Studi AISV 3

FATTORI SOCIALI E BIOLOGICI NELLA VARIAZIONE FONETICA

SOCIAL AND BIOLOGICAL FACTORS IN SPEECH VARIATION

a cura di

Chiara Bertini, Chiara Celata, Giovanna Lenoci,
Chiara Meluzzi, Irene Ricci

AISV
Associazione Italiana Scienze della Voce

FATTORI SOCIALI E BIOLOGICI NELLA VARIAZIONE FONETICA

SOCIAL AND BIOLOGICAL FACTORS IN SPEECH VARIATION

a cura di

Chiara Bertini, Chiara Celata, Giovanna Lenoci, Chiara Meluzzi, Irene Ricci

Milano 2017

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.

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: 978-88-97657-19-4

Indice

PARTE V
Lavori in corso

DOI: 10.17469/O2103AISV000000

CHIARA BERTINI, CHIARA CELATA, GIOVANNA LENOCI,
CHIARA MELUZZI, IRENE RICCI

Prefazione

Questo volume nasce da una selