funzioni che precocemente subiscono variazioni in seguito ad un processo neurodegenerativo (abilità verbali, esecutive, visuo-spaziali, attentive, mnesiche e di orientamento (Chen et al., 2001; Backman et al., 2005; Grober et al., 2008; Ismail et al., 2010)), e risultano essere le più efficaci nella discriminazione tra soggetti normali, con demenza e Mild Cognitive Impairment. In particolare sono stati somministrati:
- MMSE - Mini-Mental State Examination (Measso et al., 1993; Magni et al., 1996);
- MoCA - Montreal Cognitive Assessment (Nasreddine et al., 2005; Pirani et al., 2006);
- GPCog - General Practitioners assessment of Cognition (Brodaty et al., 2002; Pirani et al., 2010);
- CDT - Clock Drawing Test (Freedman et al., 1994; Mondini et al., 2003);
- Fluenze Verbali (fluenza fonemica e semantica (Carlesimo et al., 1996; Novelli et al., 1986));
- Paired Associate Learning (PAL, subtest del Cambridge Neuropsychological Test Automated Battery – CANTAB (Junkkila et al., 2012)).

È stato presentato anche un questionario di valutazione dell'indice di Riserva Cognitiva (Cognitive Reserve Index Questionnaire, CRI-q (Nucci et al., 2011)).

Dopo la valutazione tradizionale, i soggetti sono stati sottoposti ad una registrazione dell'eloquio spontaneo durante alcuni compiti:
1. "Descrivi questa immagine". L'immagine (tratta da Esame del Linguaggio – II (Ciurli et al., 1996)), in bianco e nero, ritrae un salotto nel quale vi sono una donna, un uomo, due bambini ed un gatto che compiono alcune semplici azioni (ad es. guardare la televisione, giocare con dei cubetti, ecc.).
2. "Descrivi una giornata lavorativa";
3. "Descrivi l'ultimo sogno che ricordi".

Oltre all'analisi linguistica, questi compiti permettono di valutare l'eventuale compromissione delle funzioni mnesiche; infatti, in tutte e tre le prove è importante ricordare ciò che si sta dicendo (memoria di lavoro), che si è già riferito (memoria episodica) o si vorrà raccontare (memoria prospettica); inoltre, è importante conoscere il significato dei termini utilizzati (memoria semantica) ed avere ricordi rispetto al passato più o meno remoto (memoria autobiografica recente e remota).

L'intera valutazione viene ripetuta a distanza di circa un anno.

3.2 Trascrizione e annotazione dei testi orali del corpus

I campioni di linguaggio parlato raccolti nel corso delle sedute di test nella forma di file .WAV (segnale mono-canale digitalizzato a 44.1KHz di frequenza, *range* dinamico di 16 bit) vengono inizialmente sottoposti a trascrizione ortografica manuale. Le trascrizioni, conformi al "Protocollo di trascrizione" definito nell'ambito del progetto e ispirato alle *guidelines* del progetto CLIPS (Albano Leoni, 2003), sono realizzate con il software *opensource* Transcriber; il programma genera un output in formato .trs (.xml con allineamento temporale degli enunciati alla waveform) e ne consente l'esportazione in formato plain-text (si veda la fig. 1).

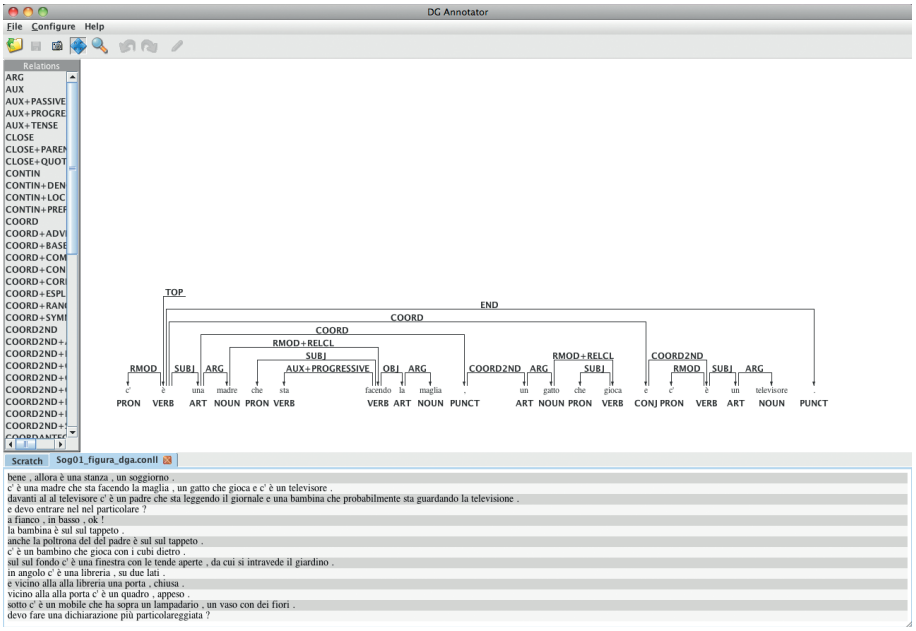
Figura 1 - *Uno screenshot del programma di trascrizione "Transcriber"*

L'unità di allineamento scelta è l'enunciato, definito e identificato sulla base di criteri intonativi. Contestualmente alla trascrizione vengono annotati anche i principali fenomeni paralinguistici (pause, disfluenze, lapsus, etc.).

Vengono quindi isolati i turni di parola del soggetto sottoposto a test: gli enunciati selezionati, emendati manualmente dalle disfluenze, sono sottoposti ad annotazione morfosintattica semi-automatica, ovvero subiscono Part of Speech (PoS) Tagging e Parsing automatico mediante il *parser* a dipendenze TULE (Lesmo, 2007), che assicura prestazioni confrontabili con lo stato dell'arte per la lingua

italiana[3]. L'annotazione viene revisionata da un linguista esperto utilizzando il software *open-source* DGA – Dependency Grammar Annotator, strumento grafico per l'annotazione sintattica di testi secondo il quadro teorico delle Grammatiche a dipendenze (fig. 2).

Figura 2 - *Uno screenshot del programma "DGA" utilizzato per la correzione manuale dell'annotazione sintattica*



Il *tagset* adottato è quello del TUT - Turin University TreeBank (Bosco et al., 2000), compatibile con il *parser* TULE. L'annotazione morfosintattica è prodotta nel formato tabulare standard CoNLL (Buchholz, Marsi, 2006).

3.3 Estrazione dei tratti linguistici

Per ciascun livello di analisi vengono calcolati indici linguistici e stilometrici descritti in letteratura (e risultati significativi negli studi condotti su lingue diverse dall'italiano) o creati ex novo. Le misure acustiche vengono calcolate direttamente sul segnale vocale, utilizzando un Voice Activity Detector[4]. I tratti lessicali e sintattici vengono invece derivati dall'annotazione automatica revisionata manualmente dal linguista (si veda la tabella 1).

[3] Si rinvia ai dati delle campagne di valutazione EVALITA, Subtask "Dependency Parsing" (http://www.evalita.it).

[4] Il VAD utilizzato nello studio è contenuto nel pacchetto software di Julius ASR (http://julius.osdn.jp/en_index.php), v. 4.3.1, denominato "adintool".

Tabella 1 - *Elenco dei tratti linguistici considerati nello studio*

Acoustic Features
Silence segments duration: mean, median and Std. Deviation (Satt et al., 2012; 2013)
Speech segments duration: mean, median and Std.Deviation (Satt et al., 2012; 2013)
Temporal regularity of voiced segment durations (Satt et al., 2012; 2013)
Verbal Rate (Singh et al., 2001; Roark et al., 2007; 2011)
Transformed Phonation Rate (Singh et al., 2001; Roark et al., 2011)
Standardized Phonation Time (Singh et al., 2001; Roark et al., 2011)
Standardized Pause Rate (Singh et al., 2001; Roark et al., 2007; 2011)
Root Mean Square energy: mean and Std.Deviation (López-de-Ipiña, 2013)
Pitch: mean and Std. Deviation (López-de-Ipiña, 2013)
Spectral Centroid, mean and Std. Deviation (López-de-Ipiña, 2013)
Higuchi Fractal Dimension, mean and Std. Deviation (López-de-Ipiña, 2013)
Lexical Features
Lexical density: Open-class ratio
Content Density (Roark et al., 2011)
PoS rate (Holmes, Singh, 1996; Bucks et al., 2000; Vigorelli, 2004; Garrard et al., 2005; Thomas et al., 2005; Peintner et al., 2008; Cantos-Gomez et al., 2009; Jarrold et al., 2010; 2014)
Reference Rate to Reality (Vigorelli, 2004)
Personal, Spatial and Temporal Deixis rate (March et al., 2006; Cantos-Gomez et al., 2009)
Relatives pronouns and negative adverbs rate
Lexical Richness: Type-Token Ratio, W - Brunét's Index and R - Honoré's Statistic (Brunét, 1978; Honoré, 1979; Bucks et al., 2000; Holmes, 1992; Holmes, Singh, 1996; Thomas et al., 2005)
Action Verbs rate (Gagliardi, 2014)
Frequency-of-use tagging (De Mauro/Paravia dictionary) (De Mauro, 1980; 2000; Barbagli et al., 2014)
Propositional Idea Density (Snowdon et al., 1996; Brown et al., 2008; Jarrold et al., 2010; Roark et al., 2011)
Syntactic Features
Number of dependent elements linked to the noun, mean
Mean Global Dependency Distance (Roark et al., 2011)
Syntactic complexity (Szmrecsanyi, 2004)
Syntactic embeddedness: maximum depth of the structure, mean and Std. Deviation
Utterance length, mean and Std. Deviation

3.4 Test statistici e classificazione automatica

L'effettiva significatività statistica per la lingua italiana degli indici e delle possibili correlazioni tra variabili è stata valutata attraverso test e metriche statistiche standard (parametriche e non parametriche) e i parametri risultati significativi sono stati ulteriormente testati mediante algoritmi di *feature selection*. Tramite metodologie di *machine learning* gli indici selezionati vengono integrati in un classificatore automatico, le cui performance sono valutate in termini di *accuracy, precision, recall* e *F-measure*.

Nella fase conclusiva del progetto i risultati del test linguistico verranno unificati in un punteggio di *scoring* sintetico, in grado di esprimere un indicatore di impoverimento del linguaggio dell'individuo sottoposto a test, da integrare nel modello OPLON di predizione del rischio di insorgenza di fragilità nell'anziano.

4. *Risultati preliminari e discussione*

L'effettiva fattibilità della metodologia è stata verificata mediante un *Proof of Concept* su un campione di 20 soggetti, 12 MCI e 8 controlli.

Tabella 2 - *I risultati di significatività statistica e le prestazione di classificazione per i tratti linguistici maggiormente significativi*

	Task "Figura"	
		KS test
Tratti selezionati	Silence segment duration (mean)	p = 0.047035
	Spectral Centroid (mean)	p = 0.009033
	Relatives pronouns and negative adverbs rate	p = 0.047035
	Utterance length (SD)	p = 0.047035
Prestazioni del Classificatore	Accuratezza = 0.9000 Precision = 1.0000 Recall = 0.8000 F1= 0.8889	
	Task "Lavoro"	
		KS test
Tratti selezionati	Higuchi Fractal Dimension (SD)	p = 0.047035
	Spectral Centroid (SD)	p = 0.028057
	De Mauro Frequency-of-use labels	p = 0.047035
Prestazioni del Classificatore	Accuratezza = 0.7000 Precision = 0.6429 Recall = 0.9000 F1= 0.7500	
	Task "Sogno"	
		KS test
Tratti selezionati	Utterance length (SD)	p = 0.009033
	Silence segment duration (mean)	p = 0.047035
	Personal Deixis rate	p = 0.076262
Prestazioni del Classificatore	Accuratezza = 0.8500 Precision = 0.8182 Recall = 0.9000 F1= 0.8571	

Per ciascun *task* del test linguistico sono state identificate le *feature* statisticamente significative nel discriminare le due popolazioni di soggetti (*p-value* < 0.05) applicando il test non parametrico di Kolmogorov-Smirnov, in ragione della ridotta

dimensione del campione, e l'algoritmo di *feature selection* mRMR (minimum-Redundancy-Maximum Relevance) (Peng et al., 2005).

Gli indici selezionati sono stati integrati in un classificatore automatico kNN 3-neighbourgs. Il classificatore esegue un campionamento *random* dei testi e 5 iterazioni delle fasi di *training* e *testing*; la dimensione relativa del training set è 80%, ovvero 16 soggetti sono stati utilizzati per l'addestramento del sistema e 4 per il test.

È evidente che la ridotta quantità di dati non permette l'applicazione produttiva di queste metodologie, se non per ricercare tendenze e informazioni al fine di predisporre metodologie più adeguate in vista del campione completo.

Le prestazioni del classificatore in relazione a ciascun *task* sono sintetizzate nella tabella 2:

4. *Conclusioni e sviluppi futuri*

I risultati presentati, ancorché preliminari, appaiono incoraggianti. Pur utilizzando un set di dati estremamente ridotto, l'esperimento ha permesso di saggiare la validità della metodologia proposta e ha fornito indicazioni preliminari sulla significatività che alcuni degli indici linguistici presentano nei testi in esame. Ad oggi non risultano presenti studi simili sull'italiano.

Nelle fasi prodromiche delle malattie dementigene è già presente un deficit linguistico, seppur subliminale; tale deficit, identificabile con strumenti automatici, non è circoscrivibile ad un singolo livello ma, come si può notare dai tratti risultati statisticamente significativi nei diversi *task*, interessa il piano acustico, lessicale e sintattico.

Per confermare tali tendenze è però assolutamente necessaria una maggior quantità di dati, organizzati in un corpus adeguatamente bilanciato. Il reclutamento dei pazienti è, al momento, ancora in corso.

Si prevede inoltre di identificare ed implementare nuovi indici linguistici, e testare nuovi algoritmi di *feature selection* e classificatori ML (ad esempio *neural network*, *Support Vector Machine*, *Quantum Classifiers*).

Bibliografia

Alegria, R., Gallo, C., Bolso, M., dos Santos, B., Prisco, C.R., Bottino, C. & Nogueira, M.I. (2013). Comparative study of the uses of grammatical categories: adjectives, adverbs, pronouns, interjections, conjunctions, and prepositions in patients with Alzheimer's disease. In *Alzheimer's & Dementia: The journal of the Alzheimer's Association*, 9 (4), 882.

Albano Leoni, F. (2003). Tre progetti per l'italiano parlato. In Maraschio, N., Poggi Salani, T. (Eds.), *Italia linguistica anno Mille. Italia linguistica anno Duemila. Atti del XXXIV Congresso Internazionale di studi della Società di linguistica italiana (SLI)*. Roma: Bulzoni, 675-683.

Alexopoulos, P., Grimmer, T., Perneczky, R., Domes, G. & Kurz, A. (2006). Progression to Dementia in Clinical Subtypes of Mild Cognitive Impairment. In *Dementia and geriatric cognitive disorders*, 22, 27-34.

Bäckman, L., Jones, S., Berger, A.K., Laukka, E.J. & Small, B.J. (2005). Cognitive Impairment in Preclinical Alzheimer's Disease: A Meta-Analysis. In *Neuropsychology*, 19 (4), 520-531.

Bambini, V. (2009). Reti metaforiche: il contributo delle neuroimmagini alla modellizzazione della pragmatica. In Ferrari, G., Benatti, R. & Mosca, M. (Eds.), *"Linguistica e modelli tecnologici di ricerca": Atti del XXXIV Congresso Internazionale di studi della Società di Linguistica Italiana (SLI)*. Roma: Bulzoni, 33-45.

Bambini, V., Ricci, I. & Bertinetto, P.M. (Eds.) (2012). *"Linguaggio e cervello - Semantica / Language and the brain - Semantics": Atti del XLII Convegno della Società di Linguistica Italiana (SLI)*. Roma: Bulzoni.

Barbagli, A., Lucisano, P., Dell'Orletta, F., Montemagni, S. & Venturi, G. (2014). Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado. In Basili, R., Lenci, A. & Magnini, B. (Eds.), *Proceeding of the First Italian Conference on Computational Linguistics (CLiC-IT 2014)*, 23-27.

Bosco, C., Lombardo, V., Vassallo, D. & Lesmo, L. (2000). Building a Treebank for Italian: a data-driven annotation schema. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC-2000)*, ELRA - European Language Resources Association.

Brodaty, H., Pond, D., Kemp, N.M., Luscombe, G., Harding, L., Berman, K. & Huppert, F.A. (2002). The GPCOG: A New Screening Test for Dementia Designed for General Practice. In *Journal of the American Geriatrics Society*, 50 (3), 530-534.

Brown, C., Snodgrass, T., Kemper, S.J., Herman R. & Covington, M.A. (2008). Automatic measurement of Propositional Idea Density from Part-of-Speech Tagging. In *Behavior Research Methods*, 40 (2), 540-545.

Brunet, E. (1978). *Le Vocabulaire de Jean Giraudoux. Structure et Evolution*. Geneve: Slatkine.

Bucks, R.S., Singh S., Cuerden J.M. & Wilcock, G.K. (2000). Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance. In *Aphasiology*, 14 (1), 71-91.

Buchholz, S., Marsi, E. (2006). CoNLL-X shared task on Multilingual Dependency Parsing. In *Proceedings of CoNLL-X*, ACL - Association for Computational Linguistics, 149-164.

Cantos-Gòmez, P. (2009). Featuring linguistic decline in Alzheimer's disease: A corpus-based approach. In Mahlberg, M., Gonzàlez Dìaz, V. & Smith, C. (Eds.), *Proceedings of the Corpus Linguistics Conference 2009 (CL2009)*.

Caramelli, P., Mansur, L.L. & Nitrini, R. (1998). Language and communication disorders in dementia of the Alzheimer type. In Stemmer, B., Whitaker, H.A. (Eds.), *Handbook of neurolinguistics*, 463-473.

Carlesimo, G.A., Caltagirone, C., Gainotti, G. & the group for the Standardization of the Mental Deterioration Battery (1996). The Mental

Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. In *European Neurology*, 36, 379-384.

Chapman, S.B., Zientz, J., Weiner, M.F., Rosenberg, R.N., Frawley, W.H. & Burns, M.H. (2002). Discourse changes in early Alzheimer disease, Mild Cognitive Impairment, and normal aging. In *Alzheimer Disease & Associated Disorders*, 16 (3), 177-186.

Chen, P., Ratcliff, G., Belle, S.H., Cauley, J.A., DeKoshy, S.T. & Ganguli, M. (2001). Patterns of cognitive decline in presymptomatic Alzheimer disease. A prospective community study. In *Archives of general psychiatry*, 58, 853-858.

Ciurli, P., Marangolo, P. & Basso, A. (1996). *Esame del Linguaggio – II*. Firenze: Organizzazioni Speciali.

De Mauro, T. (1980). *Guida all'uso delle parole*. Editori Riuniti.

De Mauro, T. (2000). *Il dizionario della lingua italiana*. Paravia.

ElvevÅg, B., GArrard, P. (2014). *Language, Computer and Cognitive Neuroscience*. In *Cortex Special Issue*, 55.

Fraser, K.C., Meltzer, J.A., Graham, N.L., Leonard, C., Hirst, G., Black, S.E. & Rochon, E. (2014). Automated classification of Primary Progressive Aphasia subtypes from narrative speech transcripts. In *Cortex*, 55, 43-60.

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K.I. & Delis, D.C. (1994). *Clock Drawing: A neuropsychological analysis*. Oxford: Oxford University Press.

Gagliardi, G. (2014). *Validazione dell'ontologia dell'azione IMAGACT per lo studio e la diagnosi del Mild Cognitive Impairment*, Tesi di dottorato, Università degli Studi di Firenze.

Garrard, P., Maloney, L.M., Hodges, J.R. & Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. In *Brain*, 128, 250-260.

Grober, E., Hall, C.H., Lipton, R.B., Zonderman, A.B., Resnick, S.M. & Kawas, C. (2008). Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. In *Journal of the International Neuropsychological Society*, 14 (2), 266-278.

Holmes, D.I. (1992). A stylometric analysis of mormon scripture and related texts. In *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155 (1), 91-120.

Holmes, D.I., Singh, S. (1996). A stylometric analysis of conversational speech of aphasic patients. In *Literary and Linguistic Computing*, 11 (3), 133-140.

Honoré, A. (1979). Some simple measures of richness of vocabulary. In *Association of Literary and Linguistic Computing Bulletin*, 7, 172-177.

Ismail, Z., Rajji, T.K. & Shulman, K.I. (2010). Brief cognitive screening instruments: an update. In *International Journal of Geriatric Psychiatry*, 25 (2), 111-120.

Jakobson, R. (2006). *Linguaggio infantile e afasia*. Torino: Einaudi.

Jakobson, R., Halle, M. (1956). *Fundamentals of Language*. Den Haag: Mouton & co.

Jarrold, W.L., Peintner, B., Yeh, E., Krasnow, R., Javitz, H.S. & Swan, G.E. (2010). Language Analytics for Assessing Brain Health: Cognitive Impairment, Depression and Pre-symptomatic Alzheimer's Disease. In Yao, Y., Sun, R., Poggio, T., Liu, J., Zhong, N. & Huang, J. (Eds.), *Brain Informatics: International Conference (BI 2010)*. Berlin - Heidelberg: Springer-Verlag, 299-307.

Jarrold, W.L., Peintner, B., Wilkins, D., Vergryi, D., Richey, C., Gorno-Tempini, M.L. & Ogar, J. (2014). Aided Diagnosis of Dementia Type through Computer-Based Analysis of Spontaneous Speech. In Resnik, P., Resnik, R. & Mitchell, M. (Eds.), *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. ACL - Association for Computational Linguistics, 27-37.

Junkkila, J., Oja, S., Laine, M. & Karrasch, M. (2012). Applicability of the CANTAB-PAL computerized memory test in identifying amnestic Mild Cognitive Impairment and Alzheimer's disease. In *Dementia and Geriatric Cognitive Disorder*, 34, 83-89.

Kempler, D, Curtiss, S. & Jackson, C. (1987). Syntactic preservation in Alzheimer's disease. In *Journal of Speech and Hearing Research*, 30 (3), 343-350.

Larrieu, S., Letenneur, L., Orgogozo, J., Fabrigoule, C., Amieva, H., Le Carret, N., Barberger-Gateau, P. & Dartigues, J. (2002). Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. In *Neurology*, 59 (10), 1594-1599.

Lehr, M., Prud'hommeaux, E., Shafran, I. & Roark, B. (2012). Fully Automated Neuropsycological Assessment for Detecting Mild Cognitive Impairment. In *Proceedings of InterSpeech 2012*.

Lesmo, L. (2007). Il parser basato su regole del Gruppo NLP dell'Università di Torino. In *Intelligenza Artificiale*, IV (2), 46-47.

Lezak, M.D. (1995). *Neuropsychological Assessment*. Oxford: Oxford University Press.

López-de Ipiña, K., Alonso, J.B., Travieso, C.M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., Ezeiza, A., Barroso, N., Ecay-Torres, M., Martinez-Lage, P. & Martinez de Lizardui, U. (2013). On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. In *Sensors*, 13, 6730-6745.

Magni, E., Binetti, G., Bianchetti, A., Rozzini, R. & Trabucchi, M. (1996). Mini-Mental State Examination: a normative study in italian elderly population. In *European Journal of Neurology*, 3, 198-202.

March, E.G., Wales, R. & Pattison, P. (2003). Language use in normal ageing and dementia of the Alzheimer type. In *Clinical Psychologist*, 7 (1), 44-49.

March, E.G., Wales, R. & Pattison, P. (2006). The uses of nouns and deixis in discourse production in Alzheimer's disease. In *Journal of Neurolinguistics*, 19, 311-340.

Measso, G., Cavarzeran, F., Zappalà, G., Lebowtz, B.D., Crock, T., Pirozzolo, F.J., Amaducci, L.A., Massari, D. & Grigoletto, F. (1993). The Mini-Mental State Examination: Normative study of an italian random sample. In *Developmental Neuropsychology*, 9 (2), 77-95.

Mondini, S., Mapelli, D., Vestri, A. & Bisiacchi, P.S. (2003). *Esame Neuropsicologico Breve (ENB). Una batteria di test per lo screening neuropsicologico*. Milano: Raffaello Cortina Editore.

Nasreddine, Z.S., Phillips, N.A., B'edirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L. & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. In *Journal of the American Geriatrics Society*, 53 (4), 695-699.

Novelli, G., Papagno, C., Capitani, E., Laiacona, M., Vallar, G. & Cappa, S.F. (1986). Tre test clinici di ricerca e produzione lessicale. Taratura su soggetti normali. In *Archivio di Psicologia, Neurologia e Psichiatria*, 4, 477-506.

Nucci, M., Mapelli, D. & Mondini, S. (2011). The Cognitive Reserve Questionnaire (CRIq): a new instrument for measuring the cognitive reserve. In *Aging clinical and experimental research*, 24, 218-226.

Peintner, B., Jarrold, W.L., Vergyriy, D., Richey, C., Gorno-Tempini, M.L. & Ogar, J. (2008). Learning diagnostic models using speech and language measures. In *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4648-4651.

Peng, H., Long, F. & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (8), 1226-1238.

Petersen, R.C., Stevens, J., Ganguli, M., Tangalos, E.G., Cummings, J.L. & DeKosky, S.T. (2001). Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). report of the quality standards subcommittee of the american academy of neurology. In *Neurology*, 56 (9), 1133-1142.

Pirani, A., Nasreddine, Z.S., Tulipani, C., Zaccherini, D., Martini, E. & Neri, M. (2007). Montreal Cognitive Assessment (MoCA): uno strumento rapido per lo screening del Mild Cognitive Impairment. Dati preliminari della versione italiana. In *Atti IV Congresso Regionale Associazione Italiana Psicogeriatria*.

Pirani, A., Brodaty, H., Martini, E., Zaccherini, D., Neviani, F. & Neri, M. (2010). The validation of the italian version of the GPCOG (GPCOG-It): a contribution to cross-national implementation of a screening test for dementia in general practice. In *International Psychogeriatrics*, 22 (1), 82-90.

Roark, B., Hosom, J.P., Mitchell, M. & Kaye, J.A. (2007). Automatically derived spoken language markers for detecting Mild Cognitive Impairment. In *Proceedings of the 2nd International Conference on Technology and Aging (ICTA)*.

Roark, B., Mitchell M. & Hollingshead, K. (2007). Syntactic complexity measures for detecting Mild Cognitive Impairment. In Cohen, K.B., Demner-Fushman, D., Frieman, C., Hirschman, L. & Pestian, J. (Eds.), *Proceedings of the Workshop BioNLP 2007: Biological, translational, and clinical language processing*. ACL - Association for Computational Linguistics, 1-8.

Roark, B., Mitchell, M., Hosom, J.P., Hollingshead, K. & Kaye, J.A. (2011). Spoken language derived measures for detecting Mild Cognitive Impairment. In *IEEE Transactions on Audio Speech, and Language Processing*, 19 (7), 2081-2090.

Satt, A., Sorin, A. & Toledo-Ronen, O. (2012). Vocal biomarkers for dementia patient monitoring. In *Proceedings of Interspeech 2012*, ISCA.

Satt, A., Sorin, A., Toledo-Ronen, O., Barkan, O., Kompatsiaris, I., Kokonozi, A. & Tsolaki, M. (2013). Evaluation of speech-based protocol for detection of early-stage dementia. In *Proceedings of Interspeech 2013, 14th annual conference of the International Speech Communication Association*, ISCA, 1692-1696.

Singh, S., Bucks, R.S. & Cuerden, J.M. (2001). An evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech. In *Aphasiology*, 15 (6), 571-583.

Snowdon, D.A., Kemper, S.J., Mortimer, J.A., Greiner, L.H., Wekstein, D.R. & Markesbery, W.R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the nun study. In *The journal of the American Medical Association*, 275, 528-532.

Szmrecs'anyi, B.M. (2004). On operationalizing syntactic complexity. In Purnelle, G., Fairon, C. & Dister, A. (Eds.), *Proceedings of the 7th International Conference on Textual Data Statistical Analysis*. Louvain-la-Neuve: Presses Universitaires de Louvain, 1031-1038.

Taler, V., Phillips, N.A. (2004). Language performance in Alzheimer's disease and Mild Cognitive Impairment: A comparative review. In *Journal of Clinical and Experimental Neuropsychology*, 30 (5), 501-556.

Thomas, C., Keselj, V., Cercone, N., Rockwood, K. & Asp, E. (2005). Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. In Gu, J., Liu, P.X. (Eds.), *Proceedings of the IEEE International Conference on Mechatronics & Automation*, vol. 3, 1569-1574.

Vigorelli, P. (Ed.) (2004). *La conversazione possibile con il malato Alzheimer*. Franco Angeli.

Visser, P.J., Verhey, F.R., Ponds, R. & Jolles, J. (2001). Diagnosis of preclinical Alzheimer's disease in a clinical setting. In *International Psychogeriatrics*, 13 (4), 411-423.

Williams, J.A., Weakley, A., Cook, D.J. & Schmitter-Edgecombe, M. (2013). Machine learning techniques for diagnostic differentiation of Mild Cognitive Impairment and Dementia. In Michalowski, M., Michalowski, W., O'Sullivan, D. & S. Wilk (Eds.), *Expanding the Boundaries of Health Informatics Using Artificial Intelligence: Papers from the AAAI 2013 Workshop*, 71-76.

DOI: 10.17469/O2101AISV000021

CLAUDIO ZMARICH, CINZIA QUARTARONE, SIMONA BERNARDINI, SABRINA BONICHINI, DANIEL COLANTONIO, GIULIA NATARELLI, DEBORA STOCCO, CATERINA PISCIOTTA

Fluenza verbale e produzione lessicale in bambini di 24 mesi d'età con familiarità alla balbuzie: confronto a scopo di prognosi

The research project "Phonetic indexes predictive of chronic stuttering in preschool children" (CNR: RSTL 995, 2007) aimed at identifying clinical phonetic and psychological indexes in preschool children with familiarity to stuttering, in order to early discriminate those at greater risk for chronic stuttering. Forty children were audio- and video recorded at 24 months of age, and those who later started to stutter (N = 14) were assessed every 3 months up to 48 months post-onset. At 24 months of age, the *stuttering-like disfluencies* (*SLD*) were assessed in a representative sample of connected speech. Lexical age data was collected through parents reports. Future stutterers exhibited a SLD percentage significantly greater than future non stutterers and showed a lexical delay, although statistically non significant.

1. *Introduzione*

Secondo l'Organizzazione Mondiale della Sanità, "la balbuzie è un disordine del ritmo della parola per cui il paziente sa cosa vorrebbe dire, ma nello stesso tempo non riesce a dirlo a causa di arresti, ripetizioni e prolungamenti del suono a carattere involontario" (Classificazione Internazionale delle Malattie, WHO, ICD, 1977; tr. di chi scrive). Yairi e Ambrose (2005) hanno chiamato questo tipo di disfluenze *stuttering-like disfluencies* (SLD), locuzione che si potrebbe tradurre con "disfluenze da balbuzie" (senza che ciò implichi che non siano presenti occasionalmente nel parlato dei normofluenti) e le hanno distinte da tutti i rimanenti tipi di disfluenza, chiamati *other disfluencies* (OD), che sono tipici del parlato normofluente ma sono presenti in egual misura anche nei balbuzienti (da qui in poi: *B*). Dunque i *B* si distinguono dai non balbuzienti (da qui in poi: *NB*) per quanto riguarda la percentuale di SLD sul numero di sillabe prodotte (per una definizione più formale di SLD e OD, vedi paragrafo 2.2). Sembra che l'analisi dell'evoluzione temporale della percentuale di SLD costituisca un importante fattore prognostico: secondo Yairi e Ambrose (2005), infatti, la produzione di un numero relativamente alto di SLD, ma non di OD, distingue i bambini *B* dai *NB*; inoltre, in chiave prognostica, se il numero di SLD si riduce in modo notevole e sistematico a partire dal secondo semestre dall'insorgenza del disturbo, i soggetti interessati avranno maggiori possi-

bilità di guarire spontaneamente rispetto a coloro per i quali la percentuale di SLD rimane alta, che invece sarebbero portati a cronicizzare.

Le ricerche di tipo epidemiologico dicono che, per il 95% dei soggetti che incominciano a balbettare, l'inizio avviene prima di compiere 4 anni, e la maggior parte degli esordi si concentra nel terzo anno di vita (Reilly, Onslow, Packman, Cini, Ukoumune, Bavin, Prior, Eadie, Block & Wake, 2013). Il tasso di prevalenza del disturbo viene stimato attorno allo 0,7% della popolazione generale (Yairi, Ambrose, 2013), mentre il tasso di incidenza generale tende verso il 10% (Yairi, Ambrose, 2013). La differenza fra i tassi di incidenza e prevalenza suggerisce che circa il 90% dei bambini che incomincia a balbettare guarisce poi spontaneamente, e per il 95% dei casi questo avviene al massimo entro il quarto anno dalla comparsa (Yairi, Ambrose, 2013). La forte componente ereditaria della balbuzie (da qui in poi: *b*) fa si che colpisca maggiormente il sesso maschile, che ha anche maggiori probabilità di cronicizzare (Kraft, Yairi, 2011). La ricerca internazionale ha inoltre messo in luce in questi ultimi 15 anni, e con una frequenza sempre maggiore, la forte componente organica del disordine, che a livello cerebrale presenta deficit sia strutturali che funzionali (per una rassegna, vedi Watkins, Smith, Davis & Howell, 2008; Cieslak, Ingham, Ingham & Grafton, 2015; Neef, Hoang, Neef, Paulus & Sommer, 2015). La *b* è poi spesso associata ad attitudini negative verso le situazioni comunicative, ma questo risultato non riveste un ruolo causale iniziale ed è anche comprensibile alla luce del ruolo importante giocato dalla capacità linguistica nella vita di una persona (Yaruss, Quesal, 2004; Zmarich, 2015b).

La scoperta delle predisposizioni organiche della *b* ha ridimensionato l'importanza che, soprattutto tra gli anni 1940 e 1960, veniva attribuita ai valori della cultura di appartenenza (Bullen, 1945; Snidecor, 1947; Lemert, 1953), che enfatizzando la capacità comunicativa, attraverso il veicolo genitoriale, avrebbero generato ansia nel bambino e creato tensione nei suoi tentativi di superare le disfluenze, rendendole per questo anomale. Studi più recenti affermano però che le differenze interlinguistiche e interculturali per la balbuzie non sono statisticamente significative (Yairi, Ambrose, 2013). Per quanto riguarda l'analisi della percentuale di SLD e OD a 24 mesi nei bambini che in seguito incominceranno a balbettare (*B*) e in quelli che non lo faranno (*NB*), non c'è letteratura scientifica, perché di solito si assume che finché un bambino non è *b* non produca un numero di SLD maggiore della norma. Per esempio, in uno dei pochissimi studi di nostra conoscenza che ha analizzato, come fatto da noi, alcune variabili comportamentali in un gruppo di bambini audioregistrati prima dell'eventuale insorgenza della *b* (ad un'età media di 3 anni e 3 mesi) e dopo (a circa un anno dalla prima registrazione), si afferma che al momento della prima registrazione nessun bambino produceva disfluenze da *b* (Kloth, Janssen, Kraaimaat & Brutten, 1995). Poiché l'esordio della *b* avviene tra i 2 e i 4 anni d'età, generalmente a seguito di un primo periodo fluente di produzione del parlato, vari ricercatori hanno messo in relazione la frequenza di occorrenza di disfluenze, lapsus e auto-correzioni con l'acquisizione della capacità sintattica di produzione dell'enunciato. Inoltre, tenendo presente che la Lunghezza Media

dell'Enunciato (LME) del bambino italiano di sviluppo tipico a 24 mesi supera già le 4 parole (Caselli, Pasqualetti & Stefanini, 2007), abbiamo ritenuto importante calcolare la percentuale di disfluenze (SLD e OD) già a 24 mesi.

1.1 Il progetto sugli indici predittivi di cronicità del CNR-ISTC di Padova

L'obiettivo del progetto di ricerca "Indici predittivi della balbuzie cronica in età prescolare" (Codice CNR: RSTL 995, 2007) è quello di individuare nel comportamento dei bambini che hanno iniziato a balbettare da poco (da meno di 6 mesi), alcuni indici di tipo clinico (cioè validi, relativamente affidabili e facili da rilevare) che permettano di discriminare i bambini a maggior rischio di cronicizzazione dai falsi positivi (cfr. Zmarich, 2015a), per riservare un trattamento precoce solo ai primi, aumentando le possibilità di successo. Gli indici studiati nel progetto sono di natura comportamentale e sono stati presentati, ma non formalizzati, da Yairi e Ambrose (2005).

A causa di restrizioni temporali e finanziarie, il progetto si è rivolto al reclutamento dei soli soggetti *B* con familiarità al disturbo (figli o fratelli di *B* o ex-*B* certificati), residenti nel Veneto (Zmarich, 2015a). Infatti, la probabilità di incominciare a balbettare per i bambini con storia familiare di *b* è circa sei volte maggiore rispetto ai bambini nati in famiglie di *NB* (Kloth et al., 1999). Potevano aderire al progetto le famiglie in cui, all'epoca del reclutamento, c'erano bambini di età compresa fra i 12 e i 24 mesi.

Incominciato nel 2008, il progetto ha reclutato 44 famiglie tramite annunci su media e locali pubblici e *email* inviate ad ex pazienti del Centro Medico di Foniatria di Padova (CMF, che collabora al progetto insieme ad A.I.BA.COM Onlus). Al momento del reclutamento, le famiglie erano debitamente informate sul contenuto del progetto (consenso informato), e, se d'accordo, firmavano l'autorizzazione al trattamento dei dati sensibili per fini di ricerca, in ottemperanza alla legge 196/2003 sulla Privacy. Al raggiungimento del 24° mese, i bambini erano sottoposti a video e audio registrazione presso il loro domicilio e alla fine della seduta i genitori erano istruiti a contattare i referenti del progetto nel caso il bambino avesse in seguito manifestato i primi sintomi della *b*. Se l'evento si verificava, il bambino era indirizzato al CMF per ricevere una diagnosi formale e per essere sottoposto ad una serie di accertamenti clinici, tra cui quello ORL e neurologico, logopedico e psicologico, allo scopo di escludere patologie importanti diverse dalla *b* e di acquisire informazioni sulle capacità cognitive, linguistiche e motorie, da correlare poi eventualmente con la *b* stessa (v. Zmarich, 2015a). Contestualmente il bambino incominciava ad essere audio e video registrato con cadenza trimestrale presso il suo domicilio da operatori formati e inviati dall'ISTC-CNR, per un periodo di 15 mesi successivi alla prima registrazione da balbuziente (in tutto 6 registrazioni), al fine di raccogliere dati sullo sviluppo fonetico (*TFPL*, Zmarich, Fava, del Monego e Bonifacio, 2012, prima dei 40 mesi; *PFLI*, Bortolini, 1995, dopo i 40 mesi); sullo sviluppo lessicale (*PVB*, Caselli et al., 2007); sulla gravità della *b* (*SSI-3*, Riley, 1994); sull'attitudine comunicativa (*KiddyCAT*, Vanryckeghem & Brutten, 2007; a partire dai 36 mesi).

Dopo 15 mesi dalla prima registrazione (da 15 a 21 mesi dopo l'esordio, a seconda del ritardo temporale della prima registrazione di ciascun soggetto dall'esordio della *b* riportato dai genitori), se il bambino stava ancora balbettando e a richiesta dei genitori, veniva iniziato il trattamento clinico presso il CMF. Da quel momento l'evoluzione naturale del disturbo era interrotta, e l'osservazione sperimentale di conseguenza cessava. I soggetti venivano comunque registrati per altre due tappe: a 3 mesi di distanza dalla sesta registrazione e a 6 mesi di distanza dalla settima (cioè a 18 e a 24 mesi dall'esordio per il bambino la cui prima registrazione da *B* era idealmente avvenuta subito dopo l'esordio). La persistenza della *b* a 36 e 48 mesi dopo la prima registrazione (per Yairi e Ambrose, 2005, i 48 mesi costituiscono l'arco temporale entro cui avviene il 95% dei recuperi spontanei) veniva ulteriormente verificata con un'intervista telefonica strutturata al genitore. Grazie al questionario, è stato perciò possibile indagare le relazioni tra alcune misure raccolte già alla tappa dei 24 mesi (prima dell'eventuale inizio della *b*), e lo status clinico finale (*B* o *NB*).

2. *Procedura Sperimentale*

2.1 Partecipanti

I bambini che sono stati registrati a 24 mesi sono 44; tre di questi si sono ritirati subito dopo e dei 41 rimasti 17 hanno cominciato a balbettare (41,46%), ma 3 di loro, dopo la prima registrazione da balbuzienti, si sono ritirati. I 14 bambini *B* rimasti, insieme ai 26 bambini che non hanno mai cominciato a balbettare, costituiscono il campione di questo studio (N = 40). Essi hanno cominciato a balbettare in media a 34,0 mesi (*range* 21-51 mesi) e sono stati sottoposti alla prima registrazione in media 3 mesi e 15 giorni dopo l'inizio (*range*: 0-15 mesi; se però si esclude un soggetto che è stato registrato 15 mesi dopo l'esordio, la media dei restanti 13 è di 2 mesi e 20 gg., *range*: 0-6 mesi). Quattro dei soggetti hanno intrapreso uno o più cicli di trattamento alla fine del periodo di osservazione. Per tutti i soggetti di questo studio, è stata condotta un'intervista telefonica per accertare il loro stato di *B/NB* a una distanza temporale media di 40 mesi (*range* 20-68 mesi) dall'inizio della balbuzie. Alla data della verifica è risultato che 4 soggetti (tutti maschi), secondo i loro genitori erano ancora *B*. Questi bambini sono gli stessi che erano stati sottoposti a terapia.

2.2 Strumenti e metodologia

Le misure indagate per la registrazione domiciliare dei 24 mesi sono state il grado di disfluenza e la dimensione del vocabolario (Richels et al., 2013). Il grado di disfluenza è stato stabilito tramite il "Profilo delle disfluenze" (Yairi, Ambrose, 2015). Per poter elicitare in bambini così piccoli una produzione lessicale non su ripetizione e non basata su parole isolate, si è fatto ricorso alla manipolazione e alla denominazione di oggetti relativi a tre scenari, come "il bagnetto", "la pappa" e "il parco-giochi", allo scopo di stimolare la verbalizzazione durante il gioco simbolico (cfr. Stoel-Gammon, 1989); inoltre si è utilizzata anche la descrizione di una vignetta

tratta dall'SSI-3. Nella stessa seduta i genitori compilavano il PVB e il questionario anamnestico.

Tutti i partecipanti sono stati audioregistrati (a 44 kHz e 16 bit con un registratore digitale *Edirol R-09HR*) e videoregistrati (*DTV Sony HDR-CX115E*, qualità del segnale *HD LP*).

Per la raccolta dei dati sono stati utilizzati i seguenti strumenti:

– *Profilo delle Disfluenze*: è un indice che si basa sulla classificazione delle disfluenze proposta da Yairi e Ambrose (2005), che le riconducono a due diverse macrocategorie: *SLD* e *OD* (per i rapporti tra le disfluenze, le malfunzioni linguistiche a breve termine o lapsus (*speech errors*) e le auto-correzioni (*speech repairs*) prodotte dai parlanti *B* e *NB*, vedi Zmarich, 2012; 2015b). Le *SLD* sono costituite da ripetizioni di parti di parola, ripetizioni di parole monosillabiche, fonazioni "disritmiche"(cfr. *dysrhythmic phonation*, che includono i prolungamenti udibili, e i "blocchi", cioè i prolungamenti non udibili; Yairi, Ambrose, 2005). Le *OD*, rilevabili anche nel parlato di persone normofluenti, includono le interiezioni, le ripetizioni di parole polisillabiche e di frasi, le revisioni e le interruzioni di frasi. Per bambini di età maggiore rispetto ai partecipanti di questo studio e per gli adulti, il Profilo delle disfluenze viene calcolato contando il numero di SLD in un campione rappresentativo di linguaggio connesso di circa 400 sillabe fluenti (cioè le sillabe dei target lessicali) ed esprimendolo in percentuale. Per esempio, in "queeesto ta-ta-ta-tavolo" contiamo 2 SLD (un prolungamento e una ripetizione di sillaba) su 5 sillabe (40%). Questo esempio rende chiaro come, nel caso delle ripetizioni, venga contato il singolo tipo e non le varie iterazioni che lo costituiscono (che nell'esempio erano tre). Per i bambini del presente studio, relativamente meno loquaci e dal lessico più limitato, il Profilo delle disfluenze è stato calcolato su 200 sillabe. La classificazione delle disfluenze si è basata sulla loro trascrizione fonetica, eseguita dai tesisti di Logopedia che avevano ricevuto un training specifico, che comprendeva anche l'uso di *PRAAT* (www.PRAAT.org), da parte del primo autore. I casi problematici di trascrizione e\o classificazione sono stati discussi insieme dal tesista e dal primo autore, e nei rari casi di permanenza del disaccordo, eliminati dalla selezione.

– *Stuttering Severity Instrument - Third Edition* (*SSI-3*; Riley, 1994): è lo strumento internazionalmente più usato per valutare il livello di gravità della balbuzie. L'SSI-3, che nella versione più recente ma pochissimo diversa dalla precedente, è diventato SSI-4, è statisticamente valido ed affidabile: su due campioni non consecutivi costituiti da almeno 200 sillabe ciascuno, l'SSI-3 valuta tre diversi parametri quali la frequenza delle disfluenze da balbuzie (percentuale di sillabe balbettate), la loro durata (media in *s* dei tre eventi di balbuzie più lunghi) e le caratteristiche fisiche concomitanti non funzionali alla comunicazione (*physical concomitants*: suoni distraenti, smorfie del viso, movimenti della testa, movimenti degli arti). I punteggi rilevati per i diversi parametri vengono trasformati, attraverso delle tabelle di conversione, in punteggi equivalenti a specifici livelli di gravità che possono essere espressi o in percentili o in giudizi qualitativi (lieve, molto-lieve, moderata, grave, molto-grave).

– *Primo Vocabolario del Bambino* (PVB; Caselli et al., 2007): la versione italiana del questionario *MacArthur-Bates Communicative Development Inventories – CDI*, nasce come un questionario da somministrare ai genitori di bambini per lo studio e la valutazione della comunicazione e del linguaggio fra gli 8 e i 36 mesi. Per fini pratici è diviso in due schede, la seconda riservata a bambini dai 18 ai 36 mesi d'età, che è quella usata per questo studio. La parte I della scheda è composta da una lista di 670 parole distribuite in 23 categorie. Il compito dei genitori consiste nel segnare con una "x", le parole che il bambino pronuncia. È stato quindi da noi calcolato il numero di parole prodotte dal bambino a 24 mesi e la relativa età lessicale, che corrisponde all'età in cui lo stesso numero di parole viene prodotto dai bambini collocati al 50° percentile del campione normativo riportato da Caselli et al. (2007). Per calcolare gli eventuali mesi di ritardo del bambino, è stata calcolata la differenza tra l'età lessicale e cronologica.
– *Questionario finale di accertamento clinico*: la persistenza della balbuzie a 36 e 48 mesi dopo la prima registrazione era verificata con un'intervista telefonica in cui il genitore doveva rispondere con SI/NO a una serie di nove domande precostituite (v. appendice 1); il rispondere affermativamente ad almeno tre domande qualificava il bambino come *B*. Inoltre il genitore doveva posizionare la *b* di suo figlio lungo una scala di tipo Likert da 1 a 7, in cui 1 rappresenta l'assenza di *b* e 7 il massimo di gravità.

3. *Risultati*

Per verificare se fosse presente una differenza significativa tra i due gruppi (*B* vs *NB*) in base allo scarto in mesi tra età lessicale ed età cronologica, è stato eseguito un t-test sulla differenza nel punteggio dei futuri *B* (media: -3,42) e *NB* (media: -1,90; Tabella 1). L'analisi statistica ha rilevato che per $t_{(31,74)} = 1,72$, $p = 0,09$, la differenza non è significativa, anche se la tendenza dei futuri *B* a produrre un minor numero di parole si avvicina alla significatività. La Tabella 1 riporta inoltre le statistiche descrittive delle percentuali di SLD e OD ricavate dall'analisi del Profilo delle disfluenze. In margine al focus principale del presente studio, è interessante osservare come a fronte di un'equivalenza tra maschi e femmine nei soggetti *NB*, ci sia una prevalenza di maschi nei *B*, che poi diventa presenza assoluta quando si passa a considerare l'esito clinico del disturbo (quattro *B* maschi persistenti a fronte di nessuna femmina, vedi fine paragrafo 2.1).

Tabella 1 - *Distribuzione dei soggetti (suddivisi per genere) in base all'outcome clinico (B e NB), Media e Deviazione Standard di età cronologica, età lessicale, differenza tra queste due misure e percentuale di SLD e OD*

	Genere		*Età cronologica*	*Età lessicale*	*Diff. less-cron.*	% SLD	% OD
	M	F	(mesi)	(m.)	(mesi)		
NB	13	13	24,59 (1,09)	22,74 (4,00)	–1,90 (4,40)	0,75 (0,69)	3,32 (2,54)
B	9	5	25,29 (3,31)	21,50 (4,68)	–3,42 (2,98)	3,01 (2,73)	4,20 (3,16)

È stata poi riscontrata una differenza tra la percentuale di SLD nei futuri *B* (M = 3,01; DS = 2,73) e quella dei *NB* (M = 0,75; DS = 0,69), che è risultata significativa al test non parametrico Mann-Whitney (U = 75,00, p = 0,002). Per quanto riguarda la percentuale delle OD, per i futuri *B* è pari a 4,20 (DS = 3,16), mentre nei bambini *NB* è pari a 3,32 (DS = 2,54). In base al test non parametrico Mann-Whitney, la differenza tra le percentuali non è significativa (U = 155,50, p = 0,45).

Per concludere, è stato fatto un confronto tra i futuri *B* che hanno poi recuperato spontaneamente (N = 10) e i futuri *B* che hanno poi cronicizzato (N=4). È stato quindi utilizzato il test *t* di Student per rilevare una possibile differenza tra i due gruppi per l'età lessicale meno l'età cronologica, dal quale emerge come, per i futuri B persistenti (M = –4,25, DS = 1,26) e i *B* che poi recupereranno spontaneamente (M = –3,10, DS = 3,44), le differenze non risultino statisticamente significative ($t_{(12)} = -0{,}637$, p = 0,536). Analizzando le percentuali di SLD attraverso il test non parametrico Mann-Whitney emerge che non ci sono differenze significative: i futuri *B* persistenti producono 2,25% SLD mentre i *B* che poi recupereranno spontaneamente producono 3,33% SLD (U = 12, p=0,256).

4. *Discussione*

Il risultato principale della presente ricerca è che già a 24 mesi d'età i futuri *B* si differenziano dai *NB* per quanto riguarda il numero di disfluenze da balbuzie (SLD) prodotte. Essi, infatti, su 100 sillabe target, producono una media di disfluenze pari a 3,01 SLD contro la media di 0,75 dei coetanei *NB*. Tale dato, significativo a livello statistico, mostra una notevole differenza, da non sottovalutare, fra i due gruppi, e potrebbe essere in qualche modo collegato a difficoltà incontrate nella nascente esigenza di costruzione dell'enunciato (si veda Zmarich, 2012). Il dato ottenuto è particolarmente rilevante ai fini di una precoce individuazione del disturbo e potrebbe essere preso in considerazione da ricerche future al fine di individuare, molto prima della manifestazione del disturbo, i bambini maggiormente a rischio, e prevenire l'insorgenza facilitandoli e assistendoli nella costruzione degli enunciati. Sarebbe inoltre interessante verificare se le ripetizioni di segmenti o sillabe che sono incluse nella percentuale prodotta dai futuri *B* siano anche tipologicamente diverse dalle ripetizioni dei coetanei *NB*, poiché sappiamo dalla letteratura (Yairi, Ambrose, 2005) che le ripetizioni delle persone che balbettano sono caratterizzate da un numero maggiore di iterazioni per ogni singolo episodio di ripetizione rispetto a quelle dei NB (per es. "ta-ta-ta-tavolo" vs "ta-tavolo"). Questo aspetto qui non è stato indagato, ma sarà fonte di un possibile approfondimento in futuro, anche in chiave di potenziale predittore clinico precoce.

Il secondo risultato interessante è che i futuri *B* hanno un lessico più limitato rispetto ai futuri *NB*, sebbene in modo statisticamente non significativo. Tale punto, all'interno della nostra ricerca, è stato esaminato attraverso i punteggi ottenuti al PVB, dal quale è stata ricavata la differenza tra l'età lessicale e l'età cronologica.

Il dato ottenuto collima con alcuni dati di letteratura (Ntorou, Conture & Lipsey, 2011), che evidenziano come i bambini *B* abbiano minori capacità nel vocabolario espressivo dei coetanei *NB*.

Per quanto riguarda il solo campione di *B*, una volta suddiviso in coloro che in seguito avrebbero cronicizzato e coloro che invece sarebbero guariti spontaneamente, sono state effettuate le stesse analisi utilizzate in precedenza per differenziare i futuri *B* dai *NB*: la differenza nella percentuale di SLD tra futuri *B* persistenti e per i *B* che recupereranno spontaneamente non risulta essere significativamente diversa, così come non lo è la differenza tra età lessicale e età cronologica.

Bibliografia

Bortolini, U. (1995). *Prove per la valutazione Fonologica del Linguaggio Infantile-PFLI*. Padova: Edit Master Srl.

Bullen, A. (1945). A cross-cultural approach to the problem of stuttering. In *Child Development*, 16, 1-88.

Caselli, M.C., Pasqualetti, P. & Stefanini, S. (2007). *Parole e frasi nel "Primo Vocabolario del Bambino"*. Milano: Franco Angeli.

Cieslak, M., Ingham, R.J., Ingham, J.C. & Grafton, S.T. (2015). Anomalous white matter morphology in adults who stutter. In *Journal of Speech, Language, and Hearing Research*, 58, 268-277.

Kloth, S.A.M., Janssen, P., Kraaimaat, F.W. & Brutten, G.J. (1995). Speech-motor and linguistic skills of young stutterers prior to onset. In *Journal of Fluency Disorders*, 20, 157-170.

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, 253-256.

Kraft, S.J., Yairi, E. (2011). Genetic Bases of Stuttering: The State of the Art 2011. In *Folia Phoniatrica et Logopaedica*, 64, 34-47.

Lemert, E.M. (1953). Some Indians who stutter. In *Journal of Speech and Hearing Disorders*, 18, 168-174.

Neef, N.E., Hoang, T.N., Neef A., Paulus, W. & Sommer, M. (2015). Speech dynamics are coded in the left motor cortex in fluent speakers but not in adults who stutter. In *Brain*, 138, 712-725.

Ntouru, K., Conture, E.G. & Lipsey, M.W. (2011). Language Abilities in Children who Stutter: A Meta-Analitical Review. In *American Journal of Speech-Language Pathology*, 20, 163-179.

Quartarone, C. (2014). *Confronto tra misure di fluenza verbale, produzione lessicale e caratteristiche socioeconomiche in un campione di bambini di 24 mesi d'età con familiarità alla balbuzie, in relazione all'eventuale insorgenza ed eventuale remissione o persistenza*, Tesi di Laurea Magistrale in Neuroscienze e Riabilitazione Neuropsicologica, Università di Padova.

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.

Richels, C.G., Johnson, K.N., Walden, T.A. & Conture, E.G. (2013). Socioeconomic status, parental education, vocabulary and language skills of children who stutter. In *Journal of Communication Disorders*, 46, 361-374.

Riley, G.D. (1994). Stuttering Severity Instrument for children and adults third edition-SSI-3. Austin, Texas: Pro-Ed.

Snidecor, J. (1947). Why the Indian does not stutter. In *Quarterly Journal of Speech*, 33, 493-495.

Stoel-Gammon, C. (1989). Language Production Scale. In Lesley, B., Olswang, D., Stoel-Gammon, C., Truman, E. & Coggins Carpenter, R.L. (Eds.), *Assessing prelinguistic and Early Linguistic Behaviors in Developmentally Young Children.* Washington: Washington Press.

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.

Yaruss, J.S., Quesal, R.W. (2004). Stuttering and the International Classification of Functioning, Disability, and Health (ICF): An update. In *Journal of Communication Disorders*, 37 (1), 35-52.

Watkins, K.E, Smith, S.M., Davis, S. & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. In *Brain*, 131 (1), 50-59.

World Health Organization, (1977). *Manual of the International statistical classification of diseases, injuries, and causes of death* (ICD), vol. 1. Geneva.

Zmarich, C. (2012). Gli aspetti psicolinguistici e fonetici della balbuzie in età prescolare: lo stato dell'arte della ricerca. In Ingenito, M.T. (Ed.), *Discutendo di ...balbuzie. Evidenze recenti in tema di diagnosi e cura in età evolutiva*. Milano: Franco Angeli, 93-113.

Zmarich, C. (2015a). 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. (2015b). Le teorie psicolinguistiche e fonetiche della balbuzie. In Tomaiuoli, D. (Ed.), *Balbuzie: fondamenti, valutazione e trattamento*. Ed. Erickson, 177-210.

Zmarich, C., Fava, I., Del Monego, G. & Bonifacio, S. (2012). *Verso un "Test Fonetico per la Prima Infanzia"*, inviato per una comunicazione al VIII Convegno AISV, Roma, 25-27 gennaio 2012.

Appendice 1 - Questionario finale di accertamento clinico

I genitori possono rispondere a queste semplici domande poste in ordine di gravità del problema. Se un genitore risponde "sì" ad almeno 3 domande questo suggerisce la possibilità che il bambino possa essere balbuziente piuttosto che disfluente.

Il bambino:

1. Ripete parti di parole piuttosto che parole intere o frasi intere?
2. Ripete suoni (*f-f-f-fame*) oppure sillabe (fa-fa-fa-fame) più di una volta ogni 8 / 10 frasi?
3. Fa almeno due ripetizioni di sillaba prima di riuscire a produrla fluentemente (*fa-fa-fame* invece di *fa-fame*)?
4. Sembra frustrato o imbarazzato quando ha difficoltà con le parole?
5. Modifica il tono di voce quando balbetta?
6. Modifica qualche caratteristica del viso quando balbetta (chiude gli occhi o guarda da un lato o mostra tensioni)?
7. Qualche volta si blocca in modo così brusco che non riesce più ad emettere suoni per parecchi secondi quando sta cercando di parlare?
8. A volte associa all'evento disfluente, come il blocco, comportamenti non verbali (battere il piede, chiudere gli occhi..) al fine di aiutarsi nel far uscire la parola?
9. A volte rinuncia a parlare, delega altri a parlare al posto suo per paura di balbettare?

Parte II – Sezione C

Codifica neurale

DOI: 10.17469/O2101AISV000022

ANNA DORA MANCA, FRANCESCO DI RUSSO, MIRKO GRIMALDI

Orderly organization of vowels in the auditory brain: the neuronal correlates of the Italian vowels

The main novelty of the study consists in using the EEG technique to elucidate the functioning of the auditory cortex in speech sound information processing. In searching for the cortical responses of two Italian vowels [a] and [i] we demonstrate that (i) as for the magnetic fields, also the modulations of the early electric N1 component is useful in exploring the mechanisms of phonetic feature extraction. The N1 amplitude and the N1 cortical generators seem to be affected by the vowel features as mapped in F2/F1 ratio so that the vowel [a] that shows a reduced inter-formant distance evokes weaker and anterior activity than [i]. The study thus, confirms that the cortical activity reflects the main spectro-acoustic dissimilarities supporting tonotopy as one of the most prominent organizing rules of the auditory cortices.

1. *Introduction*

The structures of the auditory pathway are specialized in categorizing different classes of sounds. Recent anatomical models of speech have demonstrated the existence of circuits functionally organized for the sound processing. (Hickok, Poeppel, 2007; Hickok et al., 2011; Scott, Johnsrude, 2003; Rauschecker, Tian, 2000). Cumulative findings supported the critical role of the anterior (Scott et al., 2000; Scott, Johnsrude, 2003; Obleser et al., 2006; 2010; Zatorre, Belin, 2001) and the posterior parts (Okada et al., 2010; Hickok, Poeppel, 2004; Wise et al., 2001) of the superior temporal gyrus (STG) in the phonetic and phonological encoding (Rauschecker, Tian, 2000). In general, the common assumption is that some areas along the auditory cortices show an extraordinary capability of extracting and classifying species-specific vocalizations, including the speech sounds (Alain et al., 2001; Scott et al., 2000; Obleser et al., 2010). It has been suggested that speech processing, consists of a cascade of computations that allows decoding streams of continuously varying sounds onto discrete representations. How the mapping procedures occur, it is still poorly understood. At a physical level, speech sounds are the output of the acoustic effects of the movements of specific articulator organs (Kent, 2002; Stevens, 2002). For example, vowels are described by the first three peaks of the spectral envelope, the so-called formant frequencies, i.e. F1, F2 and F3 (Delattre et al., 1952; Albano Leoni, Maturi, 2003) which are correlated with specific articulator configurations. F1 inversely correlates with the tongue height, F2 with the tongue place of articulation, and F2 with F3 are usually used for describing the lip configurations during the vowel production (Stevens, 2002). Intriguingly, the first two formants are crucial-

ly important for vowel categorization (Kent, 2002; Diesch, Luce, 2000). Yet, it seems that spectral relation between formants play a crucial role in determining the principles of functional organization of auditory cortex (Ohl, Scheich, 1997; Diesch, Luce, 1997; 2000).

In the last decade, a series of magnetoencephalographic (MEG) studies has provided new insights on the earliest steps of information processing showing tonotopy as one of the most prominent organizing rules of the human auditory cortex (Pantev et al., 1995; Eulitz et al., 1995). Specifically, it has been shown that distinctive pools of neurons, spatially located at different sites along the auditory pathway, revealed a different sensibility to the properties of the spectral structure of the linguistic stimuli (Diesch, Luce, 1997; 2000). As for vowels, for instance, the phonetic and phonological attributes seems to affect the activations of different sites of the auditory regions, even at the earliest steps of auditory information encoding (Obleser et al., 2003). The behavior of the auditory-evoked responses represent the main tool of the analysis. The N1 component and the magnetic counterpart N1m, is a broad negativity peak occurring between 80 and 150 ms after the stimulus onset, it is generally distributed over the fronto-central scalp area (Wolpaw, Perry, 1975; Woods, 1995). Its properties in terms of amplitude, latency and location of the cortical generators seem to be the cues crucial to describe the mechanisms of the sound speech encoding (Näätänen, Picton, 1987; Eulitz et al., 1995). By means of N1m, MEG data suggested a phonemotopic organization of human auditory cortices (Obleser et al., 2003; 2004; Mäkelä et al., 2003; Shestakova et al., 2004). For example, Obleser and colleagues (2003) revealed weak N1m amplitude values for /a/ because of its spectral envelope in comparing the German vowels [i] with [a]. As for the N1 latency, different studies have stressed that it is affected by acoustic correlates of speech segments peaking early for the first formant frequencies around 1 kHz (Roberts et al., 2000). For example, vowel [a] peaked earlier than vowels [i] or [u] in different studies (Roberts et al., 2000; Diesch, Luce, 1997; Poeppel et al., 1997). Again, the interaction between F1and F2 formants of the vowels, appear to guide vowel cortical mapping (Ohl, Scheich, 1997). Finally, this fact has also demonstrated that, the spatially localization of the N1 generators reflected the acoustic and the related phonological representation of the perceived sounds (Diesch, Luce, 1997; Obleser et al., 2003; 2004; Shestakova et al., 2004; Mäkelä et al., 2003).

In the present work, the cortical responses to two natural Italian vowels [a] and [i] were analyzed by exploring the auditory N1 response of the event-related response (ERP) signals. The aim of the study was twofold: on the one hand, we want to test the tonotopic organization of the auditory cortex by the N1 ERP component and on the other hand, we want to check whether, at different N1 modulations, separable cortical vowel representations can be associated or not. To do this, we analysed amplitude, latency and the cortical generators of the N1 component. It is worthy to note that, to the best of our knowledge, no study has used the electroencephalography (EEG) technique for exploring the vowel cortical mapping procedures, the magnetoencephalography (MEG) technique instead, has

usually been preferred for its high spatial accuracy in the measurements. However, some recent studies have shown that MEG offers no significant advantage over EEG in localizing a focal source (Virtanen et al., 1998; Cohen et al., 1990); moreover, both techniques seem to reveal similar differences in terms of accuracy in the spatial resolution of the cortical sources (Eulitz et al., 1997). Finally, advance in the high-density electrode montage and the EEG source analysis software provide much accurate brain localization measurements (Gevins et al., 1994; Cohen, Halghren, 2003).

Therefore, we expect that the acoustic phonetic features of the vowels shape their neuronal representation and that the electric N1 modulations show the tonotopic structure of the auditory cortex.

2. *Materials and methods*

2.1 Participants

Twelve students (7 males) with a mean age of 25±3 years took part in the experiment. They provided a written informed consent. They were consistently right-handed according to Handedness Edinburgh Questionnaire (Oldfield, 1971). None of them had any known neurological disorder or other significant health problem. The experimental procedure was approved by the local ethics committee.

2.2 Stimuli

The stimuli consisted of the two SI vowels [a] and [i]. Three native Italian male speakers (mean age 33±7 years) realized ten repetitions of each vowel in isolation, at a normal rate. The speech signal was recorded in a soundproof room with CSL 4500 and a Shure SM58-LCE microphone with a sampling rate of 44.1 kHz and an amplitude resolution of 16 bits. The stimuli were normalized for intensity (70 dB/SPL) and for duration (300ms) by using the speech analysis software Praat 5.2 (Boersma, Weenink, 2011). For every speaker (S1, S2, and S3) we selected the tokens with similar contours pitch within every vowel category. Finally, the first (F1), the second (F2) and the third (F3) formant values were measured in the vowel steady tract centered at the midpoint and the acoustic distances (F2-F1) were extracted (Table 1).

Table 1 - *Spectral compositions of the vowel stimuli (measured in Hertz) used in the experimental protocol*

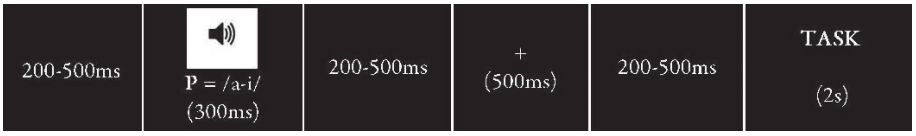
Vowel	*Speaker*	*F0*	*F1*	*F2*	*F2-F1*
/a/	S1	136	856	1344	488
	S2	122	838	1432	594
	S3	133	816	1373	557
/i/	S1	150	321	2189	1868
	S2	129	313	2470	2157
	S3	134	281	2308	2027

2.3 Experimental Procedure

A schematic representation of the experimental procedure is shown in Figure 1. In a dark and shielded room, subjects sat in front of a computer monitor and were instructed to passively listening to the Italian vowels [a] and [i] randomly presented (Perception task (P)) and then to perform a specific articulatory task when a white screen appeared. In three different sessions, subjects had to (i) produce aloud the perceived vowel i.e., Acoustic Production task (AP), (ii) produce without emitting sound the vowel perceived, i.e. Silent Production task (SP) and (iii) mentally produce the perceived vowel i.e., Articulation Imagery task (AI).

Each trial began with a random exposure to a black screen (200-500ms) followed by an audible stimulus [a] and [i] which was randomly presented for 300 ms. After a randomized time interval (200-500ms) a white fixation cross (500ms) appeared at the centre of the black screen in order to suggest subjects to concentrate and prepare for the task. Another randomized time interval (200-500ms) preceded a white screen (2sec), which triggered the onset of the production and imagination tasks. When the black screen appeared again, the next trial started. In each experimental session, vowels were presented 160 times, 80 per vowel category. Since the perception condition systematically preceded the three articulatory tasks, perceptual data were finally computed on 240 trials per vowel (80 vowels per three Tasks). The auditory interstimulus interval (ISI) ranged from 3100 to 4300ms.

Figure 1 - *Timing of the experimental conditions. The auditory N1 responses (P) preceding the articulatory tasks (AP, SP, AI) were off-line summed and computed on 240 trials per vowel category*



In the present report, only the results concerning the Perception task are considered; the whole study is reported in Manca and Grimaldi 2013.

2.4 EEG Recordings

Continuous EEG was recorded by using a 64-ch actiCAP (BrainProducts GmbH, Germany; 10-20 System) and BrainVision Recorder 1.20 (BrainProducts GmbH, Germany) with a sampling rate of 250 Hz and a band pass filter of 0.1-70 Hz. Eye movements were monitored by several electrodes: for the vertical electrooculogram (VEOG) Fp1- Fp2 and an additional electrode (O2) was attached below the right eye; for the horizontal electrooculogram (HEOG), the electrodes FT9 and FT10 were used. The online reference was at FCz and impedance was kept under 5 Ω by electrogel conductant.

3. *Data analysis*

3.1 EEG Processing

Off-line signal processing was carried out with the software package BrainVision Analyzer 2.0.1 (BrainProducts GmbH, Germany). 1200 ms ERP epochs (including 200 ms of baseline) were extracted according to the onset of the vowel stimulus and digitally filtered by a 1–30 Hz bandpass filter (48db), a notch filter (50 Hz) and re-referenced to the right and left mastoids (TP9-T910). Ocular artifacts were removed by applying the Independent Component Analysis (ICA) algorithm and, additionally, rejection criteria were set (maximum allowed absolute difference =120 μV in 1000ms). Artifact-free segments were separately averaged for each vowel and a baseline correction of 200 ms was applied. Finally, the grand averages were computed across the subjects and for each vowel category. N1 was identified as the most prominent peak between 80 and 150 ms after the stimulus onset on the central medial electrode Cz. The N1 mean peak amplitude was calculated considering an interval of 60 ms cantered at the maximum peak (Table 2).

Table 2 - *N1 latency (ms) and amplitude (μV) values of the vowels for each subject and mean values (with standard deviation)*

	Latency		*Amplitude*	
N1	*[a]*	*[i]*	*[a]*	*[i]*
Sub 1	108	109	–1.98	–2.36
Sub 2	111	104	–.89	–1.11
Sub 3	104	96	–.44	–.54
Sub 4	111	116	–3.64	–4.49
Sub 5	107	107	–1.15	–2.04
Sub 6	107	111	–2.56	–3.31
Sub 7	111	109	–1.77	–1.66
Sub 8	108	120	–2.33	–1.76
Sub 9	95	111	–.74	–1.05
Sub 10	111	111	–1.01	–1.48
Sub 11	124	120	–2.24	–3.06
Sub 12	111	116	–1.60	–1.38
mean	108 (6.6)	110 (6.8)	–1.69 (0.9)	–2.01 (1.1)

3.3 Source Analysis

Estimation of the N1 intracranial sources was carried out using the BESA 2000. We used the spatiotemporal source analysis of BESA that estimates location, orientation and time course of equivalent dipolar sources by calculating the scalp distribution obtained for a given model (forward solution). This distribution was then compared to that of the actual auditory evoked potentials (AEP). Interactive changes in source location and orientation lead to minimization of residual variance between the model and the observed spatiotemporal AEP distribution. The optimal set of parameters was found in an iterative manner by searching for a minimum in

the compound cost function. Latency ranges for fitting N1 was between 100 and 120 ms. The source generator to each vowel was calculated on the grand average N1 waveform and was modelled as a symmetric pair of equivalent current dipoles (ECD) one on each hemispere. The accuracy of the source model was evaluated by measuring its residual variance as a percentage of the signal variance, as described by the model, and by applying residual orthogonality tests (ROT; e.g., Bocker et al., 1994). The resulting individual time series for the dipole moments (the source waves) were subjected to an orthogonality test, referred to as a source wave orthogonality test (SOT; Bocker et al., 1994).

3.2 Statistical Analysis

A t-test was run to compare the N1 amplitude and latency values at the Cz electrode to the vowels [a] and [i]. Furthermore, a 2x4 repeated measure analysis of variance ANOVA having as within factors Vowel type [a, i] and Electrode Type, was run on N1 amplitudes to test hemispheric asymmetries. Analysis was carried out at the electrode sites where the components were maximal in amplitude: at C1, CP1 for the left hemisphere, and C2 and CP2 for the right hemisphere.

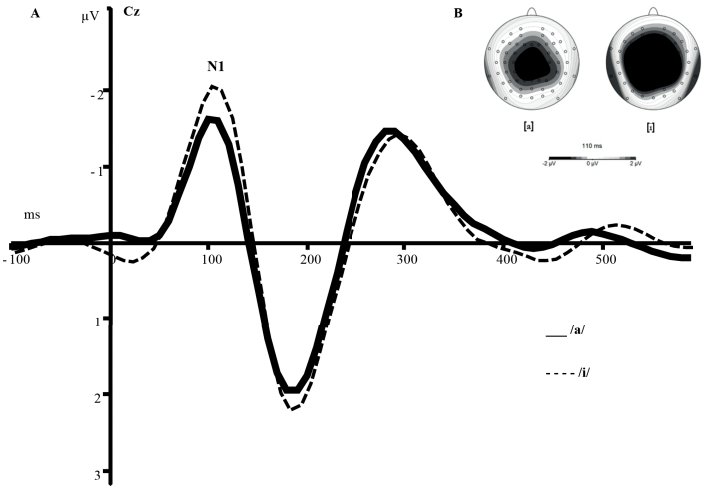
Differences in source localization within the left hemisphere (the right ECD was specular) were evaluated by means of one-way ANOVAs to test distance between the vowels on each spatial plane (anterior-posterior; x, lateral medial; y and superior-inferior; z), and the absolute Euclidean distance. All statistical results were corrected with the Greenhouse–Geisser method, whereas Bonferroni post-hoc method was applied to adjust the univariate output of repeated measurements of variances for violations of the compound symmetry assumption. In source analysis, Turkey (HSD) post-hoc method was applied. All t-statistics were evaluated for significance at the 5% level.

4. *Results*

4.1 Vowel comparison: Analysis on the N1 responses

Figure 2A shows the N1 waveforms at Cz electrode, where the activity was prominent for each vowel, in black line for [a] and in dashed line for [i]. From statistically analysis emerged that the vowels differ significantly in terms of the N1 peak amplitude at Cz electrode, $t(11)=2.40$, $p<0.05$. The vowel [i] recorded a higher value than [a]; N1 amplitude to [a] was of -1.69 μV whereas N1 to [i] had a peak of -2.01 μV. In contrast, the N1 peak latency did not statistically differ, $t(11)=0.98$, $p = 0.371$. The response to [a] had a latency of 108 ms, the response to [i] of 110 ms.

Figure 2 - *(A) Grand-averaged N1 responses (n=12 subjects) elicited by the vowel [a] (black line) and by [i] (dashed line). (B) shows the topographic maps of the N1 activities*



As showed in the topographical maps in Figure 2B, the N1 responses to [a] and [i] showed a different scalp distribution, which is more medial for [a] and more laterally distributed for [i]. Laterality ANOVA revealed a significant difference for the main effect of Vowel Type, $F(1,11)=4.44$, $p<0.05$ showing stronger negative activity in CP1 electrode for [i] relative to [a]; yet, hemisphere affected N1 modulation, $F(3,33)=3.11$, $p<0.05$, showing a larger activity in the left hemisphere for both vowels.

4.2 Source Analysis

The intracranial localization of the N1s sources for the two vowels is shown in Figure 3. The N1 response to each vowel was localized within the bilateral primary auditory cortex in Brodmann area (BA) 41, in the bilateral STG, approximately.

Figure 3 - *Intracranial source localization of N1s. In red the ECD to the vowel [i] in black ECD to the vowel [a]*

Averaging the Talairach coordinates across vowels the mean locations resulted ±52, 14, 10 (mm on x, y, z planes). Table 3 reports the Talairach coordinates within the left hemisphere.

Table 3 - *Mean (and standard deviation) Talairach coordinates of the source location of the two vowels for the N1 response in the left hemisphere*

N1	X	Y	Z
[a]	–49 (7)	–9 (2)	11 (4)
[i]	–54 (8)	–18 (4)	9 (3)
Mean	–52	–14	10

Statistical analysis showed that the vowel type influenced the N1 source location along the anterior-posterior y-axis ($F(1,11)=9.87$, $p<0.01$), the N1 generators were separated by a distance of 9 mm revealing a more anterior location for the vowel [a] relative to [i]. However, along the medial-lateral axis (x), even though the distance between vowels did not reach the statistically significance ($F(1,11)=4.21$, $p=0.06$), the N1 cortical generator to [a] was more lateral than the N1 source to [i], (5 mm). Along the superior-inferior z-axis, the difference (1 mm) was not significant ($F(1,11)<1$, ns). The Euclidean distance between the two vowels (10 mm) was significant ($F(1,11)=11.05$, $p<0.01$).

5. *Discussion*

The results of the present study suggest that the electric N1 response correlates to the functioning of the auditory cortex such that it mirrors its tonotopic structure. Therefore, N1 reveals as a suitable tool for observing the earliest steps of the brain mechanisms of extraction of linguistically relevant information. Specifically, the N1 amplitude pattern seems to index the cortical responsiveness of the auditory system to the spectro-temporal features of the natural Italian vowels [a] and [i]. The vowel [a] with close formant peaks elicited a reduction of the auditory activity at the scalp relative to vowel [i], which is characterized by a large F2-F1 distance of about 2000 Hz (cf. Table 1). A weaker auditory activity for [a] than for [i] is explained by referring to the so-called inhibitory formant frequency principle whereby the spectral relation rather than the extraction of separate formant peaks allows the vowel distinction (Ohl, Scheich, 1997; Diesch, Luce, 2000). These results are in line with previous MEG studies that have shown the effects of the acoustic structure of German (Obleser et al., 2003; 2004), Russian (Shestakova et al., 2004), and French vowels (Mäkelä et al., 2003) on the auditory mechanisms underlying speech processes.

Furthermore, the topographic differences between the vowels, showed a lateral position of the ECD source elicited by the vowels [i] relative to the ECD of the vowel [a] (Figure 2B) suggesting us the presence of different cortical generators (Näätänen, Picton, 1987). By source analysis, we found that the cortical generators of the activities evoked by the vowels were in primary auditory cortex (A1), mapped in Brodmann area BA 41, bilaterally. Imaging data has widely demonstrated the existence of a sophisticated dual-stream architecture of the auditory pathway where the core regions (BA41), work selectively during the speech information process-

ing (for a review see Hickok, Poeppel, 2004; Scott et al., 2003). From our data, we can assume that the N1 responses index the start of the sensory-based perceptual processes from the bilateral A1, (Hickok, Poeppel, 2007) thus triggering the speech recognition mechanisms (Scott et al., 2003; Formisano et al., 2008; Hickok, Poeppel, 2007; De Witt, Rauschecker, 2012). The displacement of the N1 generators along the auditory cortices gives further confirmation in that perspective. ECD to [a] results to be anterior to ECD source to [i] supporting the assumption that the depth of the source is stimulus frequency-dependent; in other word, higher frequency, deeper the location of the ECD source accounting for the N1 waves (Woods, 1995; Eulitz et al., 1995). In our study, the deep location of the ECD to [i] along the anterior-posterior axis, allows us to assume that different cortical loci are given by a different vowel acoustic structure. Actually, the spectral envelope of the vowels, as mapped in the F2-F1 acoustic space, affects the neural sensitivity of the auditory tissues (Diesch et al., 1996; Mesgarani et al., 2008; 2014). The study thus, confirms that the cortical activity reflects the main spectro-acoustic dissimilarities of the vowels, and it supports tonotopy as one of the most prominent organizing rules of the auditory cortices.

6. *Conclusion*

Exploiting for the first time the ERP N1 component for speech cortical representation, we provided further confirmation of the existence of tonotopy maps along the auditory system. Even though we are aware of some important limits of the present study such as the presentation of only two types of stimuli, it is worth noting that this is the first investigation on the Italian vowel system. The vowels [a] and [i] represent the two maximally different vowels within numerous vocalic systems therefore, they were suitable tools for the initial phases of our tonotopic inquiry. We expected that vowels differing in acoustic and articulator patterns could determine discrete cortical maps. Actually, we are working on the entire Salento Italian vowel system to establish specifically, how information on phonemes is implemented in the human brain with the aim to investigate to what extent the acoustic properties or rather the phonological features influence the cortical representation of speech sounds.

References

Alain, C., Arnott, S.R., Hevenor, S., Graham, S. & Grady, C.L. (2001). "What" and "where" in the human auditory system. In *Proceedings of the national academy of sciences*, 98 (21), 12301-12306.

Albano Leoni, F., Maturi, P. (Eds.) (1995). *Manuale di fonetica*. Roma: Carocci.

Bocker, K.B., Brunia, C.H. & Cluitmans, P.J. (1994). A spatio-temporal dipole model of the readiness potential in humans finger movement. In *Electroenchephalogry and Clinical Neurophysiology*, 91, 275-285.

Cohen, D., Cuffin, B.N., Yunokuchi, K., Maniewski, R., Purcell, C., Cosgrove, G.R., Ives, J. , Kennedy, J.G. & Schomer, D.L. (1990). MEG versus EEG localization test using implanted sources in the human brain. In *Annals of Neurology*, 28, 811-817.

Cohen, D., Halgren E. (2003). Magnetoencephalography (Neuromagnetism). In *Encyclopedia of Neuroscience*. Elsevier, 615-622.

Delattre, P., Liberman, A.M., Cooper, F.S. & Gerstman, L.J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one-and two-formant vowels synthesized from spectrographic patterns. In *Word,* 8, 195-210.

DeWitt, I., Rauschecker, J.P. (2012). Phoneme and word recognition in the auditory ventral stream. In *Proceedings of the National Academy of Sciences*, 109 (8), E505-E514.

Diesch, E., Eulitz, C., Hampson, S. & Ross, B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. In *Brain Language*, 53, 143-168.

Diesch, E., Luce, T. (1997). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. In *Psychophysiology*, 34 (5), 501-510.

Diesch, E., Luce, T. (2000). Topographic and temporal indices of vowel spectral envelope extraction in the human auditory cortex. In *Journal of cognitive Neuroscience*, 12 (5), 878-893.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S. & Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. In *Journal of Neuroscience*, 15, 2748-2755.

Eulitz, C., Eulitz, H. & Elbert, T. (1997). Differential outcomes from magneto-and electroencephalography for the analysis of human cognition. In *Neuroscience letters*, 227 (3), 185-188.

Formisano, E., De Martino, F., Bonte, M. & Goebel, R. (2008). 'Who' is saying "what"? Brain-based decoding of human voice and speech. In *Science*, 322, 970-973.

Gevins, A., Le, J., Martin, N.K., Brickett, P., Desmond, J. & Reutter, B. (1994). High resolution EEG: 124-channel recording, spatial deblurring and MRI integration methods. In *Electroencephalography and clinical Neurophysiology*, 90 (5), 337-358.

Hickok, G., Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. In *Cognition*, 92 (1), 67-99.

Hickok, G., Poeppel, D. (2007). The cortical organization of speech processing. In *Nature Reviews Neuroscience*, 8, 393-402.

Hickok, G., Houde, J. & Rong F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. In *Neuron*, 69 (3), 407-422.

Kent, R.D. (Ed.) (2002). *Acoustic analysis of speech*. Singular Publishing Group.

Mäkelä, A.M., Alku P. & Tiitinen, H. (2003). The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. In *Neuroscience Letters*, 353, 111-114.

Manca, A.D., Grimaldi, M. (2013). Perception, production, articulation and imagery articulation of Italian vowels: an ERP study. In *Proceedings of the 9th Conference of Multimodalità e Multilingualità: La sfida più avanzata della comunicazione orale.* AISV, (Eds.), Bulzoni, 213-217.

Mesgarani, N., David, S.V., Fritz, J.B. & Shamma, S.A. (2008). Phoneme representation and classification in primary auditory cortex. In *Journal of Acoustical Society of America*, 123, 899-909.

Mesgarani, N., Cheung, C., Johnson, K. & Chang, E.F. (2014). Phonetic feature encoding in human superior temporal gyrus. In *Science*, 343, 1006-1010.

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., Elbert, T., Lahiri, A. & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. In *Cognitive Brain Research*, 15, 207-213.

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.

Obleser, J., Leaver, A.M., Vanmeter, J. & Rauschecker, J.P. (2010). Segregation of vowels and consonants in human auditory cortex: evidence for a hierarchical organization. In *Frontiers of Psychology,* 1 (232).

Ohl, F.W., Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. In *Proceedings of of the National Academy of Sciences,* 94, 9440-9444.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K. & Hickok, G. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. In *Cerebral Cortex*, 20 (10), 2486-2495.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G. & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. In *Electroencephalography and clinical neurophysiology*, 94 (1), 26-40.

Poeppel, D., Phillips, C., Yellin, E., Rowley, H.A., Roberts, T.P. & Marantz, A. (1997). Processing of vowels in supratemporal auditory cortex. In *Neuroscience Letters*, 221, 145-148.

Rauschecker, J.P., Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. In *proceedings of the National Academy of Sciences*, 97 (22), 11800-11806.

Roberts, T.P., Ferrari, P., Stufflebeam, S.M. & Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. In *Journal of Clinical Neurophysiology*, 17 (2), 114-129.

Scott, S.K., Blank, C.C., Rosen, S. & Wise, R.J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. In *Brain*, 123 (12), 2400-2406.

Scott, S.K., Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. In *Trends in Neurosciences*, 26 (2), 100-107.

Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V. & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars". In *Brain Cognitive Research,* 21, 342-350.

Stevens, K.N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. In *Journal of Acoustical Society of America*, 111 (4), 1872-1891.

Virtanen, J., Ahveninen, J., Ilmoniemi, R.J., Näätänen, R. & Pekkonen, E. (1998). Replicability of MEG and EEG measures of the auditory N1/N1m-response. In *Electroencephalography and clinical Neurophysiology*, 108, 291-298.

Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K. & Warburton, E.A. (2001). Separate neural subsystems withinWernicke's area'. In *Brain*, 124 (1), 83-95.

Wolpaw, J.R., Penry, J.K. (1975). A temporal component of the auditory evoked response. In *Electroencephalography and Clinical Neurophysiology*, 39, 609-620.

Woods, D.L. (1995). The component structure of the N1 wave of the human auditory evoked potential. In *Electroencephalography and Clinical Neurophysiology-Supplements only*, (44), 102-109.

Zatorre, R.J., Belin, P. (2001). Spectral and temporal processing in human auditory cortex. In *Cerebral Cortex*, 11 (10), 946-953.

Parte II – Sezione D

Lingua dei segni e linguaggio musicale

DOI: 10.17469/O2101AISV000023

SABINA FONTANA, MARIA ROCCAFORTE

Lo strutturarsi e il destrutturarsi dei suoni nell'interazione con la lingua dei segni italiana (LIS)

Many sign languages make use of oral components in a crosslinguistically similar way. The understanding of mouth actions is strictly linked to an adequate coding system which is able to describe not only the inner characteristic but also its relation with signing without being influenced by spoken language. Mouth actions will be described in terms of visual articulatory configurations in co-occurrence with signing following on one side Kendon (2004) and McNeill's (2005) hypothesis on gesture, on the other Vogt-Svendsen's (1983) and Ajello's (2001) approach, taking into consideration the mechanism of Embodied Simulation (Gallese, Sinigaglia, 2011).

1. *Un sistema multimodale e multilineare*

Studi recenti (Cuxac, Antinoro Pizzuto, 2010) hanno evidenziato la natura multilineare e multimodale delle lingue dei segni. Si tratta di lingue multilineari poiché le unità di senso vengono veicolate simultaneamente attraverso più articolatori e multimodali poiché questi articolatori sono di natura manuale e non manuale. Gli articolatori manuali co-occorrono sistematicamente con articolatori non manuali (espressione facciale, posizione del busto, direzione dello sguardo, componenti labiali, movimenti del capo). Questi elementi sono sempre stati presenti, ma la loro funzione è stata fino ad oggi molto meno indagata rispetto all'attività delle mani, così come per le lingue vocali, solo recentemente sono state considerate non solo le componenti vocali ma anche quelle gestuali (cfr. Kendon, 2004; McNeill, 2005). Fattori neurobiologici e percettivi sono sicuramente alla base della rilevanza che, tra le componenti non manuali, possiedono quelle propriamente facciali (del resto i segnanti quando comunicano guardano il volto, non le mani), una sorta di input aggiuntivo o di anticipatore per il destinatario del messaggio di ciò che verrà segnato, come è il caso, ad esempio, degli indicatori di domanda e di negazione nella Lingua dei segni italiana o dei diversi schemi intonativi in Italiano lingua vocale.

Tra i fenomeni di multimodalità, nelle lingue dei segni è stata rilevata la massiccia presenza di componenti orali prodotte in concomitanza con le unità segniche o sintagmatiche. Sulla base della presenza/assenza di un collegamento con il parlato, sono stati distinti fenomeni di mouthing e di mouth gestures (Boyes Braem, Sutton-Spence, 2001) che d'ora in avanti definiremo rispettivamente labializzazioni e gesti labiali.

Finora è emerso che le labializzazioni sono unità lessicali o frammenti di lingua vocale, talvolta associati a vocalizzazione e articolati nella parte visivamente più rile-

vante, con funzioni diverse, tra le quali la disambiguazione di omonimi, di varianti regionali o di neologismi. Lo studio di queste componenti però non è semplice, sia perché spesso passano inosservate rispetto al livello manuale, sia perché, una volta rilevate, sono difficilmente codificabili. I gesti labiali, invece, sono vocalizzazioni o emissioni d'aria che accompagnano obbligatoriamente alcuni segni, non presentano alcun collegamento con il parlato e svolgono un ruolo importante sia al livello sub-lessicale o cheremico (possono avere un ruolo distintivo), sia al livello lessicale e morfemico (possono fungere da avverbi). Si tratta di articolazioni singole e analizzabili, la cui descrizione non può prescindere dalla considerazione di una serie di movimenti muscolari visibili che coinvolgono diversi parametri ricorrenti (Vogt-Svendsen, 1983). Tuttavia, sinora, la comprensione di queste componenti è stata ostacolata da una codifica basata su glosse e quindi su un tipo di scrittura alfabetico che ne fornisce una rappresentazione formale fortemente influenzata dalle lingue vocali e basata sull'interpretazione di chi trascrive. Pertanto, nell'analisi delle componenti orali delle lingue dei segni si pongono primariamente problemi di natura epistemologica, correlati cioè al loro riconoscimento e isolamento. In secondo luogo, la codifica di queste forme è basata più su scelte interpretative che su sistemi condivisi ed elaborati per questo scopo.

Tuttavia, la presenza sistematica di queste componenti nelle lingue dei segni richiede l'elaborazione di un modello di analisi formale che ci consenta di rappresentarle in termini di attività articolatoria senza margini interpretativi e allo stesso tempo senza forzarle nelle categorie di analisi elaborate per il parlato.

Il presente lavoro di ricerca si propone, dunque, di analizzare, in una prospettiva epistemologica e metodologica, il rapporto che intercorre tra le componenti orali e le unità segniche a partire da una prospettiva gestuale.

2. *Le componenti orali nelle lingue dei segni: Gesti labiali e Labializzazioni*

Come abbiamo visto, le similarità esistenti tra le componenti orali di diverse lingue dei segni hanno consentito una classificazione uniforme sulla base della presenza/assenza di un collegamento con il parlato. Sebbene Vogt Svendsen (2001) sostenga che sarebbe opportuno abolire tale classificazione e utilizzare il punto di vista dei segnanti, a nostro avviso, in realtà, sul piano funzionale, i segnanti sembrano utilizzare consapevolmente il collegamento con la lingua vocale nella costruzione della significazione. In particolare la scelta di una componente o dell'altra sembra essere determinata da fattori interni alla lingua (funzioni pragmatiche di disambiguazione, di parser, di imagery-metaforiche/iconiche) e da fattori esterni (situazione comunicativa, interlocutore).

Per comprendere come vengono distribuite queste componenti orali, occorre prendere in considerazione le funzioni ad esse attribuite, cominciando, in particolare, con le labializzazioni. Esse possono svolgere funzioni di tipo morfemico, poiché possono disambiguare omonimi sul piano manuale o supportare un segno manuale scarsamente standardizzato (Vogt-Svendsen, 2001). Le labializzazioni sembrano

avere alcune funzioni nel contesto delle produzioni segniche, sebbene siano fortemente influenzate da variabili esterne di natura sociolinguistica – educazione, contesto, interlocutore – (Ajello, Mazzoni & Nicolai, 1988; Boyes Braem, 2001; Hohenberger, Happ, 2001; Sutton-Spence, Day, 2001).

Poiché non sembra esserci una particolare ragione per cui un dato segno debba obbligatoriamente co-occorrere con la corrispondente labializzazione, se non ha omonimi manuali, se non è un neologismo o un segno diatopicamente marcato, occorrerebbe liberarsi da categorie predeterminate inquadrando le labializzazioni in modo dinamico, ovvero in modo funzionale al discorso segnico. (Ebbinghaus, Hessman, 2001). È indubbio però che le produzioni segnate senza labializzazioni (Fontana, Raniolo, 2015) non sono considerate appropriate da parte dei segnanti e spesso non risultano essere del tutto comprensibili. In quest'ottica, sarebbe opportuno esplorare l'uso delle labializzazioni nel discorso e valutarle, come suggerisce Pizzuto (2003), non solo in riferimento alle ricerche sulle caratteristiche percettive e articolatorie del parlato (Ajello et al., 2001), ma anche tenendo presenti le ricerche compiute sul gesto coverbale (Kendon, 2004, Mc Neill, 2005, in termini di gesto cosegnico (Fontana 2008; 2009). Le variabili esterne (educazione, contesto, interlocutore) agiscono piuttosto sulla forma delle labializzazioni determinando delle trasformazioni che si possono collocare lungo un continuum di globalità/discretezza. Si potrebbe affermare che le labializzazioni, infatti, oscillino da una gestalt labiale di elementi foneticamente rilevanti di una data parola (ma che non sono analizzabili sulla base di criteri fonologici), ad una parola che è interamente articolata con possibili alterazioni morfologiche (cfr. Fontana, Raniolo, 2015). In ogni caso, generalmente, le labializzazioni non sono sempre riconducibili all'articolazione di una parola della lingua orale, ma alla riproduzione della forma labiale del gruppo fonetico rilevante per i sordi in termini articolatori, cioè quello che risulta più facilmente visibile mediante la lettura labiale[1]. La parola non è sempre pronunciata per intero: generalmente vengono omesse le parti non giudicate rilevanti dal punto di vista della decodifica (tra cui per esempio le informazioni morfologiche della lingua vocale). Per esempio, gli elementi nominali e aggettivali, che sono tra quelli più labializzati, compaiono tendenzialmente nella rispettiva forma singolare, mentre i verbi co-occorrono con una labializzazione solitamente all'indicativo presente, non sempre flessa correttamente per persona e numero, infine le labializzazioni rispecchiano il movimento e il ritmo del segno (Fontana, Fabbretti, 2000). Può esserci un'articolazione ed emissione di voce presumibilmente non intenzionale connessa ad obiettivi enfatici, per quanto non sia rilevante per la comunicazione nel caso in cui gli interlocutori siano sordi. Ad esempio, per il segno LAVORO viene labializzato normalmente solo 'lavo' e spesso enfatizzato il gruppo fonetico iniziale 'la'.

I gesti labiali sono invece forme di oralità prive di rapporto con la lingua vocale. Sebbene siano tradizionalmente considerati parte del sistema delle lingue dei segni, la loro codifica appare piuttosto problematica poiché consistono prevalentemente

[1] Da qui la definizione *borrowed word-pictures* tradotta in italiano come 'immagini di parole prestate' (cfr. Franchi, 1987).

in suoni prodotti con emissioni d'aria, protrusioni della lingua e delle labbra. Tra i gesti labiali della LIS è possibile distinguere tre sottogruppi (Fontana, 2009):

– gesti che sono considerati iconici come per esempio GELATO, poiché la componente orale rispecchia effettivamente l'azione compiuta nell'atto di mangiare.
– gesti che sono considerati arbitrari poiché non esiste alcun collegamento significativo con la componente manuale. Ad esempio: IMPOSSIBILE come incapacità oggettiva con la componente orale "papapa";
– gesti che possono essere considerati metaforici come l'emissione d'aria che veicola l'informazione relativa al segno 'diminuzione dell'ansia'.

In tutti questi casi il gesto labiale non può essere omesso, poiché è parte integrante del segno.

Occorre qui precisare che le componenti orali non possono essere analizzate isolatamente ma in relazione ad altre componenti manuali e non manuali come l'espressione facciale (rigonfiamento delle guance aggrottamento della fronte, innalzamento dell'arcata sopracciliare ecc.) o una data postura. Ad esempio il segno COSA, oltre che dalla presenza di labializzazione, appare caratterizzato da una particolare espressione facciale e da un avanzamento del capo.

Le componenti orali nelle lingue dei segni presentano, dunque, una duplice natura: da una parte sono di natura gestuale e si strutturano lungo le coordinate dell'embodied simulation, dall'altra si collocano in continuità con la lingua orale, a cui ormai quasi tutti i sordi vengono educati sin dalla prima infanzia e presentano gradi variabili di discretezza. Occorre pertanto elaborare un approccio che consenta di codificare questa duplice natura senza prendere in prestito modelli strutturati sulle lingue vocali che finirebbero per offuscare la natura del fenomeno.

3. *Dall'embodied simulation al gesto labiale*

Il presente studio si basa sull'ipotesi che le componenti orali delle lingue dei segni svolgano una funzione paragonabile al gesto co-verbale nelle lingue vocali e che possa essere pertanto produttivo inquadrarle come gesto co-segnico (Fontana, 2008; 2009), cioè come una forma ulteriore di *visible bodily action* (Kendon, 2004). D'altra parte, se le lingue dei segni e le lingue vocali sono modalità equivalenti sotto il profilo funzionale, anche il gesto, co-verbale nelle LS e co-segnico nelle LV, pur avendo forme ontologicamente diverse, potrebbe funzionare in modi analoghi nelle due lingue. Le componenti orali delle lingue dei segni condividerebbero cioè le caratteristiche descritte per il gesto co-verbale:

> [...] there is a wide range of ways in which visible bodily actions are employed in the accomplishment of expression that, from a functional point of view, are similar to, or even the same as expression in spoken language. At times they are used in conjunction with spoken expressions, at other times as complements, supplements, substitutes or as alternatives to them. These are utterance uses of visible action and it is these uses that constitutes the domain of 'gesture' [...] (Kendon, 2004: 1-2)

In sintesi, un gesto è un'unità produttiva, con un grado di convenzionalità variabile, che veicola una dimensione di significato assente dal parlato (McNeill, 2005) che contribuisce al segno al livello iconico, metaforico e pragmatico (nei termini di parser per esempio). Infatti, sia nelle lingue dei segni che nelle lingue vocali gli articolatori sono coinvolti in altre attività come il bere o il respirare per gli organi fonatori (Fonagy, 1983) e tutta una serie di prassi quotidiane per le mani.

I suoni delle lingue orali possono veicolare altri messaggi attraverso l'uso di tratti paralinguistici (una certa intonazione) e di tratti linguistici (allungamento delle vocali, scelte prosodiche enfatiche, ad esempio). Tutto ciò determina un 'double econdage' che imprime nella materia significante il valore reale del significato codificato linguisticamente attraverso quella che Fònagy definisce "performance articulatoire expressive".

> L'encodage secondaire qui consiste dans une déformation sémiotique (significative) du message primaire serait donc plutot la tache d'un modulateur qui n'est pas incorporé à la grammaire. Selon le modèle proposé, tous les phonèmes passent nécessairement par le modulateur. Ceci revient à dire que tous les sons concrets sont expressifs. (Fònagy, 1983: 19)

Questa ulteriore codifica avviene ad opera di componenti sovrasegmentali quali ad esempio la prosodia, l'intonazione e il timbro che sfruttano le potenzialità insite nei singoli fonemi per evocare ulteriori valori significativi. Nella definizione del valore del 'double encodage', il modulatore ed il contesto appaiono centrali.

Le componenti orali si collocano al centro di due sistemi di valori: da una parte, infatti, si ricollegano ad un vocabolario motorio che si fonda su pattern ricorrenti attraverso cui il nostro corpo si muove nello spazio, manipola ed interagisce con gli oggetti, come nel caso del gesto orale che occorre con GELATO che riproduce l'azione di mangiare il gelato; dall'altra, fanno riferimento al sistema fonetico o fonosimbolico di una data lingua orale che viene percepito, tuttavia, attraverso un apparato sensoriale il cui senso vicario è la vista. A questo sistema si ricollegherebbero gesti labiali come 'bababa' per CANE-CHE-ABBAIA e 'papapa' per PISTOLA-CHE-SPARA che sembrano assoggettarsi, come tutti i fenomeni onomatopeici, al sistema fonologico di una lingua orale. La natura significativa delle componenti orali, tuttavia, emerge soltanto in rapporto all'unità segnica, confermando ancora una volta l'ipotesi che i due livelli siano strettamente interdipendenti. Il 'double encodage' è condivisibile soltanto nella misura in cui il linguaggio è 'embodied' al livello interno con il meccanismo dell'embodied simulation e al livello esterno attraverso i gesti (Margethis, Bergen, 2014). Vari studi (Glenberg, Kaschak, 2002; Glenberg et al., 2008; Kemmerer et al., 2008; Sato et al., 2008,) hanno evidenziato il coinvolgimento del sistema sensomotorio nella comprensione linguistica. Comprendere la frase 'Luca calcia la palla' significa attivare le aree del cervello collegate all'azione del piede anche se l'azione non viene effettivamente compiuta. Questo vale anche per le descrizioni linguistiche della percezione e dell'emozione. L'attivazione dei circuiti neurali in assenza dell'azione, della percezione o dell'emozione corrispondente è stata definita Embodied Simulation (Gallese, Sinigaglia, 2011).

Vari studi (vd. Marghetis and Bergen, 2014 per una rassegna) hanno mostrato che l'Embodied Simulation svolge una funzione sia nella produzione che nella comprensione del gesto. In altre parole, il meccanismo dell'Embodied Simulation implica, contemporaneamente, sia la dimensione dell'embodiment, sia la dimensione cognitiva poiché si tratta di un meccanismo che funziona al livello del cervello e del corpo (Cuccio, 2015).

Nelle componenti orali delle lingue dei segni sono pertanto riconoscibili diverse dimensioni, ciascuna delle quali è portatrice di un 'double encodage' che rinvia al meccanismo dell'embodied simulation da una parte e dall'altra all'attività motoria associata alla fonazione. Ad esempio, la lingua viene utilizzata in espressioni di disgusto, l'uso di occlusive si associa ad eventi improvvisi e l'emissione continua di un soffio ad un'azione continua. Per poter descrivere e codificare proficuamente le componenti orali, dunque, occorre prendere in considerazione le dimensioni che le costituiscono.

3.1 La ricerca

Alla luce delle nostre riflessioni epistemologiche sulla natura delle componenti orali, il presente studio si pone il triplice obiettivo di:

- modificare sulla base dell'analisi del corpus Fabbretti e delle considerazioni sopra esposte il modello Vogt-Svendsen (1983) per la classificazione delle componenti orali;
- testare la sostenibilità del modello Vogt-Svendsen adattato sulla base dell'analisi del corpus Fabbretti e delle considerazioni sopra esposte;
- verificare la presenza di configurazioni articolatorie ricorsive embodied perché legate ad attività prassiche che coinvolgono gli organi fonatori, utilizzati anche nella nutrizione e nella respirazione.

L'indagine si muove a partire dai dati del corpus Fabbretti (1997). Il corpus in oggetto è una raccolta di videoregistrazioni delle produzioni in segni di dieci sordi segnanti che raccontano in Lis una sequenza video della durata di 05:05 minuti del film muto *Muraglie* del 1931 con i noti attori comici Stanlio e Ollio. Tutti i soggetti sono segnanti nativi, in quanto hanno entrambi i genitori sordi, sono membri attivi della comunità dei sordi e utilizzano l'Italiano parlato e la LIS nella loro vita quotidiana. Tale corpus è tratto da un campione di dati più ampio raccolto con l'obiettivo di indagare la relazione tra la produzione scritta e flusso segnico nelle persone sorde (Fabbretti, 1997).

I partecipanti erano invitati a narrare in segni quanto avevano visto nel filmato ad un interlocutore segnante. In una prima fase, le storie videoregistrate sono state trascritte in glosse (in maiuscolo) a cui è stata aggiunta l'informazione sulla produzione labiale co-occorrente (in minuscolo).

3.2 Risultati

Da subito emerge che il sistema di scrittura alfabetico si rivela del tutto inadeguato alla descrizione di queste componenti che finiscono con l'essere trattate in riferimento alla lingua vocale e non descritte secondo le loro specificità.

La tabella che segue risponde al primo dei tre punti (modifiche al modello Vogt-Svendsen) e riporta tutte le componenti orali rintracciate nel corpus Fabbretti e catalogate a partire dal preesistente modello Vogt-Svendsen (1983), ripreso anche da Ajello, Mazzoni, Nicolai (2001) categorizzato secondo criteri fisiologici e articolatori.

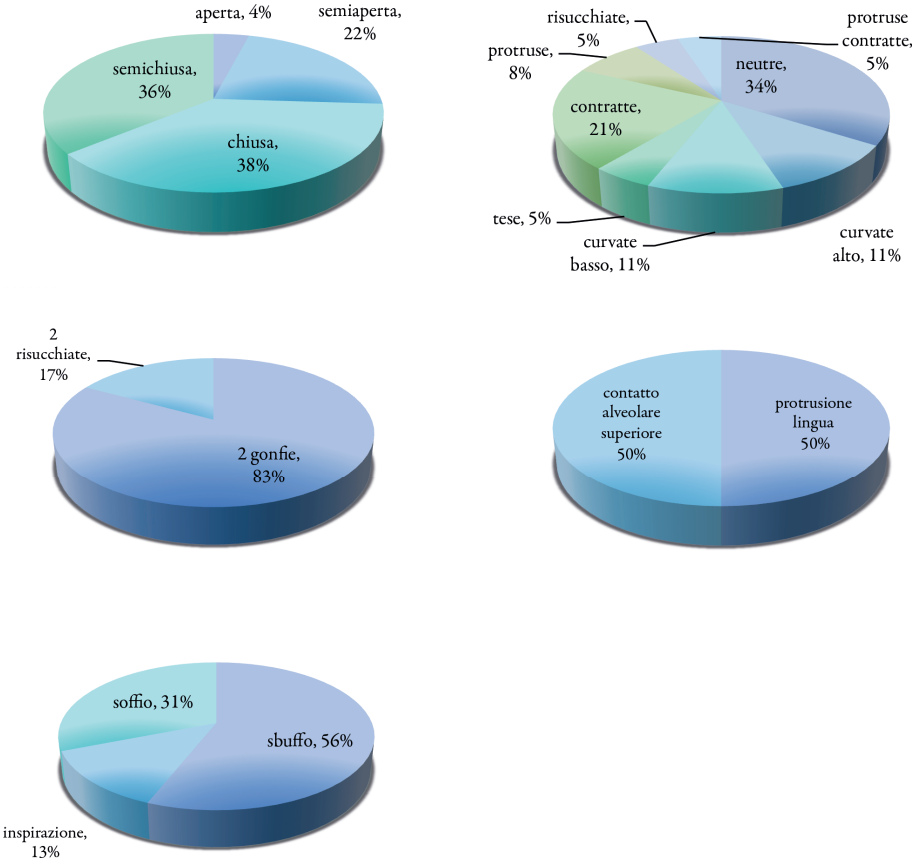
Figura 1 - *Adattamento del modello Vogt-Svendsen, ripreso da Ajello con proposte di aggiunte*

1- Apertura della mandibola	2- Posizione delle labbra	3- Posizione delle guance	4- Posizione della lingua	5- Uso dell'aria
a. chiusa	a. neutre	a. 2 gonfie	a. protrusione punta	a. sbuffo
b. semichiusa	b. curvate alto	b. dx gonfia	b. alveolare sup.	b. soffio
c. semiaperta	c. curvate basso	c. sx gonfia	c. alveolare inf.	c. inspirazione
d. aperta	d. tese	d. risucchiate	d. dentale sup.	d. vibrazione
	e. contratte		e. labiale inf.	
	f. protruse		f. labiale sup.	
	g. risucchiate			
	h. inferiore teso			
	i. inferiore su arcata dent superiore			
	l. tese all'insù			
	m. tese all'ingiù			
	n. protruse contratte			

Per rispondere al secondo quesito (cioè testare la sostenibilità del modello Vogt-Svendsen adattato) è stato necessario concentrarsi su quali e quante fossero le configurazioni articolatorie più frequenti e come si strutturassero in relazione al flusso segnico e destrutturassero rispetto al parlato). Dai nostri dati una preferenza dei punti di articolazione più visibili e maggiormente esterni quali le bilabiali, le labiodentali e le alveolari e, per quanto riguarda i modi di articolazione, le plosive e le fricative. Per ciò che concerne invece la struttura, abbiamo rilevato una preferenza, sia nei gesti labiali che nelle labializzazioni, per produzioni monosillabiche o bisillabiche, per l'uso di flussi d'aria prevalentemente egressivi e per l'uso marcato delle guance e della lingua.

Anche nel caso delle labializzazioni tendenzialmente non è l'intera parola della lingua vocale ad essere articolata, ma il gruppo fonetico rilevante per i sordi, cioè del gruppo che risulta più facilmente visibile nella lettura labiale, inoltre vengono omesse le parti non giudicate salienti dal punto di vista della decifrazione e vengono apportate alterazioni morfologiche per interfacciare la labializzazione con il flusso segnato (quali la velocità, il ritmo, la durata ecc.).

Figura 2 - *Nell'ordine: 1. Apertura mandibola, 2. Posizione labbra, 3. Posizione guance, 4. Posizione lingua, 5. Uso dell'aria*



Per rispondere al terzo quesito (cioè verificare la presenza di configurazioni articolatorie ricorsive embodied), abbiamo preso in considerazione tutti i segni che presentano un gesto labiale embodied ovvero iconico e collegato in un certo senso al significato (Pierce) e abbiamo osservato quanti segnanti vi facessero ricorso all'interno del corpus (numero riportato tra parentesi).

STANLIO E OLLIO (4) indicati come FACCIA-GROSSA E FACCIA MAGRA

SPARARE (10)

FOLLA CHE SCAPPA (10)

FOLLA CHE SI MUOVE (8)

MANGIARE (1)

ESSERE DISORIENTATO (4)

Le occorrenze embodied indicate veicolano informazioni complementari al flusso segnico e mostrano come la selezione delle configurazioni articolatorie co-occorrenti con il segno avvenga sulla base di criteri embodied, cioè legati ad un vocabolario motorio. In questa ottica nel nostro corpus abbiamo rilevato alcune configurazioni articolatorie preferenziali: rispetto al modo di articolazione, per esempio, l'apertura della mandibola, la fricazione e la plosione: l'aria viene fatta fluire nella rappresentazione della folla che scappa per veicolare sul piano del suono un'informazione visiva della folla concitata che fugge, o ancora l'aria viene bloccata nel segno "sparare" per veicolare l'informazione relativa al suono caratteristico della sparatoria. Per quanto riguarda i punti di articolazione, è evidente che siano privilegiati sia nei gesti labiali che nella labializzazione quelli collocati nella parte anteriore dell'apparto fonoarticolatorio e cioè suoni labializzati, labiodentali e alveolari.

4. *Conclusioni*

L'analisi quantitativa e qualitativa conferma la centralità delle componenti orali nella Lingua dei segni italiana e mostra la necessità di un approccio che inquadri l'unità segnica nelle sue dimensioni multimodali. Inoltre dai nostri dati sembrerebbe che, sul piano articolatorio, labializzazioni e gesti labiali si comportino in modo analogo poiché luoghi e punti di articolazioni rilevati sono ricorrenti in entrambe le categorie.

Il modello Vogt-Svendsen applicato ai nostri dati si è rivelato produttivo, mostrando che le diverse dimensioni articolatorie dei gesti labiali e delle labializzazioni finiscono con l'essere completamente annullate da codifiche basate sulla scrittura o su glosse. Per comprendere le componenti orali, è necessario, dunque, partire da una prospettiva gestuale intendendole come visible bodily action nella loro duplice funzione sia al livello cognitivo, nel modellare il discorso, attraverso il meccanismo dell'embodied simulation, sia nella comunicazione integrando le informazioni veicolate nel flusso segnico.

La modalità linguistica di espressione (visivo-gestuale anziché acustico-vocale) determina una forma diversa del gesto e sembra incidere sulla natura dell'equilibrio tra le due componenti (Fontana, 2008). Seguendo l'ipotesi formulata da Johnson (2007), la nostra idea è quella di considerare la totalità del segno come un cluster articolatorio che comprende sia il piano gestuale che quello non manuale.

Ulteriori analisi saranno necessarie per avvalorare queste tesi sulla base di un ampliamento del campione più rappresentativo della popolazione sorda su cui applicare il modello fin qui teorizzato.

Bibliografia

Ajello, R., Mazzoni, L. & Nicolai, F. (1988). Gesti linguistici: la labializzazione in LIS. In *Quaderni della sezione di glottologia e linguistica, Università "G. D'Annunzio"*, Chieti, 5-45.

Ajello, R., Mazzoni, L. & Nicolai, F. (2001). Linguistic gestures: Mouthing in Italian Sign Languages (LIS). In Boyes Braem, P., Sutton-Spence, R. (Eds.) (2001), *The Hands are the Head of the Mouth.* Hamburg: Signum Verlaeg, 231-246.

Boyes Braem, P., Sutton-Spence, R. (Eds.) (2001). *The Hands are the Head of the Mouth*. Hamburg: Signum Verlaeg.

Cuccio, V. (2015). The notion of representation and the brain. In *Phenomenology and Mind,* 7, 247, 258.

Cuxac, C., Antinoro Pizzuto, E. (2010). Emergence, norme et variation dans les langues des signes: vers une redéfinition notionnelle. In *Langage et Societé,* 131, 37-53.

Ebbinghaus, H., Hessmann, J. (2001). Sign language as multidimensional communication: Why manual signs, mouthings, and mouth gestures are three different things. In Boyes Braem, P., Sutton-Spence, R. (Eds.) (2001), *The Hands are the Head of the Mouth*. Hamburg: Signum Verlaeg, 133-151.

Fabbretti, D. (1983). *Scrivere e segnare: la costruzione del discorso nell'italiano scritto e nella lingua dei segni delle persone sorde*. Tesi di dottorato di ricerca, Roma, Università Sapienza, Dipartimento di Psicologia dei processi di sviluppo e socializzazione.

Fonagy, I. (1983). *La vive voix: Essays de psycho-acoustic*. Paris: Payot.

Fontana, S. (2008). "Mouth Actions as Gestures in Sign Language". In Kendon, A., Russo Cardona, T. (Eds.), *Dimensions of Gestures: Special Issue of Gesture*, 8:1. Amsterdam: John Benjamins Publishing Company, 104-123.

Fontana, S., Fabbretti, D. (2000). "Classificazione e Analisi delle forme labiali della LIS in storie elicitate". In Bagnara, C., Chiappini, G. et al., *Viaggio nella città invisibile Pisa*. Edizioni Del Cerro, 103-111.

Fontana, S. (2009). *Linguaggio e Multimodalità Oralità e gestualità nella lingua dei segni italiana e nella lingua vocale*. Pisa: Edizioni ETS.

Fontana, S., Raniolo, E. (2015). Interazioni tra oralità e unità segniche: uno studio sulle labializzazioni nella Lingua dei Segni Italiana. In Schneider, G.M., Janner, M.C. & Bénédicte, E. (Eds.), *Voix et silence – Voce e silenzio – Voces y silencio. VII. Dies Romanicus Turicensis.* Avec la coedition de – Sulla coeditizione di – Bajo la coedición de Albizu, C., Dekens, J., Della Costanza, M.A., Dubois, C., Frei, V., Meisner, F.C., Schwarzenbach, M. Berna: Peter Lang, 241-258.

Gallese, V., Sinigaglia, C. (2011). What is so special about embodied simulation? In *Trend Cogn Sci*, 15 (11), 512-9.

Glemberg, A.M., Kaschak, M.P. (2002). Grounding language in action. In *Psycon Bull Rev*, 9 (3), 558-65.

Glemberg, A.M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D. & Buccino, G. (2008). Processing abstract language modulates motor system activity. In *Q.J. Exp Psychol (Hove)*, 61 (6), 905-19.

Hoemberger, N., Happ, D. (2001). The linguistic primacy of signs and mouth gestures over mouthings; Evidence from language production in German Sign Language (DGS). In Boyes Braem, P., Sutton-Spence, R. (Eds.) (2001), *The Hands are the Head of the Mouth*. Hamburg: Signum Verlaeg, 153-190.

Johnson, M. (2007). *The meaning of the body*. Chicago: Chicago University Press.

Kemmerer, D., Castillo, J.G., Talavage, T., Patterson, S. & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: evidence from fMRI. In *Brain Lang,* 107 (1), 16-43.

Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Marghetis, T., Bergen, B. (2014). Embodied meaning, inside and out: The coupling of gesture and mental simulation. In Müller, C. et al. (Eds.), *Body-Language-Comunication.* New York: Mouton de Gruyter, 9-40.

McNeill, D., (2005). *Gesture and Thought.* Chicago: Chicago University Press.

Pizzuto, E. (2003). Coarticolazione e multimodalità nelle lingue dei segni: dati e prospettive di ricerca dallo studio della lingua dei segni italiana (LIS). In Marotta, G., Nocchi, N. (Eds.), *La coarticolazione: Atti della XIII Giornata di Fonetica Sperimentale.* Pisa: Edizioni ETS, 59-77.

Sato, M., Mengarelli, M., Riggio, L., Gallese, V. & Buccino, G. (2008). Task related modulation of the motor system during language processing. In *Brain Lang*, 105 (2), 83-90.

Vogt-Svendsen, M. (1983). Positions and movements of the mouth in Norwegian Sign Language (NSL). In Kyle, J., Woll, B. (Eds.), *Language in sign.* London: Croom Helm, 85-96.

Vogt-Svendsen, M. (2001). A comparison of mouth gestures and mouthings in Norwegian Sign Language (NSL). In Boyes Braem, P., Sutton-Spence, R. (Eds.) (2001), *The Hands are the Head of the Mouth.* Hamburg: Signum Verlaeg, 9-40.

DOI: 10.17469/O2101AISV000024

MARIA ROCCAFORTE

Componenti orali e riconoscibilità nella Lingua dei Segni Italiana (LIS)

In contrast to the single-articulatory system of spoken languages, sign languages employ multiple articulators, including manual articulators and oral components. Oral components includes: voiced or voiceless parts of words of spoken language and voiced or voiceless sounds without any links to spoken language. In this paper, we present evidence based on the analysis of an Italian Sign Language (LIS) corpus gathered through two different requests: a naming task and a spontaneous production task. Analyses were made on 2.143 signs (naming task) and 68 minutes (spontaneous production task). The aims were to verify the statistical frequency of the oral components in signs in naming production task and in spontaneous production task and to test if the presence of mouthing facilitate the recognition of the sign in deaf signers.

1. *Le componenti orali della Lingua dei Segni*

Non tutti i sordi conoscono e usano la lingua dei segni, ma tutti i sordi sono esposti alla lingua verbale nazionale sia in contesto guidato che spontaneo, questo perché alcune famiglie decidono di non far conoscere al proprio figlio sordo la lingua dei segni, ma al contrario, non esistono o sono estremamente rare famiglie che decidano di educare il proprio figlio solo alla lingua dei segni (Russo, 2007). Stando così le cose si potrebbe affermare che, contrariamente ai sordi educati solo oralmente, ovvero solo alla lingua verbale, i sordi segnanti, seppur con diversi livelli di competenza linguistica nelle due lingue, possano essere considerati bilingui (Fabbretti, Tomasuolo, 2006).

Si tratta, tuttavia, di un bilinguismo speciale dovuto alla particolare situazione sociolinguistica della comunità sorda che di fatto vive in costante contatto con la comunità udente e possiede anche la lingua verbale nel suo repertorio linguistico (Fontana, 2009). In questi numerosissimi casi la competenza linguistica nella lingua verbale e quella nella lingua dei segni coesistono, facendo sì che i due sistemi, ontologicamente così diversi, entrino in contatto e si fondano (Ajello et al., 1998).

Il fenomeno di cui ci occuperemo in questo studio è probabilmente la manifestazione più evidente di questa contaminazione linguistica tra lingua dei segni italiana (LIS) e l'Italiano lingua verbale, che viene definito dalla letteratura internazionale con il termine di "mouthing" (tradotto in italiano "labializzazioni[1]") e consiste

[1] Il termine labializzazione non va confuso con l'accezione che ha in ambito fonologico di arrotondamento delle labbra, ma come traduzione del termine originario nella letteratura internazionale

nell'articolare con la bocca frammenti di lingua vocale con o senza vocalizzazione (Boyes Braem, Sutton-Spence, 2001).

In passato, in Italia, il fenomeno del mouthing è stato lungamente trascurato. A partire dal Congresso di Milano che nel 1880 apriva le porte a una forte tradizione oralista nel nostro paese e bandiva l'uso della LIS nell'educazione dei sordi, i primi studi sulla lingua dei segni, sviluppatisi alla fine degli anni '70, erano tutti vòlti a dimostrare che la LIS non fosse solo uno strumento in più nell'educazione dei sordi, ma anche una lingua storico-naturale, degna di essere studiata come tutte le altre lingue (Russo, 2007). In questa prospettiva le componenti orali nelle lingue dei segni fornivano un supporto proveniente dalla lingua verbale e potevano far ritenere che, da sola, la LIS non fosse in grado di veicolare contenuti di ogni tipo. D'altra parte, però, proprio questa tradizione oralista, conseguenza del veto sulla lingua dei segni, ha portato molti sordi ad utilizzare entrambe le lingue in circostanze diverse: la lingua verbale in contesti ufficiali e la lingua dei segni in contesti familiari.

Oggi gli studi sulle lingue dei Segni vanno sempre più in direzione di una osservazione linguistica oggettiva, scevra da pregiudizi e stilemi sociopolitici, e invece orientata a rilevare quelle caratteristiche e peculiarità che la rendono un fenomeno linguistico unico ed estremamente interessante.

2. *Obiettivi e Metodo*

Se è vero che, come per le lingue verbali, nella LIS è soprattutto il contesto che aiuta a distinguere tra segni identici o molto simili, è vero anche che la componente labiale è stata spesso descritta come elemento che conferisce caratteristiche distintive nelle lingue dei segni. Mentre, però nelle lingue verbali abbiamo un gran numero di studi che supportano questa ipotesi, lo stesso non si può dire per la Lingua dei Segni. Per questo il presente lavoro si pone due obiettivi che saranno sviluppati individualmente. Il primo consiste nell'osservare all'interno di un corpus in lingua dei segni, la frequenza del fenomeno del mouthing sia un compito di naming che un compito di produzione spontanea, il secondo consiste nello sperimentare se e in quale misura la presenza del mouthing faciliti la riconoscibilità del segno in un gruppo di informanti sordi segnanti per dimostrare se sia ragionevolmente attribuibile al mouthing una funzione di disambiguazione.

Per raggiungere il primo obiettivo è stato utilizzato un corpus[2] che raccoglie le videoregistrazioni di 30 sordi segnanti da cinque città italiane (6 da Genova, 6 da Gorizia, 6 da Roma, 6 da Foggia e 6 da Caltanissetta) a cui sono state mostrate 108 immagini di alimenti con la richiesta di produrre il segno corrispondente (compito di naming). Infine è stato chiesto a ciascun segnante di raccontare cosa avessero mangiato nel corso della giornata precedente (compito di produzione spontanea).

'mouthing'.

[2] Corpus Perea Costa raccolto dalla ricercatrice Maria de la Luz Perea Costa nell'ambito di una borsa Marie Curie e conservato nell'archivio dell'ISTC-CNR.

Le immagini fanno tutte capo a un'area semantica particolarmente soggetta a fenomeni di omonimia: quella alimentare. Immagini di cibi, dunque, mostrati a sordi provenienti da diverse parti d'Italia, videoregistrati e conservati su dvd.

Le analisi sono state effettuate su un totale di 2.143 segni (compito di naming) e 68 minuti totali di registrazione (compito di produzione spontanea).

Per raggiungere il secondo obiettivo sono stati estratti dal Corpus Perea Costa i video di 24 segni standard (scelti equamente tra le diverse città d'Italia) e sono stati sottoposti random a 36 sordi ai quali è stato chiesto di indicare l'immagine corrispondente al segno visto nel video.

Nell'analizzare il corpus si è notato che i segni che si presentavano erano fondamentalmente di due tipi:

– segni molto simili sul piano manuale ma non nel mouthing (segni potenzialmente ambigui);
– segni molto diversi sia sul piano manuale che sul piano del mouthing (segni potenzialmente riconoscibili).

Sono stati quindi selezionati dal corpus:

– 12 segni potenzialmente ambigui a coppie (Es. farina e zucchero Fig. 1 e burro e marmellata Fig. 2);
– 12 segni potenzialmente riconoscibili (Es. Fig. 3 arancia, mela, fungo e cioccolato).

Figura 1 - *Farina e zucchero*

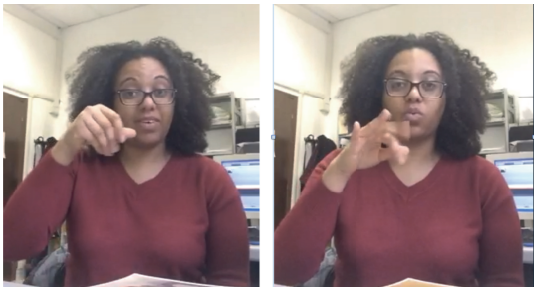


Figura 2 - *Burro e marmellata*

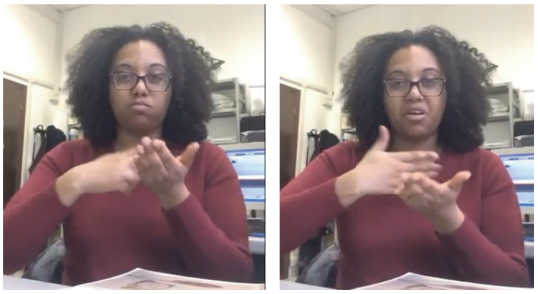


Figura 3 - *Arancia, mela, fungo e cioccolato*

Quando parliamo di segni 'potenzialmente riconoscibili' e segni potenzialmente ambigui facciamo riferimento ad una somiglianza fonologica che, nelle lingue dei segni, si traduce in una somiglianza cheremica. I segni 'farina' e 'zucchero', per esempio possono essere considerati una coppia minima, in quanto tutte le unità che li compongono (luogo, orientamento, movimento) sono identici, tranne una (la configurazione, ovvero la forma che assume la mano). Possono quindi essere considerati una coppia minima secondo il modello proposto da William Stokoe (1960), al pari della coppia minima in Italiano 'cane' e 'pane'. Consideriamo invece 'segni potenzialmente riconoscibili' quei segni che si distinguono per tutte le unità costituenti

I video dei 24 segni proposti hanno però una caratteristica: in tutti i segni è presente la componente labiale, ma, all'informante veniva dapprima mostrato il video del segno con l'area corrispondente alla bocca del segnante schermata (vd. Fig. 4) per impedirgli di riconoscere il segno attraverso la lettura labiale, e, solo successivamente, il video del segno gli veniva mostrato con la componente labiale perfettamente visibile.

Figura 4 - *Esempio di video con la bocca del segnante oscurata*

A questo punto i 48 segni sono stati sottoposti random a 12 sordi segnanti adulti ai quali è stato chiesto di indicare l'immagine dell'alimento corrispondente al segno visto nel video, oppure rispondere 'non so', indicando l'immagine del punto interrogativo (vd. Fig. 5). Le quattro immagini di cibi sono state scelte secondo i criteri che seguono:

– Scelta corretta.
– Immagine visivamente affine[3].
– Immagine fonologicamente affine.
– Immagine non affine né visivamente né fonologicamente.
– Non so.

Figura 5 - *Videata con la quale è stato chiesto agli informanti di selezionare la figura corrispondente al significato del segno mostrato nel video*

3. *Risultati*

I risultati della prima indagine hanno dato la misura della frequenza del fenomeno sia nel compito di naming che in quello di produzione spontanea, quelli della seconda parte dello studio hanno rivelato l'indice di riconoscimento del segno da parte dei sordi segnanti in relazione all'oscuramento o meno della bocca. Dunque con i primi risultati è possibile dare una risposta relativamente alla portata del fenomeno del mouthing e alla presenza di emissione di suoni in contemporanea al segno, supportati da un corpus piuttosto consistente e particolarmente interessante per i motivi sopra esposti (§ 2); con i risultati della prova sperimentale è possibile dare una risposta, per quanto ancora parziale e che necessita di un ampliamento del campione, relativamente alla funzione di discriminazione del mouthing in parole fuori contesto.

3.1 Risultati dell'analisi su Corpus

I dati sull'analisi del corpus Perea Costa dimostrano che la presenza del mouthing è, nel compito di naming, significativamente più alta rispetto alla presenza del mouthing in produzione spontanea (vd. Fig. 6 e 7).

[3] Con fonologico si intende in questo caso il cherema del movimento "sbriciolamento".

Figura 6 - *Percentuale della presenza e assenza di mouthing nel compito di naming*

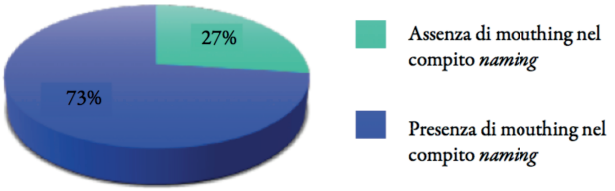
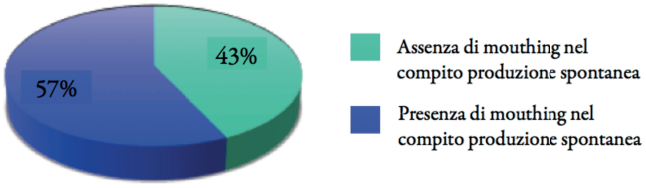
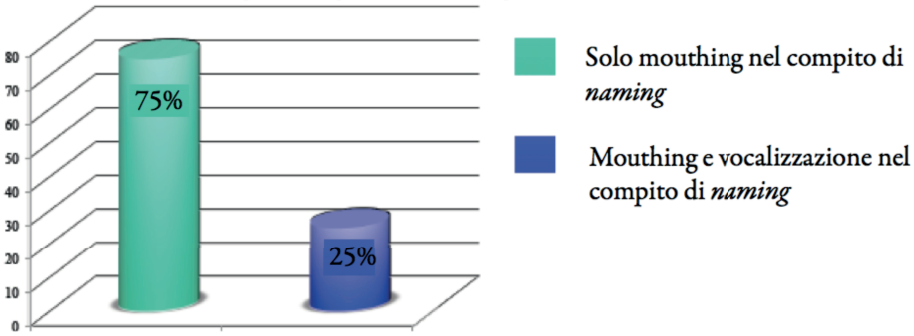


Figura 7 - *Percentuale della presenza e assenza di mouthing nel compito di produzione spontanea*



I dati mostrano inoltre che solo nel 25% dei casi il mouthing è accompagnato da vocalizzazion, e ovvero da emissione di suoni contestualmente alla componente orale (vd. Fig. 8).

Figura 8 - *Confronto tra la percentuale di segni accompagnati da mouthing senza vocalizzazione e con vocalizzazione*



3.2 Risultati della prova sperimentale

I risultati della prova sperimentale relativamente alla funzione di disambiguazione della componente orale mostrano che nessuno dei 24 segni presentati senza oscuramento della bocca hanno ingenerato errori nei 12 segnanti e che nessuno dei 12 segni considerati potenzialmente riconoscibili hanno ingenerato errori nei 12 segnanti. Gli errori di riconoscimento si presentano solo con segni potenzialmente ambigui e solo a bocca schermata, nella percentuale del 19% (vd. Fig. 9) e tra questi: il 72% delle risposte è 'non so', il 20% degli errori è di natura fonologica, il restante 8% ha scelto l'item visivamente affine (vd. Fig. 10).

Nessuno ha scelto il distrattore né fonologicamente, né visivamente affine.

Figura 9 - *Percentuale degli errori di riconoscimento su segni potenzialmente ambigui con bocca schermata*

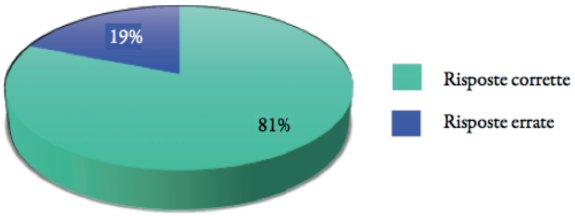
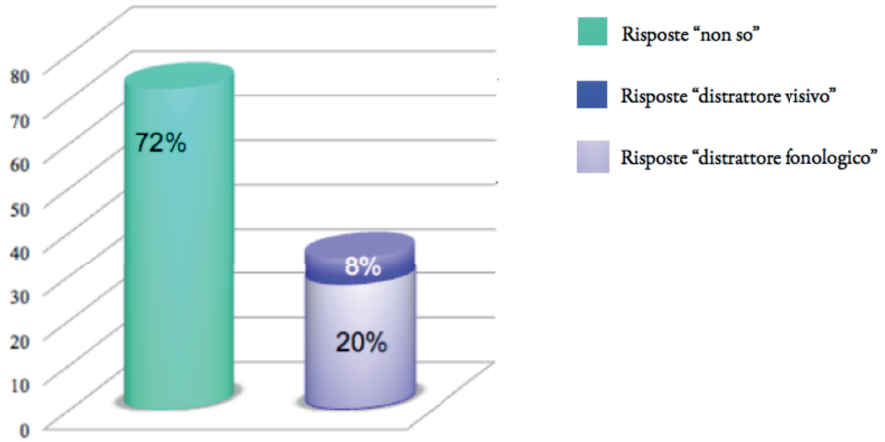


Figura 10 - *Classificazione degli errori sui segni potenzialmente ambigui con bocca schermata*



4. *Conclusioni e discussione*

Alla luce dei risultati sul corpus e sul pilot di prova sperimentale possiamo affermare che nei segni relativi all'alimentazione la percentuale di incidenza della componente orale è decisamente alta, molto più nel compito di naming che nella produzione spontanea, il che potrebbe essere collegato all'intenzione comunicativa del segnante che nell'atto di nominare un alimento vuole, con il mouthing, enfatizzarne la nominalizzazione.

La funzione di disambiguazione si è rivelata solo in piccola percentuale e solo nelle parole fuori contesto, nonostante le scelte siano state tutte rivolte a rendere arduo il compito di riconoscimento: sia per la scelta di segni ambigui, sia per l'oscuramento della zona della bocca, sia per la scelta delle opzioni tra cui i segnanti hanno dovuto selezionare quella corretta. I risultati sin qui raggiunti, relativi alle particolari alterazioni della qualità e della quantità in relazione alle variabili esaminate, cioè alla natura dell'input, sono prove a sostegno della ipotesi che le labializzazioni possano essere considerate una forma di gesti co-segnici paragonabili ai gesti co-verbali, ai marcatori discorsivi o alle caratteristiche soprasegmentali che accompagnano le lingue verbali e che svolgono funzioni di natura semantica, pragmatica e metadiscorsiva assicurando una maggiore stabilità e intellegibilità nella comunicazione.

Potranno in futuro essere ampliati estendendo il campione le catalogandolo in base all'età di acquisizione/apprendimento della lingua dei segni.

Il fenomeno del mouthing, dunque, merita senz'altro di essere ulteriormente indagato sia perché la sua funzione resta, ancora, per buona parte, da definire sia per la rilevanza del fenomeno ai fini di ricerche e di indagini nel campo delle tecnologie basate su avatar e facce parlanti e destinate a un'utenza sorda.

Bibliografia

Ajello, R., Mazzoni, L. & Nicolai, F. (1988). Gesti linguistici: la labializzazione in LIS. In *Quaderni della sezione di glottologia e linguistica, Università "G. D'Annunzio"*, Chieti, 5-45.

Boyes Braem, P., Sutton-Spence, R. (Eds.) (2011). *The Hands are the Head of the Mouth*. Hamburg: Signum.

Caselli, M.C., Maragna, S., Pagliari Rampelli, L.C. & Volterra, V. (1994). *Linguaggio e Sordità. Parole e segni nell'educazione dei sordi*. Firenze: La Nuova Italia.

Fabbretti, D., Tomasuolo, E. (2006). *Scrittura e Sordità. Parole e segni nell'educazione dei sordi*. Roma: Carocci.

Fontana, S. (2009). *Linguaggio e Multimodalità Oralità e gestualità nella lingua dei segni italiana e nella lingua vocale*. Pisa: Edizioni ETS.

Pizzuto, E. (2003). Coarticolazione e multimodalità nelle lingue dei segni: dati e prospettive di ricerca dallo studio della lingua dei segni italiana (LIS). In Marotta, G., Nocchi, N. (Eds.), *La coarticolazione: Atti della XIII Giornata di Fonetica Sperimentale*. Pisa: Edizioni ETS, 59-77.

Maragna, S., Roccaforte, M. & Tomasuolo, E. (2014). *Una didattica innovativa per l'apprendente sordo.* Roma: Franco Angeli.

Russo Cardona, T., Volterra, V. (2007). *Le lingue dei segni*. Roma: Carocci.

Stokoe, W. (1960). *Sign Language Structure*. Silver Spring, Maryland: Linstok Press.

DOI: 10.17469/O2101AISV000025

PAOLO BRAVI

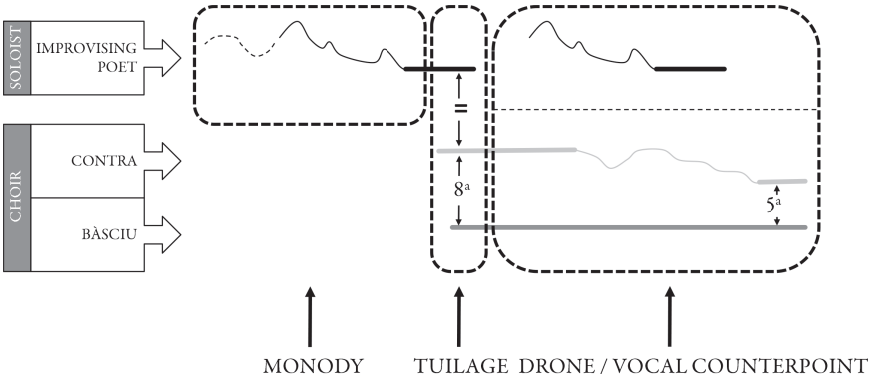
The Construction of the Individual Style of Singing in the Sung Poetry of Southern Sardinia: Reflections upon a Critical Case

In this paper I describe a case in which the very first steps in the construction of the singing style of a young Sardinian improviser encountered some difficulties and, in particular, caused some problems for the two-part accompanying choir during a poetry contest. I also report the protagonists' comments and opinions on the topic, analyze musical aspects of the performance and propose some keys for the interpretation of the observed anomalies. These keys are based on both the peculiar situation of the improviser, who emigrated from Sardinia when he was 24, and of his native village, Burcei, which shows influences from the Barbagia area especially as far as poetical styles are concerned.

1. *Introduction*

Southern Sardinia is host to a long-standing tradition of contests of improvised poetry, known as *cantadas campidanesas.* On these occasions, a number of *cantadoris* (improvising poets), usually four, confront each other in poetry using two different verse types: the *mutetu longu* (which is longer and more difficult and used in the first part of the *cantada*) and the *versu* (which is shorter and easier and adopted in the final part of the contest). These performances are usually held in summer, on occasion of the Festivals organized to celebrate the Patron Saints of the various villages. The improvisers appearing onstage in these performances are 'professionals': they have another job, but receive a significant reward for their participation in the poetry contest. When they perform the complex metrical structure of the *mutetu longu* they are accompanied through non-measured singing by a two-part vocal choir made up of a *bàsciu* ('bass') and a *contra* (Bravi, 2010: 317-332). The former sings a drone one octave below the *tonus finalis* of the *cantadori* at the end of each section of the *mutetu longu* and is characterized by a typical guttural emission that gives rise to the acoustic phenomenon of 'period doubling' (Bailly, Heinrich & Pelorson, 2010). The *contra,* a part that intervenes a little before the *bàsciu* in the polyphonic part of the *mutetu,* sings a melody that usually starts at the same pitch as the *tonus finalis* of the *cantadori* and then, after a short melodic movement, ends one fifth above the *bàsciu's* note (figure 1).

Figure 1 - *Schematic graphical representation of the basic musical structure of the* a mutetus longus *singing style. To be noted in particular, the standard tonal relations among the three voices in the central section, where the polyphonic section starts in the form of a "tuilage"*[1] *between the improvising poet on the one hand and the* contra *and the* bàsciu *on the other*



2. *Case-Study: the first trial of the improviser Severino Monni*

It is common usage that a poet who aspires to become a '*cantadori*', i.e. a (semi) professional improviser, is first "tested" by other professionals during an informal poetry contest. The case-study here considered is that of the Campidanese improviser Severino Monni's (from now on: SM) first trial ('*collaudu*'). He sings at a private poetry contest together with three professional *cantadoris* (Antonello Orrù, Roberto Zuncheddu, Daniele Filia) and with the professional accompanying singers Filippo Orrù (*contra*) and Rossano Cardia (*bàsciu*). The performance was held in Serdiana (CA) in Southern Sardinia on 24.04.2012, in the presence of a few listeners, mostly SM's relatives and aficionados of Sardinian improvised poetry, and was documented by a home video recording. In this *cantada* one can hear that, when SM is singing, the accompanying choir (made up of two young semi-professional singers) rarely manages to start its intervention in the right tonality (*crai*, 'key') (Figure 2).

Indeed, as regards the music, the poetic trial appeared to be quite an ordeal for the new would-be *cantadori*. Nevertheless, his expert colleagues generally appreciated the quality of his improvisation, and he was acknowledged as being able to take part in public exhibitions in the future. As a matter of fact, this actually did come about in the following years. Then again, as SM himself told me in one of our interviews, this is a common problem for improvising poets who are starting out on their careers:

[1] "Tuilage" indicates the temporary overlapping of parts that may be realized during the alternation between two independent musical parts (cfr. Agamennone, 1998: 27). In the case of the *a mutetus longus* sung poetry, this happens when the vocal parts of the *contra* (before) and the *bàsciu* (later) start accompanying the voice of the improviser, who can then – but not always – perform another line, singing above the two-part choir (Mossa, 1999).

specialmente all'inizio, un improvvisatore si concentra sul testo di quello che deve dire, quindi la voce la lascia dov'è e questo, visto dal punto di vista de *sa contra*, è un grosso problema [tr.: 'especially in the beginnings, an improviser concentrates on the text that he has to say, therefore he does not give attention to his voice and this, from *sa contra*'s point of view, is a big problem'] (Monni, 2014).

Figure 2 - *Spectrograms (window length: 0.05 s.) of two example cases of* tuilage *between the improviser (SM)'s* tonus finalis *and the* contra's tonus initialis*. On the left, we can see the spectrogram showing the only 'correct' tonal juncture in the performance (*mutetu *n. 6, section 1), i.e. the improviser's* finalis *has the same* f_o *as the* contra's initialis *reached through a glissando; on the right, the spectrogram displays one 'incorrect' tonal juncture, with the two notes at different pitch levels performed by the improviser and the* contra *singer*

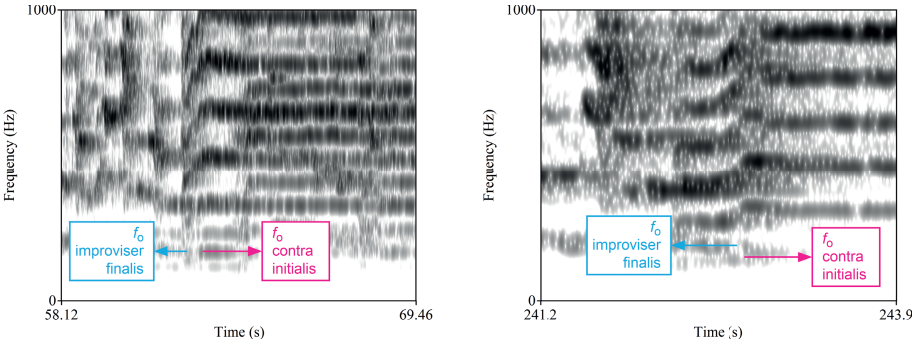


Figure 3 - *The cantadori Severino Monni (left) and the contra singer Filippo Urru, together with Paolo Bravi, (right) during the interviews*

The topic was discussed with the two protagonists and examined by means of a 'bi-focal' approach (Bravi, 2012): on the one hand, I interviewed the poet SM and the *contra* singer Filippo Urru in order to record their memories about that event and to capture their perspectives and opinions on the issue, while, on the other hand, I carried out a music analysis of some aspects relevant to the tonal disorders emerging in the performance.

With regard to the musical problems that emerged during this performance, three main aspects have been considered here. These were both analyzed and dis-

cussed with the singers in order to provide some prospective explanations of the tonal disorientation between the improviser and the two-part accompanying choir that occurred during the performance in question.

2.1 Lowering of the tonal centre

The *contra* always starts singing (*tonus initialis*) at a lower pitch with respect to the last note (*tonus finalis*) sung by the improviser SM.

Figure 4 - *Differences between the* finalis *(f) of SM's melodic stanzas and the* initialis *(i) of the* contra *intervention. Note that, apart from the last* mutetu *(n. 6), not all the stanzas (*torradas*) of each* mutetu longu *were recorded, so a different number of pairs (*cantadori*'s* finalis */* contra*'s* initialis*) are reported in the six* mutetus *sung by SM during the performance*[2]

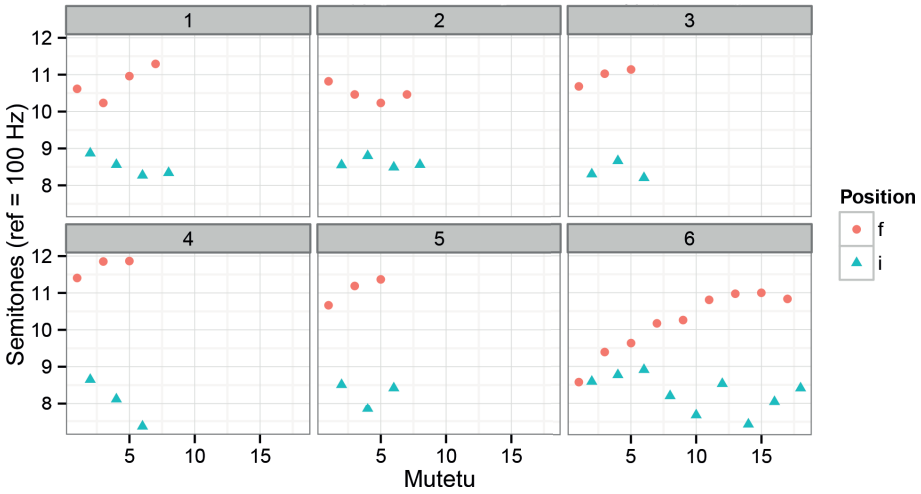


Figure 4 shows that, with the exception of the first intervention of *mutetu* n. 6, there is always a low melodic jump (ranging from -0.63 st to -4.51 st) between the two notes which, according to the music rules of the genre, should instead have the same pitch.

[2] This partial and 'selective' recording is due to the fact that, according to specific rules and based on the mechanism of *retrogradatio* (in Sardinian-Campidanese: *arretroga*), many textual repetitions are present in *a mutetus longus* poetical contests (cfr. Zedda, 2008: XXII-XXVIII). These repetitions are not usually documented in amateur recordings of this kind of improvised poetry (Lutzu, 2012).

Figure 5 - *The inverse relationship between SM (the* cantadori*)'s final and the* contra's *initial pitch*

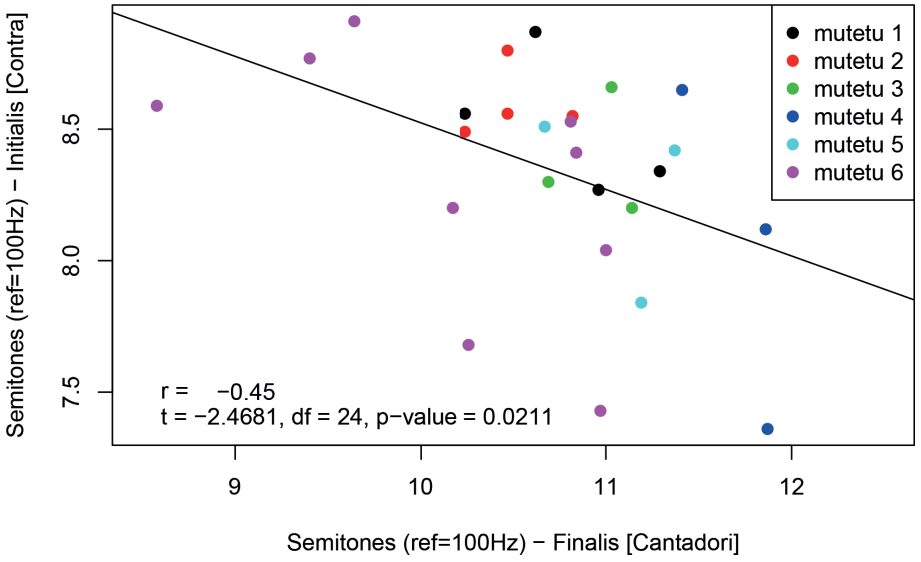
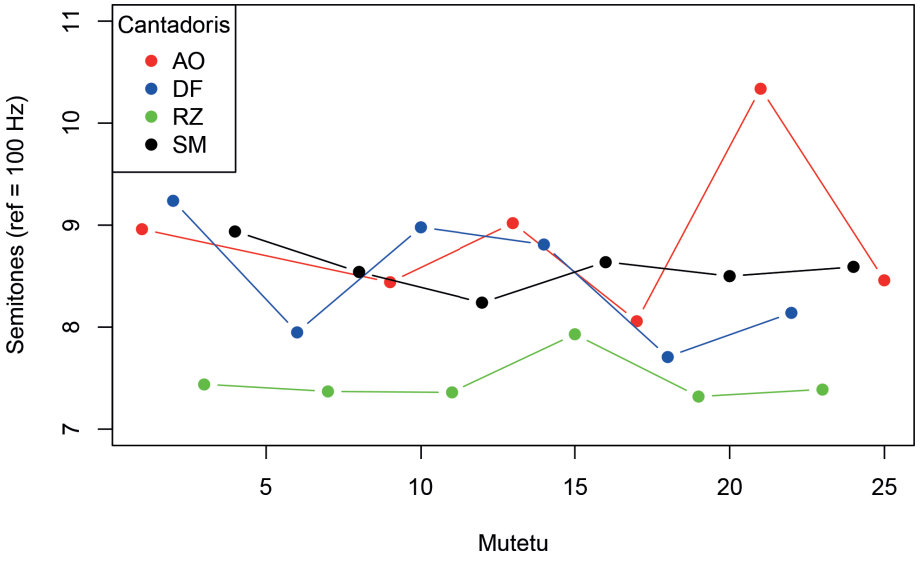


Figure 6 - *The improviser Roberto Zuncheddu (RZ)'s singing usually has a low and quite stable key*



In general, it can be observed that, with respect to a frequent rise of SM's *toni finales* within each *mutetu*, the *contra*'s *toni initiales* tend to be less unstable and, in some cases, show a pitch-lowering tendency. On the whole, these opposite tendencies have a moderate correlation (Figure 5), which fits well with what the *contra* singer Filippo Urru states, i.e. that he likes improvisers singing in a low tonality and who are able to maintain the established key throughout the entire performance. In fact, Roberto Zuncheddu, the improviser who is especially appreciated by the *contra*

Filippo Urru, uses a quite stable and low tonality in this performance, with a tonal center around 155 Hz (~E^b3; Figure 6):

> Io con Roberto [Zuncheddu] mi trovo benissimo. Mi trovo un po' male con Marco quando usa troppo l'acuto perché Marco, a differenza degli altri [improvvisatori], parte con quest'acuto, e quello ti fa stonare. Però poi anche lui si rende conto e allora abbassa il tono e va cadenzando come gli altri [improvvisatori] [tr.: 'I get on very well with Roberto [Zuncheddu]. I don't get on with Marco [Melis] when he uses a high pitch because Marco, unlike the other [improvisers], starts with a high note, and that makes you go out of tune. But then he realizes and therefore he lowers the pitch and keeps on singing as the others [improvisers]'] (Urru, 2014).

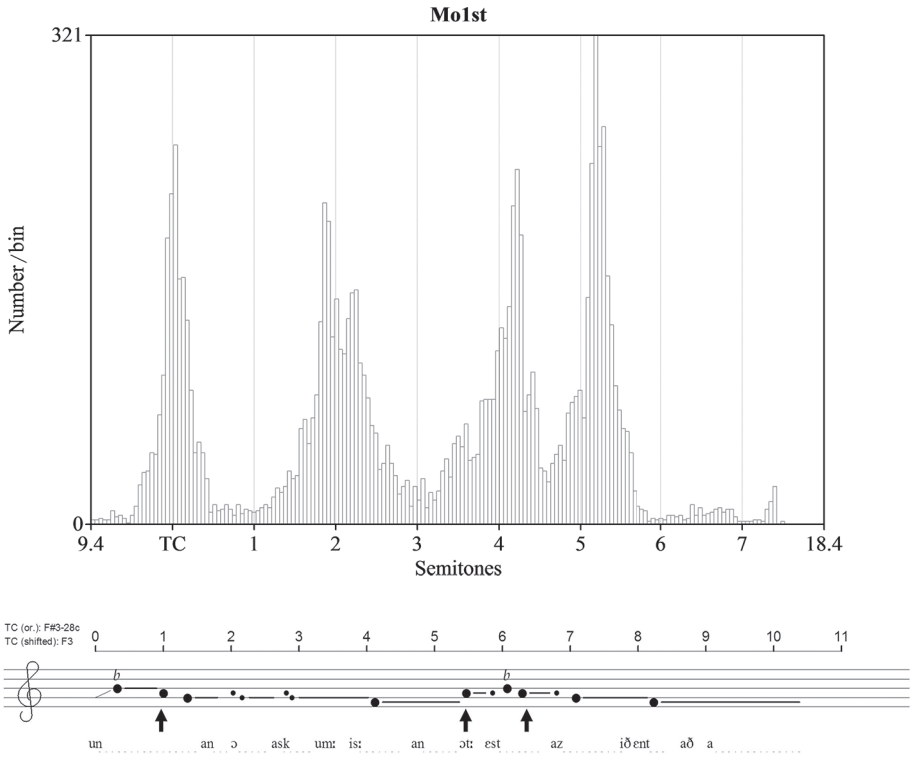
2.2 A peculiar third degree

Just as happens for other cultural aspects in the mountain village of Burcei, SM's singing style seems to be unintentionally influenced by Logudorese improvised and sung poetry (Pillonca, 2002; Manca, 2009; Pilosu, 2012). What follows is what SM has to say about this aspect of his village's culture and about the historical relationship between Burcei and Barbagia, the mountainous part of Sardinia where Logudorese poetry is widespread and currently practiced:

> Burcei è un paese nell'entroterra, è un paese che è nato intorno al 1600 ed è un paese che è nato dalla transumanza dei pastori della Barbagia. C'è un legame molto forte fra Burcei e la poesia barbaricina, la poesia logudorese, ecco perché ancora oggi Burcei è una roccaforte anche dei [poeti] logudoresi: ci sono due gare all'anno dei [poeti] logudoresi a Burcei. Il legame fra Burcei e la poesia penso che sia un legame secolare e forte proprio perché i pastori cantavano quando facevano queste transumanze del bestiame [tr.: 'Burcei is a backcountry village that sprang up around 1600, and it is a village born from the transhumance of the shepherds from Barbagia. There is a very strong connection between Burcei and the poetry from Barbagia, that is to say Logudorese poetry, this is the reason why Burcei is a stronghold of Logudorese [poets] even today: in Burcei there are two poetry contests a year of Logudorese [poets]. I think that the connection between Burcei and poetry is centuries old and strong precisely because the shepherds used to sing while they moved their herds'] (Monni, 2014).

One of the differences between the singing styles of the Campidanese improvisers and the Logudorese ones is the different intonation of the third degree, basically close to a major third from the tonal centre in Logudorese poetry (ex. C – E; cfr. Sassu, 2010 or. ed. 1973), whereas it is usually quite close to a minor third from the tonal centre in Campidanese *a mutetus longus* poetry (ex. C – E^b; cfr. Bravi, 2010, ch. 24). In this performance, SM seems to sing aiming to tune the third degree of his vocal melody towards the major degree, instead of the minor one, as would usually be the case for this singing genre (Figure 7).

Figure 7 - *At the top: fo histogram of* mutetu *n. 1 of SM (bins: 10 cents), where the pinnacle corresponding to the third degree is clearly placed over a distance of four semitones from the tonal centre (labelled as "TC" in the X axis), i.e. close to an interval of major third; at the bottom: musical transcription of the first line of the same* mutetu *(arrows indicate three A notes, at a distance of a major third with respect to the* tonus finalis *of the line (and tonal centre of the song) F*



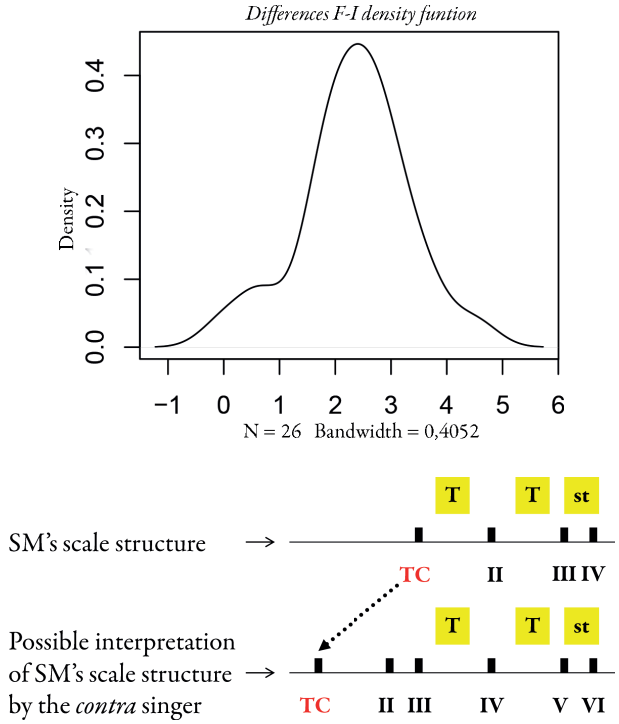
This musical peculiarity of SM's singing style may find its explanation in his education and practice in the first years of his apprenticeship. In this period, SM, just like all the other people living in the village of Burcei, used to listen to Logudorese sung poetry (octaves, *modas* etc.). This practice most likely established some kind of unaware imprinting on his musical conception that shows its effects on what he does in the poetical and musical Campidanese system.

> Io andavo e vado pazzo per *is modas* specialmente di Pazzola, tipo ce n'era una sulla Sardegna e una su Eleonora D'Arborea che quand'ero piccolo le canticchiavo sempre. Probabilmente inconsciamente, non è stata mai una cosa voluta, hanno influenzato la prima parte del mio apprendistato. Però è stata una cosa inconscia, io non ho voluto però il nostro background rientra in tutto quello che facciamo [tr.: 'I was and still am nuts about *is modas*, beyond the octaves, particularly Pazzola's [a Logudorese improviser]. There was one [moda] about Sardinia and one about Eleonora D'Arborea which I always used to sing when I was young. Probably unconsciously – never on

purpose – they influenced the first part of my apprenticeship. I wasn't aware of it, I didn't want it but our background always affects everything we do'] (Monni, 2014).

SM's tendency to tune his third degree close to the major third interval might be due to his early habit of listening to Logudorese sung poetry and singing some renowned songs from this repertoire on his own. As a hypothesis, one could argue that this could contribute to explain the lowering in *contra* tonality described in the par. 2.1. This lowering of the pitch is widely variable but, in a number of cases, close to three semitones. If we consider the equivalence between the I-IV degrees of a major scale and the III-VI degrees of a minor one, it is possible that an overall misunderstanding of the function associated to the intervals used by the improviser might have occurred, i.e. that the *contra* interpreted – of course with no theoretical awareness – the *finalis* as an anomalous and unexpected third degree and lowered the tonality according to his perception of the tonal relations in the improviser's singing (Figure 8).

Figure 8 - *In the top panel, a plot of the density function of the intervals between SM's* finalis *and the* contra*'s* initialis *expressed in semitones, where the maximum value is close to three semitones, i.e. a minor third interval; in the bottom panel, a prospective musical interpretation of the* contra*'s approach to SM's singing*



2.3 Training alone

Some biographical notes and some specific information about SM's experience and practice in the field of improvised poetry are of great interest in explaining what happened in the *cantada* analyzed here.

Firstly, as has already been said, his first experiences and approach to Sardinian poetry in his youth were tied to his family's attitudes and shared passion and involved not only Campidanese, but also Logudorese poetry.

> Mio padre ha sempre registrato le gare, girava il Campidano, girava il Sulcis e quindi io sono sempre andato con lui. Tutta la nostra famiglia si spostava per seguire le gare. [...] Contemporaneamente è nato l'interesse per un altro tipo di poesia che era la poesia logudorese, specialmente Zizi e Remundu Piras, perché c'era un mio zio che era appassionatissimo di poesia logudorese. E la poesia logudorese mi è piaciuta subito, forse ancora più della campidanese, in quel momento [tr.: 'My father always recorded poetry contests, [for these reasons] he travelled the Campidano [South Sardinian sub-region], he travelled around Sulcis [South-West Sardinian sub-region], and so I always went with him. All my family travelled around to attend poetry contests. [...] At the same time my interest for another type of poetry was born, Logudorese poetry, and especially for [Bernardo] Zizi and Remundu Piras, because my uncle was a big fan of Logudorese poetry. And I liked Logudorese poetry at once, perhaps even more than the Campidanese one, at that moment'] (Monni, 2014).

Secondly, SM had emigrated along with his parents and younger brother to Piedmont (a region in the North-Western part of Italy) in 2000. His father was the principal inspiring muse and mentor for him and his younger brother, Simone, who is now also doing improvisation. To get to work, all three of them have to travel by car every day. While they are on the road, they are used to improvising poetry, challenging one another:

> Con me c'è mio fratello che canta [improvvisa], e c'è anche mio padre che canta. [...] Quindi quando siamo in tre generalmente cantiamo *a mutetus longus* quando siamo in macchina. [...] Quando sono solo [in macchina], perfeziono di più la voce, quando invece sono in compagnia cerco di perfezionare l'improvvisazione, l'argomentazione [tr.: 'My brother sings [improvises] with me, and then there's my father who sings as well. [...] So when the three of us are together we usually sing in [the style of] *mutetus longus* when we're in the car [...] When I'm on my own [in the car], I do all I can to improve my vocal ability, whereas when I am with others I try to improve my ability to improvise, to debate'] (Monni, 2014).

The absence of a constant exchange of opinions with a community of improvisers and expert listeners and the practice of singing without a real accompanying choir represents an uncommon training process for a Campidanese improviser. *Cantadoris* usually start their experience as improvisers by taking part in informal meetings where amateurs and professionals often face each other in poetry contests (Bravi, 2010: 60-63). The scarcity of this kind of poetic confrontation within a community of practitioners and the lack of discussions and exchange of views that are common therein are significant factors that could have a role in the anomalies

in SM's first trial which has been considered here. The absence of a "preventive censure of the community" – to use the concept introduced by Roman Jakobson and Pëtr Bogatyrëv in their well-known 1929 article (Jakobson, Bogatyrëv, 1966 or. ed. 1929) – can explain, to a certain extent, what happened during the performance. In particular, the lack of any reciprocal knowledge between SM and the singers of the accompanying choir is acknowledged by SM himself as a reason for the musical difficulties that occurred during the performance:

> [In quella *cantada*] la mia voce ha creato delle difficoltà a *sa contra* ... uno, perché *sa contra* non mi conosceva; due, perché io non conoscevo *sa contra* e sostanzialmente io non avevo la cosiddetta *crai*, quella che si chiama *sa crai de sa contra*, che può essere bassa o alta [tr.: '[in that *cantada*] my voice caused some difficulties for *sa contra* ... firstly, because sa contra didn't know me; secondly, because I didn't know *sa contra*, and basically I didn't have the so-called 'key', what they call the '*contra* key', which may be low or high'] (Monni, 2014).

3. *Final considerations*

The construction of a personal style of singing (*tràgiu*, or *cadentza* to use the village of Burcei's terminology) is a long-lasting process that usually starts from the conscious or unconscious imitation of other improvisers' styles. From this early stage, a singing style usually evolves through a process of personalization and adaptation to one's personal vocal means, specific esthetic goals and needs as an improviser. SM describes the starting point of the process of construction of his singing in this way:

> Il mio *tràgiu* è un *tràgiu* solamente abbozzato per adesso e quindi il cammino è tutto da fare. Quello che ho cercato di fare è dare un'impronta originale [al mio stile di canto]. Probabilmente quello che ho fatto io dall'inizio è unire o mescolare il mio background, quello che avevo imparato fino ad allora, e metterlo in voce. [...] Qualcuno mi ha detto che un po' assomiglia a delle cadenze logudoresi, qualcuno mi ha detto che è influenzato dai miei ascolti di musica leggera addirittura, cantautori ecc. ecc. Non so se è vero o no. In realtà io ho cercato di fare qualcosa che riprendesse un *tràgiu* [di altri *cantadoris*] e ho cercato di modificarlo in modo che fosse il mio. Ancora non ci sono riuscito, ci sto ancora provando. Dei miglioramenti ci sono stati, però il processo è lungo. – [*PB*]: Qual è il tuo modello da personalizzare? – [*SM*] Il mio modello è un incrocio fra Urru [Raffaele] e Zuncheddu [Roberto], almeno questa era l'idea [...] Ci sono determinati *tragi* che danno la possibilità all'improvvisatore di fermarsi quando vuole durante *sa sterrina* [la prima sezione del *mutetu longu*]. Questo è quello che cercavo io: di poter fare le pause come le faceva Urru, come le fa Vargiu [Eliseo]; invece i *tragi* molto veloci non lo permettono, tipo quello di Pani [Antonio], quello di Falqui [Pierpaolo] ed altri [improvvisatori] [tr.: 'My *tràgiu* [singing style] is just a rough *tràgiu* for now and so there's a long way to go. What I have tried to do is to give an original mark [to my singing style]. Probably what I have done from the beginning is to join or mix my previously learnt background and put it into my voice. [...] Somebody told me that it seems a bit like the Logudorese style, others have said that it is even influenced by my listening to popular music, etc. I

> don't know whether this is true or not. Actually I tried to do something that conjures up the singing style [of other *cantadoris*] and I tried to modify it to adopt it as mine. I haven't managed to do it yet, I am still trying. There have been some improvements, but the process is still a long one. – [*PB*]: What was your model? – [*SM*] My model is a mixture between Urru [Raffaele] and Zuncheddu [Roberto], at least this was my idea. [...] There are singing styles which give the improviser the chance to stop when he wants during *sa sterrina* [the first section of the *mutetu longu*]. This was what I was looking for: to have the chance to make pauses as Urru did, as Vargiu [Eliseo] does; instead, quick singing styles do not allow you to do this, like that of Pani [Antonio], or Falqui [Pierpaolo] and other [improvisers]'] (Monni, 2014).

At the moment, for the improviser SM the evolutionary process of the construction of a personal and completely satisfying singing style is still in embryo. He is maturing as a poet thanks to his strong passion for poetry, his study of the masters of Campidanese improvisation from the past and present, his frequent trips to Sardinia where he can meet other improvisers and experts, as well as his active participation on social media sites, where he takes part in 'virtual' *cantadas* and publishes the video recordings of his poetry contests. As far as his voice is concerned, as Filippo Urru says, there is still room for improvement:

> Comenti a boxi, issu si depit fai ancora una boxi po cantai, perou cussu gei ddu at a fai cun su tempus. Cantendi unu s'aderetzat sa boxi, agoa stabilit cali est su tonu suu, chi ddi andat a gustu puru. Issu nci depit traballai ancora, poita no s'imparat in d-un'annu ma s'imparat in d-una carriera [tr. from Sardinian: 'As for his voice, he still has to build a voice for singing, but this needs time. With the experience of singing one refines his own voice, then establishes his style, which satisfies him. He still has to work on it, because one doesn't learn to do this in a year, but one learns it in a career'] (Urru, 2014[3]).

Although the voice is not the most important gift of an improviser, since one primarily needs a sharp and brilliant mind and a particular and well-trained memory, it is however an important feature and the hallmark with which he presents himself to the public.

> La prima cosa che si sente di un improvvisatore, ancora prima del vero discorso poetico, è la voce, e la voce è un marchio di fabbrica: ognuno deve avere il suo, [anche se]

[3] Given the level of linguistic fragmentation in Sardinia and the presence of two principal varieties of Sardinian – Logudorese and Campidanese – the issue of the search for a common Sardinian orthographic standard has been largely debated in the last decades (Contini, 1993; Argiolas, Serra, 2001; Grimaldi, Mensching, 2004; Comitau "Abbia a unu sardu comunu", 2004; Pittau, 2005; Corongiu, 2009, among others). As a consequence of this discussion and of the publication of both regional and national laws aiming at the safeguarding and the promotion of minority languages (Sardinian Regional Law n. 26/1997 and Italian National Law n. 482/1999), the Autonomous Region of Sardinia published two proposals for a standard system of orthographic transcription of Sardinian for administrative usage at different times (R.A.S., 2001; R.A.S., 2006). In the present case, the transcription of the interview in the Sardinian-Campidanese language was made according to the orthographic standard defined in Cardia et al., 2009, a collective work realized under the auspices of the Province of Cagliari and officially adopted by the same Institution on 17.03.2010.

> in realtà poi tutti i più grandi improvvisatori hanno ripreso cadenze precedenti [tr.: 'The first thing that one hears from an improviser, before his poetic argumentation, is his voice, and his voice is a trade-mark: everyone must have his own, [even if] as a matter of fact the greatest improvisers borrowed preceding singing styles']. (Monni, 2014).

Just like any other young poet, SM is on the road to reaching this goal, armed with the attention and diligence provided by his strong determination and passion for poetry.

Acknowledgements

I wish to thank Severino Monni and Filippo Urru for being so kind as to discuss this delicate issue with me in a number of interviews. My thanks also go to Sebastiano Pilosu, Andrea Mameli, and Daniela Mereu for having discussed with me aspects regarding the peculiarity of Burcei village from where both Monni and Urru come.

Bibliography

Agamennone, M. (Ed.) (1998). *Polifonie. Procedimenti, tassonomie e forme: una riflessione «a più voci»*. Roma: Bulzoni.

Argiolas, M., Serra, R. (Eds.) (2001). *Limba lingua language. Lingue locali, standardizzazione e identità in Sardegna nell'era della globalizzazione*. Cagliari: CUEC.

Bailly, L., Heinrich, N. & Pelorson, X. (2010). Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling. In *J. Acoust. Soc. Amer*., 127 (5), 3212-3222.

Bravi, P. (2010). *A sa moda campidanesa. Pratiche, poetiche e voci degli improvvisatori nella Sardegna meridionale*. Nuoro: ISRE.

Bravi, P. (2012). The practice of ornamentation in the multipart vocal music of Southern Sardinia. A 'bifocal' perspective in ethnomusicological analysis. In Macchiarella, I. (Ed.), *Multipart music: a specific mode of musical thinking, expressive behaviour and sound*. Udine: Nota, 467-490.

Cardia, A., Cherchi, S., Dessì, N., Madrigale, M., Madrigale, M., Maxia, F. et al. (2009). *Arrègulas po ortografia, fonètica, morfologia e fueddàriu de sa Norma Campidanesa de sa Lingua Sarda*. Quartu S.E., IT: Alfa Editrice.

Comitau "Abbia a unu sardu comunu" (2004). *Limba de Mesania po s'unidade de sa limba sarda.* Retrieved from: http://www.condaghes.com/public/docs/ldm.pdf.

Contini, M. (1993). *Parlare e scrivere il sardo*. Sassari: EDES.

Corongiu, G. (2009). *Una limba comuna cun milli limbàgios*. Sestu (CA): Domus de Janas.

Grimaldi, L., Mensching, G. (Eds.) (2004). *Su sardu. Lingua de Sardigna e lingua de Europa.* Cagliari: CUEC.

Jakobson, R., Bogatyrëv, P. (1966 or. ed. 1929). Folklore as a Special Form of Creation. In Jakobson, R., *Selected Writings* (O'Hara J. M., Trad.). The Hague: Mouton, 1-15.

Lutzu, M. (2012). Il ruolo degli appassionati nella poesia improvvisata campidanese. In Bravi, P., Mereu, D. & Murgia, I. (Eds.), *A campu. Archivi e ricerche di poesia orale a Villasimius*. Quartu S. Elena (CA): Alfa, 55-68

Manca, M. (2009). *Cantare in poesia per sfidare la sorte. Un approccio antropologico alla gara poetica logudorese in Sardegna*. Nuoro: ISRE.

Monni, S. (2014, 12 28). Interview. (P. Bravi, Interviewer).

Mossa, M. (1999). L'improvvisazione poetica nel Campidano. Rilievi etnomusicologici. In *Quaderni oristanesi*, 43-44, 57-79.

Pillonca, P. (2002). *Chent'annos.* Selargius: Domus de Janas.

Pilosu, S. (Ed.) (2012). *Enciclopedia della musica sarda, vol. 13: Poesia improvvisata – vol. 1*. In Casu, F., Lutzu, M. (Eds.). Cagliari: L'Unione Sarda.

Pittau, M. (2005). *Grammatica del sardo illustre*. Sassari: Delfino.

R.A.S. (2001). *Limba Sarda Unificada. Sintesi de sas Normas de base: ortografia, fonètica, morfolozia, lèssicu.* Cagliari: Regione Autonoma della Sardegna - Assessorato della Pubblica Istruzione, Beni Culturali, Informazione, Spettacolo e Sport.

R.A.S. (2006). *Limba Sarda Comuna. Norme linguistiche di riferimento a carattere sperimentale per la lingua scritta dell'Amministrazione regionale*. Retrieved from: Regione Autònoma de Sardigna - Regione Autonoma della Sardegna: http://www.regione.sardegna.it/documenti/1_72_20060418160308.pdf.

Sassu, P. (2010 or. ed. 1973). Le strutture musicali. In Carpitella, D., Sole, L. & Sassu, P., *Musica sarda*. Udine: Nota, 68-95.

Urru, F. (2014, 12 26). Interview. (P. Bravi, Interviewer).

Zedda, P. (2008). *L'arte de is mutetus*. Iesa (SI): Gorèe.

Zedda, P. (2009). The Southern Sardinian Tradition of the Mutetu Longu: A Functional Analysis. In *Oral Tradition*, 24 (1), 3-40.

PARTE III

TRATTAMENTO AUTOMATICO DELLE LINGUE NATURALI

DOI: 10.17469/O2101AISV000026

RODOLFO DELMONTE

Expressivity in TTS from Semantics and Pragmatics

In this paper we present ongoing work to produce an expressive TTS reader that can be used both in text and dialogue applications. The system called SPARSAR has been used to read (English) poetry so far but it can now be applied to any text. The text is fully analyzed both at phonetic and phonological level, and at syntactic and semantic level. In addition, the system has access to a restricted list of typical pragmatically marked phrases and expressions that are used to convey specific discourse function and speech acts and need specialized intonational contours. The text is transformed into a poem-like structures, where each line corresponds to a Breath Group, semantically and syntactically consistent. Stanzas correspond to paragraph boundaries. Analogical parameters are related to ToBI theoretical indices but their number is doubled. In this paper, we concentrate on short stories and fables.

1. *Introduction*

In this paper we present ongoing work to produce an expressive TTS reader that can be used both in text and dialogue applications. The system called SPARSAR has been used to read (English) poetry so far but it can now be applied to any text provided it has been previously fully analyzed both at phonetic and phonological level, and at syntactic and semantic level. In addition, the system has access to a restricted list of typical pragmatically marked phrases and expressions that are used to convey specific discourse function and speech acts and need specialized intonational contours. In this paper, we concentrate on short stories and fables.

Current TTS systems are dull and boring and characterized by a total lack of expressivity. They only take into account information coming from punctuation and in some cases, from tagging and syntactic constituency. Few expressive synthetic speech synthesizers are tuned to specific domains and are unable to generalize. They usually convey specific emotional content linked to a list of phrases or short utterances – see below. In particular, comma is a highly ambiguous punctuation mark with a whole set of different functions which are associated with specific intonational contours. In general, question and exclamative marks are used to modify the prosody of the previous word. We use the word "expressivity" in a specific general manner which includes sensible and sensitive reading that can only be achieved once a complete syntactic and semantic analysis has been provided to the TTS manager.

From a general point of view, the scientific problem can be framed inside the need to develop models that are predictive for a speech synthesizer to be able to sound natural and expressive, getting as close as possible to human-like performance. This can only be achieved manipulating prosody so that the text read aloud

sounds fully natural, informative and engaging or convincing. However, in order to achieve something closer to that, text understanding should be attained or some similar higher level semantic computation. As Xu (2011) puts it, "It will probably be a long time before anything close to that is developed, of course" (*ibi*, 94). Similar skeptical or totally negative opinions are expressed by Marc Huckvale (2002), when summarizing work he and his group have been carrying out for a long period over the project for an articulatory TTS called ProSynth. In fact, he bitterly criticizes current approaches at using machine learning and statistical processing to attain commercially viable results, as the one based on Unit Selection, because in this way no scientifically useful information will ensue towards the goal of increasing our knowledge of how human speech production works (*ibi*, 1261). The goal of speech synthesis, in this perspective would be that of "understanding how humans talk" rather than the one of replicating a human talker. In his opinion,

> We are not in any sense close to replicating how humans or HAL use speech because we are nowhere close to building a system with intentionality, with the need to communicate. The lack of expression or interest or emotion in the speech of current systems is due to the fact that the systems don't actually understand what they are saying, see no purpose to the communication, nor actually have any desires of their own. It makes no sense to add "expressiveness" or "emotion" to voice output from a mindless system (*ibi*, 1262).

This is what the majority of the systems are currently trying to do "Adding emotion to a mindless system", and they do that by focusing on specific words or expressions. Research on emotion from a perceptive and productive way has started long time ago (see Cahn, 1990), but exploded lately, say from 2000 onward. The bibliography is extensive and regards the six basic emotions (Disgust, Happiness, Sadness, Anger, Fear, Surprise, Neutral) as well as richer sets including Boredom, Despair, Anxiety, Contempt, Shame; Indignant, Pleasant. Emotion Generation is mainly achieved by statistical methods using different approaches: Formant and Articulatory Synthesis; Parametric Synthesis (HMMs); Diphone Synthesis; Unit Selection Synthesis.

Synthesis of expressive sounds is easy and can be extended easily to whole phrases by unit selection methods. This can be done by collecting separate databases for different emotions as suggested by Lakshmi et al.. This latter method seems the most promising in that it allows to address units of various size, starting from Phones, Halfphones but also Words, Phrases and entire Utterances, depending on the domain. Drawbacks are the need to improve on prosodic modeling and also to account for fine-grained differences in emotional content as clearly indicated by Nick Campbell in the conclusion of his talk for SSW6 (Speech Synthesis Workshop 6).

More linguistically based work on emotions has been documented by the group working at Loquendo (now acquired by Nuance). They report their approach based on the selection of Expression which is related to a small inventory of what they call "speech acts" – rather more Conversational and Argumentative or Dialogue acts. They implemented the acoustic counterpart of a limited, but rich, set of dialogue acts, including: refuse, approval/disapproval, recall in proximity, announce, request

of information, request of confirmation, request of action/behaviour, prohibition, contrast, disbelief, surprise/ astonishment, regret, thanks, greetings, apologies, and compliments. In total, they managed to label and model accordingly some 500 different (expressive) utterances that can be used domain and context independently.

Work related to what we are trying to do is to be found in the field of storytelling and in experiments by the group from Columbia University working at MAGIC a system for the generation of medical reports. Montaño et al. present an analysis of storytelling discourse modes and narrative situations, highlighting the great variability of speech modes characterized by changes in rhythm, pause lengths, variation of pitch and intensity and adding emotion to the voice in specific situations. In their experiment on storytelling TTS they consider it as a first step towards modeling pitch, intensity and tempo. They found the following important possible triggers for prosody adjustment: a wide variety of speech variability; continual rhythm changes; including pauses of different duration; sometimes adding emotion to the voice. These have had as a consequence, variation of pitch and intensity, and the need to treat dialogue structures correctly. In other words, in order to make TTS for storytelling more natural and expressive they had to address those triggers and model rhythm variability, pauses and voice quality.

However, the most interesting approach related to ours is the one by the group of researchers from Columbia University, where we can find Julia Hirschberg and Kathy McKeown. In the paper by S. Pan, K. McKeown & J. Hirschberg (1999) they highlight the main objectives of their current work, i.e.

> The production of natural, intelligible speech depends, in part, on the production of proper prosody: variations in pitch, tempo and rhythm. Prosody modelling depends on associating variations of prosodic features with changes in structure, meaning, intent and context of the language spoken. This process involves selecting information that has potential to influence prosody, identifying correlations between this information and prosodic parameters through data exploration, and using learning algorithms to build prosody models from these data (*ibi*, 1420).

In fact, their attempt at using machine learning for building prosody models has been only partially achieved as we will see in detail below. Now the important part of their work is the one related to the concept-to-speech manager. As reported in their paper,

> The content planner uses a presentation strategy to determine and order content. It represents discourse structure, which is a hierarchical topic structure in MAGIC, discourse relations, which can be rhetorical relations, and discourse status, which represents whether a discourse entity is given, new or inferable and whether the entity is in contrast with another discourse entity. Most of the features produced at this stage have been shown to have influence on prosody: discourse structure can affect pitch range, pause and speaking rate (Grosz, Hirschberg, 1992); given/new/ inferable can affect pitch-accent placement (Hirschberg, 1993); a shift in discourse focus can affect pitch-accent assignment (Nakatani, 1998); and contrastive entities can bear a special pitch accent (Prevost, 1995) (*ibi*, 1422).

Further work towards predicting prosodic structure was published by Bachenko & Fitzpatrick, 1990, Delmonte & Dolci, 1991, and Wang & Hirschberg, 1992. The objective of their experiment was modeling ToBI prosody features, i.e. pitch accents, phrase accents, boundary tones and break indices. Given the fact that there are six pitch-accent classes, five break-index classes, three phrase-accent classes, and three boundary-tone classes, they come up with a total of 17 different features organized in four separate classes. The experiment was carried out on a corpus of spontaneous speech with some 500 dialogues on medical issues, which ended up by being reduced to 250 annotated dialogues.

In fact, the features they managed to annotate are just surface syntactic and semantic ones, and they are: (1) ID: the ID of a feature vector; (2) Lex: the word itself; (3) Concept: the semantic category of a content word; (4) SynFunc: the syntactic function of a word; (5) SemBoundary: the type of semantic constituent boundary after a word; (6) SemLength: the length, in number of words, of the semantic constituent associated with the current SemBoundary; (7) POS: the part-of-speech of a word; (8) IC: the semantic informativeness of a word, where in particular, the latter is – in our opinion – wrongly computed as a "semantic feature", being constituted by the logarithm of the relative frequency of a term in the corpus. The most disappointing fact was that they attempted to carry out a complete annotation but didn't succeed. In the paper they report their annotation efforts on the spontaneous-speech corpus which was automatically annotated with POS information, syntactic constituent boundaries, syntactic functions, and lexical repetitions, using approximations provided by POS taggers and parsers. It was also manually labelled with given/new/inferable information. But when it comes to semantic and discourse level information:

> We are still working on manually labelling discourse structure, discourse relations, and semantic abnormality... We are currently annotating the speech corpus with features closely related to meaning and discourse (*ibi*, 1426).

No further publication reports experiments with the complete annotation. And this is clearly due to difficulties inherent in the task. Now, this is what our system allows us to do, i.e. using discourse structure and relations to instruct the prosody manager to introduce the appropriate variation of prosodic parameters. According to ToBI features, this implies the ability to achieve:

- Juncture placing prediction;
- Phrase boundary tone prediction;
- Prominence prediction;
- Intonational Contour movement prediction.

To be more specific, given an input text the "Ideal System" will read it aloud using naturally sounding prosody, where:

- Phrasing is fully semantically consistent;
- Intonation varies according to structural properties of clauses in discourse and speaker intention;

- Prominence is assigned on the basis of novelty of topics and related events;
- Expressivity is intended to convey variations of attitude and mood as they are derived from deep subjective and affective analysis.

Our reformulation of ToBI features from general/generic into concrete and implemented analogical parameters for natural and expressive TTS will be achieved at the end of the paper. The correspondence between prosodic features and linguistic representation is the issue to cope with and will be presented here. Levels of intervention of syntactic-semantic and pragmatic knowledge include:

- syntactic heads which are quantified expressions;
- syntactic heads which are preverbal SUBJects;
- syntactic constituents that starts and ends an interrogative or an exclamative sentence;
- distinguish realis from irrealis mood;
- distinguish deontic modality including imperative, hortative, optative, deliberative, jussive, precative, prohibitive, propositive, volitive, desiderative, imprecative, directive and necessitative etc.;
- distinguish epistemic modality including assumptive, deductive, dubitative, alethic, inferential, speculative etc.;
- any sentence or phrase which is recognized as a formulaic or frozen expression with specific pragmatic content;
- subordinate clauses with inverted linear order; distinguishing causal from hypotheticals and purpose complex sentences;
- distinguishing parentheticals from appositives and unrestricted relatives;
- Discourse Structure to tell satellite and dependent clauses from main;
- Discourse Structure to check for Discourse Moves - Up, Down and Parallel;
- Discourse Relations to tell Foreground Relations from Backgrounds;
- Topic structure to tell the introduction of a new Topic or simply a Change at relational level.

There is general consensus on the need to provide "Natural Language Understanding" abilities to generate expressive TTS. We dub NLU as the ability to represent the semantics and pragmatics of an utterance, a text or a dialogue. What kind of semantics and pragmatics is then needed to produce expressive TTS? Do there exist state of the art systems that can generate such a representation?

We produced a number of computational models for a linguistically driven TTS in our research work for Digital Equipment in the '80s aiming at the Italian version of DecTalk (Delmonte et al., 1984; 1986). The scheme of interaction between linguistic processing and parameterization for TTS was fairly general at the time and included phonetics, phonology and syntax. No semantics was considered for lack of computer power and memory storage (Delmonte, 1982; 1985).

In our prosodic computational model (1986, figg. 1-4), all interactions between the two levels – the phonological and the grammatical one – are already clearly

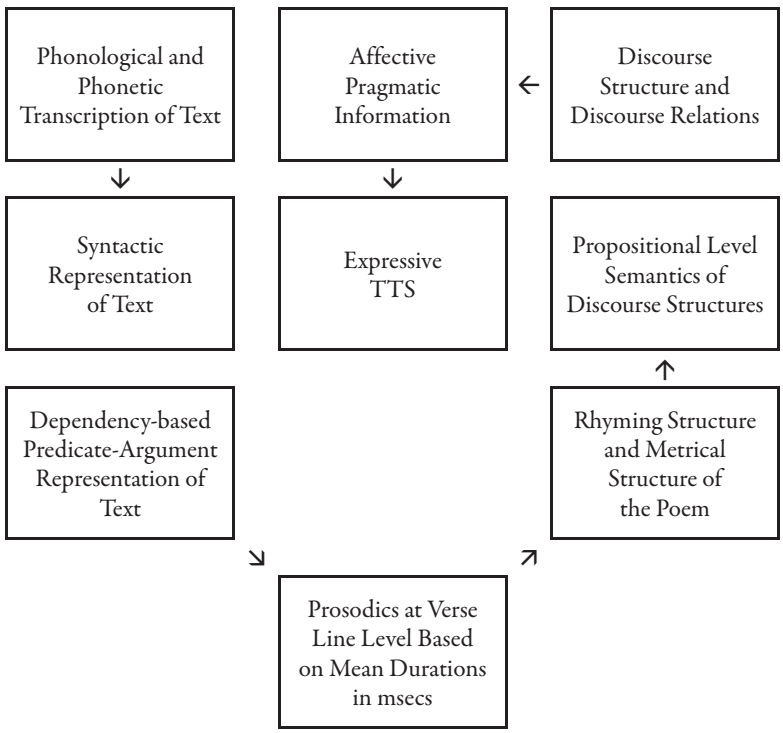
indicated. However, no semantic information would be made available except for syntactic structure computed with the help of a subcategorization lexicon of Italian and an ATN parser (1986, fig. 6).

Our current system, on the contrary, develops a fully specified semantic representation which is then responsible for the great majority of focus displacements and intonational movements. As is clarified above, semantics is an essential component and contributes most of the information needed for a TTS to be perceived as "expressive".

2. *Semantic Representation for TTS*

Systems that can produce an appropriate semantic representation for a TTS are not many at an international level but they can be traced from the results of a Shared Task organized by members of SigSem and are listed here below in the corresponding webpage http://www.sigsem.org/w/index.php?title=STEP_2008_ shared_ task:_comparing_semantic_representations. State of the art semantic systems are based on different theories and representations, but the final aim of the workshop was reaching a consensus on what constituted a reasonably complete semantic representation, as can be gathered from the questionnaire that has been used to grade each system.

Figure 1 - *System Architecture Modules for SPARSAR*



Semantics in our case not only refers to predicate-argument structure, negation scope, quantified structures, anaphora resolution and other similar items, it refers essentially to a propositional level analysis, then to a text and discourse level analysis. Propositional level semantic representation is the basis for discourse structure and discourse semantics contained in discourse relations. It also paves the way for a deep sentiment or affective analysis of every utterance, which alone can take into account the various contributions that may come from syntactic structures like NPs and APs where affectively marked words may be contained. Their contribution needs to be computed in a strictly compositional manner with respect to the meaning associated to the main verb, where negation may be lexically expressed or simply lexically incorporated in the verb meaning itself.

In Figure 1 we show the architecture of our deep system for semantic and pragmatic processing, in which phonetics, prosodics and NLP are deeply interwoven.

The system does the usual low level analyses before semantic modules are activated, that is tokenization, sentence splitting, multiword creation from a large lexical database. Then chunking and syntactic constituency parsing which is done using a rule-based recursive transition network: the parser works in a cascaded recursive way to include always higher syntactic structures up to sentence and complex sentence level. These structures are then passed to the first semantic mapping algorithm that looks for subcategorization frames in the lexica made available for English, including VerbNet, FrameNet, WordNet and a proprietor lexicon with most frequent verbs, adjectives and nouns, containing also a detailed classification of all grammatical or function words. This mapping is done following LFG principles, where c-structure is turned into f-structure obeying uniqueness, completeness and coherence. The output of this mapping is a rich dependency structure, which contains information related also to implicit arguments, i.e. subjects of infinitivals, participials and gerundives. It also has a semantic role associated to each grammatical function, that is used to identify the syntactic head lemma uniquely in the sentence. Finally it takes care of long distance dependencies for relative and interrogative clauses.

Now that fully coherent and complete predicate argument structures have been built, pronominal binding and anaphora resolution algorithms can be fired. Also coreferential processed are activated in the semantic level: they include a centering algorithm for topic instantiation and memorization that we do using a three-place stack containing a Main Topic, a Secondary Topic and an Potential Topic. In order to become a Main Topic, a Potential Topic must be reiterated and become persistent in the text.

Discourse Level computation is done at propositional level by building a vector of features associated to the main verb of each clause. They include information about tense, aspect, negation, adverbial modifiers, modality. These features are then filtered through a set of rules which have the task to classify a proposition as either objective/subjective, factual/nonfactual, foreground/ background. In addition, every lexical predicate is evaluated with respect to a class of discourse relations. Eventually, discourse structure is built, according to usual criteria of clause dependency where a clause can be classified either as coordinate or subordinate.

We have a set of four different moves to associate to each clause: root, down, level, up. We report here below semantic and discourse structures related to the poem by Sylvia Plath "Edge" which you can find here, http://www.poetryfoundation.org/poem/178970, and in the *Appendix*.

Figure 2 - *Propositional semantics for Edge*

Clause No.	Funct/ Role	View	Factivity	Change	Relevance	Aspect	Pred	Tense	Disc_ Rel_	Subject_
33	coord/ prop	external	factive	null	background	activity	crackle	pres	narration	objective
32	coord/ prop	external	factive	null	background	activity	drag	pres	narration	objective
26	main/ prop	external	factive	culmintd	foregrnd	activity	use	perf	cause	objective
21	main/ prop	external	factive	null	background	activity	be	pres	narration	objective
20	main/ prop	external	factive	null	background	activity	have	pres	narration	objective
16	coord/ prop	external	factive	culmintd	foregrnd	activity	bleed	past	narration	objective
15	main/ prop	external	factive	null	background	activity	close	pres	narration	objective
14	main/ prop	external	factive	null	background	activity	fold	perf	narration	objective
11	main/ prop	external	factive	null	background	activity	bare	pres	narration	objective
10	main/ prop	external	factive	null	background	activity	be	pres	narration	objective
9	main/ prop	external	factive	null	background	activity	flow	pres	result	objective
8	main/ prop	external	factive	null	background	activity	have	pres	narration	objective
7	main/ prop	external	factive	null	background	activity	have	pres	narration	objective
6	main/ prop	external	factive	culmintd	foregrnd	activity	perfect	perf	result	objective
5	main/ prop	external	factive	culmintd	foregrnd	activity	say	past	narration	objective
4	main/ prop	external	factive	null	background	activity	seem	pres	narration	objective
3	main/ prop	external	factive	null	background	activity	wear	pres	narration	objective
1	xcomp/ prop	internal	factive	null	background	state	edge	nil	setting	objective

In Figure 2, clauses governed by a copulative verb like BE report the content of the predication to the subject. The feature CHANGE can either be set to NULL, GRADED or CULMINATED: in this case Graded is not used seen that there no progressive or overlapping events.

Figure 3 - *Discourse level Semantics for Topic Hierarchy*

Topic Type	Clause No.	Pred	Semant_ Id_	M-Feats Per, Gen, Num	Semantic Inherent Feats	Semantic Role
main	1	edge	id1	[3, neu, sing,	[abstrct, legal, nquant, objct],	theme_bound]
poten	3	illusion	id2	[3, nil, nil,	[abstrct, inform, danger],	theme_bound]
poten	3	scroll	id3	[3, mas, sing,	[abstrct, tecno],	goal]
poten	3	foot	id4	[3, nil, nil,	[animat, body_part, objct],	theme_bound]
poten	3	smile	id5	[3, mas, sing,	[activ, inform],	goal]
poten	3	toga	id6	[3, nil, nil,	[body_part, objct],	theme_bound]
poten	3	dead_body	id7	[3, mas, sing,	[objct, hum],	goal]
poten	3	necessity	id8	[3, nil, nil,	[place, inform, state],	theme_bound]
poten	3	accom-plishment	id10	[3, mas, sing,	[abstrct, chang, state],	goal]
main	3	woman	id2	[3, fem, sing,	[any, relat, social, hum],	theme]
second	15	garden	id11	[3, neu, plur,	[instit, objct, instrum],	agent]
poten	15	child	id12	[3, neu, sing,	[any, activ, body_part, objct, relat, social, instrum, hum],	actor]
poten	15	serpent	id13	[3, neu, sing,	[animt, objct, instrum],	theme]
poten	15	throat	id14	[3, neu, plur,	[body_part, objct, instrum, hum],	loc_origin]
poten	15	stiffen	id16	[3, neu, plur,	[instit],	goal]
poten	15	body	id7	[3, neu, sing,	[abstrct, activ, body_part, inform, instit, place, objct, instrum, hum],	loc_direct]
poten	15	pitcher	id15	[3, mas, sing,	[activ, inform, nquant, objct, relat, social, instrum, hum],	specif]
poten	15	milk	id17	[3, neu, sing,	[body_part, edible, objct, hum],	specif]
poten	15	petal	id18	[3, neu, plur,	[plant],	agent]
poten	15	flower	id19	[3, neu, sing,	[plant, time],	theme]
poten	15	night	id20	[3, neu, sing,	[state, time],	specif]
main	21	hood	id21	[3, mas, sing,	[objct, instrum, hum],	loc_origin]
poten	21	moon	id22	[3, neu, sing,	[event, place, objct, time],	experiencer]
poten	29	sort_of	id23	[3, nil, nil,	[abstrct, activ, inform, relat, social, state, tecno, hum],	attr]

In Figure 3. below, we see topics of discourse as they have been computed by the coreference algorithm, using semantic indices characterized by identifiers starting

with ID. Every topic is associated to a label coming from the centering algorithm: in particular, WOMAN which is assigned ID id2 reappears as MAIN topic in clauses marked by n. 15. Also BODY reappears with id7. Every topic is associated to morphological features, semantic inherent features and a semantic role. Eventually, the final computation which concerns Discourse Structure is this one:

Figure 4 - *Discourse Semantics for Discourse Structures*

Sent_ No.	Clause No.	Subject_	Disc_ Rel_	Tense	Pred	Relevance	Disc_ Move	Disc_Struct_ Attach_Node
edge_7	39	objective	narration	pres	crackle	background	level	down(31-39))
edge_7	38	objective	narration	pres	drag	background	level	down(31-39))
edge_6	31	objective	cause	perf	use	foreground	up	to(1-31))
edge_5	25	objective	narration	pres	moon	background	level	down(18-25))
edge_5	24	objective	narration	pres	have	background	level	down(18-24))
edge_5	23	objective	narration	pres	stare	foreground	down	down(18-23))
edge_4	18	objective	narration	past	bleed	foreground	up	to(1-18))
edge_4	17	objective	narration	perf	fold	background	level	level(11-17))
edge_4	15	objective	circumstance	pres	stiffen	background	level	level(11-16))
edge_3	11	objective	narration	pres	it	background	level	level(7-11))
edge_3	10	objective	narration	pres	come	foreground	level	level(7-10))
edge_2	7	objective	result	pres	flow	background	down	down(1-7))
edge_2	5	objective	narration	past	say	foreground	up	to(1-5))
edge_2	4	objective	narration	pres	seem	background	level	level(1-4))
edge_2	3	objective	narration	pres	wear	background	level	level(1-3))
edge_2	6	objective	narration	perf	perfect	foreground	down	down(1-6))
edge_1	1	objective	setting	nil	edge	background	up	down(nil-1))

Movements in the intonational contours are predicted to take place when FOREGROUND and UP moves are present in the features associated to each clause. At discourse level computation, Background and Foreground are associated to opposing semantic analyses. In the case of Foreground, a new event is taking place or a new character appears on the scene. Thus Foreground is based on temporal and aspectual analysis, anaphoric and coreference analysis. These in turn are based on propositional level analysis as explained below. In addition, discourse structures and relations are responsible for a choice between Back and Foreground, that is the meaning of the governing predicates of each clause.

3. *From Poetry to Story Reading*

We referred to a poem in the previous section because in fact we will be using rules associated to poetry prosodic mapping in our work on story reading. We assume that reading a story aloud requires the reader to organize pauses in such a way that expressivity and meaning is preserved. This process is usually referred to as Breath Group organization: a breath group is a well-formed group of words conveying a concept or a meaning, and we would like to compare it to a line in a poem. Poems are organized into lines and stanzas, while stories usually have punctuation to mark main concepts and introduce pauses. Punctuation however is not sufficient in itself and does not always guarantee meaning coherence. We already noted that Commas are highly ambiguous and may be used for a whole set of different functions in discourse. So eventually what we can actually trust are Breath Groups. Continuing our comparison with poems, lines may be end-stopped or enjambed when they run on the following line or stanza. The same may happen with Breath Groups, they may be end-stopped or enjambed and require a different prosodic setup.

We will then define Breath Groups as syntactically and semantically coherent units coinciding with an Intonation Phrase in ToBI terms: IPs are characterized by different tones, possible boundary tones and break indices. Pitch Accents are associated to word accents which are present in our phonetic representation: except that only syntactic heads are associated with Pitch Accents, dependents are demoted.

3.1 Implementing the Rules for Expressive TTS

Let's now look at one example, a short story by Aesop, "Bellying the Cat" that can be found here, http://www.taleswithmorals.com/aesop-fable-belling-the-cat.htm. At first we show the decomposition of the story into Breath Groups and then the mapping done by the Prosodic Manager (hence PM):

Table 1 - *Decomposition of the text into Breath Groups*

long_ago ß
the mice had a general council ß
to consider what measures they could take to outwit their common enemy ß
the cat ß
some said this ß
and some said that ß
but at_last a young mouse got_up ß
and said he had a proposal ß
to make ß
which he thought would meet the case ß
ß
you will all agree ß
said he ß
that our chief danger consists in the sly ß
and treacherous manner ß
in which the enemy approaches us ß

ß
now ß
if we could receive some signal of her approach ß
we could easily escape from her ß
i venture ß
therefore ß
to propose that a small bell be procured ß
and attached by a ribbon round the neck of the cat ß
by_this_means ß
we should always know when she was about ß
and could easily retire ß
while she was in the neighborhood ß
ß
this proposal met with general applause ß
until an old mouse got_up ß
and said ß
ß
that is all very_well ß
but who is to bell the cat ß
ß
the mice looked at one_another ß
and nobody spoke ß
then the old mouse said ß
ß
it is easy ß
to propose impossible remedies ß

A first set of the rules to map the story into this structures are reported below. The rules are organized into two separate sets: low level and high level rules. Here are low level ones:

3.1.1 Breath Group Creation Rules

- Follow punctuation first, but check constituent length;
- Rule for Coordinate Structures;
- Rule for Subordinate Clauses;
- Rule for Infinitival Complements;
- Rule for Complement Clauses;
- Rule for Relative Clauses;
- Rule for Between Subject and VerbPhrase;
- Rule for AdverbialPhrase but only beginning of Clause;
- Rule for Obligatory complements followed by adjuncts – with long constituents.

The high level corresponds to the recursive level. Recursive rules are associated with complex sentences and with Coordinate, Subordinate and Complement clauses. These rules need to be conceived in their recursive structure when implemented in a real algorithm, in particular, they will reorganized as follows.

Here below is the mapping of the text above into analogical phonetic acoustic correlates of pitch, speaking rate and intensity, and pauses. Here PBAS is used for

F0 rise/fall instructions; RATE for speaking rate; VOLM for energy or intensity; SLNC stands for silence, hence pauses and RSET for reset all parameter to the default value. However, since we intend to refer to ToBI framework for prosody annotation, we will then match every new feature we inserted in the text with the inventory made available by the original ToBI manual – see below. This analogical version can be copy/pasted into a TextEdit file and spoken aloud by Apple TTS.

[[pbas 38.000; rate 160; volm +0.5]]Bellying the cat, a story by Aesop . [[slnc 400]],[[rset 0]] [[pbas 44.000; rate 140; volm +0.3]][[pbas 36.000; rate 110; volm -0.2]]Long ago [[rset 0]] , the mice had a general council to [[pbas 36.000; rate 110; volm +0.5]]consider[[slnc 50]],[[rset 0]] what measures they could [[pbas 36.000; rate 110; volm +0.5]]take[[slnc 30]],[[rset 0]] to [[slnc 100]][[pbas 40.000; rate 150; volm +0.5]]outwit their common [[pbas 38.000; rate 130; volm +0.3]]enemy[[slnc 200]],[[rset 0]] , the [[pbas 38.000; rate 130; volm +0.3]]cat[[slnc 200]],[[rset 0]] .
Some [[pbas 36.000; rate 110; volm +0.5]]said[[slnc 50]],[[rset 0]] [[pbas 38.000; rate 130; volm +0.3]]this[[slnc 200]],[[rset 0]] , [[slnc 100]]and some said [[pbas 38.000; rate 130; volm +0.3]]that[[slnc 200]],[[rset 0]] ;
[[pbas 36.000; rate 110; volm +0.5]]but[[slnc 30]],[[rset 0]] at_last a young [[rate 130; volm +0.5]]mouse got_up [[slnc 100]]and said he had a proposal to [[pbas 38.000; rate 130; volm +0.3]]make[[slnc 200]],[[rset 0]] , which he thought [[slnc 100]][[pbas 50.000; rate 120; volm +0.5]]would meet the [[pbas 38.000; rate 130; volm +0.3]]case[[slnc 200]],[[rset 0]] .
" You will [[rate 130; volm +0.5]]all agree " , said [[pbas 38.000; rate 130; volm +0.3]] he[[slnc 200]],[[rset 0]] , " that our chief danger consists in the [[pbas 36.000; rate 110; volm -0.2]]sly and treacherous [[rset 0]] manner in which the enemy approaches [[pbas 38.000; rate 130; volm +0.3]]us[[slnc 200]],[[rset 0]] " .
Now , if we [[slnc 100]][[pbas 50.000; rate 120; volm +0.5]]could receive some signal of [[rate 130; volm +0.5]]her approach , we could easily escape from [[pbas 38.000; rate 130; volm +0.3]]her[[slnc 200]],[[rset 0]] .
I [[pbas 38.000; rate 130; volm +0.3]]venture[[slnc 200]],[[rset 0]] , [[pbas 38.000; rate 130; volm +0.3]]therefore[[slnc 200]],[[rset 0]] , to [[pbas 36.000; rate 110; volm +0.5]]propose[[slnc 50]],[[rset 0]] that a small bell [[slnc 100]][[pbas 50.000; rate 120; volm +0.5]]be procured , [[slnc 100]]and attached by a ribbon round the neck of the [[pbas 38.000; rate 130; volm +0.3]]cat[[slnc 200]],[[rset 0]] .
by this means we should always know [[slnc 100;pbas 48.000; rate 150; volm +0.3]] when she was [[pbas 38.000; rate 130; volm +0.3]]about[[slnc 200]],[[rset 0]] , [[slnc 100]]and could easily [[pbas 36.000; rate 110; volm +0.5]]retire[[slnc 50]],[[rset 0]] while she was in the [[pbas 38.000; rate 130; volm +0.3]]neighborhood[[slnc 200]],[[rset 0]] " .
This proposal [[slnc 100]][[pbas 50.000; rate 120; volm +0.5]]met with general [[pbas 38.000; rate 130; volm +0.3]]applause[[slnc 200]],[[rset 0]] , until an old mouse [[slnc 100]][[pbas 50.000; rate 120; volm +0.5]]got up [[slnc 100]]and [[pbas 38.000; rate 130; volm +0.3]]said[[slnc 200]],[[rset 0]] : [[pbas 48.000; rate 130; volm +0.9]]
" That is [[rate 130; volm +0.5]]all very_well , but [[pbas 54.000; rate 170; volm +0.3]]who is to bell the cat[[slnc 300; pbas 54.000; rate 170; volm +0.3]] ? [[rset 0]]
" the mice looked at one another [[slnc 100]]and [[rate 110; volm +0.3]]nobody[[slnc 100]],[[rset 0]] [[pbas 38.000; rate 130; volm +0.3]]spoke[[slnc 200]],[[rset 0]] .

[[pbas 54.000; rate 170; volm +0.3]]Then the old mouse said : [[pbas 48.000; rate 130; volm +0.9]]
" it is [[pbas 40.000; rate 140; volm +0.3]]easy[[slnc 30]],[[rset 0]] to propose [[pbas 36.000; rate 110; volm -0.2]]impossible remedies [[rset 0]] " .

3.2 ToBI features Implemented

We will now discuss the use of Pierrehumbert's inventory of Tones and Break Indices, in relation to its actual application in real texts reading. We tried to map the analogical features required by the prosodic manager into ToBI features but we soon discovered that more features were needed. This is different from trying to predict ToBI indices on a statistical basis, as can be found in extended number of works in the literature – we refer here to one such paper, by Lee, Kim & Lee, 2001. We shall start from Break Indices which amount to 5, starting from 0 to 4 included. Here below the list of features from AME (American) ToBI inventory and their interpretation according to the original manual which can be found at http://www.speech.cs.cmu.edu/tobi/:

0: word boundary apparently erased;
1: typical between-word disjuncture within a phrase;
2: mismatched inter-word disjuncture within a phrase;
3: end of an intermediate phrase;
4: end of an intonational phrase.

We assume that in fact BI 0 is a very special and peculiar index and covers an aspect of prosody which has no general practical application. As for BI 2 we will use it to cover one of the phenomena indicated in the manual, that the idea to indicate a stronger sense of disjuncture than 1, for careful deliberation as in the comment reported in the footnote. From what we can see, we come up with two types of BIs: 3 and 4 are also intonationally related and regard phrase and sentence level prosody. BI 1 and B 2 are to be regarded as pertaining to word level and to possible internal constituent or internal phrase disjuncture. The latter BIs have no effect on the intonational contour. In terms of our analogical parameterization, the former two indices require a ***reset*** at the end that accompany the ***silence***, the latter two have no ***reset***. However, when we completed our experimentation, we discovered that many more disjunctures needed to be marked and for different purposes than the ones indicated in the original ToBI manual. Here is the list we have drawn at the end of our annotation task.

[[slnc 300]],[[rset 0]]	**BI-4**
[[slnc 200]],[[rset 0]]	**BI-3**
[[slnc 100]]	**BI-2**
[[slnc 30]],[[rset 0]]	**BI-32**
[[slnc 50]],[[rset 0]]	**BI-33**
[[slnc 100]],[[rset 0]]	**BI-23**
[[slnc 300]]	**BI-22**
[[slnc 400]]	**BI-44**
[[rate 110; volm +0.3]]	**<slow down**
[[rate 130; volm +0.5]]	**<slow down**

There are different 3 breaks, the reason for that is due to the use of the break in presence of end of Breath Group, with punctuation (BI 3) and without punctuation. The latter case is then split into two subcases, one in which the last word – a syntactic head – is followed or not by a dependent word, hence 33 and 32 respectively are the indices we used. We also use 44 for the separation of the title from the text. Finally 23 is a break with a reset between constituents of a specific type, quantifiers. Then we have two slow down commands, where again one precedes quantifiers, and the other for all syntactic heads, indicating end of Breath Groups. Coming now to tones and accents, the list from the manual includes 14 different combinations. Just to recollect the meaning of some specific label like *downstep*, we report their comment: "downstepped H star: a downstepped H pitch accent indicates that the tone of the prominent syllable is realized by a perceptually lower f0 than that of an immediately preceding High tone: the tone has 'stepped down' from the preceding High."

It follows that since this is High tone realized with a Lowering movement there is no downstepped L star index. Also consider Bitonal pitch accents which are two tone as part of the same pitch accent where the use of + indicates that the two tones are associated and form a single unit: a complex bitonal accent. We assume that a bitonal pitch accents can only be used for special cases of surprise or for expressing contrast. So they would be rather rare in real texts and dialogues. We then decided to port it to more common structures and will show one such usage below which however does not appear in the original list of the manual.

In our new ToBI list, we find downstepped tones to be rather common. They would be necessary every time a sequence of two or more clauses are expressed in a sequence and are interconnected by some discourse relation, usually through the use of a discourse marker.

We report here below a table which includes the new classification where we associate Analogical Parameters used for TTS with their symbolic representation in ToBI.

Table 2 - *Description of Tones Pitch Accents with their corresponding Analogical Parameterization*

Description	**Analogical Pramaterization**	**ToBI**
Beginning of Text for the title	pbas 38.000; rate 160; volm +0.5	**H*-L**
End of Breath Group sentence internally	pbas 38.000; rate 130; volm +0.3	**H*-L%**
	slnc 200, rset 0	**BI 3**
Beginning of sentence when expressing an up movement and a foreground relevance in discourse structure	pbas 44.000; rate 140; volm +0.3	**H*-H**
Beginning of sentence when expressing an up movement and a foreground relevance in discourse structure and it is preceded by a paragraph boundary	pbas 54.000; rate 170; volm +0.3	**H*-H 1**
Sentence internal and a Breath Group boundary	pbas 40.000; rate 140; volm +0.3	**H*-L% 1**

Description	Analogical Pramaterization	ToBI
End of Breath Group and syntactic head	pbas 36.000; rate 110; volm +0.5	**L-L%**
	slnc 50, rset 0	**BI 33**
Sentence internal and a foreground relevance in discourse structure	pbas 40.000; rate 150; volm +0.5	**H*-L**
Sentence internal and an adjunct clause with foreground relevance in discourse structure	pbas 50.000; rate 120; volm +0.5	**H-H* 2**
Sentence internal and an adjunct clause with background relevance in discourse structure	pbas 40.000; rate 120; volm +0.5	**H-H* 4**
	pbas 38.000; rate 130; volm +0.3	**H*-L% 2**
	slnc 200, rset 0	**BI 3**
Direct speech Breath Group boundary + Exclamative	pbas 54.000; rate 170; volm +0.3	**BI 44**
		H*-H%
SAD affective tone associated to a phrase or word	pbas 36.000; rate 110; volm -0.2	**L*-L%**
	rset 0	
Sentence internal and a discourse marker indicates beginning of a subordinate clause	pbas 48.000; rate 150; volm +0.3	**H*-H 3**
	pbas 44.000; rate 140; volm +0.3]	**H-!H* 2**
Sentence internal and a coordinate clause with foreground relevance in discourse structure	pbas 50.000; rate 120; volm +0.5	**H*-H 2**
	pbas 44.000; rate 140; volm +0.3	**H-!H* 2**
Direct Speech + Elaboration or Explanation	pbas 54.000; rate 170; volm +0.3	**H*-H 1**
	pbas 50.000; rate 160; volm +0.5	**H-!H* 1**
Sentence internal and a discourse marker indicates beginning of a subordinate clause	pbas 48.000; rate 150; volm +0.3	**H-!H***
	pbas 44.000; rate 140; volm +0.3]	
Sentence internal and a coordinate clause with foreground relevance in discourse structure	pbas 50.000; rate 120; volm +0.5	**H-!H***
	pbas 44.000; rate 140; volm +0.3	
Declarative sentence with a Resultative Infinitival	slnc 100; pbas 40.000; rate 150; volm +0.5	**H-!L***
	slnc 100; pbas 38.000; rate 150; volm +0.5	
Split Exclamative	pbas 54.000; rate 170; volm +0.3	**H*+L%**
	pbas 36.000; rate 110; volm -0.2	
	rset 0	
Exhortative	pbas 57.000; rate 170; volm +0.5	
	pbas 36.000; rate 170; volm +0.5	**H*+L-**

Description	Analogical Pramaterization	ToBI
	pbas 24.000; rate 130; volm +0.5	
	pbas 60.000; rate 150; volm +0.5	**!L+H*%**
	slnc 100, rset 0	**BI 23**

Here below we explain some of the linguistic pragmatic description we associated to a combination of parameters like Exhortative. This is associated with the complex combination of tones used to express the specific meaning associated to utterances like "Come on, baby", which is to be regarded as an affective invitation. Here below is the set of instructions starting from the command that looks for the expression in the input string, the first sequence of parameters constituted by a bitonal pitch accent made of a very high tone followed by a low one, H*+L- placed on the expression "Come on" and a following tail made by the word "baby" where on the contrary we have a very special case of !L+H*% where the Low tone is upstepped with respect to the previous low.

```
checkfrozen(W,W1,Sent),
['[[pbas 57.000; rate 170; volm +0.5]]',W,
'[[pbas 36.000; rate 170; volm +0.5]]',,,W1,,']
H*+L-

Sent=[Dear,,|Rest],
checkdear(Dear),
'[[pbas 24.000; rate 130; volm +0.5]]',Dear,
!L+H*%
'[[pbas 60.000; rate 150; volm +0.5]]',
'[[slnc 100]],[[rset 0]]']
```

We think in this way to have increased the inventory organized theoretically by the ToBI group, which has shown to be useful when it comes to reorganizing the information needed in a real implementation.

3.3 One specific case: downstepped Direct Speech

Consider now the case of another of the fables by Aesop we worked on – The Fox and the Crow, that can be found here, http://www.taleswithmorals.com/aesop-fable-the-fox-and-the-crow.htm. In this story the main character introduced by the narrator, starts his speech with an exclamative sentence and then continues with some explanation and elaborations. These discourse structures need to be connected to the previous level of intonation. This requires receiving information at clause level from the discourse level, in order to allow for the appropriate continuation. In particular, this is done by:

– detecting the presence of Direct Speech by both verifying the presence of a communication verb governor of a sentence started by the appropriate punctuation mark, inverted commas. This same marker will have to be detected at the end of direct speech. The end may coincide with current sentence or a number of

additional sentences might be present as is the case at stake. The current reported speaker continues after the exclamative with a series of apparently neutral declarative sentences, which can be computed as explanations and elaborations. But they all depend from the exclamative and need to be treated accordingly at discourse level.

To work at discourse level, the system has a switch called "point of view" which takes into account whose point of view is reported in each sentence. The default value for a narrative text would be the "narrator" whenever the sentence is reported directly with no attribution of what is being said. When direct speech appears, the point of view is switched to the character whom the sentence has been attributed to. This switch is maintained until the appropriate punctuation mark appears. So eventually, it is sufficient for the Prosodic Manager (PM) to take the current point_of_view under control. If it is identical to the previous one, nothing happens. If it changes to a new holder and it is marked with direct speech, the algorithm will be transferred to a different recursive procedure which will continue until point_of_view remains identical. This new procedure allows the PM to assign downstepped intonational contours as shown here below. In this fragment, we also mark the presence of a word – HUE – which is wrongly pronounced by Apple synthesizer and requires activating the exceptional phonetic conversion.

"What a noble bird I see **BI-3** above me **BI-22 H*-H-1 ! BI-2 H-!H*-1** Her beauty is without **H*-L%** equal **BI-3** , **H*-L** the [[inpt PHON]]hUW[[inpt TEXT]] of her plumage **H*-H-3** exquisite **BI-2 . H-!H*-1** If only her voice is **BI-2** as sweet **BI-2** as her **BI-2 H-!H*-1** looks are **H*-L** fair **BI-3** , she **BI-2 H-H*-2** ought **L*-L%** without doubt [[rset 0]] to be Queen of the **H*-L%-2** Birds **BI-3** ."

In case this information was not made available to the PM, the result would have been the following.

"What a noble bird I see **BI-3** above me **BI-22 H*-H-1 ! BI-2 H-!H*-1! H*-L%** equal **BI-3** , **H*-L** the [[inpt PHON]]hUW[[inpt TEXT]] of her plumage **H*-H-3** exquisite **BI-2** .
If only her voice is **BI-2** as sweet **BI-2** as her **BI-2 H-!H*-1** looks are **H*-L** fair **BI-3** , she **BI-2 H-H*-2** ought **L*-L%** without doubt [[rset 0]] to be Queen of the **H*-L%-2** Birds **BI-3** ."

5. *Evaluation and Conclusion*

The system has undergone extensive auditory evaluation by expert linguists. It has also been presented at various demo sessions always receiving astounded favourable comments (see Delmonte, Bacalu, 2013; Delmonte, Prati, 2014; Delmonte, 2015). The evaluation has been organized in two phases, at first the story is read by Apple TTS directly from the text. Then the second reading has been done by the system and a comparison is asked of the subject listening to it. In the future we intend to produce an objective evaluation on a graded scale using naïve listeners English

native speakers. We will be using the proposal in Xu (2011: 95), called MOS, or Mean Opinion Score, with a five-level scale: 5-Excellent, 4-Good, 3-Fair, 2-Poor, 1-Bad, with the associated opinions: 5-Imperceptible, 4-Perceptible but not annoying, 4-Slightly annoying, 2-Annoying, 1-Very annoying.

In this paper we presented a prototype of a complete system for expressive and natural reading which is fully based on internal representations produced by syntactic and semantic deep analysis. The level of computation that is mostly responsible for prosodic variations is the discourse level, where both discourse relations, discourse structures, topic and temporal interpretation allow the system to set up an interwoven concatenation of parameters at complex clause and sentence level. Pragmatically frozen phrases and utterances are also fully taken into account always at a parameterized level. Parameters have been related to ToBI standard set and a new inventory has been proposed. The system is currently working on top of Apple TTS but will be soon ported to other platforms. It is available for free download at a specific website: sparsar.wordpress.com.

References

Bachenko, J., Fitzpatrick, E., (1990). A computational grammar of discourse-neutral prosodic phrasing in English. In *Comp. Ling.* 16, 155-170.

Balyan, A., Agrawal, S.S. & Dev, A. (2013). Speech Synthesis: A Review. In *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, Issue 6, 57-75.

Cahn, J.E. (1990). The generation of affect in synthesized speech. In *Journal of the American Voice I/O Society*, 8, 1-19.

Campbell, N. (2002). Towards a grammar of spoken language: incorporating paralinguistic information. In *7th ISCA International Conference on Spoken Language Processing*, Denver, Colorado, USA, September 16-20, 2002.

Campbell's, N., Conclusion for SSW6 TALK on Towards Conversational Speech Synthesis; "Lessons Learned from the Expressive Speech Processing Project". http://www.isca-speech.org/archive_open/archive_papers/ssw6/material/ssw6_022/ssw 6_022.ppt.

Delmonte, R. (1982). Automatic Word-Stress Patterns Assignment by Rules: a Computer Program for Standard Italian. In *Proc. IV F.A.S.E. Symposium*, 1, ESA, Roma, 153-156.

Delmonte, R. (1986). A Computational Model for a text-to-speech translator in Italian. In *Revue - Informatique et Statistique dans les Sciences humaines*, XXII, 1-4, 23-65.

Delmonte, R. (2015). SPARSAR - Expressivity in TTS and its Relations to Semantics. In *Invited Talk at AISV 2015*, Bologna.

Delmonte, R., Bacalu, C. (2013). SPARSAR: a System for Poetry Automatic Rhythm and Style AnalyzeR, *SLATE 2013 - Demonstration Track*, Grenoble.

Delmonte, R., Dolci, R. (1991). Computing Linguistic Knowledge for text-to-speech systems with PROSO. In *Proceedings 2nd European Conference on Speech Communication and Technology*, ESCA, Genova, 1291-1294.

Delmonte, R., Mian, G.A. & Tisato, G. (1984). A Text-to-Speech System for the Synthesis of Italian. In *Proceedings of ICASSP '84*, San Diego(Cal), 291-294.

Delmonte, R., Mian, G.A. & Tisato, G. (1986). A Grammatical Component for a Text-to-Speech System. In *Proceedings of the ICASSP '86, IEEE*, Tokyo, 2407-2410.

Delmonte, R., Prati, A.M. (2014). SPARSAR: An Expressive Poetry Reader. In *Proceedings of the Demonstrations at the 14th Conference of the EACL*, Gotheborg, 73-76.

Grosz, B., Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In *Proc. Int. Conf. Spoken Language Processing, Banff, Canada, 1986,* vol. 1, 429-432.

Hamza, W., Bakis, R., Eide, E.M., Picheny, M.A. & Pitrelli, J.F. (2004). The IBM Expressive Speech Synthesis System. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP),* Jeju, South Korea, October, 2004.

Hirschberg, J. (1993). Pitch accent in context: predicting intonational prominence from text. In *Artificial Intell*. 63, 305-340.

Huckvale, M. (2002). Speech synthesis, speech simulation and speech science. In *Proceedings of the International Conference on Speech and Language Processing 2002*, 1261-1264.

Lee, J-S., Kim, B. & Lee, G.G. (2001). Automatic Corpus-based Tone Prediction using K-ToBI Representation. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*. ACL. downloadable at https://aclweb.org/anthology/W/W01/W01-0517.pdf.

Montaño, R., Alías, F. & Ferrer, J. (2013). Prosodic analysis of storytelling discourse modes and narrative situations oriented to Text-to-Speech synthesis. In *8th ISCA Speech Synthesis Workshop*.

Murray, I.R., Arnott, J.L. (1993). Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. In *Journal of the Acoustic Society of America*, 93 (2), 1097-1108.

Nakatani, C. (1998). Constituent-based accent prediction. In *Proc. COLING-ACL '98, Montreal, Canada*, 939-945.

Pan, S., Mckeown, K. (1997). Integrating language generation with speech synthesis in a Concept-to-Speech system. In *Proc. of ACL/EACL '97 Concept to Speech Workshop*, Madrid, Spain.

Pan, S., Mckeown, K. (1998). Learning intonation rules for concept to speech generation. In *Proc. of COLING/ACL '98*, Montreal, Canada.

Pan, S., Mckeown, K. (1999). Word informativeness and automatic pitch accent modeling. In *Proc. of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*.

Pierrehumbert, J., Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In Cohen, P., Morgan, J. & Pollack, M. (Eds.), *Intentions in Communication*. Cambridge, Mass.: MIT Press, 271-311.

Polzehl, T., Möller, S. & Metze, F. (2011). "Modeling speaker personality using voice". In *Proc. INTERSPEECH*, ISCA.

Prevost, S. (1995). A semantics of contrast and information structure for specifying intonation in spoken language generation. PhD thesis, University of Pennsylvania, Philadelphia, PA, USA.

Saheer, L., Potard, B. (2013). Understanding Factors in Emotion Perception. In *8° Speech Synthesis Workshop*.

Shaikh, M.A.M., Molla, M.K.I. & Hirose, K. (2008). Assigning suitable phrasal tones and pitch accents by sensing affective information from text to synthesize human-like speech. In *Proceedings of InterSpeech*, 326-329.

Shaikh, M.A.M., Prendinger, H. & Ishizuka, M. (2008). Sentiment assessment of text by analyzing linguistic features and contextual valence assignment. In *Applied Artificial Intelligence*, vol. 22, issue 6, 558-601. Taylor & Francis.

Silverman, K., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C., Price, P. & Hirschberg, J. (1992). ToBI: A standard scheme for labelling prosody. In *Proceedings of the International Conference on Speech and Language Processing* 1992.

Steidl, S., Polzehl, T., Bunnell, T., Dou, Y., Muthukumar, P., Perry, D., Prahallad, K., Vaughn, C., Black, A. & Metze, F. (2012). Emotion identification for evaluation of synthesized emotional speech. In *Proc. Speech Prosody*.

Wang, M., Hirschberg, J. (1992). Automatic classification of intonational phrase boundaries. In *Comp. Speech Lang.*, 6, 175-196.

Zovato, E., Tini-Brunozzi, F. & Danieli, M. (2008). Interplay between pragmatic and acoustic level to embody expressive cues in a Text to Speech system. In *AISB Proceedings - Affective Language in Human and Machine*, vol. 2, 88-91.

Appendix: Edge *by Sylvia Plath*

The woman is perfected.
Her dead

Body wears the smile of accomplishment,
The illusion of a Greek necessity

Flows in the scrolls of her toga,
Her bare

Feet seem to be saying:
We have come so far, it is over.

Each dead child coiled, a white serpent,
One at each little

Pitcher of milk, now empty.
She has folded

Them back into her body as petals
Of a rose close when the garden

Stiffens and odors bleed
From the sweet, deep throats of the night flower.

The moon has nothing to be sad about,
Staring from her hood of bone.

She is used to this sort of thing.
Her blacks crackle and drag.

DOI: 10.17469/O2101AISV000027

PIERO COSI, GIULIO PACI, GIACOMO SOMMAVILLA, FABIO TESSER

KALDI: Yet Another ASR Toolkit? Experiments on Italian children speech

In this paper, the KALDI ASR engine adapted to Italian is described and the results obtained so far on some children speech ASR experiments are reported. We give a brief overview of KALDI, we describe in detail its DNN implementation, we introduce the acoustic model (AM) training procedure and we end describing some ASR experiments on Italian children speech together with the final test procedures.

1. *Introduction*

Deep Neural Networks (DNNs) are the latest hot topic in speech recognition. Since around 2010, many papers have been published in this area, and some of the largest companies (e.g. Google, Microsoft) are starting to use DNNs in their production systems. Moreover, many different Automatic Speech Recognition frameworks have been developed for research purposes during the last years and, nowadays, various open-source automatic speech recognition (ASR) toolkits are available to even small research laboratories. Systems such as HTK (Young et al., 2009), SONIC (Pellom, 2001), (Pellom, Hacioglu, 2003), SPHINX (Lee et al., 1990), (Walker et al., 2004), RWTH (Rybach et al., 2009), JULIUS (Lee et al., 2001), KALDI (Povey et al., 2011), the more recent ASR framework SIMON (Simon, WEB), and the relatively new system called BAVIECA (Bolaños, 2012) are a simple and probably not exhaustive list.

Indeed new systems such as KALDI (Povey et al., 2011) demonstrated the effectiveness of easily incorporating "Deep Neural Network" (DNN) techniques (Bengio, 2009) in order to improve the recognition performance in almost all recognition tasks.

In this paper, the KALDI ASR engine adapted to Italian is described and the results obtained so far on some children speech ASR experiments are reported. We give a brief overview of KALDI, and in particular of its DNN implementation, we introduce the acoustic model (AM) training procedure and we end describing some experiments on Italian children speech together with the final test procedures.

2. *KALDI*

As written in his official web site (http://KALDI.sourceforge.net), the KALDI ASR environment should be mainly taken into consideration for the following simple reasons:

- it's "easy to use" (once you learn the basics, and assuming you understand the underlying science);
- it's "easy to extend and modify";
- it's "redistributable": unrestrictive license, community project;
- if your stuff works or is interesting, the KALDI team is open to including it and your example scripts in our central repository: more citation, as others build on it.

Another main reason for choosing KALDI is represented by the fact that complete open source recipes for building speech recognition systems, that work from widely available databases such as those provided by the ELRA or Linguistic Data Consortium (LDC), are easily available for testing the system.

3. *DNN (Deep Neural Networks) in KALDI*[1]

KALDI currently contains two parallel implementations for DNN training. Both of these recipes are deep neural networks where the last (output) layer is a softmax layer whose output dimension equals the number of context-dependent states in the system (typically several thousand). The neural net is trained to predict the posterior probability of each context-dependent state. During decoding the output probabilities are divided by the prior probability of each state to form a "pseudo-likelihood" that is used in place of the state emission probabilities in the HMM.

The first implementation (Kaldi, WEB-a), (Vesely et al., 2013) supports Restricted Boltzmann Machines (RBM) pre-training (Hinton et al., 2012), (Dahl et al., 2012), (Seide et al., 2011), stochastic gradient descent training using NVidia graphics processing units (GPUs), and discriminative training such as boosted MMI (Povey et al., 2008) and state-level minimum Bayes risk (sMBR) (Gibson, Hain, 2006), (Povey, Kingsbury, 2007).

The second implementation of DNNs in KALDI (Kaldi, WEB-b), (Zhang et al., 2014), (Povey et al., 2015) was originally written to support parallel training on multiple CPUs, although it has now been extended to support parallel GPU-based training and it does not support discriminative training.

One is located in code sub-directories nnet/ and nnetbin/[2], and is primarily maintained by Karel Vesely. The other is located in code subdirectories nnet2/ and nnet2bin/, and is primarily maintained by Daniel Povey (this code was originally

[1] Most of the text used to briefly describe Kaldi in this section is taken from "Deep Neural Networks in Kaldi" (http://kaldi.sourceforge.net/dnn.html) with permission from the Author (Daniel Povey).
[2] See the code at: https://svn.code.sf.net/p/kaldi/code/trunk/egs.

based on an earlier version of Karel's code, but it has been extensively rewritten). Neither codebase is more "official" than the other. Both are still being developed in parallel.

In the example directories (referring to the HKUST Mandarin Telephone Speech, Resource Management, Switchboard, Timit, and Wall Street Journal corpora)[3] neural net example scripts can be found. Before running those scripts, the first stages of "run.sh" in those directories must be run in order to build the systems used for alignment.

Regarding which of the two setups you should use:

– Karel's setup (nnet1) generally gives somewhat better results but it only supports training on a single GPU card, or on a single CPU which is very slow.
– Dan's setup generally gives slightly worse results but is more flexible in how you can train: it supports using multiple GPUs, or multiple CPU's each with multiple threads. Multiple GPU's is the recommended setup. They don't have to all be on the same machine.

The reasons for the performance difference is still unclear, as there are many differences in the recipes used. For example, Karel's setup uses pre-training but Dan's setup does not; Karel's setup uses early stopping using a validation set but Dan's setup uses a fixed number of epochs and averages the parameters over the last few epochs of training. Most other details of the training (nonlinearity types, learning rate schedules, etc.) also differ.

3.1 Karel's DNN training implementation

The implementation of DNNs from Karel Vesely (Kaldi, WEB-a), (Vesely et al., 2013) uses the following techniques:

– layer-wise pre-training based on RBMs (Restricted Boltzmann Machines);
– per-frame cross-entropy training;
– sequence-discriminative training, using lattice framework, optimizing sMBR criterion (State Minimum Bayes Risk).

The systems are built on top of MFCC, LDA, MLLT, fMLLR with CMN[4] features – see (Rath et al., 2013) for all acronyms references – obtained from auxiliary GMM (Gaussian Mixture Model) models. Whole DNN training is run in a single GPU using CUDA (Compute Unified Device Architecture, the parallel computing architecture created by NVidiaTM). However, cudamatrix library is designed to also run on machines without a GPU, but this tends to be more than 10x slower.

The script for standard databases such wsj is located at: "egs/wsj/s5/local/run_dnn.sh" and it is split into several stages. At first the 40-dimensional features

[3] such as egs/hkust/s5b, egs/rm/s5, egs/swbd/s5, egs/timit/s5/, and egs/wsj/s5/. Karel's example scripts can be found in local/run_dnn.sh or local/run_nnet.sh, and Dan's example scripts can be found in local/run_nnet2.sh.

[4] MFCC: Mel-Frequency Cepstral Coefficients; LDA: Linear Discriminant Analysis; MLTT: Maximum Likelihood Linear Transform; fMLLR: feature space Maximum Likelihood Linear Regression; CMN: Cepstral Mean Normalization.

(MFCC, LDA, MLLT, fMLLR with CMN) are stored to disk in order to simplify the training scripts.

3.1.1 Pre-training phase

The implementation of layer-wise RBM (Restricted Boltzmann Machine) pre-training is following "www.cs.toronto.edu/~hinton/absps/guideTR.pdf" (Hinton, 2010).

The training algorithm is Contrastive Divergence with 1-step of Markov Chain Monte Carlo sampling (CD-1). The hyper-parameters of the recipe were tuned on the 100 hours Switchboard subset. If smaller databases are used, mainly the number of epochs N needs to be set to 100 hours/set_size. The training is unsupervised, so it is sufficient to provide single data-directory with input features.

When training the RBM with Gaussian-Bernoulli units, there is a high risk of weight-explosion, especially with larger learning rates and thousands of hidden neurons. To avoid weight-explosion, a mechanism has been implemented, which compares the variance of training data with the variance of the reconstruction data in a minibatch. If the variance of reconstruction is > 2x larger, the weights are shrinked and the learning rate is temporarily reduced.

3.1.2 Frame-level cross-entropy training

In this phase a DNN which classifies frames into triphone-states is trained. This is done by mini-batch Stochastic Gradient Descent[5]. The default is to use Sigmoid hidden units, Softmax output units and fully connected layers AffineTransform. The learning rate by default is 0.008, size of mini-batch 256; no momentum or regularization is used. The optimal learning-rate differs with type of hidden units, the value for sigmoid is 0.008, for tanh 0.00001.

The input_transform and the pre-trained DBN (i.e. Deep Belief Network, stack of RBMs) is passed into to the script using the options '–input-transform' and '–dbn'; only the output layer is initialized randomly. An early stopping criterium is used to prevent over-fitting: for this, the objective function on the cross-validation set (i.e. held-out set) is measured. Therefore two pairs of feature-alignment dirs are needed to perform the supervised training.

3.1.3 Sequence-discriminative training

In this phase, the neural network is trained to classify correctly the whole sentences, which is closer to the general ASR objective than frame-level training. The objective of sequence-discriminative training is to maximize the expected accuracy of state labels derived from reference transcriptions: lattice framework to represent competing hypothesis is used. The training is done by Stochastic Gradient Descent (SGD) with per-utterance updates, low learning rate: 1e-5 kept constant is used and

[5] A good summary paper on DNN training is "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups" by Geoffrey Hinton's group (Hinton et al., 2012).

3-5 epochs are done. Faster convergence when re-generating lattices after 1st epoch are observed.

MMI, BMMI, MPE and sMBR[6] training are all supported – (see Rath et al., 2013 for all acronyms references). All the techniques perform very similarly, even if sMBR is a little bit better. In sMBR optimization silence frames are excluded from accumulating approximate accuracies[7] (Vesely et al., 2013).

3.2 Povey's DNN training implementation

For the full documentation that covers Dan Povey's version of the deep neural network code in KALDI one could refer to the following web link (Kaldi, WEB-b), and to the following papers (Zhang et al., 2014; Povey et al., 2015).

In its last implementation stage, as indicated in (Povey et al., 2015), where the Dan's DNN is used on a speech recognition setup called Fisher English, which is English language conversational telephone speech, sampled at 8kHz, for a total amount of training data of 1600 hours, the "... DNN system uses speaker adapted features from a GMM system, so it requires a first pass of GMM decoding and adaptation ...".

> ... The GMM system is based on MFCC features, spliced across ±3 frames and processed with LDA+MLLT to 40-dimensional features, then adapted with feature-space MLLR (fMLLR) in both training and test time. See (Povey, Ghoshal, 2011) for an explanation of these terms and the normal system build steps. All these systems used the same phonetic context decision tree with 7880 context-dependent states; the GMM system had 300000 Gaussians in total ...
> ... The 40-dimensional features from GMM are spliced across ±4 frames of context and used as input to the DNN. DNN is a p-norm DNN (Zhang et al., 2014), with 5 hidden layers and p-norm (input, output) dimensions of (5000, 500) respectively, i.e. the nonlinearity reduces the dimension tenfold. In this framework 15000 "sub-classes" are used – see Section C.3 of (Povey et al., 2015) for explanation – and the number of parameters is 19.3 million. The system is trained for 12 epochs with learning rate varying from 0.08 to 0.008, trained with 8 parallel jobs with online natural gradient SGD (NG-SGD) and, for this DNN system, K=400000 samples per outer iteration for each machine are used for training...

As for the TIMIT recipe (s5), Dan's DNN is much simpler and adopts a classic Hybrid Training and Decoding framework using a simple deep network with tanh nonlinearities. Moreover, also system combination using minimum Bayes risk decoding is used, and in this case a lattice combination is used to create a union of lattices normalized by removing the total forward cost from them and using the resulting lattice as input to the last decoding step.

---

[6] MMI: Maximum Mutual Information; BMMI: Boosted MMI; MPE: Minimum Phone Error; sMBR: State-level Minimum Bayes Risk.

[7] More detailed description is at: http://www.danielpovey.com/files/2013_interspeech_dnn.pdf.

4. *KALDI on Italian*

In this section, the adaptation of KALDI to Italian is described, starting from the TIMIT recipe, and the results obtained so far on some children speech ASR experiments are reported (see Table 1).

In the experiments described below, the Italian FBK ChildIt Corpus (Gerosa et al., 2007) was taken into consideration. This is a corpus that counts almost 10 hours of speech from 171 children; each child has read about 60 children literature sentences; the audio was sampled at 16 kHz, 16 bit linear, using a Shure SM10A head-worn mic.

Table 1 - *Preliminary results obtained in the experiments executed on the CHILDIT corpus in various configurations adapting the KALDI's TIMIT-recipe scripts (see text for the definition of all keywords)*

	Training & Decoding	%PCR	%SUB	%DEL	%INS	%PER	%SER
MonoPhone	mono	81.6	10.5	7.9	2.5	20.8	99.6
Delta + Delta-Deltas	tri1	87.5	7.1	5.4	1.6	14.1	97.6
LDA + MLTT	tri2	88.7	6.2	5.1	1.4	12.7	96.7
LDA + MLTT + SAT	tri3	91.0	4.9	4.1	1.2	10.2	93.4
sgmm2_4: SGMM2	sgmm2_4	92.1	4.1	3.8	1.0	8.9	91.0
MMI + SGMM2 (iteration n.1)	sgmm2_4_mmi_b0.1	92.8	4.0	3.2	1.5	8.7	90.8
MMI + SGMM2 (iteration n.2)	sgmm2_4_mmi_b0.2	92.9	4.0	3.1	1.6	8.7	89.9
MMI + SGMM2 (iteration n.3)	sgmm2_4_mmi_b0.3	92.9	4.0	3.1	1.6	8.7	90.1
MMI + SGMM2 (iteration n.4)	sgmm2_4_mmi_b0.4	92.9	4.0	3.1	1.6	8.7	90.2
DNN Hybrid (Dan's)	tri4-nnet	90.6	4.8	4.5	1.8	11.1	94.1
SGMM2 + DNN Hybrid (Dan's) (it. 1)	combine_2 (1)	92.9	4.0	3.1	1.3	8.4	89.7
SGMM2 + DNN Hybrid (Dan's) (it. 2)	combine_2 (2)	92.9	4.0	3.1	1.3	8.4	89.9
SGMM2 + DNN Hybrid (Dan's) (it. 3)	combine_2 (3)	92.8	4.0	3.3	1.1	8.3	89.0
SGMM2 + DNN Hybrid (Dan's) (it. 4)	combine_2 (4)	93.0	4.0	3.0	1.3	8.4	89.5
DNN Hybrid (Karel's)	dnn4_pretrain-dbn_dnn	92.7	3.9	3.4	1.4	8.6	90.6
DNN Hybrid (Karel's), sMBR training (it. 1)	dnn4_pretrain-dbn_dnn_smbr (1)	92.9	3.8	3.3	1.2	8.3	89.8
DNN Hybrid (Karel's), sMBR training (it. 6)	dnn4_pretrain-dbn_dnn_smbr (6)	93.2	3.7	3.0	1.4	8.1	88.6

Various experiments have been carried out in many configurations and in all cases, training and test materials have been kept separate. In all the experiments described in this work, the standard MFCC features were considered, setting reasonable defaults, even if other options could be exploited in the future. The results, shown in Figure 1, are the best obtained with KALDI on the CHILDIT corpus and they are the best obtained so far in comparison with those obtained by similar experiments reported in the literature (Gerosa et al., 2007; Giuliani, Gerosa, 2003; Cosi, Pellom, 2005; Cosi, 2008; Cosi, 2009; Cosi, Hosom, 2000; Cosi et al., 2014).

Phone Error Rate (PER) was considered for computing the score of the recognition process. The PER is defined as the sum of the deletion (DEL), substitution (SUB) and insertion (INS) percentage of phonemes in the ASR outcome text with respect to a reference transcription. PCR and SER refers to the Phone Correct Rate and Sentence Error Rate respectively.

Ideally, a hand-labelled reference would have been preferred, because it would have been corrected at the phonetic level to take into account children's speech pronunciation mistakes. Since this was not available for the CHILDIT corpus, the automatic phonetic sequences obtained by a Viterbi alignment of the word-level or-

thographic transcription have been used. The reference test transcriptions were created with a SONIC-based aligner using a previously trained children speech Italian acoustic model (Cosi, Pellom, 2005). This method was chosen because it allowed for automatically selecting the best pronunciation for each word in the training data among the alternative choices available in the 400,000-word Italian lexicon available.

The results shown in Table 1 refer to the various training and decoding experiments – see (Rath et al., 2013) for all acronyms references:

- MonoPhone (mono);
- Deltas + Delta-Deltas (tri1);
- LDA + MLLT (tri2);
- LDA + MLLT + SAT (tri3);
- SGMM2 (sgmm2_4);
- MMI + SGMM2 (sgmm2_4_mmi_b0.1-4);
- Dan's Hybrid DNN (tri4-nnet),
- system combination, that is Dan's DNN + SGMM (combine_2_1-4);
- Karel's Hybrid DNN (dnn4_pretrain-dbn_dnn);
- system combination that is Karel's DNN + sMBR (dnn4_pretrain-dbn_dnn_1-6).

In the Table, SAT refers to the Speaker Adapted Training (SAT), i.e. train on fMLLR-adapted features. It can be done on top of either LDA+MLLT, or delta and delta-delta features.

If there are no transforms supplied in the alignment directory, it will estimate transforms itself before building the tree (and in any case, it estimates transforms a number of times during training). SGMM2 refers instead to SGMM training (Huang, Hasegawa-Johnson, 2010) with speaker vectors. This training would normally be called on top of fMLLR features obtained from a conventional system, but it also works on top of any type of speaker-independent features (based on deltas+delta-deltas or LDA+MLLT).

On this difficult phonetic ASR task, Karel's DNN looks slightly better than Dan's DNN and it outperforms all other non-DNN configurations.

5. *Conclusions*

At its early stage, the KALDI toolkit supported modeling of context-dependent phones of arbitrary context lengths, all commonly used techniques that can be estimated using maximum likelihood, and almost all adaptation techniques available in the ASR literature. At present, it also supports the recently proposed Semi-supervised Gaussian Mixture Model (SGMMs) (Huang, Hasegawa-Johnson, 2010), (Povey et al., 2011), discriminative training, and the very promising DNN hybrid training and decoding (Kaldi, WEB-a; Vesely et al., 2013; Kaldi, WEB-b; Zhang et al., 2014; Povey et al., 2015). Moreover, developers are working on using

large language models in the FST framework, and the development of KALDI is continuing.

The adaptation of KALDI to Italian, and in particular to the ChildIt corpus, was indeed quite straightforward, and results are truly exceptional with respect to those previously obtained on the same data (Gerosa et al., 2007; Giuliani, Gerosa, 2003; Cosi, Pellom, 2005; Cosi, 2008; Cosi, 2009; Cosi, Hosom, 2000; Cosi et al., 2014). Indeed, this is mainly because all the very last ASR techniques, including DNNs, could be easily implemented by adapting to Italian the already available downloadable scripts written by a quite large number of various developers around the world for a similar task in English based on TIMIT.

Acknowledgements

We would like to acknowledge and thank Daniel Povey for his invaluable work on KALDI, indeed a great ASR Toolkit, and for his great help and kindness in letting us use some of the text of the KALDI's web pages describing the software. Moreover, we would like to acknowledge Karel Vesely for his great work on his very effective version of DNN and for his great help and kindness in supporting me on Italian implementation.

References

Bengio, Y. (2009). Learning Deep Architectures for AI. In *Foundations and Trends in Machine Learning*, vol. 2, n. 1, 1-127.

Bolaños, D. (2012). The Bavieca Open-Source Speech Recognition Toolkit. In *Proceedings of IEEE Workshop on Spoken Language Technology (SLT)*, December 2-5, 2012, Miami, FL, USA.

Cosi, P. (2008). Recent Advances in Sonic Italian Children's Speech Recognition for Interactive Literacy Tutors. In *Proceedings of 1st Workshop On Child, Computer and Interaction (WOCCI-2008)*, Chania, Greece.

Cosi, P. (2009). On the Development of Matched and Mismatched Italian Children's Speech Recognition Systems. In *Proceedings of INTERSPEECH 2009*, Brighton, UK, 540-543.

Cosi, P., Hosom, J.P. (2000). "High Performance General Purpose Phonetic Recognition for Italian". In *Proceedings of ICSLP 2000*, Beijing, 527-530.

Cosi, P., Nicolao, M., Paci, G., Sommavilla, G. & Tesser, F. (2014). Comparing Open Source ASR Toolkits on Italian Children Speech. In *Proceedings of Workshop On Child, Computer and Interaction (WOCCI-2014)*, Satellite Event of INTERSPEECH 2014, Singapore, September 19, 2014.

Cosi, P., Pellom, B. (2005). Italian Children's Speech Recognition For Advanced Interactive Literacy Tutors. In *Proceedings of INTERSPEECH 2005*, Lisbon, Portugal, 2201-2204.

Dahl, G.E., Dong, Yu, D., Deng, L. & Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. In *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, n. 1, 30-42.

Gerosa, M., Giuliani, D. & Brugnara, F. (2007). Acoustic Variability and automatic recognition of children's speech. In *Speech Communication*, vol. 49.

Giuliani, D., Gerosa, M. (2003). Investigating Recognition of Children's Speech". In *Proceedings of ICASSP, Hong Kong*.

Gibson, M., Hain, T. (2006). Hypothesis Spaces For Minimum Bayes Risk Training In Large Vocabulary Speech Recognition. In *Proceedings of INTERSPEECH*.

Hinton, G., (2010). *A Practical Guide to Training Restricted Boltzmann Machines*. UTML TR 2010:003, Momentum, 9:1.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Abdelrahman, M., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. In *Signal Processing Magazine, IEEE*, vol. 29, n. 6, 82-97.

Huang, J.T., Hasegawa-Johnson, M. (2010). Semi-Supervised Training of Gaussian Mixture Models by Conditional Entropy Minimization. In *Proceedings of INTERSPEECH 2010*, 1353-1356.

Kaldi-WEBa - Karel's DNN implementation: http://KALDI.sourceforge.net/dnn1.html.

Kaldi-WEBb - Dan's DNN implementation: http://KALDI.sourceforge.net/dnn2.html.

Lee, A., Kawahara, T. & Shikano, K. (2001), JULIUS - an open source real-time large vocabulary recognition engine. In *Proceedings of INTERSPEECH 2001*, 1691-1694.

Lee, K.F., Hon, H.W. & Reddy, R. (1990). An overview of the SPHINX speech recognition system. In *IEEE Transactions on Acoustics, Speech and Signal Processing* 38.1, 35-45.

Pellom, B. (2001). "SONIC: The University of Colorado Continuous Speech Recognizer". In *Technical Report TR-CSLR-2001-01*, Center for Spoken Language Research, University of Colorado, USA.

Pellom, B., Hacioglu, K. (2003). Recent Improvements in the CU SONIC ASR System for Noisy Speech: The SPINE Task. In *Proceedings of ICASSP*.

Povey, D., Burget, L. et al. (2011). The subspace Gaussian mixture model-A structured model for speech recognition. In *Computer Speech & Language*, vol. 25, n. 2, 404-439.

Povey, D., Ghoshal, A. et al. (2001). The KALDI Speech Recognition Toolkit. In *Proceedings of ASRU*.

Povey, D., Kanevsky, D., Kingsbury, B., Ramabhadran, B., Saon, G. & Visweswariah, K. (2008). Boosted MMI for Feature and Model Space Discriminative Training. In *Proceedings of ICASSP*.

Povey, D., Kingsbury, B. (2007). Evaluation of proposed modifications to MPE for large scale discriminative training. In *Proceedings of ICASSP*.

Povey, D., Zhang, X. & Khudanpur, S. (2015). Parallel Training of DNNs with Natural Gradient and Parameter Averaging (*under review as a conference paper at ICLR 2015*).

Rath, S.P., Povey, D., Vesely, K. & Cernocky, J. (2013). Improved feature processing for Deep Neural Networks. In *Proceedings of INTERSPEECH 2013*, 109-113.

Rybach, D., Gollan, C., Heigold, G., Hoffmeister, B., Lööf, J., Schlüter, R. & Ney, H. (2009). The RWTH Aachen University Open Source Speech Recognition System. In *Proceedings of INTERSPEECH 2009*, 2111-2114.

Seide, F., Li, G. & Yu, D. (2011). Conversational speech transcription using context-dependent deep neural networks. In *Proceedings of INTERSPEECH, 2011*, 437-440.

Simon-WEB: http://www.simon-listens.com.

Vesely, K., Ghoshal, A., Burget, L. & Povey, D. (2007). Sequence - Discriminative Training of Deep Neural Networks. In *Proceedings of INTERSPEECH 2013*, 2345-2349.

Vesely, K., Ghoshal, A., Burget, L. & Povey, D. (2013). Sequence-discriminative training of deep neural networks. In *Proceedings of INTERSPEECH 2013*.

Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., Wolf, P. & Woelfel, J. (2004). *Sphinx-4: A flexible Open Source Framework for Speech Recognition*. Sun Microsystems Inc., Technical Report SML1 TR2004-0811.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V. & Woodland, P. (2009). *The HTK Book* (version 3.4). Cambridge Univ. Eng. Dept.

Zhang, X., Trmal, J., Povey, D. & Khudanpur, S. (2014), Improving Deep Neural Network Acoustic Models Using Generalized Maxout Networks. In *Proceedings of ICASSP 2014*.

DOI: 10.17469/O2101AISV000028

MASSIMILIANO TODISCO, FABIO POROLI, MAURO FALCONE

Uno strumento per la prototipizzazione rapida di "dialoghi pratici"

By the term "practical dialogue" we mean what is necessary to a voice interaction with-out particular limitations regarding the naturalness and freedom of the language used, but on the fulfillment of a specific goal. The tool, or better the tools, that arise refer to two well-defined and specific phases of development of a speech dialogue: a first step having the purpose of evaluation of the tools and characterization of dialogue and voice interaction with the user through the simulation of a hypothetical and ideal system simulated in part or in whole by an operator with the technique of the Wizard of Oz; a second which allows, quickly and easily, to provide a system of voice dialog exploiting third-party technologies for both voice technologies and language processing.

1. *Introduzione*

L'utilizzo delle tecnologie di riconoscimento vocale (*ASR Automatic Speech Recogni-tion*) e di sintesi vocale da testo (*TTS Text to Speech Synthesis*) trova oramai un'ampia serie di possibili applicazioni in servizi e automatizzazioni. L'ambizione ultima è certamente quella di realizzare un qualcosa con cui si possa dialogare e che riesca a capirci ed aiutarci nella risoluzioni di richieste o compiti più o meno complessi. Se questo obiettivo è stato per anni mitizzato nella comune opinione, basti ricordare le scene di dialogo con il computer HAL del film di Kubrick "2001 Odissea nello spazio" del 1968, solo nel 2012 con il rilascio da parte di Apple del "iPhone 4S" con il software SIRI si è per la prima volta osato fornire a livello di generico consumatore "*an intelligent assistant that helps you get things done just by asking. Siri understands context allowing you to speak naturally when you ask it questions* (Apple, 2011)". A onore del vero dobbiamo ricordare che sofisticati sistemi ad interazione vocale erano già utilizzati con successo nell'automatizzazione di servizi telefonici ed in particolare nell'ambito del CRM (*Customer Relationship Management*) (Minker, 2011), ma questi operavano su compiti ed in scenari specifici. La barriera tecnologica e concettuale da oltrepassare era proprio quella di un 'unico assistente virtuale tuttofare' a cui ci si possa rivolgere parlando normalmente. Nell'arco di pochi anni, si è visto come i principali attori interessati stiano dotando i loro sistemi di 'assistenti' che possano interagire vocalmente con l'utente. Siri per Apple, Cortana per Microsoft, Nina per Nuance sono solo alcuni esempi, vedi figura 1, molte altre soluzioni sono fornite da società o centri di ricerca che propongono i loro 'assistenti vocali' per specifici compiti o, con grande ambizione, totalmente aperti a qualsiasi tipo di interazione.

Figura 1 - *Sistemi di dialogo vocale*

La 'lampada' che ospita questa specie di 'géni' al nostro servizio e con cui possiamo interagire con la voce è, nei casi citati, un semplice *smartphone*, ma potrebbe essere anche un *tablet*, un *personal computer* o altro dispositivo che utilizziamo comunemente per svolgere altre attività. È comunque una soluzione di tipo software che va ad aggiungersi come un servizio o una funzionalità supplementare. Tuttavia l'importanza di un 'assistente vocale' sta crescendo sempre più negli scenari tecnologi, fino a divenire esso stesso l'elemento fondante. Sparisce quindi ogni altro elemento interferente e l'assistente vocale si materializza semplicemente in entità vagamente antropomorfe (Jibo, 2014), o in una specie di neutro totem familiare (Amazon, 2014), sino a dematerializzarsi del tutto come auspicato da soluzioni ancora in fase di studio (Dirha, 2014), e che ricordano, citando nuovamente un immaginario collettivo, quanto disponibile nella serie televisiva "*Star Trek: The Next Generation*".

Auspicare che tali soluzioni potessero essere realtà, è stato possibile solo dopo che la ricerca nella sintesi vocale e nel riconoscimento del parlato avessero raggiunto prestazioni ragguardevoli e comunque sufficienti per affrontare l'ultimo è più difficile problema: il dialogo vocale. Sebbene la ricerca sul riconoscimento del parlato e sulla sintesi vocale non si siano concluse (anzi, ad esempio, il riconoscimento sta vivendo una interessante rinascita con i paradigmi di "*deep learning*" (Yu, 2015), oggi è certamente il dialogo il problema che aziende e centri di ricerca devono risolvere. Se da un lato sono chiari gli obiettivi da raggiungere, ovvero naturalezza del dialogo e ampiezza del dominio di dialogo, dall'altro non è ancora evidente la strategia da adottare: rimangono infatti sulla cresta dell'onda sistemi basati su regole e sull'analisi del testo trascritto dal segnale vocale, ma al contempo si spera che, come è successo per il riconoscimento del parlato, il problema possa essere risolto con un approccio totalmente statistico sulla base di un apprendimento di una grande mole di dati di esempio (Thomson, 2013). Quest'ultima soluzione, forse più interessante, richiede tuttavia risorse che, come abbiamo visto per il riconoscimento vocale, solo grandi aziende come Google, Amazon, etc. possono permettersi. È quindi facile profetizzare che soluzioni di questo tipo nella risoluzione del dialogo vocale potranno a breve essere fornite da

queste o da analoghe aziende, similmente a quanto, ad esempio, Google ha fatto per le tecnologie di ASR e di TTS.

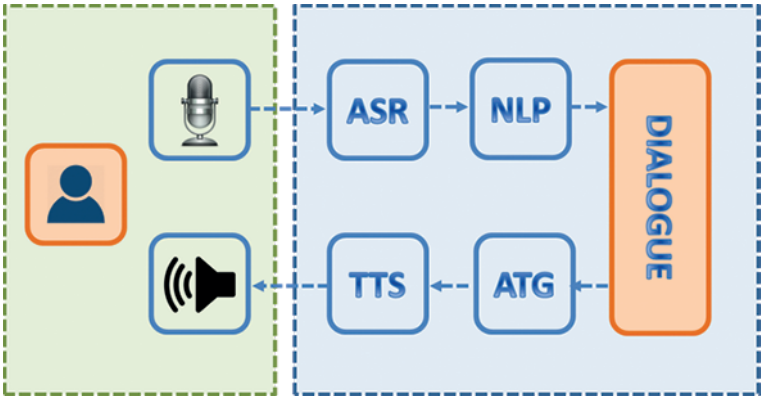
Chi voglia oggi trovare una soluzione per un sistema basato su interazione vocale, può quindi utilizzare con soddisfazione soluzioni di riconoscimento e di sintesi vocale di terzi, ma dovrà risolvere autonomamente il problema dialogo o utilizzare, qualora lo ritenga opportuno, sistemi di sviluppo specifici e vincolati a soluzioni proprietarie come nei casi già citati. Qualora invece si voglia ricorrere a strumenti terzi a supporto dello sviluppo dell'interfaccia di dialogo, il panorama delle soluzioni offerte non è così promettente (McTear, 2004), specialmente se consideriamo il caso piuttosto frequente dove si voglia realizzare, con uno sforzo ragionevole o meglio minimo, un dialogo vocale per comandare un sistema già esistente (ad esempio un televisore), o si voglia trasformare un servizio esistente in un servizio fruibile con la sola interazione vocale (ad esempio la prenotazione di un treno, di un aereo, di un albergo, etc.), o ancora si voglia sostituire un operatore di "*call center*" con un sistema automatico, ad esempio un centro della Pubblica Amministrazione che informi e guidi l'utente relativamente a nuove disposizioni e/o regolamentazioni, etc.

Precisamente in questo ambito si inquadra il sistema in oggetto: avere la possibilità di sviluppare, ottimizzare e realizzare un dialogo vocale reale per compiere uno specifico compito sfruttando, quando possibile, le tecnologie di riconoscimento, di sintesi, di elaborazione del testo riconosciuto (NLP *Natural Language Processing*) e di generazione automatica del testo per le risposte (ATG *Automatic Text Generation*) disponibili da terze parti. Questa attività è stata condotta nell'ambito del progetto 'Speaky-Acutattile' di "Industria 2015", il cui obiettivo è la realizzazione di una piattaforma a supporto di utenti 'deboli' (anziani, ipovedenti, non vedenti, disabili motori, etc.) anche attraverso l'acceso tramite dialogo vocale a servizi, applicazioni e sistemi.

2. *Disegno, valutazione e realizzazione del dialogo*

Premesso che lo scopo ultimo è un sistema di dialogo vocale che permetta all'utente di raggiungere il completamento di un obiettivo in uno specifico 'dominio', dove con questo termine intendiamo l'utilizzo di un servizio, il controllo di un sistema, l'accesso a delle informazioni, la guida o indirizzamento ('*steering*') dell'utente in ambienti o servizi complessi, etc., è evidente che il raggiungimento di questo obiettivo deve, o comunque può, passare attraverso una serie di specifiche fasi.

Figura 2 - *Schema a blocchi della interfaccia di dialogo vocale*



In particolare qualora ci si trovi ad affrontare il problema della realizzazione di un dialogo utilizzando risorse tecnologiche nuove, nel senso da noi mai utilizzate prima, o senza aver ancora maturato sufficiente esperienza nel disegno del dialogo, è bene avere strumenti che ci permettano il controllo e la valutazione dei singoli elementi, dialogo compreso.

Insomma ogni qual volta nel nostro complesso sistema entrino in gioco elementi che non abbiamo sperimentato prima, ad esempio un nuovo sistema di ASR, oppure più semplicemente un diverso tipo di interfaccia microfonica, e così via, oppure dove la classe di utenti è caratterizzabile da elementi ben noti, ad esempio turisti con scarsa conoscenza della lingua italiana, ed in generale ogni qual volta abbiamo delle variabili, ovvero delle parti del sistema, si veda lo schema di figura 2, che possono influenzare le prestazioni finali del nostro sistema e di cui non abbiamo sufficienti informazioni sulle prestazioni, è bene avere uno strumento per valutare al meglio ciascuna di queste variabili, le loro mutue influenze e infine l'impatto sull'intero sistema.

Per meglio chiarire questa problematica supponiamo di avere un sistema di dialogo 'ideale' e che vogliamo studiare l'impatto che un approccio all'acquisizione del segnale vocale di tipo "*push to talk*" (in questo caso l'utente deve premere un pulsante per attivare il microfono e inviare il segnale vocale al riconoscitore) rispetto a uno "*open mic*" (in questo caso invece il microfono è sempre attivo e quindi il sistema acquisisce continuamente il segnale vocale) ha sul dialogo. Oppure supponiamo di voler valutare come due o più sistemi ASR si comportano, o meglio impattano sul dialogo. Avremo quindi bisogno di configurare un sistema di dialogo dove solo alcune componenti sono effettivamente quelle che vogliamo valutare, ovvero le nostre variabili, mentre tutto il resto deve essere 'ideale', teoricamente ad errore zero (o comunque non variabile), e questo può essere simulato solo attraverso l'ausilio di un operatore umano che svolge, in modalità nascosta all'utente, questa funzionalità ideale a errore zero. Questa tecnica è ben nota nella valutazione dei sistemi, vocali e non, ed è comunemente detta del 'Mago di Oz' (WOz *Wizard of Oz*) dall'omonimo romanzo di Frank Baum del 1900 dove i personaggi del romanzo, raggiunto il paese

di Oz, si trovano a parlare con una 'entità astratta' che credono impersonale, ma che in realtà non era altro che uno uomo nascosto dietro una tenda.

Il primo software realizzato è Speaky-WOz, un sistema che permette ad un operatore remoto di controllare il dialogo pratico di un sistema dove ogni singolo elemento può essere controllato o sostituito dall'operatore. Tipicamente l'operatore sostituisce il sistema di riconoscimento e comprensione del parlato in modo che non vi siano errori di comprensione, mentre la sintesi vocale è operata con messaggi preregistrati o generati online, eventualmente anche con l'ausilio di un avatar per rendere l'interazione più gradevole. Il dialogo invece è codificato secondo una strategia di "*frame-box filler*" dove ogni compito è scomposto in singoli "*frame*" elementari, ciascuno dei quali si considera risolto solo dopo che una sequenza di "*box*" sono state correttamente riempite. Il dialogo vocale, si veda un esempio in figura 3, opera in maniera tale che l'utente possa fornire tutte le informazioni per risolvere un "*frame*", quindi decide come procedere, saltando al successivo o ad altro "frame", ed eventualmente operando delle azioni attuative reali o simulate se lo stato del dialogo lo richiede. È questa una strategia procedurale ben nota, e che possiamo esemplificare come riportato di seguito.

Figura 3 - *Un esempio di interazione vocale con il sistema*

```
COMPITO: Prenotazione voli
FRAME01: [CITTA_PARTENZA] [CITTA_ARRIVO]

U: salve vorrei prenotare un volo {per [Milano Malpensa]} per la
    prossima settimana
S: da che città vuole partire?
U: ah... si... ehm parto da casa... {da [Roma]}
...
```

Una volta disegnato il dialogo e valutato il sistema con Speaky-WOz, possiamo operare le necessarie modifiche al dialogo e scelte tecnologiche, e quindi valutare nuovamente l'intero insieme in modo iterativo fino ad arrivare ad una soluzione soddisfacente del nostro sistema, ovvero del disegno del dialogo e delle tecnologie da utilizzare. A questo punto dobbiamo mettere in atto i frutti della nostra ricerca sperimentale e realizzare un reale sistema di dialogo che, sulla base delle prestazioni e delle interazioni studiate e messe a punto in precedenza e utilizzando quelle tecnologie già valutate nel contesto di nostro interesse, sia in grado di realizzare un sistema ad interazione vocale reale e non più simulato tramite l'ausilio di un operatore, ovvero tramite il Mago di Oz. Per fare questo abbiamo realizzato un secondo software Speaky-RD (*Speaky-Real Dialogue*), che lavora con la medesima strategia di "*frame-box filler*" per quanto riguarda la codifica e gestione del dialogo, e che utilizza risorse di ASR, di TTS e di NLP terze, disponibili in rete o che è possibile portare localmente sul nostro sistema.

Figura 4 - *Schema a blocchi della interfaccia di dialogo vocale "distribuita in rete"*

Chiaramente qualora si voglia fisicamente attuare il risultato di un dialogo, ad esempio il controllo di un sistema di domotica o di intrattenimento, sarà necessario sviluppare del software specifico aggiuntivo. Vedremo successivamente come procedere in tal senso e come sia stato possibile realizzare il controllo di un sistema "*media-center*" con uno sforzo minimo. Il precedente schema generale è stato da noi realizzato utilizzando risorse distribuite in rete e quindi meglio schematizzabile secondo la figura 4.

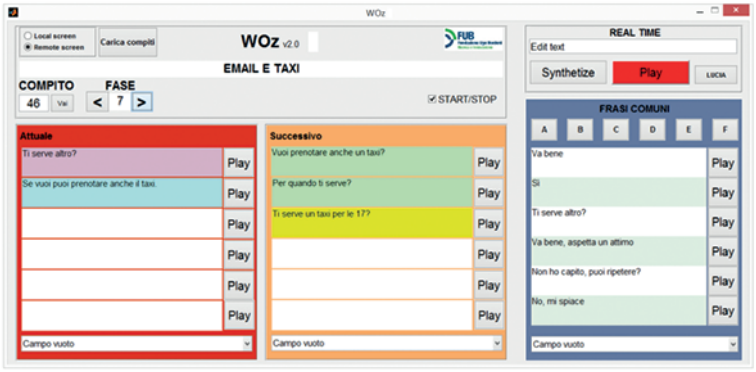
2.1 Speaky-WOz: simulare per valutare

Figura 5 - *Un esempio dell'interfaccia lato utente del sistema (domotica)*

Il software Speaky-WOz è sviluppato in linguaggio MatLab ed è in pratica un 'cruscotto di lavoro' dal quale l'operatore controlla il dialogo vocale in maniera autonoma, ma seguendo rigidamente quanto definito negli schemi di dialogo del tipo "*frame-box filler*".

Il cruscotto comanda remotamente il terminale, un semplice personal computer dotato di video e microfono, con cui interagisce l'utente. È sufficiente quindi una connessione di rete di media/buona qualità, e il WOz può gestire l'interazione vocale anche con utenti situati in città diverse da quella in cui si trova (Poroli, 2014).

Figura 6 - *Un esempio della di interfaccia lato "WOz" del sistema*



Sul terminale di interazione con l'utente è possibile inviare sia immagini, sia video, sia messaggi audio. Nel nostro caso, tipicamente, si inviava una immagine video a descrizione dello scenario di lavoro e un filmato relativo alla interazione vocale con un avatar. Nella figura 5 è mostrato un esempio del terminale che l'utente si trova davanti nel caso di una simulazione di casa domotica, uno degli scenari da noi utilizzati. Il WOz invece si trova a lavorare utilizzando un cruscotto come quello riportato nella figura 6.

In un file Excel sono codificati tutti i passaggi della interazione ovvero del dialogo. Possono raggrupparsi in un unico file un numero qualsiasi di 'compiti', ovvero schemi di dialogo con l'utente per raggiungere uno specifico obiettivo più o meno semplice, dalla prenotazione di un viaggio o di un ristorante, all'inserimento di un farmaco nei promemoria, ad un consulto medico e successiva prenotazione di visita, e così via. Ogni compito è quindi diviso in 'sotto compiti' o "*frame*" di più semplice gestione dove attraverso il dialogo è necessario riempire le "*box*" ovvero risolvere le variabili, tipicamente da due a quattro, dello specifico "*frame*". Sul cruscotto è evidenziato, al centro dello schermo, l'argomento del compito, in alto a destra il numero del compito ed il numero del "*frame*" in cui ci troviamo. Il WOz può avanzare o in genere navigare tra i "*frame*" con un semplice click. A schermo compariranno le possibili frasi che il WOz può, secondo gli schemi di dialogo prestabiliti nel file Excel, presentare all'utente nel "*frame*" corrente e in quello successivo in modo che il WOz abbia visione di come evolverà il dialogo. In funzione di quanto pronunciato dell'utente il WOz può scegliere la risposta premendo il bottone "*play*" corrispondente. A questo punto sul terminale dell'utente saranno trasmesse le immagini e l'audio corrispondente a quella scelta. Poiché ovviamente non è possibile prevedere e codificare tutte le possibili risposte da dare all'utente, in alcuni casi è ne-

cessario fornire delle risposte al di fuori dello schema prestabilito. A tal fine il WOz ha a disposizione, a sinistra del cruscotto, una sezione con cui interagire in tempo reale ovvero dove può gestire situazioni di errore, di attesa e altro o, se necessario, realizzare *on line* risposte specifiche da dare all'utente in modo da essere in grado di gestire qualsiasi tipo di situazione. Nella nostra esperienza (Poroli, 2013) l'utilizzo di questa sezione del cruscotto si è rilevata piuttosto rara, ma molto efficace quando necessaria. Il sistema registra in un file di testo lo storico di tutti i passaggi, ovvero dei tempi e delle scelte che il WOz effettua. Ovviamente sarà necessario non solo definire nel file Excel tutti i compiti e i "*frame*" relativi, ma anche le immagini e i le risposte audio (o audio-video nel caso di utilizzo di avatar) per ciascuna delle possibili scelte che il WOz può operare. Il frutto di questo lavoro è un sistema che simula un'interazione vocale di alto livello, come mostrato ad esempio nel video (CSP, 2013), e che permette la raccolta di dati relativi a interazioni vocali in diverse città italiane da un unico punto di controllo come descritto in (Poroli, 2013). Una volta effettuate le simulazioni e le relative valutazioni, sarà quindi possibile avere maggior confidenza nel disegnare i dialoghi per una reale interfaccia vocale per un sistema o un servizio.

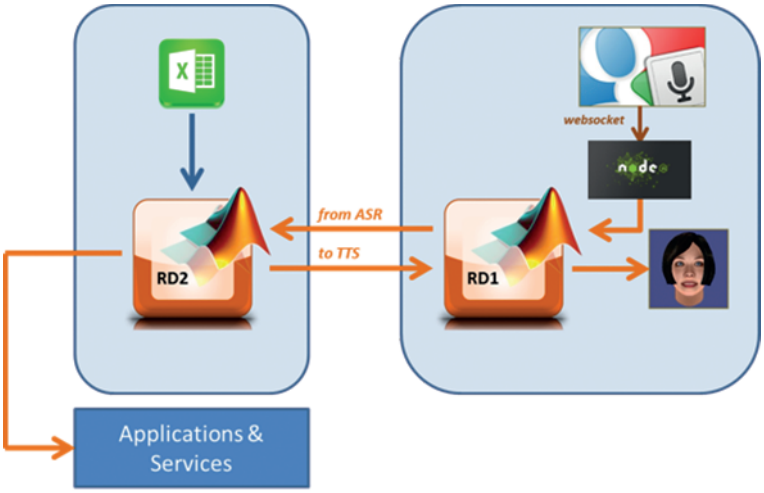
2.2 Speaky-RD: per la prototipizzazione rapida

Il software Speaky-RD è composto, come vedremo, da una serie di moduli, i principali e più complessi in MatLab. Il suo obiettivo è emulare il sistema "Speaky-WOz" senza che vi sia l'operatore umano, e in tempo reale. La codifica dei dialoghi, delle azioni da intraprendere e quanto altro è sempre contenuta in file Excel, il che rende di semplice e immediata realizzazione l'implementazione dei dialoghi pratici.

3. *L'infrastruttura e le scelte operate*

L'infrastruttura del sistema deve ora includere l'utilizzo di un sistema ASR e di un sistema di TTS che comunicano con il modulo di gestione del dialogo, che a sua volta può utilizzare moduli di NLP e attuare fisicamente dei comandi su dispositivi esterni o alternativamente simularli. Tutto questo implica che i vari moduli comunichino tra loro su rete secondo protocolli ben definiti. Nel nostro caso il sistema di ASR prescelto è quello offerto da Google, mentre il sistema di sintesi vocale è quello Microsoft nella declinazione della voce femminile di "Silvia", le stringhe riconosciute dal sistema ASR vengono inviate al modulo di dialogo attraverso un '*socket TCP*', analogamente il sistema di dialogo comunica con il modulo di sintesi inviando il testo da pronunciare attraverso un '*socket TCP*'. I sistemi ASR e TTS sono stati raggruppati in un unico modulo software Speaky-RD1 da noi sviluppato e configurabile, se necessario, con altri sistemi di riconoscimento e sintesi. Tutti gli altri moduli sono raggruppati in Speaky-RD2 che acquisisce le informazioni per gestire il dialogo e i comandi attuativi da i file Excel.

Figura 7 - *Schema a blocchi del sistema di dialogo vocale*



Nel dettaglio, RD1 apre una pagina html nel browser Chrome che è in grado di acquisire il segnale audio dal microfono del computer e di operare la trascrizione di quanto pronunciato. Il microfono può rimanere sempre aperto (*default*) o aprirsi solo quando il sistema di sintesi non stia parlando.

Questa ultima modalità di funzionamento da un lato facilita il riconoscimento del parlato, ma al contrario è mal accettata dagli utenti in quanto non naturale nei normali dialoghi parlati. Per motivi di sicurezza la pagina html può comunicare solo attraverso la nuova tecnologia di "*websocket*", pertanto si è dovuto realizzare un livello intermedio di comunicazione attraverso un semplice programma in linguaggio "*node js*" che legge i messaggi da un protocollo e li rispedisce nell'altro al sistema, come delineato in figura 7. Aver sviluppato il modulo RD1 permette di sostituire il sistema di ASR o di TTS facilmente senza dover operare alcuna modifiche nei file Excel o negli altri moduli. Riassumendo, possiamo dire che i moduli RD1 e RD2 si scambiano in RD1 invia a RD2, le stringhe relative a quanto riconosciuto dal sistema di ASR, mentre nel verso contrario, RD2 invia a RD1, le stringhe relative ai messaggi di testo da far pronunciare al sistema di TTS o all'avatar. In aggiunta il sistema RD2 potrà inviare stringhe a dispositivi per attuare comandi o fruire di servizi (e.g. accendere una luce, ascoltare un programma radio o televisivo, etc.).

3.1 Speaky-RD1: gestione del "ASR" e del "TTS"

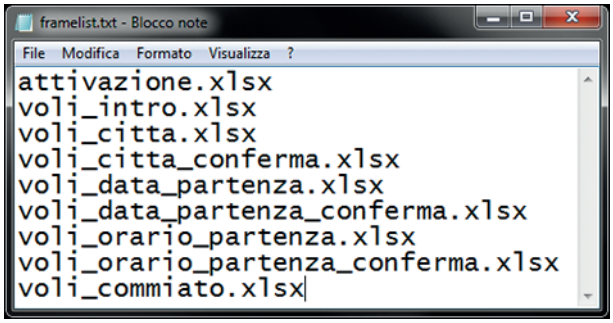
Questa parte del sistema si occupa della gestione del "*front-end*" audio che si presenta all'utente e della integrazione dei sistemi di riconoscimento del parlato e di sintesi. Il riconoscimento del parlato è operato attraverso una pagina html che sfrutta il riconoscitore di Google per parlato continuo. Una volta eseguita la pagina html, tutto quanto acquisito dal microfono viene spedito al sistema di riconoscimento che inizia a elaborare il segnale e propone una serie di ipotesi di riconoscimento che evolvono dinamicamente sino a quando il segnale audio non presenti una pausa

sufficiente ed il sistema di riconoscimento abbia risolto la trascrizione con determinato grado di affidabilità. Solo a questo punto il segnale audio viene 'consolidato' e la relativa trascrizione viene comunicata attraverso "*websocket*" all'applicativo "*node js*" che a sua volta lo comunica al pro-gramma MatLab, che lo invierà al modulo di gestione del dialogo. Sebbene il sistema di riconoscimento in questione funzioni molto bene, abbiamo riscontrato delle situazioni che necessitano una particola attenzione. Un problema che è subito risultato evidente consiste nel fatto che il riconoscitore può attendere anche diversi secondi quando deve consolidare segnali audio di brevissima durata, come ad esempio per la frase "si" che gli utenti utilizzano spesso per confermare. In questi casi infatti il sistema riconosce immediatamente il segnale, ma tarda nel consolidamento o addirittura fallisce nel consolidamento. Per eliminare questo problema abbiamo modificato il sorgente html derivato da quelli dei dimostrativi di Google, e abbiamo intercettato il testo riconosciuto ma non consolidato, e se questo corrisponde ad un determinato messaggio (al messaggio "si" nel nostro caso) il testo viene forzatamente marcato come consolidato simulando quanto avrebbe dovuto fare il sistema. Strategie diverse e più sofisticate per la forzatura del consolidamento del testo, ad esempio basate sul tempo di attesa o di non operabilità del riconoscitore, potrebbero anche essere operate, ma per ora questa semplice operazione è stata sufficiente a risolvere la maggior parte delle situazioni. Il secondo problema incontrato è anche legato al consolidamento e si manifesta quando vi sia un forte rumore di fondo o quando l'utente continui a parlare anche dopo aver comunicato con il sistema, ad esempio con dei commenti tra sé e sé oppure parlando con altre persone presenti. Il sistema di riconoscimento non è ovviamente in grado di capire se l'utente stia parlando al sistema o con altri, né tantomeno è in grado di capire se il segnale audio che arriva al microfono debba essere analizzato dal riconoscitore o no (come nel caso di rumore di fondo). Essendo il sistema nel nostro caso sempre in ascolto, tutto quanto arriva al microfono viene mandato al riconoscitore che cercherà di trascrivere fino a che non si riconosca una pausa di silenzio opportuna. Se per i motivi sopra elencati tale pausa di silenzio non si verifica il sistema tarderà a consolidare quanto pronunciato dall'utente. Purtroppo una soluzione tecnica a questo tipo di problema non è di banale realizzazione, e quindi si è cercato di evitare il problema, piuttosto che risolverlo, utilizzando il sistema solo in ambienti chiusi di ufficio e con un microfono di prossimità o comunque mai posizionato a distanze maggiori di qualche decina di centimetri dal parlatore. Per quanto riguarda il sistema di sintesi da testo questo può considerarsi, tipicamente, come una risorsa interna al computer che ospita RD1. Nel nostro caso si sono esaminate diverse sintesi da testo (Microsoft, Loquendo, etc.) e l'avatar 'Lucia' sviluppato dal CNR (Cosi, 2003) ed utilizzato in Speaky-WOz. Altre risorse di sintesi da testo, ad esempio la sintesi da testo offerta da Google, possono facilmente essere integrate in Speaky-RD1. Come abbiamo detto, un sistema da sintesi da testo che permetta di conoscere l'istante il cui il sistema ha terminato la comunicazione può risultare utile nel caso di un più stretto controllo dell'acquisizione audio, a svantaggio ovviamente di maggior naturalezza del dialogo.

3.2 Speaky-RD2: gestione del dialogo e attuazione

Speaky-RD2 è il modulo che controlla il dialogo con l'utente e che permette l'attuazione o la simulazione di semplici operazioni che realizzano quanto richiesto dall'utente. Il testo generato dal sistema di ASR arriva a RD2 che come prima cosa lo normalizza eliminando eventuali caratteri spuri e di punteggiatura, riportandolo tutto a minuscolo ed altro, secondo una serie di regole che possono facilmente essere modificate. A questo punto una prima elaborazione del testo consiste in una ulteriore e più forte normalizzazione riportando ciascuna parola riconosciuta al lemma corrispondente. Si crea quindi, con l'ausilio di un 'lemmatizzatore' (Zanchetta, 2005), un vettore di termini dello stesso ordine di quello relativo al testo trascritto ma i cui elementi corrispondono ai lemmi delle corrispondenti parole. Sebbene analisi ed elaborazioni del testo più complesse (Pianta, 2008) possano essere operate per facilitare la comprensione del testo, la riduzione in lemmi è risultata sufficiente, come vedremo, alla gestione di dialoghi pratici. Il sistema legge da un semplice file di testo la lista dei file Excel che contengono i dialoghi pratici, o meglio le singole parti di uno o più dialoghi pratici. In assenza di un comando, ad esempio richiesto dall'utente, che forza il dialogo a saltare in un preciso stato, il programma esegue linearmente la lista dei file Excel, che quindi si dovrà aprire sempre con un file di attivazione, ad esempio su 'parola chiave', e di introduzione. Ad esempio il file lista per un dialogo pratico per un ipotetico compito legato all'utilizzo di un servizio di prenotazioni voli, potrà essere il seguente, riportato in figura 8.

Figura 8 - *Un esempio di file di comando per il sistema*



L'attivazione fa sì che il sistema sia operativo solo dopo che l'utente abbia pronunciato una parola chiave. Nel nostro caso l'utente era invitato ad attivare il sistema con il comando 'ciao speaky', ma la parola chiave utilizzata era, per ovvi motivi, solamente 'ciao', in quanto la parola 'speaky' non è contenuta in un vocabolario standard dell'italiano. Nel file 'attivazione' possono essere definite una serie di parole o frasi chiave diverse per attivare il sistema. Il file 'voli_intro', analogamente a 'voli_commiato', non prevede l'interpretazione del parlato pronunciato dall'utente, ma comunica semplicemente dei contenuti di saluto, istruzione, etc. a seconda dei casi, per lo specifico compito. Gli altri file Excel contengono i dialoghi per acquisire le informazioni su città di partenza e di arrivo, giorno e mese di partenza, orario di partenza, e

per ciascuna di queste informazioni è possibile avere conferma con un ulteriore file Excel. Nel caso di risposta positiva il dialogo prosegue, tipicamente con file successivo ma potrebbe puntare ad altro se necessario, mentre nel caso di risposta negativa il dialogo prosegue puntando, tipicamente, al file Excel che richiede nuovamente i dati non confermati dall'utente ma anche in questo caso potrebbe puntare ad altro se necessario. Ciascun file Excel è composto da quattro fogli: "*grammar*", "*tts*", "*jump*" e "*action*", che devono essere opportunamente riempiti a seconda delle funzionalità che si vogliono realizzare.

Nel foglio "*grammar*", figura 9, sono definite le varie "*box*" che devono essere riempite per completare correttamente la parte di dialogo pratico corrispondente al "*frame*", ovvero al file Excel, in elaborazione. Per ciascuna "*box*" avremo due colonne: una prima corrispondente al testo che deve trovarsi nella stringa riconosciuta dal trascrittore, nella seconda il contenuto o comando da associare a quanto riconosciuto ed individuato nella grammatica.

Figura 9 - *Un esempio di file Excel per la definizione di una grammatica*

S-	[BOX1] grammar	[BOX1] command	[BOX2] grammar	[BOX2] command
DEFAULT VALUE				
	roma	roma	roma	roma
	da roma	roma	a roma	roma
	aosta	aosta	per roma	roma
	da aosta	aosta	aosta	aosta
			a aosta	aosta
			ad aosta	aosta
			per aosta	aosta

Figura 10 - *Un esempio di file Excel per la definizione delle risposte per l'utente*

[BOX1] status	[BOX2] status	Answer1	Answer2
0	0	dimmi città di partenza e città di arrivo	per iniziare dimmi da che città vuoi partire
0	1	la destinazione è £2£ , ora dimmi da dove vuoi partire	per il tuo volo a £2£ , ora dimmi da che città vuoi partire
1	0	la città di partenza è £1£, ora dimmi dove vuoi andare	hai scelto un volo con partenza da £1£, ora scegli la città di destinazione
1	1	mi confermi partenza da £1£ e destinazione £2£ ?	partenza da £1£ ed arrivo a £2£, è corretto?

Se vi sono 'n *box*' nella grammatica, lo stato del dialogo potrà essere in 2^n condizioni diverse a seconda che ciascuna singola *box* sia stata risolta o no. Queste condizioni sono enumerate nel foglio "*tts*" dove per ciascuna situazione, da zero a (2^n)-1, sono riportate le frasi che il sistema deve rivolgere all'utente per completare il dialogo pratico, si veda figura 10. Per ciascuno stato è possibile avere un numero arbitrario di possibili risposte: il sistema ne sceglierà casualmente una tra quelle disponibili. Al fine di rendere il dialogo più naturale è possibile inserire nel testo da sintetizzare il contenuto delle *box* già risolte utilizzando sia il valore della cella di grammatica, sia quella di comando.

Il foglio "*jump*", si veda figura 11, interpreta un particolare contenuto dei comandi associati alle grammatiche.

Figura 11 - *Un esempio di file Excel per la gestione dei flussi di dialogo condizionati*

N	GRAMMAR A	COMMAND A
DEFAULT VALUE		
	si	>1
	bene	>1
	va bene	>1
	confermo	>1
	ok	>1
	no	>2
	sbagliato	>2
	errato	>2

JUMP
DEFAULT VALUE
voli_data_partenza.xlsx
voli_citta.xlsx

Quando nel foglio delle grammatiche la casella associata alla parola riconosciuta, ovvero la casella di '*command*', è del tipo numerico, preceduto dal simbolo '>' questo viene interpretato dal sistema come un salto condizionato e più specificatamente al file di grammatiche Excel contenuto nel foglio "*jump*" ed indicizzato dal valore associato alla casella del foglio "*grammar*" riconosciuta. In questo caso si possono, ad esempio, predisporre nel dialogo dei momenti di verifica e/o di conferma che fanno procedere il dialogo a seconda della risposta dell'utente, come nell'esempio in cui se si ha una risposta affermativa (">1") il dialogo prosegue con la richiesta della data di partenza, altrimenti nel caso di risposta negativa (">2") si torna a chiedere la città di partenza e di destinazione.

Un formalismo del tutto analogo vale per il foglio "*action*", si veda figura 12.

Figura 12 - *Un esempio di file Excel per la attuazione di semplici comandi*

ACTION
sys:RAI1.vbs
sys:RAI2.vbs
sys:RAI3.vbs
sys:RAI-RADIO1.vbs
....

Sempre sulla base dei valori numerici associati nel riconoscimento e codificati nel foglio "*grammar*" oltre a saltare secondo quanto definito nel foglio "*jump*" appena descritto, il sistema esegue il comando corrispondente del foglio "*action*". Ad esempio con il suffisso "*sys:*" il sistema interpreta la stringa seguente come un comando di sistema, e nel nostro caso, predisponendo specifici file di comando in *visual basic* per il controllo di un *media-center*, è stato possibile attuare delle scelte operate vocalmente dall'utente direttamente sul dispositivo televisivo.

Utilizzando opportunamente le funzionalità sopra descritte e limitatamente alla realizzazione di dialoghi pratici, è stato possibile realizzare interfacce vocali per dei test in cui l'utente, dopo aver attivato vocalmente il sistema con una parola chiave, entrava in un menù che permetteva: la prenotazione di un biglietto aereo su tratta nazionale, servizio esclusivamente vocale tipo IVR (*Interactive Voice Response*); il controllo di una casa domotica, simulata su schermo; il controllo reale di un *media-center*, dove l'utente poteva accedere ai canali tv e radio, a contenuti multimediali audio e video. Come abbiamo visto un ruolo centrale è svolto dal foglio "*grammar*" a cui è demandata l'analisi della stringa di testo riconosciuta del sistema ASR e opportunamente elaborata da un semplice sistema di NLP, un lemmatizzatore nel nostro caso. La ricerca dei termini contenuti nelle varie colonne di '*grammar*' procede dapprima cercando nella stringa le caselle che contengono il più alto numero di termini nell'intero insieme della grammatica (ad esempio si cerca prima la casella "per roma fiumicino", poi "per roma" e infine "roma" nel caso ovviamente tutti e tre le caselle siano presenti nella grammatica). Il secondo termine di ricerca è ovviamente dalla prima colonna di grammatica a sinistra all'ultima a destra. Una volta individuata una corrispondenza, questa viene catalogata opportunamente e il testo della stringa viene epurato del testo relativo. In realtà è possibile definire una strategia specifica su come operare sulla stringa riconosciuta immettendo un codice nella prima casella del foglio "*grammar*" (ad esempio il codice "S-" nella figura 9). In particolare è possibile comandare il sistema in modo che i termini delle grammatiche riconosciute siano cancellati o no dalla stringa, che la stringa sia ereditata da un foglio Excel al successivo, che la stringa al contrario sia azzerata prima di procedere con la nuova acquisizione e analisi e altro ancora come dallo schema di figura 13.

Figura 13 - *Liste delle possibili modalità di controllo del testo riconosciuto*

codice	azione operata sulla stringa dell'ARS
N	Non elabora la stringa in ingresso e non ne salva una nuova in uscita
D	Non elabora la stringa in ingresso e la cancella in uscita
S+	Non elabora la stringa in ingresso e la salva completa in uscita
S-	Non elabora la stringa in ingresso e la salva priva delle *keyword* in uscita
P	Elabora la stringa in ingresso e non ne salva una nuova in uscita
PD	Elabora la stringa in ingresso e la cancella in uscita
PS+	Elabora la stringa in ingresso e la salva completa in uscita
PS-	Elabora la stringa in ingresso e la salva priva delle *keyword* in uscita

Ovviamente sarà cura del '*dialog designer*' sfruttare le potenzialità e le funzionalità offerte dai fogli Excel appena descritti, al fine di progettare al meglio i dialoghi pratici di interesse. Considerando che quella qui presentata è la prima versione del sistema da noi realizzato, questo si è mostrato già piuttosto efficace e ci ha permesso agilmente di realizzare una serie di diversi dimostrativi sviluppati in breve tempo e facilmente. Futuri studi saranno mirati alla formalizzazione delle attuali funzionalità, all'integrazione di altre funzioni per supportare il disegno dei dialoghi o per sfruttare sistemi di NLP più sofisticati.

4. *Conclusioni*

Si sono sviluppati due sistemi a supporto della realizzazione e valutazione di interazioni vocali in linguaggio naturale attraverso dialoghi pratici. I software sviluppati permetto l'utilizzo e la facile integrazione di risorse terze per il riconoscimento vocale e per la sintesi vocale, in modo che chiunque possa personalizzare il sistema a proprio piacimento. Anche l'infrastruttura scelta per la realizzazione del sistema è completamente aperta e può pertanto adattarsi a qualsiasi configurazione: da un sistema su un singolo computer, a un sistema distribuito su rete locale o pubblica. Il primo sistema si basa sulla tecnica del Mago di Oz e permette ad un operatore di controllare il dialogo in modo che si possano valutare le strategie più opportune e caratterizzare i comportamenti dell'utenza durante l'interazione vocale. Il secondo sistema permette la realizzazione di reali interfacce vocali in tempo reale, basate su dialogo naturale per gestire a voce dei servizi o per controllare, sempre con il dialogo vocale, sistemi o dispositivi. La codifica delle parole chiave, del dialogo, e della eventuale attuazione di comandi sui dispositivi è predisposta in dei normali file Excel. Il primo sistema è stato utilizzato per raccogliere un corpus di dati sul territorio italiano su un campione di circa 80 parlatori che dovevano svolgere un numero limitato tra quaranta compiti diversi. Il secondo sistema è stato utilizzato su un campione di circa 20 soggetti che dovevano operare compiti quali la prenotazione di un volo, il controllo di un media center, la gestione di una casa domotica.

Bibliografia

Amazon (2014). *Introducing Amazon Echo*. http://www.amazon.com/oc/echo.

Apple (2011). Apple Press Info. *Apple Launches iPhone 4S, iOS 5 & iCloud*. https://www.apple.com/pr/library/2011/10/04Apple-Launches-iPhone-4S-iOS-5-iCloud.html.

Cosi, P., Fusaro, A. & Tisato, G. (2003). LUCIA a new italian talking-head based on a modified Cohen-Massaro's labial coarticulation model. In *Proceedings of Eurospeech 2003*, Geneve, Switzerland.

CSP (2013). *Progetto Speaky Acutattile*. https://www.youtube.com/watch?v=8sIOorZ7w7c.

Dirha (2014). *Distant-speech Interaction for Robust Home Applications*. https://dirha.fbk.eu/sites/dirha.fbk.eu/files/docs/Dirha_Brochure_def.pdf

Jibo (2014). *Welcome to the Jibo Blog on Social Robots*. https://www.jibo.com.

McTear M. (2004). *Spoken dialogue technology: toward the conversational user interface*. London: Springer-Verlag.

Minker, W. et al. (2011). *Spoken dialogue systems technology and design*. London: Springer Science Business Media, Springer-Verlag.

Pianta, E., Girardi, C. & Zanoli, R. (2008). *The TextPro tool suite*. In *Proceedings of LREC, 6th edition of the Language Resources and Evaluation Conference*, 28-30 May 2008, Marrakech, Morocco.

Poroli, F., Paoloni, A. & Todisco, M. (2013). Gestione degli errori in un corpus di dialogo uomo-macchina: strategie di riformulazione. In *Atti del X convegno nazionale AISV*, Torino, gennaio 22-24, 2014.

Poroli, F. et al. (2013a). Prime indagini su un corpus di dialogo uomo-macchina raccolto nell'ambito del progetto Speaky Acutattile. In *Atti del IX convegno nazionale AISV*, Venezia, gennaio 11-13, 2013.

Poroli, F. et al. (2014). Il corpus Speaky. In *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014*, 9-11 December, Pisa.

Thomson, B. (2013). *Statistical Methods for Spoken Dialogue Management.* London: Springer Theses, Springer-Verlag.

Yu, D., Deng, L. (2015). *Automatic speech recognition a deep learning approach*. London: Signals and Communication Technology, Springer-Verlag.

Zanchetta, E., Baroni, M. (2005). *Morph-it! A free corpus-based morphological resource for the Italian language*, proceedings of Corpus Linguistics 2005, University of Birmingham, Birmingham, UK.

DOI: 10.17469/O2101AISV000029

PIERO COSI, GIULIO PACI, GIACOMO SOMMAVILLA, FABIO TESSER

Building Resources for Verbal Interaction Production and Comprehension within the ALIZ-E Project

The European FP7 project ALIZ-E ("Adaptive Strategies for Sustainable Long-Term Social Inter- action") has recently concluded and its main goal was to study children-robot interactions in the field of healthcare. ISTC CNR, UOS Padova has been the partner of the ALIZ-E project responsible of carrying out studies in the field of speech technologies. One of its main achievements has been the collection of three new Italian children's speech annotated corpora, made up by read sentences, spontaneous utterances and recordings from a listen and repeat experiment.

1. *Introduction*

The European FP7 project ALIZ-E ("Adaptive Strategies for Sustainable Long-Term Social Interaction") (Belpaeme et al., 2013) has recently concluded. Its main goal was to study children-robot interactions in the field of healthcare. The Padova Institute of Cognitive Sciences and Technologies (ISTC) of the National Research Council (CNR) has been the partner of the ALIZ-E project responsible of carrying out studies in the field of speech technologies, as described in (Tesser et al., 2013) and (Paci et al., 2013).

One of its main achievements has been the collection of three new Italian children's speech annotated corpora, made up by read sentences, spontaneous utterances and recordings from a listen and repeat experiment.

2. *Data Collection*

Three new Italian children's speech annotated corpora, made up by read sentences, spontaneous utterances and recordings from a listen and repeat experiment were collected during the European FP7 project ALIZ-E.

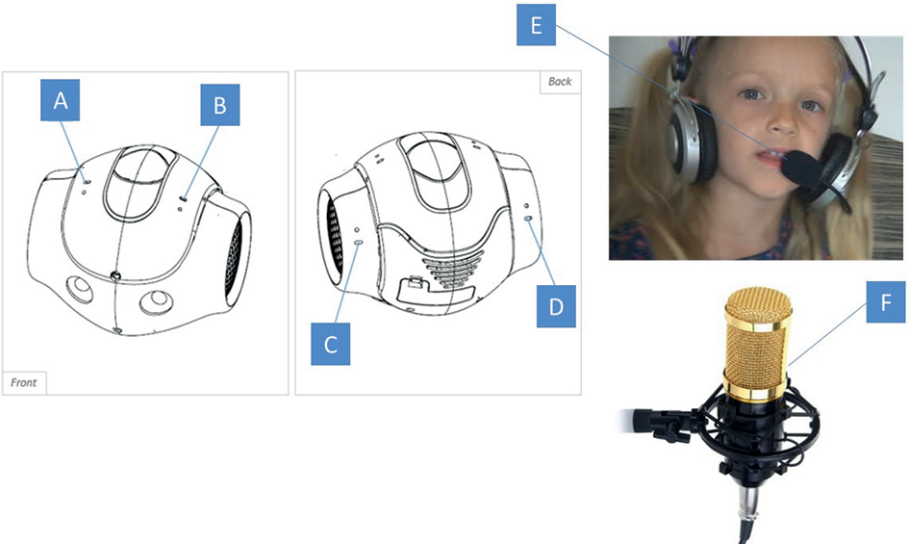
2.1 Read Speech

With respect to read speech data collection, prompts from the FBK ChildIt corpus (Gerosa et al., 2007) have been used in order to extend the training material of children speech ASR systems. They are simple phonetically balanced sentences, selected from children's literature[1].

[1] Typical sentences: "il fabbro usa la tenaglia con forza", "nella zona avvenne uno scoppio", ecc.

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.

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 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.

2.2 Semi-Spontaneous Speech

While read Speech is very useful for increasing the size of speech training data to be used in the AM training procedure, on the other hand it is not well suited for building a reliable test set for ASR, since it does not represent well a real case of recorded audio from an interaction of the ALIZ-E project.

In order to build a proper test set, speech data recorded during a project's experiments in a Wizard of Oz modality have been collected using the same recording framework adopted for read speech (see Figure 1) and manually transcribed and an-

notated. The interactions took place at the "San Raffaele" hospital and at *Summer Schools* for diabetic children, The experiments consisted in a "Quiz Game", with the robot and the young user posing questions to each other.

Transcription and annotation of semi-spontaneous speech recordings of the above mentioned experiments have been manually performed by means of the "Transcriber" tool. The ALIZ-E project partners agreed to perform a special level of transcription: speech events relevant to Natural Language Understanding (NLU) components (such as questions and answers) have been annotated.

Also non-verbal sounds occurring specifically in such interactions (for example robot's speech and motor noise) have been collected and classified. Globally, the 76 ALIZ-E verbal interactions have been annotated at the word level, totaling about 20 hours of total audio, containing more or less 3.6 hours of actual children's speech. This semi-spontaneous speech annotated corpus has allowed to build a test set for ASR that is almost ten times larger than the one used initially in the ALIZ-E project. It comprises 540 sentences from 46 speakers, totaling 5423 words.

2.3 Listen and Repeat Speech

Another data collection session has been set up at "*Istituto Comprensivo Berna*" (Mestre, Italy) as part of "ITACIE" (Italian TTS Adult Children Intelligibility Experiment).

Children aged 7-11 were asked to listen to about 40 sentences generated by the Italian ALIZ-E TTS system played by either a robot's loudspeakers or by earphones and to repeat them aloud. These recordings served in a TTS intelligibility experiment. Usually, in standard intelligibility tests the participants are asked to write down or type what they understand. This kind of tasks is quite tedious and tiring for children. For this reason a listen and repeat experiment has been set up. It has proven to be a convenient and fast technique to collect annotated speech data: it is more reliable than making the children read a text, and it is much faster to transcribe than recording spontaneous speech. The text prompts have been built using Semantically Unpredictable Sentences (SUS)[2] (Benoît et al., 1996). The aim is to prevent the child to be able to "guess" the words in sentences thanks to the semantic context. For this reason, the prompts were randomly generated by an automatic program that provides grammatically correct but semantically anomalous sentences. In order to avoid the problem that the sentences were too difficult to pronounce for the children, a lexicon with the most common words used by Italian children has been applied. Words with CV (Consonant-Vowel) patterns have been favored, and only a few common words containing CC (Consonant-Consonant) patterns, like "scuola", have been allowed. While the recording session was running, the annotations of the recorded audio sentences were automatically generated, assuming that the child repeated exactly what the TTS system has pronounced. If the child

[2] Typical SUS (grammatically correct but semantically anomalous) sentences: "la parola piena beve la scuola", "la pace marrone odia la figlia", ecc.

mispronounced some words, however, right after the end of the utterance, the experimenter could modify the transcription accordingly (or tag the sentence to be corrected afterwards). Ninety-five children aged 7-11 have been recorded and almost three and a half hours of children's speech have been collected.

3. *Final Remarks*

ISTC-CNR intends to make all data collected available to the research community. ISTC-ChildIT (read speech) or CHILDIT-2 will be released soon, probably in 2015. ISTC-CNR is currently discussing with ALIZ-E partner "Fondazione San Raffaele del Monte Tabor" about the possibility of publishing the Spontaneous Speech audio data.

Acknowledgements

This work was entirely supported by the EU FP7 "ALIZ-E" project (grant number 248116).

References

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuay'ahuitl, H., Kiefer, B. et al. (2013). Multimodal Child-Robot Interaction: Building Social Bonds. In *Journal of Human-Robot Interacion*, vol. 1, n. 2, 33-53.

Benoît, 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*, vol. 18, n. 4, 381-392.

Gerosa, M., Giuliani, D. & Brugnara, F. (2007). Acoustic variability and automatic recognition of children's speech. In *Speech Communication*, vol. 49, 847-860.

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.

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.

DOI: 10.17469/O2101AISV000030

ANTONIO ORIGLIA, FRANCESCO CANGEMI, FRANCO CUTUGNO

WORDEN: a PYTHON interface to automatic (non)word generation

WORDEN is a PYTHON implementation of a knowledge engine containing data on the phonological structure of Italian words and on their frequency. These data are imported from the PHONITALIA database and used with a rules base to provide a formal description of the operations needed to generate words, nonsense words and phonological neighbourhoods for speech production and perception studies. The knowledge engine supports a wide number of queries from the command line but can also be used via a specifically designed graphical user interface. In this paper, we report the formal rules included in WORDEN and provide usage examples to illustrate the capabilities of the tool.

1. *Introduction*

A knowledge engine, sometimes called an *expert system*, is a software module that, given a set of known *facts* (knowledge base) and a set of inference *rules* (rules base), is able to:

- provide yes/no answers to user *queries* (e.g. "Is <pane> a word?");
- find all the solutions that satisfy a pattern described in the rules (e.g. "Which words have a CVCV structure?").

Logic programming is not new in the field of Natural Language Processing. Nugues (2014) provides a deep overview of the application of PROLOG to NLP in general. In computational phonology, logic programming has been applied, for example, to build computational models of Autosegmental Phonology (Bird, Klein, 1990; Bird, Ellison, 1994) but its usage in the field is documented in the '80s, too (see Grishman, 1986). In the last years, logic programming has found applications in the field of computational semantics (Sowa, 2014; Stone et al., 2014) for the exploration of complex semantic networks. Usage of logic programming is typically considered as opposed to nowadays popular machine learning approaches, where supervised analysis of large corpora is used to obtain a set of probabilistic models to classify objects. The two approaches are, in truth, complementary. Logic programming allows the researcher to define and easily test formal theories but it does not support data analysis during the definition of such theories, which can be very complex in the case of partially known phenomena. Machine learning, on the other hand, allows to easily analyse and detect regularities in large datasets to obtain statistical models. These, however, are often hard to interpret and based on probabilities, thus being less suitable for the definition of formal theories.

In the following, we present WORDEN, a logic programming-based engine that assists linguists in finding suitable stimuli for production and perception experiments. In Section 2, we provide some examples of the research areas in linguistics and psycholinguistics that can benefit from such an engine. The details of the engine (knowledge base, rules base, graphical user interface) are presented in Section 3. Section 4 offers concluding remarks and sketches avenues to future expansion.

2. *Applications*

Much experimental work in linguistics and psycholinguistics requires subjects to utter words or sentences (production) or to react to stimuli presented orthographically or auditorily (perception). Both lines of work require a careful selection of the stimuli to be used.

In intonation research, for example, the experimenter is often concerned with the analysis of fundamental frequency contours extracted from speakers' recordings. However, f0 contours are highly sensitive to segmental perturbations (Lehiste, 1970; Di Cristo, 1985). At best, such microprosodic effects make the collected material highly dependent on the stimuli used in the elicitation task (e.g. vowels have different intrinsic pitch); at worst, they obscure the phenomena under scrutiny (e.g. voiceless segments do not allow for f0 estimation). For these reasons, many intonologists working with utterances elicited from scripted material devote great attention to the selection of experimental stimuli, e.g. by using sonorant segments only, or by restricting stressed vowels to one single type. The following test sentences for data elicitation, drawn from three studies on Neapolitan Italian, exemplify such practices:

1. Vedrai la bella mano di Mammola domani? (D'Imperio, 2001)
   Will you see Mammola's beautiful hand tomorrow?
2. Il benevolo manovale beveva vino (Petrone, 2008)
   The benevolent labourer drank wine

3a. Danilo vola da Roma (Cangemi, D'Imperio, 2013)
   Danilo takes the flight from Rome

3b. Serena vive da Lara
   Serena lives at Lara's

3c. $[\text{CV.C\`V.CV}]_{S}\,[\text{C\`V.CV}]_{V}\,[\text{CV}][\text{C\`V.CV}]_{IO}$

The examples show that it is indeed possible to minimize microprosodic perturbations by design (i.e., by choosing appropriate target words). However, they also show that this outcome has a cost – either in terms of the semantic and pragmatic plausibility of the stimuli (example 1), of the familiarity of the lexemes employed (example 2) or of occasional lapses in constraint satisfaction (example 3b, with its initial voiceless fricative). Examples 3a-b illustrate the case in which the corpus to elicit features various sentences sharing phonological and syntactic make-up (3c), i.e. a trisyllabic paroxitone noun as subject, followed a bisyllabic paroxitone verb, and by a trisyllabic paroxitone prepositional phrase as indirect object. In this

case, the researcher has to find actual words in the lexicon to fill in the slots in the sentence template. Such a template-based stimulus search is particularly costly for the researcher, if performed manually. WORDEN is a first step towards the automatization of such searches.

Careful stimulus selection is also required for psycholinguistic studies investigating spoken-word recognition (e.g. Luce, Pisoni, 1998) and the structure of the mental lexicon. In the past twenty years, research in this vein has documented the importance of phonological neighbourhood effects in speech production and perception. Neighbourhoods can be understood as "conglomeration of words that are highly similar to one another along a critical characteristic" (Marian et al., 2012: 1). Phonological neighbours are thus the words which can be constructed by substituting, inserting or deleting one or more phonemes from a source word (or similar revisions of Coltheart et al., 1977 *N* metric). For example, close phonological neighbours of Eng. <cat> /kæt/ include <bat> /bæt/, <scat> /skæt/ and <at> /æt/.

The individuation of phonological neighbours for words in any given language is a task that can only be performed using computerized tools. Such tools exist for Dutch, English, French, German and Spanish words (CLEARPOND, see Marian et al., 2012), and for Basque, Dutch, English, French, German, Serbian, Spanish, and Vietnamese nonwords (WUGGY, see Keuleers, Brysbaert, 2010) but, to our knowledge, not for Italian words or nonwords. This state of affairs greatly limits the possibility of exploring the structure of the mental lexicon in Italian through the prism of phonological neighbourhood.

A recent example of such limitations is provided by the study on phonological analogy in Neapolitan Italian by Cangemi & Barbato (2014). The study tested the hypothesis that mid-vowel selection in nonce words is guided by the existence of phonologically similar actual word in the lexicon. For example, given the existence of the Neapolitan Italian words /ɛrba/ ('grass') and /aʃɛrba/ ('unripe', fem.), subjects were expected to utter /tɛrba/ (rather than /terba/) when asked to read aloud the nonword <TERBA>. Crucially, in the absence of a tool for the calculation of phonological neighbours, the notion of *phonologically similar words* had to be operationalized as *words with the same (poetical) rhyme*. While this choice made it possible to use an invert dictionary to find the target words, it also motivated some unexpected results which had to be reinterpreted post-hoc. This is the case of the nonword <PERMO>, which was expected to be uttered as /permo/ under the influence of /fermo/ 'closed', and /ermo/ 'lonely, remote' (which is however a relatively infrequent word). <PERMO> was however most often uttered as /pɛrmo/, probably under the influence of /pɛrno/ 'pivot', /pɛrdo/ 'to lose', and the very frequent /pɛrso/ 'lost' – words which do not have the same (poetical) rhyme as the stimulus, but are among its phonological neighbours.

By using a syllabic template to generate (non)words, and by listing their phonological neighbours along with their frequency of usage, WORDEN provide linguists and psycholinguists with a powerful tool to select stimuli for their experiments.

3. *WORDEN*

WORDEN is a library of functions to generate words and non-sense words based on a PYTHON module for knowledge engines development called *Python Knowledge Engine* (PYKE) (Frederiksen, 2008). We chose an declarative programming solution integrated in Python, rather than relying on external engines, to improve portability. It is also capable of providing insight about their phonological neighbours and the associated frequency information. In this section, we describe how the knowledge and the rules bases used in WORDEN were built. Examples on the queries the system can answer are provided: for each set of facts and rules, its formal definition is reported together with examples of queries and the system's output. We provide here a short summary of how logic programming works and how the examples are provided. Assume the following situation:

– Bill is Tom's parent;
– Bill is Will's sibling;
– Will is Richard's parent.

This is represented by a set of *facts* in the knowledge base. Genealogy states that the children of two siblings are cousins and that if one person is sibling of another the opposite is also true. This is represented by *rules* that can be used by a knowledge engine to infer the fact that Tom and Richard are cousins. We then query the system to ask it if

– Will is Bill's sibling (answer = yes);
– Who are Tom's cousins, if any (answer = Richard);
– Which parents are known by the system (answer = Bill and Will).

From the syntactic point of view, a *free variable* (an object the system can assign values to provide solutions to queries) is represented by a name starting with an uppercase character (e.g. X, Prefix, etc.). The underscore character represents a "don't care" variable, to which any possible value can be assigned during the solution of the query. The graphical representation of the example described above is organized as follows. In the left column, we display both facts (top) and rules (bottom). On the right side, lines starting with "?" indicate a user's query. Lines starting with ">>" indicate the system's feedback.

parent(Bill, Tom)	*? sibling(Will, Bill)*
parent(Will, Richard)	>> yes
sibling(Bill, Will)	
	? cousin(Tom, X)
sibling (X, Y) : - sibling(Y, X)	>> X= Richard
cousin (X, Y) : - parent(Z, X),	
parent(K, Y),	*? parent(X, _)*
sibling(Z, K)	>> *X= Bill*
	>> *X= Will*

Normally, an expert system is queried by command line to allow wide expressivity to the user. To help users access a closed set of functionalities related to words and nonsense words generation without using the command line, WORDEN also comes with a Graphical User Interface.

3.1 Knowledge base

The current version of WORDEN contains a knowledge base describing Italian only. This has been automatically generated starting from the PHONITALIA database (Goslin et al., 2013, http://www.phonitalia.org/). For each term in PHONITALIA, the grammatical categories it is assigned to are imported as part of the *wordGramCat* fact together with the frequency of occurrence of the word with that particular usage. The symbols used to represent grammatical categories are the same used in PHONITALIA and the syllabic structure of the word is imported as part of the *word* fact together with its phonological representation. For example, the word *abate* (Eng. 'abbot') is represented in WORDEN as follows (see the PHONITALIA guidelines for the meaning of the labels describing the grammatical categories). When asked, the system knows *abate* is a word.

wordGramCat(abate, S, 8)	*? word(abate, _, _)*
wordGramCat(abate, E, 4)	>> yes
wordGramCat(abate, E_IN_Ea, 2)	
wordGramCat(abate, S_IN_Ea, 1)	*? word(argine, _, _)*
word(abate, (a,ba,te), (a,b,a,t,e))	*>> Cannot prove*

The reader should note that, if asked about an atom being a word, if the knowledge base does not contain any information about that particular atom being a word, the system's output specifies that, with the current knowledge, it is not possible to prove that the atom is a word, implying that the answer may still be *yes,* should more knowledge be provided. In WORDEN, asking the system if an atom is a word actually means asking if the word is present in PHONITALIA. Thus, given the knowledge base represented in the left column of the example, the query on whether *argine* is a word yields the feedback *Cannot prove*.

The knowledge base of WORDEN also imports data concerning lemmas from PHONITALIA. However, in order to reduce the amount of data to be loaded, we only imported the words that are not lemmas of themselves. This way, it is later possible to introduce a rule stating that, if a word does not have a lemma in WORDEN's knowledge base, its lemma is the word itself.

	? lemma(X, abati)
lemma(abate, abati)	*>> X= abate*
lemma(X, X) :- not(lemma(_, X))	*? lemma(X, abate)*
	>> X= abate

In order to manage queries related to the words' CV structures, WORDEN's knowledge base contains information differentiating the symbols that can form a word in PHONITALIA between consonants and vowels.

vowel(a)	*? vowel(a)*
...	>> yes
consonant(t)	
...	*? consonant(9)*
	>> Cannot prove

3.2 Rules base

Rules are the most important part of WORDEN. A small set of rules can give the system the capability of inferring a large set of facts. Rules are often defined inductively: first, a simple base case is defined and, then, an induction step is introduced attempting to subdivide a complex problem into smaller ones while checking if it is possible to reach the base case.

In WORDEN's knowledge base we have imported the phonological representation of PHONITALIA words and we have introduced knowledge about PHONITALIA symbols representing consonants or vowels. This allows WORDEN to infer the CV structure of lists of PHONITALIA symbols using the following rules:

– The CV structure of an empty sequence is empty, too (base case)
– The CV structure of a sequence of symbols is composed by the symbol representing whether the first symbol of the list is a consonant or a vowel followed by the CV structure of the rest of the list.

In the following examples, we will indicate a list as [*Head, *Rest*], where *Head* is a variable containing the first symbol in the list while **Rest* is a list containing the same symbols of the whole list minus the first one. [] indicates an empty list.

cvStructure	*([], [])*	
cvStructure	*([Head, *Rest],*	
	*[CvHead, *CvRest]) :-*	
	CvHead = 'V'	*? cvStructure([l, a], X)*
	vowel(Head)	*>> X= [C, V]*
	*cvStructure(*Rest, *CvRest)*	
cvStructure	*([Head, *Rest],*	*? cvStructure(X, [C, V])*
	*[CvHead, *CvRest]) :-*	*>>X=[ba, da, ..., bi, di...]*
	CvHead = 'C'	
	consonant(Head)	
	*cvStructure(*Rest, *CvRest)*	

To better illustrate the example, let's consider the list of symbols representing the article *la*.

1. Of course, [l, a] does not represent an empty list so the first rule cannot be applied and the answer cannot be [];
2. The second rule checks if the first element of the list is a vowel: /l/ is not a vowel so it not possible to say that the answer is [V] followed by the structure of the list [a];
3. The third rule checks if the first element of the list is a consonant: /l/ is a consonant so it is possible to say that the answer is [C] followed by the structure of the list [a];
4. To complete the solution at point 3, the same set of rules is applied to the sublist [a]: since the first element is a vowel, the answer to the subproblem is [V] followed by the structure of an empty list [];
5. The CV structure of [] is [] by the first rule, so the solution to the subproblem of establishing the structure of [a] is [V] prepended to [], thus [V];
6. Since the solution to the problem of establishing the structure of [l, a] is [C] prepended to the structure of [a], the answer to the original query is [C, V] and is provided as output.

This way, in WORDEN a formal definition of the concept of *CV structure* is present. This provides the following advantages:
– It is not necessary to specify the structure of every word;
– It is possible to get the structure of any list of symbols (not only words in the knowledge base);
– Given an inductive definition of the concept, the knowledge engine is able to use the same rules to find all the possible solutions that lead to a specific structure.

In the example shown above, we query the system to get all the possible sequences of consonants followed by vowels. It is also possible to use the CV rules to get all the words or all the nonsense words with a specific CV structure. This is accomplished by adding two rules covering the different situations.

```
    cvNonce([CvHead, *CvRest],
[Head, *Rest]) :-
cvStructure([Head, *Rest],
[CvHead, *CvRest])
    not(word(_, _, [Head, *Rest]))

    cvWord([CvHead, *CvRest],
[Head, *Rest]) :-
cvStructure([Head, *Rest],
[CvHead, *CvRest])
    word(_, _, [Head, *Rest])
```

```
? cvNonce([C, V], X)
>> X= [[k, u], [w, o], [z, u]...]

? cvWord([C, V], X)
>> X= [[k, i], [s, e], [k, e]...]
```

An additional set of rules allows WORDEN to consider also optional symbols in the CV structure during generation so that, by considering the sequence [C, (C), V], the system outputs all the possible CV and CCV sequences of symbols.

In order to be able to manage phonological neighbourhoods, it is necessary to introduce in the rules base a formal description of the operations of *insertion* (a symbol is introduced in the starting sequence), *deletion* (a symbol is removed from the starting sequence) and *substitution* (a symbol becomes another one in the sequence). As these rules should be used in inductive procedures to generate phonological neighbourhoods, it is sufficient to define the three operations in terms of the first element, as we did for the first example in this section. The three rules should describe that:

- A sequence is obtained from another sequence by substituting its first element if all the symbols in the two sequences are the same but the first. In both cases the head symbols must be PHONITALIA symbols;
- A sequence is obtained by inserting a specific symbol to the head of another sequence if that symbol is prepended to the original sequence;
- A sequence obtained by deleting the head symbol of another sequence is the original one minus its head, regardless of what it was.

```
substitution([Start, *Rest1],
[Altstart, Rest2]) :-
    symbol(Altstart)
    symbol(Start)
    Start != Altstart

insertion([Start, *Rest], Ins, [Ins,
Start, *Rest]) :-
    symbol(Ins)
    symbol(Start)

deletion([_, *Rest], Rest)
```

```
? substitution([o,r,a], [a,r,a])
>> yes

? substitution(X, [a,r,a])
>> X= [[t,r,a], [f,r,a],...]

? insertion([m,a,r,i], a, [a,m,a,r,i])
>> yes

? substitution([p,a,r,c,o], [p,a,r,t,o])
>> yes
```

At this point, it is possible to introduce a set of rules to verify if a sequence can be obtained by applying one of the possible operations to another sequence. We will refer to the transformation of a string to another by insertion, deletion or substitution of a symbol as *perturbation.* During the exploration of the possible strategies to obtain a target sequence starting from an initial one it is necessary to consider that skipping a symbol is an additional operation. The rules should describe that:

1. A sequence is a perturbation of itself;
2. If two sequences share the same head symbol, this can be skipped;
3. If two sequences have different heads, it is possible to consider a substitution, deletion or insertion operation and check if the rest of the sequence is the target one.

This set of rules covers the case where one operation is allowed to reach a target sequence from a starting one. In this paper, we will cover this situation only. Future work will extend the system to include more complex planning to include sequences of operations.

```
    perturbation([Head, *Rest], [Head,
*Rest])

    perturbation([Head, *Rest1], [Head,
*Rest2]) :- perturbation(Rest1, Rest2)

    perturbation([Head1, *Rest1], [Head2,
*Rest2]) :-
substitution([Head1, *Rest1], [Head2,
*Rest2])

    perturbation([Head1, *Rest1], [Head2,
*Rest2]) :-
    insertion([Head1, *Rest1], [Head2,
*Rest2])

    perturbation([Head1, *Rest1], [Head2,
*Rest2]) :-
    deletion([Head1, *Rest1], [Head2,
*Rest2])
```

```
    ? perturbation([p,a,r,c,o],
[p,a,r,t,o])
    >> yes

    ? perturbation([p,a,r,c,o],
[p,a,r,c,o])
    >> yes

    ? perturbation([p,a,r,c,o],
[p,a,r, e, t, i])
    >> Cannot prove
```

At this point, it is possible to define a phonological neighbour of a sequence of symbols as another sequence of symbols that is a perturbation of the initial sequence and corresponds to a word. The starting sequence may be either a word or a nonsense word.

```
    phonNeighbour([Head1, *Rest1],
[Head2, *Rest2]) :-
    word(_, _, [Head2, *Rest2])
    perturbation([Head1, *Rest1],
[Head2, *Rest2])
```

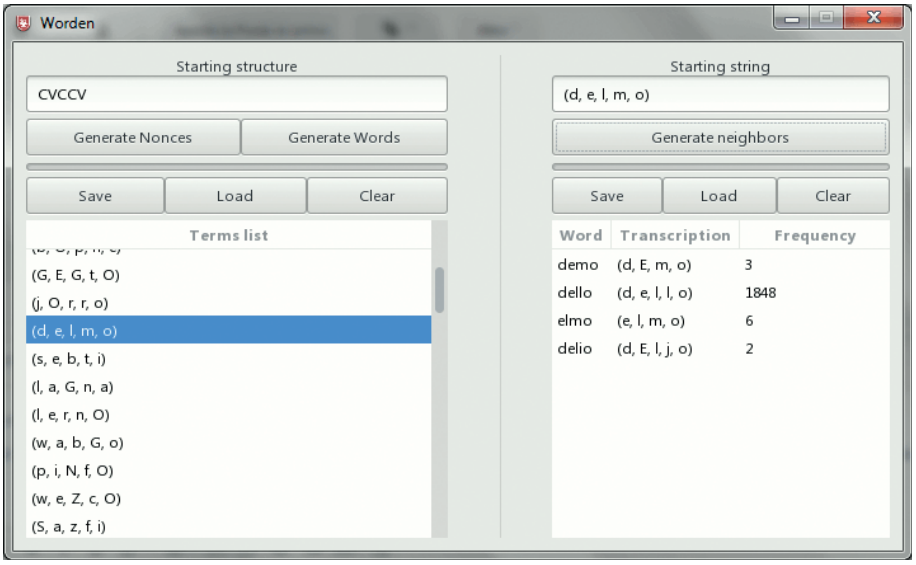
```
    ? phonNeighbour([p,a,r,c,o],
[p,a,r,t,o])
    >> yes

    ? phonNeighbour([p,e,r,m,o],
[f,e,r,m,o])
    >> yes

    ? phonNeighbour([p,e,r,m,o], X)
    >> X= [[f,e,r,m,o], ...]
```

3.3 Graphical User Interface

The WORDEN knowledge engine can be queried from command line to submit a variety of *questions* to the system. However, to support researchers working in the specific field of speech production and perception, we created a Graphical User

Interface to submit specific queries to the system. In Figure 3.1, we show a screenshot of the GUI.

Figure 1 - *Screenshot of the WORDEN Graphical User Interface*



The window is organized in two vertical frames. The frame on the left is dedicated to word and nonsense words generation starting from a given CV structure possibly including optional symbols. The frame on the right is dedicated to the generation of phonological neighbourhoods starting from given sequences of symbols. As phonological neighbourhoods only include words in PHONITALIA, the corresponding frequency is also given as output. It is possible to move a generated sequence to the field for phonological neighbourhood generation by double clicking the generated word/nonsense word. Functions to clear and load/save to text files both lists are provided.

4. *Conclusions*

We have presented WORDEN, a knowledge engine including a knowledge base extracted from the PHONITALIA database combined with rules providing a formal definition of the operations needed to generate words and nonsense words. WORDEN also contains a formal definition of phonological neighbourhood starting from arbitrary sequences of symbols.

Using knowledge engines allows to represent formal theories that can be automatically proved or disproved by the engine. Moreover, given a formal description of the elements involved in a theory, knowledge engines remove the need to store a large amount of data that can be inferred at runtime. Words, nonsense words and phonological neighbourhoods, in WORDEN, are generated rather than stored.

This allows future extension of the system to include more complex theories building on the top of the already provided concepts. While knowledge engines can answer a wide set of queries on the basis of the provided facts and rules sets, we developed a GUI to let researchers interact more easily with the system. While the GUI limits the possible queries to the ones that are connected to the buttons, it is always possible to submit more complex queries from the command line by using PYKE. Being implemented in PYTHON, the system is cross platform as long as PYTHON is installed. To make it easier for WINDOWS users to access WORDEN, a platform specific binary is provided.

Future work will concentrate on extending the concepts included in the WORDEN knowledge. This can both be done by including more languages from databases other than PHONITALIA and by extending the rules set to allow more complex theories to be represented. Our current orientation, in this sense, is to include the possibility of building perturbation plans involving sequences of operations based on sub- or supra-segmental features and introducing more specific rules on symbols assembling (i.e. the sonority scale principle) to automatically syllabify input sequences of symbols. The tool can be obtained by contacting the authors.

References

Bird, S., Klein, E. (1990). Phonological Events. In *Journal of Linguistics,* 26, 33-56.

Bird, S., Ellison, T.M. (1994). One level phonology autosegmental representations and rules as finite automata. In *Computational Linguistics*, 20, 55-90.

Cangemi, F., Barbato, M. (2014). A laboratory approach to acquisition and change: Italian mid-vowels in nonce words. Talk presented at the *7th Laboratory Approaches to Romance Phonology (LARP) Conference.*

Cangemi, F., D'Imperio, M. (2013). Tempo and the perception of sentence modality. In *Journal of the Association for Laboratory Phonology*, 4 (1), 191-219.

Coltheart, M., Davelaar, E., Jonasson, J.T. & Besner, D. (1977). Access to the internal lexicon. In *Attention and Performance*, VI, 535-555.

D'Imperio, M. (2001). Focus and tonal structure in Neapolitan Italian. In *Speech Communication*, 33 (4), 339-356.

Di Cristo, A. (1985). *De la microprosodie à l'intonosyntaxe*. Aix-en-Provence: Publications Université de Provence.

Frederiksen, B. (2008). *Applying expert system technology to code reuse with Pyke*. In Proc. of PyCon [Online: pyke.sourceforge.net/PyCon2008- paper.html].

Goslin, J., Galluzzi, C. & Romani, C. (2013). Phonitalia: a phonological lexicon for Italian. In *Behavior Research Methods*, 46 (3), 872-886.

Grishman, R. (1986). *Computational Linguistics: An introduction*. Cambridge University Press.

Keuleers, E., Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. In *Behavior Research Methods*, 42 (3), 627-633.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge: The MITPress.

Luce, P.A., Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. In *Ear and Hearing*, 19, 1-36.

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).

Nugues, P.M. (2014). *Language Processing with Perl and Prolog: Theories, Implementation and Application*. New York: Springer-Verlag.

Petrone, C. (2008). *Le rôle de la variabilité phonétique dans la représentation des contours intonatifs et de leur sens*, PhD Thesis, Université Aix-Marseille I.

Stone, M. (2014). Semantics and computational semantics. In Aloni, M., Dekker, P. (Eds.), *Cambridge Handbook of Semantics*. Cambridge: Cambridge University Press.

DOI: 10.17469/O2101AISV000031

ANDREA PAOLONI, MASSIMILIANO TODISCO

Calcolo del rapporto di verisimiglianza tra parlanti in prove soggettive di ascolto

Voice identification using measurements of physical parameters, generally made through instrumental techniques, combines a significant performance, especially in the case of automatic methods, with an easy control of the data obtained. However, the possibility of working in this field with subjective investigation criteria on the basis of human listening should not be underestimated. In the present work we illustrate a method for obtaining a numerical value based on the likelihood, for subjective tests in speaker recognition tasks. This method was finally validated through a number of experiments made by two different types of listener: non-experts and technical experts.

1. *Introduzione*

È esperienza comune riconoscere una persona dalla voce quando la ascoltiamo parlare senza vederla, perché in un ambiente diverso o perché la ascoltiamo attraverso il telefono. Peraltro, il riconoscimento vocale dell'identità di una persona basato sul mero ascolto è stato l'unico disponibile per molti millenni ed è stato adottato, nelle sue diverse connotazione in numerosi casi giudiziari, anche celebri[1].

In ambito forense però questa metodica di riconoscimento del parlante è stata sostanzialmente abbandonata in favore dell'identificazione della voce tramite parametri acustici, in quanto queste misure erano in grado di supportare il parere dell'esperto con dati numerici verificabili (Paoloni, 1997; 2014). Inoltre queste metodiche uniscono al vantaggio della parziale o totale automazione dei processi operativi una notevole velocità di esecuzione delle misure, soprattutto nel caso dei metodi automatici.

Per quanto detto, sembrerebbe logico l'abbandono dell'indagine uditiva a favore delle tecniche automatiche. Vogliamo però richiamare l'attenzione sul fatto che, allo attuale della conoscenza, anche le più sofisticate metodologie automatiche (Campbell, 2014) (Drygajlo, 2012) non utilizzano, al fine del riconoscimento del parlante, tutte le caratteristiche della sua voce, caratteristiche che nella loro interezza, sono valutate dall'orecchio e dal sistema di percezione di un ascoltatore e a volte costituiscono un elemento decisivo di identificazione. Le considerazioni sopra riportate suggeriscono quindi l'opportunità di utilizzare entrambe le metodologie. Nel presente lavoro illustriamo una via per ottenere, anche dalle prove soggettive, un dato numerico verificabile nella forma oggi ritenuta migliore, ossia nel rapporto di verisimiglianza. Il rapporto di veri-

[1] La voce del sequestratore del figlio del famoso trasvolatore Lindbergh fu da lui riconosciuta in Tribunale come quella che aveva chiamato Condon al cimitero la notte della consegna del riscatto.

simiglianza (LR – Likelihood Ratio) è il rapporto tra la similarità dei campioni a confronto, ovvero quanto sia simile il campione del noto rispetto a quello dell'anonimo, e la tipicità, ovvero quanto il campione all'esame sia comune nella popolazione di riferimento. Supponendo che i valori di similarità siano dello stesso ordine di grandezza, il rapporto dipende da quanto il campione sia comune all'interno della popolazione.

Il presente lavoro è organizzato nel seguente modo: nel paragrafo 2 verranno illustrate i criteri per le prove soggettive di ascolto; nel paragrafo 3 la metrica utilizzata e la definizione dell'esperimento; i risultati ottenuti sono riportati nel paragrafo 4 e infine il paragrafo 5 presenterà le conclusioni.

2. *Le prove soggettive di ascolto*

Nell'identificazione del parlante attraverso l'ascolto si possono definire due diverse situazioni. La prima è quando la voce da identificare è nota all'ascoltatore (*familiar voices*) e sfrutta la capacità dell'uomo di memorizzare archetipi della caratteristiche peculiari di un suono, sia esso una voce, sia esso una musica. Questa capacità, che è quella che ci consente di riconoscere la voce dei nostri congiunti e dei nostri amici, è quella che secondo la letteratura fornisce le migliori prestazioni, con un tasso di riconoscimento superiore al 90%.

La seconda è quella dagli esperti negli esperimenti di "ricognizione" e si basa sul confronto tra due o più voci non note all'ascoltatore (*unfamiliar voices*) rese contemporaneamente disponibili da un sistema di registrazione audio. In questo caso l'ascoltatore confronta coppie di voci in successione (o anche contemporaneamente), cercando elementi di diversità o di somiglianza. Questa metodica ha prestazioni lievemente inferiori alla precedente, con un tasso di riconoscimento intorno all'85% (Nolan, 1983). Nel caso che la situazione sia quella di un esperto estraneo ai fatti (e non può perciò trattarsi di voce familiare) le metodologie più frequentemente impiegate per il confronto uditivo delle voci possono essere ricondotte ai seguenti due criteri.

Un primo criterio basato su un reiterato ascolto, da parte dell'esperto, dei campioni di voce in esame al fine di individuare eventuali elementi di natura linguistica, fonatoria o acustica comuni alle due voci. L'esperto, sulla base degli elementi recepiti, esprimerà un giudizio sulla attribuzione o meno ad uno stesso parlatore delle voci ascoltate. Un secondo criterio è quello basato sul confronto delle voci effettuato da una squadra di ascoltatori, anche non esperti. Il materiale fonico in questo caso è costituito da un insieme di voci comprendenti la voce da identificare, le voci anonime ed eventualmente alcune voci estranee (popolazione di riferimento), prelevate da parlatori aventi caratteristiche fonatorie simili a quelle delle voci in esame. Questi campioni vengono poi ascoltati a coppie che saranno, se presentate tutte agli ascoltatori, $M(M\text{-}1)/2$, dove M è il numero di campioni al confronto. Ciascun operatore dopo l'ascolto di ogni coppia dovrà esprimere un giudizio di similarità o meno delle voci a confronto. L'elaborazione statistica dei giudizi espressi dagli ascoltatori consente di giungere a conclusioni di tipo sostanzialmente qualitativo (Anil, 2005). È tuttavia possibile, utilizzando questo secondo metodo, stimare un rapporto di verisimiglianza (LR) che sarà dato dal rapporto tra il

valore medio della somiglianza attribuita alla coppia o alle coppie voce anonima-voce del sospettato (voce nota) e il valore medio della somiglianza attribuita alle coppie voce del sospettato-voci della popolazione di riferimento.

3. *Definizione dell'esperimento*

La nostra preferenza, tra i metodi di ascolto citati, va all'ultimo dei criteri illustrati che si rifà ad una collaudata metodica utilizzata nelle valutazioni di qualità in campo telefonometrico. Questo metodo consente la stima della verisimiglianza relativa alla particolare prova eseguita e pertanto consente una valutazione dell'attendibilità del risultato rispettando i criteri richiesti in ambito giudiziario. Per verificare quali valori di LR siano ottenibili con questo tipo di prove, è stato proposto un test disponibile in rete che mette a confronto 14 campioni di popolazione (p_1,p_2,...,p_{14}) + 2 campioni anonimi (a_1,a_2) con 2 campioni della voce nota (n_1,n_2) per un totale di 32 confronti divisi in due sessioni di test (16 + 16). L'ascoltatore potrà ascoltare quante volte desidera ciascuno dei 32 confronti ed esprimerà, per ciascuno di esso, un giudizio di somiglianza a 7 livelli: somiglianza 0 (nessuna), somiglianza 1 (minima), somiglianza 2 (limitata), somiglianza 3 (qualche somiglianza), somiglianza 4 (abbastanza simile) somiglianza 5 (molto simile), somiglianza 6 (identità). La *LR* (rapporto di verisimiglianza) è calcolata secondo la formula

$$LR = \frac{1}{K}\sum_{k=1}^{K} lr_k$$

dove *K* è il numero di ascoltatori e *lr* è dato da

$$lr = \frac{\dfrac{1}{N_n N_a}\displaystyle\sum_{i=1}^{N_n}\sum_{j=1}^{N_a} S(n_i, a_j)}{\dfrac{1}{N_n N_p}\displaystyle\sum_{i=1}^{N_n}\sum_{j=1}^{N_p} S(n_i, p_j)}$$

dove *S* è il valore di somiglianza tra le coppie di voci, N_n, N_a e N_p sono, rispettivamente, il numero totale delle voci note, delle voci anonime e delle voci della popolazione di riferimento. In particolare, nel nostro esperimento $K = 26$, $N_n = 2$, $N_a = 2$ e $N_p = 14$.

Si è ritenuto altresì utile verificare se le prestazioni di ascoltatori esperti fossero migliori di quelle di soggetti che non hanno particolare addestramento nel campo riconoscimento delle voci e pertanto il test è stato sottoposto a 9 esperti e 17 non esperti.

4. *Risultati*

I risultati delle misure soggettive d'identificazione del parlante sono riportati nella Tabella 1 e 2.

Nella Tabella 1 sono riportate le LR di 9 esperti, mentre nella Tabella 2 quelle di 17 non esperti.

Nella Tabella 3 sono invece riportare le medie e le varianze relative ai due esperimenti, a1 e a2, degli esperti e non esperti.

Tabella 1 - *Likelihood Ratio dei 9 Esperti*

		ESPERTI																
		n1	n2	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	LR
E1	a1	2	0	1	0	0	0	0	0	1	2	0	1	0	0	0	0	5.60
	a2	0	2	0	0	0	1	0	1	1	0	1	0	1	1	0	2	3.50
E2	a1	6	0	1	0	1	0	1	0	0	1	0	1	0	0	0	1	14.00
	a2	0	6	1	0	0	0	0	0	0	1	0	0	0	0	0	1	28.00
E3	a1	3	0	2	0	0	0	0	0	0	4	0	3	0	0	0	0	4.67
	a2	0	5	3	0	0	3	1	2	4	4	0	2	1	3	0	3	2.69
E4	a1	4	4	4	0	1	3	0	1	2	2	0	1	0	0	0	4	3.11
	a2	0	4	2	0	5	5	1	1	3	2	2	1	1	5	0	2	1.87
E5	a1	4	5	5	5	0	2	1	1	0	0	0	2	1	0	0	0	3.29
	a2	4	6	3	6	6	5	1	5	1	2	5	5	4	2	1	5	1.65
E6	a1	3	0	1	1	0	0	1	0	0	3	0	1	0	0	0	0	6.00
	a2	0	4	0	1	0	0	0	0	1	1	1	0	0	1	0	0	11.20
E7	a1	4	0	4	1	0	0	0	0	0	0	0	1	0	1	0	1	7.00
	a2	0	5	5	1	0	5	5	4	4	0	1	2	1	4	0	0	2.19
E8	a1	4	1	1	0	0	0	0	1	0	2	0	0	0	0	0	0	14.00
	a2	0	4	0	0	5	1	0	-1	1	1	0	0	1	5	0	0	4.31
E9	a1	5	0	3	2	0	0	5	0	0	0	0	4	1	0	0	0	4.67
	a2	0	5	0	1	5	4	0	4	5	0	1	1	2	3	0	1	2.59

Tabella 2 - *Likelihood Ratio dei 17 non Esperti*

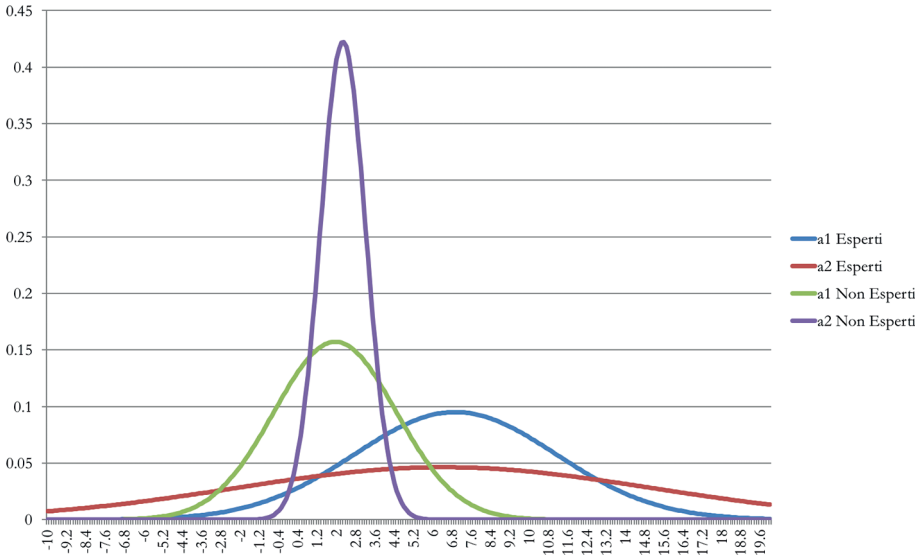
		NON ESPERTI																
		n1	n2	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	LR
I1	a1	6	6	5	0	3	4	3	3	4	4	0	1	0	2	0	0	2.90
	a2	2	6	2	0	1	6	0	-1	6	0	0	3	1	5	0	1	3.50
I2	a1	1	0	3	1	4	2	0	4	4	5	3	0	5	4	1	5	0.34
	a2	4	4	4	4	5	1	2	3	1	1	0	0	4	0	1	1	2.07
I3	a1	5	0	5	0	0	0	0	0	0	2	0	0	0	0	0	0	10.00
	a2	0	5	2	0	2	1	1	2	0	4	0	0	0	5	0	0	4.12
I4	a1	1	0	3	2	1	0	1	0	1	3	2	1	1	1	1	1	0.78
	a2	1	4	4	1	3	3	2	2	1	4	1	1	2	2	0	0	2.15
I5	a1	2	4	5	5	0	2	4	2	5	5	1	2	1	0	2	3	0.76
	a2	4	3	5	2	2	4	0	3	5	3	2	5	4	6	4	5	0.84
I6	a1	6	4	6	0	6	4	0	4	2	5	0	5	4	6	0	6	1.75
	a2	0	5	2	0	6	0	4	4	6	4	2	5	3	5	0	5	1.52
I7	a1	1	2	1	2	1	2	6	1	2	4	4	3	2	2	2	3	1.20
	a2	3	6	2	3	4	5	1	1	4	2	2	3	0	2	3	3	2.40
I8	a1	0	4	4	0	0	0	0	0	2	0	1	4	0	0	0	4	0.00
	a2	0	2	4	2	3	3	0	4	3	1	3	4	3	0	0	1	0.90
I9	a1	2	3	1	2	0	0	0	2	1	1	3	3	1	0	0	3	1.65
	a2	2	3	3	2	3	3	1	2	0	1	3	1	0	1	0	2	1.91
I10	a1	4	3	0	2	1	0	5	1	1	3	1	1	0	0	0	0	3.73
	a2	0	6	5	1	5	5	4	1	4	3	0	2	1	1	2	0	2.47
I11	a1	2	2	4	1	3	5	1	2	1	2	1	4	4	2	1	4	0.80
	a2	2	5	4	2	5	4	6	5	5	2	3	4	2	4	1	4	1.37
I12	a1	6	0	6	0	0	0	4	0	1	6	0	1	0	0	1	1	4.20
	a2	0	5	1	0	3	6	0	1	5	2	5	5	0	6	0	5	1.79
I13	a1	1	0	1	0	0	0	0	0	0	3	0	0	0	0	0	2	2.33
	a2	0	3	1	0	2	3	1	0	2	0	0	1	0	2	0	1	3.23
I14	a1	4	0	6	6	1	0	1	3	1	3	1	3	0	2	0	6	1.70
	a2	0	6	4	1	5	6	3	4	4	0	3	2	5	4	2	3	1.83
I15	a1	0	0	0	0	5	1	1	0	1	5	0	5	1	1	5	4	0.00
	a2	2	6	1	0	4	5	5	1	6	0	1	1	0	5	3	4	2.33
I16	a1	2	4	2	2	1	2	4	1	2	4	4	3	2	2	2	3	0.82
	a2	2	5	2	1	3	3	1	1	4	2	2	3	0	2	3	3	2.33
I17	a1	0	0	4	0	0	0	0	0	0	0	0	4	0	0	0	0	0.00
	a2	0	6	6	0	1	0	0	0	0	5	0	5	0	6	0	0	3.65

Tabella 3 - *Medie e le varianze relative ai due esperimenti, a1 e a2, degli esperti e non esperti*

	MEAN ESPERTI	STD ESPERTI	MEAN NON ESPERTI	STD NON ESPERTI
a1	6.93	4.19	1.94	2.54
a2	6.44	8.60	2.26	0.94

In Figura 1 sono stati altresì riportati i valori dei test d'identificazione del parlante a1 e a2 eseguiti da esperti e non esperti

Figura 1 - *Risultati dei test d'identificazione del parlante a1 e a2 eseguiti da esperti e non esperti*



5. *Conclusioni*

Obiettivo del presente lavoro era proporre un metodo che consenta di oggettivizzare i risultati delle prove soggettive d'ascolto utilizzando il rapporto di verisimiglianza così come avviene per i sistemi automatici di riconoscimento del parlante. L'esperimento ha permesso anche di verificare che le prestazioni dell'esperto sono senz'altro migliori di quelle di soggetti che non abbiano particolare addestramento nel campo riconoscimento delle voci. Per quanto attiene ai valori di LR, che confermano l'identità delle voci di interesse, si osserva una compressione della scala, nel senso che è lecito ritenersi che anche in casi evidenti, ossia di grande somiglianza delle voci a confronto e di significativa differenza di quelle della popolazione di riferimento, difficilmente si riuscirà a ottenere valori di verisimiglianza elevati in quanto un soggetto tende a non fornire dati che escludano la somiglianza delle voci "diverse", ponendo tale valore a zero, e dare il valore massimo alle voci provenienti alla stessa sorgente. Il fatto quindi che il rapporto di verisimiglianza abbia valori

così compressi potrebbe far ritenere opportuno considerare l'impiego di una scala logaritmica come in altre fenomenologie percettive.

Bibliografia

Anil, A., Dessimoz, D., Botti, F. & Drygajlo, A. (2005). "Aural and Automatic Forensic Speaker Recognition in Mismatched Conditions". In *The International Journal of Speech, Language and the Law*, vol. 12, 214-234.

Campbell, J.P. (2014). Speaker Recognition for Forensic Applications. Keynote Address at Odyssey 2014: *The Speaker and Language Recognition Workshop*. Joensuu, Finland.

Drygajlo, A. (2007). "Forensic Automatic Speaker Recognition". In *IEEE Signal Processing Magazine*, 24 (2), 132-135.

Drygajlo, A. (2012). Automatic Speaker Recognition for Forensic Case Assessment and Interpretation. In Neustein, A., Patil, H.A. (Eds.), *Forensic Speaker Recognition, Law Enforcement and Counter-Terrorism*. Berlin: Springer, 21-39.

Falcone, M., De Sario, N. (1994). "A PC speaker Identification System for Forensic Use: IDEM". In *ESCA Workshop on Automatic Speaker Recognition, Identification, and Verificatio*, Martigny, Switzerland.

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Nolan, F. (1997). Speaker recognition and forensic phonetics. In Hardcastle, W., Laver, J. (Eds.), *A Handbook of Phonetic Science*. Oxford: Blackwell, 744-767.

Koenig, B (1993). "Selected Topics in Forensic Voice Identification". In *Crime Laboratory Digest*, vol. 20, n. 4, 78-81.

Paoloni, A. (1997). Il riconoscimento del parlatore. In *Detective&Crime Magazine / Criminalistica - Le indagini fonetiche*.

Paoloni, A. (2014). Sul riconoscimento del parlante in ambito forense. In *Sicurezza e Giustizia*, Anno IV, n. 3.

Paoloni, A. (2003). Note sul riconoscimento del parlante nelle applicazioni forensi con particolare riferimento al metodo parametrico IDEM. In *Riv. Italiana di Acustica*, vol. 27 n. 3-4.

Rose, P. (2002). *Forensic Speaker Identification*. Taylor & Francis.

Wolf, J.J. (1972). "Efficient acoustic parameters for speaker recognition". In *J.A.S.A.*, vol. 51, n. 6, 2044-2056.

DOI: 10.17469/O2101AISV000032

ALESSANDRO VIETTI, VITTORIO ANSELMI, LORENZO SPREAFICO

Verso un sistema di riconoscimento automatico del parlato tramite immagini ultrasoniche

We provide a proposal for a silent speech interface based on Ultrasound Tongue Imaging (UTI) data. The model we present appears to be the first proposal for a silent speech interface based on UTI for the Italian language. The model is fed with ultrasonic data only (e.g. lips movements are not included in the model) and the extraction of information from ultrasound images is accomplished via a model of artificial neural network instead of standard hidden Markov Model. In the paper, we illustrate the phases of acquisition, filtering and modeling of articulatory data. Then, we describe the training of the neural network, with particular emphasis on the problems associated with the process of identification and extraction of salient information. Eventually, we report the success rate for the model when recognizing Italian real words.

1. *Introduzione*[1]

Nel contributo sono presentate le procedure seguite per formulare un primo modello di interfaccia silente basata su immagini ecografiche. Anzitutto vengono illustrate le fasi di acquisizione, filtraggio e modellizzazione del dato articolatorio e, successivamente, sono descritte le modalità di addestramento della rete neurale, mettendo in particolare rilievo i problemi e le difficoltà legate al processo di individuazione ed estrazione di informazioni salienti per il funzionamento dell'interfaccia silente.

Le interfacce silenti (in inglese: *silent speech interfaces*, SSI) sono sistemi costituiti da dispositivi fisici e virtuali che permettono anche agli individui impossibilitati a produrre un segnale acustico udibile di comunicare. Le interfacce silenti possono essere impiegate sia come ausilio per persone con disabilità riconducibili a disturbi centrali della motricità del distretto fono-articolatorio e/o a cause organiche (malformazioni congenite o acquisite) o chirurgiche (p.e.: laringectomia, glosseectomia parziale) che determinano alterazioni del tratto vocale; sia come supporto funzionale per persone operanti in ambienti estremamente rumorosi o, al contrario, estremamente silenziosi che devono però poter comunicare.

Le interfacce silenti operano intercettando segnali di varia natura in uno qualunque degli stadi neuromiofonofisiologici di produzione del parlato, interpretandoli e convertendoli in una qualche forma di rappresentazione che possa essere succes-

[1] Sebbene il contributo sia il risultato dell'attività di ricerca congiunta dei tre autori, la responsabilità della stesura dell'articolo è stata così ripartita: §1, §2 LS; §3 LS e AV; §4 VA e AV; §5 AV.

sivamente impiegata per la sintesi artificiale della voce o del messaggio linguistico. Come analiticamente illustrato in Denby et al. (2010), i dati interpretabili e sfruttabili per la sintesi del parlato possono avere natura radicalmente differente. È quindi possibile distinguere tra loro sistemi che intercettano l'informazione a livello encefalico e sistemi che intercettano l'informazione a livello articolatorio. I sistemi del primo tipo interpretano segnali elettroencefalografici (EEG) associati all'articolazione (Wester, 2006) o al pensiero di parole (Porbadnigk, 2009), oppure segnali provenienti da microelettrodi impiantati nella corteccia motoria mirati a monitorare l'attività dei neuroni coinvolti nella produzione del parlato (Brumberg et al., 2010). I sistemi che intercettano l'informazione a livello articolatorio, invece, possono interpretare segnali assai più diversificati: per fare un esempio che riguarda la laringe, dati elettromiografici di superficie (sEMG) relativi al movimento dei muscoli (Jorgensen, Dusan, 2010); dati laringografici (EGG) relativi all'attività glottidale integrati con dati sui movimenti dei tessuti durante l'articolazione ricavati tramite sensori di moto (GEMS: *general electromagnetic motion sensor*; Quatieri et al., 2006); dati acustici relativi ai cosiddetti mormorii non udibili (NAM: *non-audible murmur*) intercettati con microfoni stetoscopici (Heracleous, 2007); dati elettromagnetoarticolatografici (EMA: *electromagnetic articulography*) relativi al movimento di punti fissi degli articolatori orali (Wang et al., 2012); dati ultrasonici relativi al movimento dell'intera lingua (UTI: *ultrasound tongue imaging*).

Scopo di questo articolo è presentare un modello di interfaccia silente basata su dati ultrasonici linguali e finalizzata al riconoscimento di un numero limitato di parole della lingua italiana. La tecnica, non invasiva e clinicamente sicura, consiste nel visualizzare in tempo reale una sezione sagittale della lingua del parlante grazie al posizionamento di una sonda ecografica sotto al suo mento. Le informazioni utili per inferire quali parole siano state pronunciate o silenziosamente articolate dal parlante sono quindi ricavate dall'osservazione della posizione e della forma della sezione sagittale della lingua.

L'impiego della tecnica UTI per la costruzione di un'interfaccia silente non è una novità e, anzi, tra le alternative riportate sopra è sicuramente quella più e meglio esplorata, soprattutto in virtù del suo potenziale. Infatti basandosi sull'osservazione dei soli movimenti linguali e non prevedendo né attività glottidale, né passaggio d`aria nella cavità orale, la tecnica si presta all'applicazione in ciascuno dei campi d'impiego elencati inizialmente, inclusi i parlanti laringectomizzati e con la sola esclusione dei parlanti glottectomizzati (peraltro non assoluta, cfr. infatti la ricerca di Stone et alii, 2013 sulla riconfigurazione dei gesti linguali in parlanti che abbiano subito resezioni parziali della lingua).

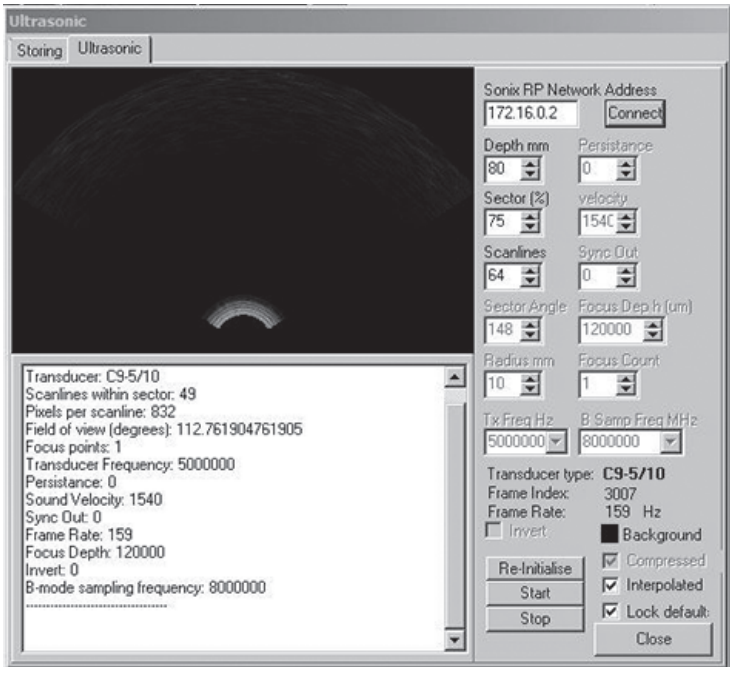
Il modello di interfaccia silente qui presentato differisce da quelli proposti altrove per tre motivi almeno: (i) anzitutto perché ci risulta essere la prima proposta di interfaccia silente su base UTI per la lingua italiana (le altre applicazioni note riguardano anzitutto l'inglese: Denby, Stone, 2004; Denby et al., 2006; Hueber et al., 2007a,b,c; 2008a,b; 2010; 2011; 2012; e, in misura più ridotta, il francese: Hueber, 2013); (ii) quindi perché vengono impiegati esclusivamente dati ultrasonici e non,

come nella maggior parte della altre proposte, dati ultrasonici integrati da dati relativi alla posizione delle labbra (Hueber et al., 2007a,b,c; 2008a,b; 2010); infine (iii) perché l'estrazione delle informazioni dalle immagini ecografiche viene operata ricorrendo a un modello di rete neurale artificiale e non, come in altri approcci, a un modello di Markov nascosto (cfr. Hueber et al., 2007a,b,c; 2008a,b; 2010; 2011; 2012). Prima di discutere di quest'ultimo aspetto presentiamo la metodologia di elicitazione dei dati impiegata.

2. *Acquisizione del dato articolatorio*

Per le finalità del progetto i dati articolatori linguali sono stati elicitati ricorrendo a un sistema *Ultrasonix™ SonixTablet™* accoppiato ad una sonda microconvessa C9-5/10 operante a 5MHz. Ecografo e sonda sono stati interfacciati al sistema di sincronizzazione, acquisizione e analisi di dati ultrasonici *Articulate Assistant Advanced™ 2.14* prodotto dalla *Articulate Instrument™*. I dati sono stati campionati a 159 Hz con l'impostazione di fuoco di 8 cm e di un angolo di vista di 112° (Figura 1).

Figura 1 - *Impostazioni di acquisizione dei dati*



La sonda è stata fissata sotto il mento degli informanti impiegando un elmetto stabilizzatore (Scobbie et al., 2008; cfr. Figura 2).

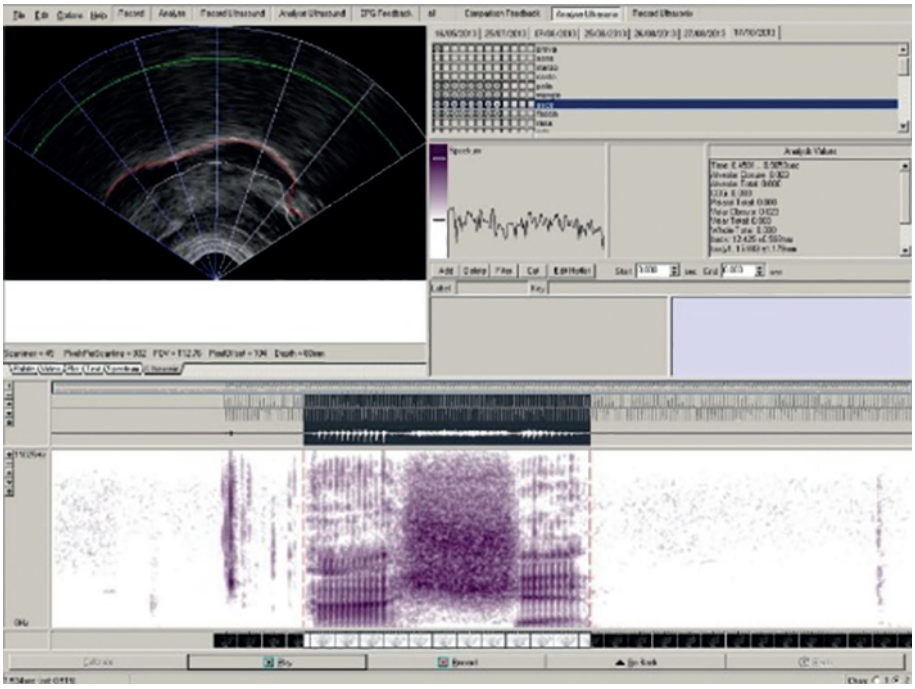
Figura 2 - *Stabilizzazione della sonda ecografica*

Nel campione sono stati inclusi tre soggetti maschili di età compresa tra 35 e 45 anni, parlanti nativi di italiano provenienti da regioni diverse dell`Italia settentrionale (Lombardia, Piemonte, Veneto). Ai parlanti è stato chiesto di leggere a voce alta una lista di 28 parole: la lista conteneva elementi differenziati tra loro per proprietà fonologiche e includeva coppie minime differenziate per luogo di articolazione di un fono come in: /kanto ~ tanto; aʃʃe ~ asse; mandzo ~ mandʒo; martso ~ martʃo; kanna ~ kaɲɲa; palla ~ paʎʎa/; per modo di articolazione di un fono come in: /sana ~ tana ~ rana ~ lana; fattʃa ~ faʃʃa/); o per sonorità di un fono come in: /kala ~ gala; rido ~ rito/. La lista di parole conteneva inoltre elementi quali /si; no; dɛstra; sinistra/ funzionali al progetto di costruzione di un`interfaccia silente per il riconoscimento di comandi elementari.

La lista di parole, presentate in isolamento e non contenute in frasi, è stata ripetuta a voce normale per quattro volte da ciascun parlante. I gesti articolatori così elicitati sono quelli relativi al parlato non silente.

3. *Analisi del dato articolatorio*

Figura 3 - *Identificazione dei fotogrammi ultrasonici*



I dati articolatori linguali sono stati elaborati in AAA™2.14 con la finalità di estrarre i profili linguali sagittali per ogni ripetizione di parola. Una volta identificate inizio e fine della parola sulla base delle informazioni acustiche (oscillogramma e spettrogramma), i profili linguali vengono tracciati automaticamente dal programma per l'intera parola (Figura 3). La procedura automatica di tracciatura prevede l'individuazione dei punti più luminosi lungo una serie di 42 raggi posti a ventaglio sull'immagine ecografica e successivamente l'unione dei punti così ottenuti tramite una *spline*, la linea che congiunge tutti i punti individuati (Wrench, Scobbie, 2011).

In questo caso, non si sono operate correzioni manuali e le coordinate della *spline* sono state successivamente esportate in MATLAB™ come una matrice di 42 coppie di coordinate per ciascun fotogramma ecografico nel tempo della parola (~80 frame per parola). Per ciascuna ripetizione di parola così elaborata ed esportata nel piano cartesiano si è selezionato un singolo punto sull'asse delle ascisse al fine di tracciare l'andamento del profilo linguale lungo l'asse delle ordinate e misurare quindi la dislocazione di quella sezione di lingua per ciascun istante nel tempo[2] (per una rappresentazione grafica v. Figura 4).

[2] Nel campo dell'ultrasonografia corrisponde grosso modo alla modalità di produzione dell'onda e rappresentazione dell'immagine conosciuta come *M-mode*.

Figura 4 - *Identificazione di un punto sul profilo linguale derivato dal segnale ultrasonico*

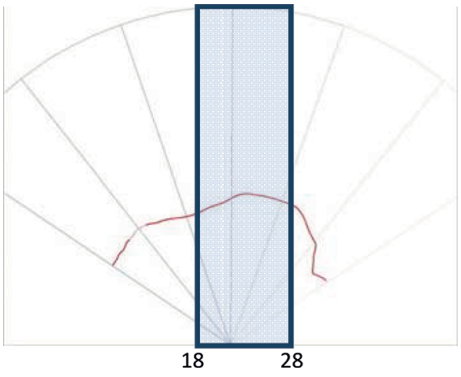
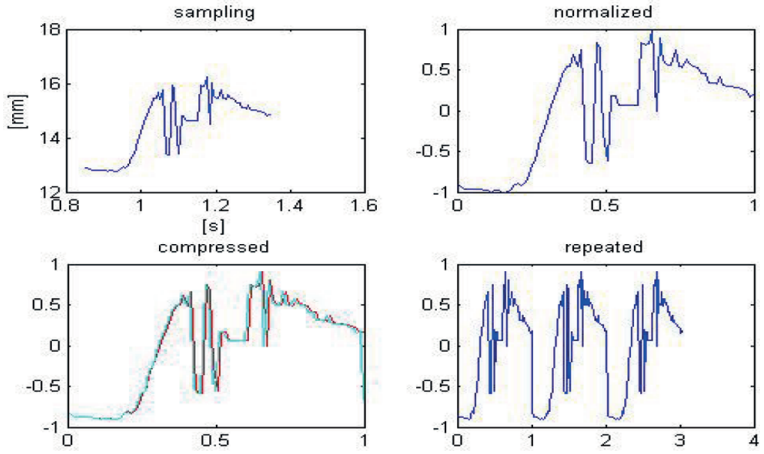


Figura 5 - *Normalizzazione della parola mangio.*
Il grafico in alto a sinistra riproduce l'andamento nel tempo del valore di y (ampiezza) per il punto considerato; quello in alto a destra riproduce l'andamento del valore di y dopo la normalizzazione; quello in basso a sinistra rappresenta il segnale compresso che viene ripetuto tre volte nel grafico in basso a destra



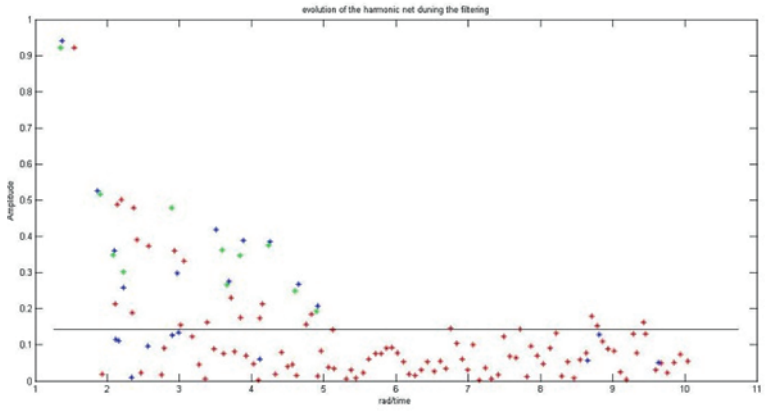
Il punto sull'asse delle ascisse è stato selezionato all'interno di uno spazio funzionale compreso tra la regione postdorsale e quella apicale della lingua che, da ripetute verifiche, è risultato essere quello più rappresentativo. Qualora la tracciatura del movimento lungo l'asse y fosse impedita dall'assenza di un valore in virtù di discontinuità nell'immagine ecografica dettata da un contatto tra la lingua e un altro articolatore, in particolare il palato, oppure da ombre gettate sull'immagine da strutture ossee come lo ioide e la mandibola (Stone, 2005; Spreafico, Vietti, 2013), è stato computato il valore di y per il punto x più prossimo a quello selezionato. I dati ottenuti sono stati successivamente convertiti in una matrice dal

formato (time, $x_{selectedpoint}$, $y_{timecorresponding}$), dove per ogni istante di tempo si ha un valore spaziale.

I valori così campionati sono stati normalizzati in ampiezza e durata. I valori di normalizzazione per l'ampiezza sono stati compresi tra ±0.9. Quelli per la durata tra 0 e 1 (Figura 5).

Il segnale campionato e normalizzato è stato utilizzato per addestrare una rete neurale armonica e per estrarre le prime cento armoniche in una serie di Fourier. Dopo 1000 epoche di addestramento il segnale è stato filtrato per escludere le armoniche con ampiezza inferiore al valore soglia arbitrariamente definito di 0.15 (Figura 6, punti rossi). La procedura di addestramento e filtraggio è stata ripetuta per altre 2000 epoche sul segnale così ripulito (Figura 6, punti blu).

Figura 6 - *Evoluzione della rete armonica durante il filtraggio del segnale*



La procedura di addestramento e filtraggio è stata replicata per ulteriori 3000 epoche (oppure finché non fosse raggiunto un errore quadratico medio di 10^{-6}) sul segnale ripulito per la seconda volta (Figura 7). In questo modo è stato generato un segnale di *output* (Figura 8).

Figura 7 - *Tre fasi di filtraggio del segnale per la parola "mangio": input normalizzato (linea scura), prima fase di addestramento (linea azzurra); seconda fase di addestramento (linea verde)*

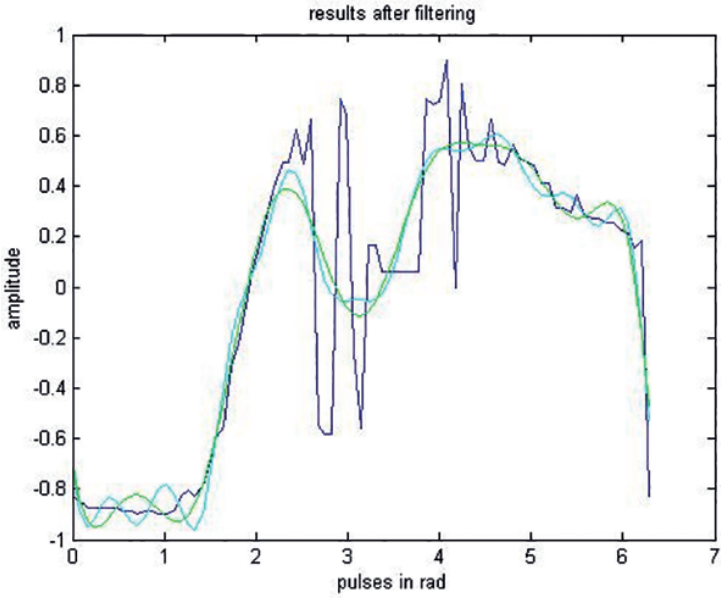
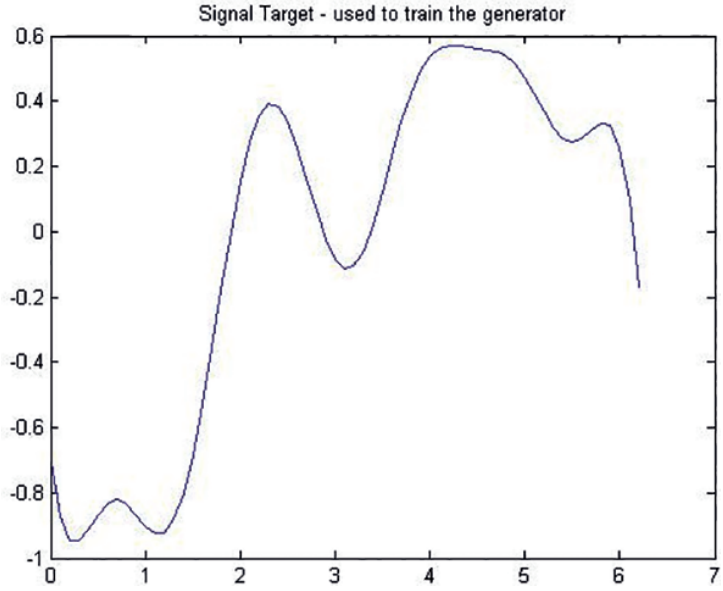


Figura 8 - *Segnale finale usato per l'addestramento per la parola "mangio"*



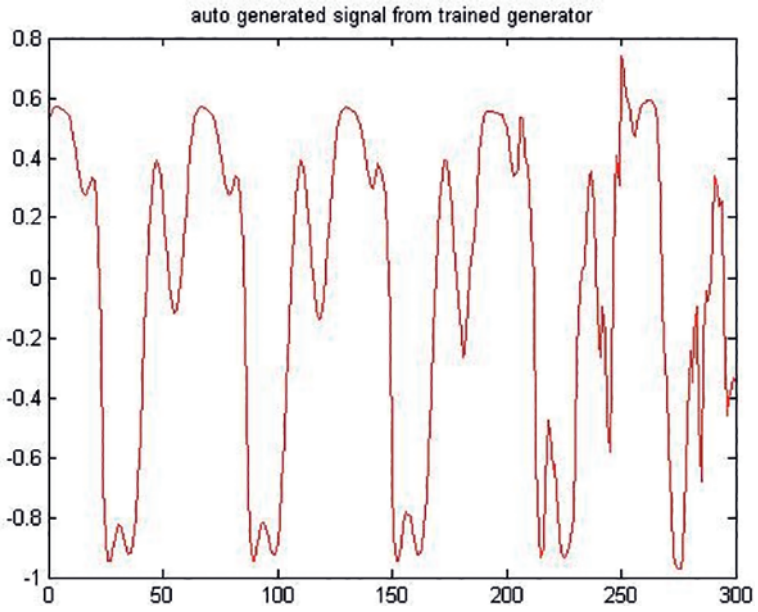
4. *Addestramento e verifica della rete neurale*

Per la programmazione del modello di rete neurale artificiale viene impiegato un algoritmo di apprendimento supervisionato provvisto di una regola di *error-backpropagation* (Rumelhart, Hinton & Williams, 1986; Chichocki, Unbehauen, 1993). Il nostro modello preliminare prevede due strati, uno di ingresso e uno di uscita, in grado di classificare uno stimolo a seconda dell'appartenenza o meno di questo alla classe desiderata, nel nostro caso particolare una data parola. Prima di immetterlo nello strato di ingresso del neurone, il segnale è replicato 9 volte.

Per lo strato in ingresso della rete sono state utilizzate 40 unità di ingresso a partire dal segnale filtrato (v. Figura 8) e replicato, ovvero il neurone utilizza i primi 40 punti sul segnale come informazioni in ingresso e formula una previsione per il 41° punto.

Per esempio, il segnale per la parola /aʃʃe/ è composto dai valori per 63 punti (corrispondenti ai 63 fotogrammi ecografici) e la rete utilizza come unità di ingresso le informazioni contenute nei primi 40 punti. Dopo aver predetto l'elemento 41, il punto generato viene inserito come primo del nuovo ciclo e il neurone si muove lungo la serie alla previsione delle coordinate del punto 42. L'operazione viene ripetuta finché l'intero segnale viene ricreato dal neurone. L'addestramento viene ripetuto per 10000 epoche a meno che si raggiunga una soglia di errore quadratico medio fissata arbitrariamente a 10^{-6}. Il segnale in uscita (attrattore) è costituito dai primi 40 punti del segnale originale e dai nuovi punti generati dal neurone (Figura 9).

Figura 9 - *Segnale completamente ri-generato dalla rete neurale*



Nel modello qui presentato per ogni parola della lista è stato addestrato e creato un neurone. Allo scopo di testare preliminarmente le capacità di riconoscimento del nostro sistema, ogni parola nel set lessicale è stata sottoposta alla procedura di analisi e filtraggio dei dati. Il segnale filtrato fornito nell'*output* è stato confrontato con attrattori delle varie parole. Per determinare a quale categoria-parola corrisponda l'elemento di stimolo in ingresso, sono stati calcolati e confrontati gli errori quadratici medi. In sintesi, "vince" l'attrattore che presenta il valore più basso dell'errore.

Sulla base di queste indicazioni abbiamo compiuto un semplice test esplorativo rappresentato nella Tabella 1 per il raggio 19 (cfr. Figura 4).

Tabella 1 - *Test preliminare della capacità di riconoscimento*

x=19	asce 0.1465	asce 0.60618	asce 0.18871	asce 0.42244	asce 0.28451	asce 0.11725	asce 0.56	asce 0.23899	asce 0.33147	asce 0.13195
	cala 0.097659	cala 0.091811	cala 0.21903	cala 0.18699	cala 0.096241	cala 0.11046	cala 0.012317	cala 0.1481	cala 0.14953	cala 0.22222
	destra 0.11974	destra 0.5388	destra 0.22983	destra 0.38952	destra 0.29603	destra 0.14368	destra 0.60395	destra 0.25412	destra 0.29218	destra 0.12757
	faccia 0.15754	faccia 0.37802	faccia 0.2943	faccia 0.070078	faccia 0.12756	faccia 0.27672	faccia 0.39493	faccia 0.34323	faccia 0.4452	faccia 0.19652
	mangio 0.11306	mangio 0.43325	mangio 0.36748	mangio 0.14948	mangio 0.068518	mangio 0.14367	mangio 0.5868	mangio 0.084485	mangio 0.20962	mangio 0.12863
	marcio 0.1803	marcio 0.20559	marcio 0.35999	marcio 0.41826	marcio 0.21254	marcio 0.27368	marcio 0.12763	marcio 0.25411	marcio 0.37375	marcio 0.39004
	paglia 0.045741	paglia 0.13745	paglia 0.107	paglia 0.088636	paglia 0.090478	paglia 0.086386	paglia 0.022968	paglia 0.092925	paglia 0.11761	paglia 0.097603
	palla 0.16903	palla 0.45437	palla 0.22574	palla 0.25208	palla 0.15767	palla 0.13855	palla 0.53588	palla 0.037018	palla 0.24297	palla 0.20237
	pasta 0.13763	pasta 0.88461	pasta 0.25547	pasta 0.1187	pasta 0.24605	pasta 0.12014	pasta 1.0279	pasta 0.10135	pasta 0.19473	pasta 0.14467
	sinistra 0.36584	sinistra 0.22485	sinistra 0.5912	sinistra 0.28512	sinistra 0.45567	sinistra 0.29982	sinistra 0.12276	sinistra 0.34387	sinistra 0.5292	sinistra 0.48191

Da questa prova risulta una capacità di corretta classificazione delle parole in 4 casi su 10: in verde si possono osservare le corretta assegnazione dei segnali alle parole *cala*, *faccia*, *mangio* e *palla*, mentre in giallo si possono osservare le classificazioni errate delle parole *asce*, *destra*, *marcio*, *paglia*, *pasta* e *sinistra*.

5. *Conclusioni*

Data la relativa semplicità dell'approccio proposto, i risultati del test eseguito devono essere considerati come più che promettenti per la realizzazione di un sistema di riconoscimento di parole basato su dati fisiologici cinematici elicitati tramite la tecnica ecografica.

Dalla rianalisi del processo di formulazione del modello, alcuni elementi si sono rivelati come particolarmente problematici e dovranno perciò essere affrontati e discussi nei prossimi stadi della ricerca.

In primo luogo, si è evidenziato un problema ben noto nella ricerca linguistica che utilizza UTI, ovvero quello della stabilità o robustezza del segnale attraverso diverse sessioni dello stesso parlante e/o di parlanti diversi. Ciò significa, in pratica, che risulta molto complesso da un lato riprodurre le medesime condizioni di acquisizione (anche per lo stesso parlante) del dato e dall'altro rintracciare strategie simili di produzione delle parole tra parlanti simili (p.e. in uno spazio-tempo normalizzato). In altri termini la riduzione delle fonti di variabilità e la parallela individuazione delle componenti stabili e caratterizzanti il segnale (di un dato parlante che produce una particolare parola) si pongono come temi di estremo interesse sia in fase di acquisizione che in fase di analisi e filtraggio. La possibile soluzione di questi problemi porterebbe non soltanto dei vantaggi nella prospettiva applicativa, ma presenterebbe importanti risvolti anche per la ricerca teorica sull'articolazione del linguaggio.

È infatti auspicabile un'integrazione dell'approccio computazionale di analisi del segnale con le conoscenze fattuali del fonetista sperimentale, in particolare riguardo allo specifico strumento di misurazione (Cleland et al., 2011; 2013; Scobbie et al., 2013) e alle cinematiche articolatorie (Hardcastle, Hewlett, 1999; Iskarous et al., 2013; Lammert et al., 2013). Un esempio emblematico riguarda la presenza di improvvise e (troppo) profonde cadute del segnale che indicano in modo evidente un contatto tra articolatori, per esempio tra la lingua e il palato (Epstein, 2005).

Una seconda possibilità di riduzione dell'ambiguità del segnale deriva non solo dalla rimozione delle fonti di rumore, ma anche dall`aumento della complessità complessificazione del modello. In questa direzione già nel presente contributo si è proceduto alla selezione di un punto all'interno dello spazio cartesiano (rappresentazione dello spazio orale) che fosse portatore dell'informazione "migliore"[3] per caratterizzare una parola. Come era prevedibile un singolo punto non è sufficiente poiché genera un segnale troppo poco distintivo, ciò nonostante la prestazione è stata positiva. Per questi motivi sarà necessario aggiungere al modello ulteriori punti con altrettanti neuroni in grado procedere parallelamente per consentire una migliore prestazione complessiva del modello. Anche in questo caso la scelta di quanti e quali punti è una questione che interessa tanto l'ingegnere/informatico quanto il linguista/fonetista (Stone, 1990)[4].

Un ultimo versante di intervento è di tipo metodologico e pertiene più direttamente il linguista. Si tratta infatti della progettazione di una lista di stimoli che non siano semplicemente distinti per opposizioni tra singoli segmenti-fonemi, ma che si oppongano in termini di configurazioni gestuali o se vogliamo di tratti fonologici linguali. Anche in questo caso un esempio può rendere può chiara la natura del problema. Nel pianificare l'insieme di stimoli per questo primo esperimento ci si è concentrati, in modo ingenuo, soprattutto sulle principali opposizioni fonematiche tra consonanti dell'italiano. In questo modo si sono generate alcune comuni coppie minime come p.e. /sana ~ tana/ senza tuttavia considerare che sul piano dei tratti linguali queste parole sono costituite da sequenze articolatorie molto simili che prevedono un'alternanza di movimenti di innalzamento e abbassamento dell'apice della lingua (accoppiati con movimenti di leggero arretramento della radice della lingua per la produzione di [a]). Appare perciò evidente come un'area di ulteriore riduzione di ambiguità tra i segnali vada rintracciata nella fase di definizione degli stimoli. In questo stadio di *experiment design* si dovrà tenere in maggiore considerazione una visione gestuale (Browman, Goldstein, 1992) della fonologia con lo scopo di massimizzare le distinzioni basate su caratteristiche linguali (in particolare riguardo all'opposizione dorsale/coronale).

[3] Ovvero in grado di assegnare correttamente gli stimoli alle categorie corrispondenti in misura maggiore rispetto ad altri.

[4] Basti pensare alle varie proposte avanzate in ambito computazionale (v. il recente Berry, 2012), in quello fonetico (teorie del controllo neuro-motorio, cfr. Perrier et al., 1996) e fonologico (p.e. *articulatory phonology* in termini articolatori linguali).

Bibliografia

Berry, J. (2012). *Machine learning methods for articulatory data. Ph.D. dissertation*, University of Arizona.

Browman, C., Goldstein, L. (1992). Articulatory Phonology: An Overview. In *Phonetic*, 49, 155-180.

Brumberg, J.S, Nieto-Castanon, A., Kennedy, P.R. & Guenther, F.H. (2010). Brain-computer interfaces for speech communication. In *Speech Communication*, 52 (4), 367-379.

Chichocki, A., Unbehauen, R. (1993). *Neural networks for optimization and signal processing*. Wiley.

Cleland, J., McCron, C. & Scobbie, J. (2013). Tongue reading: Comparing the interpretation of visual information from inside the mouth, from electropalatographic and ultrasound displays of speech sounds. In *Clinical Linguistics and Phonetics*, 27 (4), 299-311.

Cleland, J., Wrench, A., Scobbie, J. & Semple, S. (2011). Comparing articulatory images: An MRI / Ultrasound Tongue Image database. In *Proceedings of the 9th International Seminar on Speech Production*, 2161-2164.

Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. & Brumberg, J. (2010). Silent speech interfaces. In *Speech Communication*, 52, 270-287.

Epstein, M., Stone, M. (2005). The tongue stops here: Ultrasound imaging of the palate. In *Journal of the Acoustical Society of America*, 118 (4), 2128-2131.

Hardcastle, W., Hewlett, N. (Eds.) (1999). *Coarticulation: Theory, data and techniques*. Cambridge: Cambridge University Press.

Heracleous, P., Kaino, T., Saruwatari, H. & Shikano, K. (2007). Unvoiced speech recognition using tissue-conductive acoustic sensor. In *EURASIP*, 1, 1-11.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P. & Stone, M. (2007a). Eigentongue feature extraction for an ultrasound-based silent speech interface. In *ICASSP07*, Honolulu, vol. 1, 1245-1248.

Hueber, T., Chollet, G., Denby, B., Stone, M. & Zouari, L. (2007b). Ouisper: corpus based synthesis driven by articulatory data. In *Internatational Congress of Phonetic Sciences*, Saarbrücken, Germany, 2193-2196.

Hueber, T., Chollet, G., Denby, B., Dreyfus, G. & Stone, M. (2007c). Continuous-speech Phone recognition from ultrasound and optical images of the tongue and lips. In *Interspeech*, Antwerp, Belgium, 658-661.

Hueber, T., Chollet, G., Denby, B., Dreyfus, G. & Stone, M. (2008a). Phone recognition from ultrasound and optical video sequences for a silent speech interface. In *Interspeech*, Brisbane, Australia, 2032-2035.

Hueber, T., Chollet, G., Denby, B. & Stone, M. (2008b). Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application. In *International Seminar on Speech Production*, Strasbourg, France, 365-369.

Hueber, T., Benaroya, E.-L., Chollet, G., Denby, B., Dreyfus, G. & Stone, M. (2010). Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. In *Speech Communication*, 52 (4), 288-300.

Hueber, T. (2013). Ultraspeech-player: Intuitive visualization of ultrasound articulatory data for speech therapy and pronunciation training. In *Proceedings of Interspeech*, Lyon, France, 752-753.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C.H., Saltzman, E. & Whalen, D.H. (2013). The Coarticulation/Invariance Scale: Mutual Information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. In *Journal of the Acoustical Society of America*, 134 (2), 1271-1281.

Lammert, A., Goldstein, L., Narayanan, S. & Iskarous, K. (2012). Statistical methods for estimation of direct and differential kinematics of the vocal tract. In *Speech Communication*, 55 (1), 147-161.

Jorgensen, C., Dusan, S. (2010). Speech interfaces based upon surface electromyography. In *Speech Communication*, 52 (4), 354-366.

Perrier, P., Loevenbruck H. & Payan Y. (1996). Control of tongue movements in speech: the Equilibrium Point Hypothesis perspective. In *Journal of Phonetics*, 24, 53-75.

Porbadnigk, A., Wester, M., Calliess, J. & Schultz, T. (2009). EEG-based speech recognition impact of temporal effects. In *Biosignals*, Porto, Portugal, 376-381.

Quatieri, T.F., Messing, D., Brady, K., Campbell, W.B., Campbell, J.P., Brandstein, M., Weinstein, C.J., Tardelli, J.D. & Gatewood, P.D. (2006). Exploiting nonacoustic sensors for speech enhancement. In *IEEE*, 14 (2), 533-544.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning representations by backpropagating errors. In *Nature*, 323, 533-536.

Rumelhart, D.E., McClelland, J. (1987). *Parallel distributed processing*. MIT Press.

Scobbie, J., Punnoose, R. & Khattab, G. (2013). Articulating five liquids: a single speaker ultrasound study of Malayalam. In Spreafico, L., Vietti, A. (Eds.), *Rhotics: New Data and Perspectives.* Bolzano: BU Press, 99-124.

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. In *Journal of the Acoustical Society of America*, 87, 2207-2217.

Stone, M., Rizk, S., Woo, J, Murano, E., Chen, H. & Prince J., (2013). Frequency of Apical and Laminal /s/ in Normal and Post-glossectomy Patients. In *Journal of Medical Speech Language Pathology*, 20 (4), 106-111.

Wang, J., Samal, A., Green, J.A. & Rudzicz, F. (2013). Sentence recognition from articulatory movements for silent speech interfaces. In *ICASSP 2012*, 4985-4988.

Wester, M. (2006). *Unspoken speech - speech recognition based on electroencephalography*. Institut fur Theoretische Informatik Universit at Karlsruhe (TH), Karlsruhe, Germania.

Wrench, A., Scobbie, J. (2011). *Very high frame rate ultrasound tongue imaging*. In *ISSP9*, 155-162.

Autori

Vittorio Anselmi – Libera Università di Bolzano.
vittorio.anselmi@iveco.com

Cinzia Avesani – CNR-ISTC Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, Padova.
cinzia.avesani@pd.istc.cnr.it

Daniela Beltrami – Alma Mater Studiorum, Università di Bologna e Arcispedale S. Maria Nuova, Reggio Emilia.
daniela.beltrami3@unibo.it

Camilla Bernardasci – Universität Zürich.
camilla.bernardasci@uzh.ch

Simona Bernardini – ABC Balbuzie, Torino.

Catherine T. Best – MARCS Institute for Brain, Behaviour and Development, and School of Humanities & Communication Arts, Western Sydney University.
C.Best@westernsydney.edu.au

Manuela Bollani – Università di Pisa.
bollanimanu@yahoo.it

Sabrina Bonichini – Università di Padova.
s.bonichini@unipd.it

Paolo Bravi – Conservatorio di Musica "G.P. Palestrina", Cagliari.
pa.bravi@tiscali.it

Silvia Calamai – Università degli Studi di Siena.
silvia.calamai@unisi.it

Laura Calzà – Alma Mater Studiorum, Università di Bologna.
laura.calza@unibo.it

Francesco Cangemi – IfL Phonetik, University of Cologne.
fcangemi@uni-koeln.de

Valeria Caruso – Università degli Studi di Napoli "L'Orientale".
vcaruso@unior.it

Sonia Cazzorla – Universität Zürich.
sonia.cazzorla@uzh.ch

Maria Assunta Ciardullo – Università della Calabria.
maria_assunta.ciardullo@unical.it

Daniel Colantonio – Università di Padova.

Piero Cosi – CNR-ISTC Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, Padova.
piero.cosi@pd.istc.cnr.it

Franco Cutugno – Università di Napoli "Federico II".
cutugno@unina.it

Rodolfo Delmonte – Università di Venezia "Ca' Foscari".
delmont@unive.it

Anna De Meo – Università degli Studi di Napoli "L'Orientale".
ademeo@unior.it

Bruno Di Biase – MARCS Institute for Brain, Behaviour and Development, and School of Humanities & Communication Arts, Western Sydney University.
B.DiBiase@westernsydney.edu.au

Francesco Di Russo – University of Rome "Foro Italico".
francesco.dirusso@uniroma4.it

Lorenzo Filipponio – Universität Zürich.
filippon@rom.uzh.ch

Gloria Gagliardi – Alma Mater Studiorum, Università di Bologna.

Vincenzo Galatà – Libera Università di Bolzano e ISTC-CNR, Padova.
vincenzo.galata@unibz.it

Enrico Ghidoni – Arcispedale S. Maria Nuova, Reggio Emilia.

Barbara Gili Fivela – Università del Salento, Lecce.
barbara.gili@unisalento.it

Mauro Falcone – Fondazione "Ugo Bordoni", Roma.
falcone@fub.it

Sabina Fontana – Università di Catania, sede di Ragusa.
sabina.fontana@unict.it

Mirko Grimaldi – Università del Salento, Lecce.
mirko.grimaldi@unisalento.it

Massimiliano M. Iraci – Università del Salento, Lecce.
massimiliano.iraci@unisalento.it

Norina Marcello – Arcispedale S. Maria Nuova, Reggio Emilia.

Giovanna Lenoci – Scuola Normale Superiore, Pisa.
giovanna.lenoci@sns.it

Anna Dora Manca – Università del Salento, Lecce.
annadora.manca@unisalento.it

Giulia Natarelli – Università di Padova.

Rosalba Nodari – Scuola Normale Superiore, Pisa.
rosalba.nodari@sns.it

Antonio Origlia – Università di Napoli "Federico II".
antonio.origlia@unina.it

Giulio Paci – CNR-ISTC Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, Padova.
giulio.paci@pd.istc.cnr.it

Valentina Paggini – Università degli Studi di Siena.
valefrin@gmail.com

† Andrea Paoloni – Fondazione "Ugo Bordoni", Roma.

Elisa Pellegrino – Università degli Studi di Napoli "L'Orientale".
epellegrino@unior.it

Francesca Pinto – Università di Roma "Sapienza".
francesca.pinto87@gmail.com

Caterina Pisciotta – Centro Medico di Foniatria, Padova e A.I.BA.COM Onlus.

Marianne Pouplier – Institut für Phonetik und Sprachverarbeitung, LMU, München.
pouplier@phonetik.uni-muenchen.de

Fabio Poroli – Fondazione "Ugo Bordoni", Roma.
fabioporoli@gmail.com

Typhanie Prince – University of Nantes, Lling (UMR6310).
typhanie.prince@univ-nantes.fr

Cinzia Quartarone – Università di Padova.

Daniel Recasens – Departament de Filologia Catalana, Universitat Autònoma de Barcelona, Bellaterra.
daniel.recasens@uab.es

Maria Roccaforte – Università di Roma "Sapienza".
maria.roccaforte@uniroma1.it

Luciano Romito – Università della Calabria.
luciano.romito@unical.it

Rema Rossini Favretti – Alma Mater Studiorum, Università di Bologna.

Stephan Schmid – Universität Zürich.
stephan.schmid@uzh.ch

Giacomo Sommavilla – CNR-ISTC Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, Padova.
giacomo.sommavilla@pd.istc.cnr.it

Lorenzo Spreafico – Libera Università di Bolzano.
lorenzo.spreafico@unibz.it

Debora Stocco – Centro Medico di Foniatria, Padova.

Fabio Tamburini – Alma Mater Studiorum, Università di Bologna.
fabio.tamburini@unibo.it

Yilun Tang – Shanghai International Studies University.

Andrea Tarasi – Università della Calabria.
tarasiandrea17@gmail.com

Fabio Tesser – CNR-ISTC Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, Padova.
fabio.tesser@pd.istc.cnr.it

Graziano Tisato – CNR-ISTC Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, Padova.
tisato@tin.it

Massimiliano Todisco – Fondazione "Ugo Bordoni", Roma.
mtodisco@fub.it

Ottavia Tordini – Università di Pisa.
o.tordini@studenti.unipi.it

Mario Vayra – Alma Mater Studiorum, Università di Bologna e ISTC-CNR, Padova.

Alessandro Vietti – Libera Università di Bolzano.
alessandro.vietti@unibz.it

Marilisa Vitale – Università degli Studi di Napoli "L'Orientale".
marilisavitale@gmail.com

Sarah Wachter – Universität Zürich.
sarah.wachter@gmx.ch

Ying Xu – Tianjin Foreign Studies University.

Claudio Zmarich – CNR-ISTC Consiglio Nazionale delle Ricerche - Istituto di Scienze e Tecnologie della Cognizione, Padova.
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.

Mario Vayra è Professore Associato presso la Scuola di Lettere e Beni Culturali dell'Università di Bologna dove insegna Glottologia e Comunicazione Parlata. A lungo Visiting Researcher presso gli Haskins Laboratories (USA), è vice-presidente dell'AISV. I suoi interessi di ricerca vertono intorno al tema delle basi fonetiche della struttura fonologica, segmentale e prosodica. Lavora inoltre sull'interfaccia intonazione/struttura informativa, sullo sviluppo fonetico e sulla percezione della fonologia non-nativa.

Cinzia Avesani è Primo Ricercatore presso l'Istituto di Scienze e Tecnologie della Cognizione del CNR. Dopo il Perfezionamento in Linguistica presso la Scuola Normale Superiore di Pisa ha lavorato come Linguistic Consultant al Linguistic Department dei Bell Laboratories (USA). Vice-presidente dell'AISV dal 2010 al 2016, i suoi interessi di ricerca si incentrano sugli aspetti prosodici del linguaggio e si estendono alla sociofonetica e alla fonologia di L2.

Fabio Tamburini è Professore Associato presso la Scuola di Lettere e Beni Culturali dell'Università di Bologna dove tiene numerosi corsi di Linguistica Computazionale, Linguistica dei Corpora e Linguistica Applicata. I suoi principali interessi di ricerca vanno dall'elaborazione della lingua parlata al trattamento automatico delle lingue.

AISV - Associazione Italiana Scienze della Voce
sito: www.aisv.it
email: aisv@aisv.it| redazione@aisv.it
ISBN: 978-88-97657-11-8

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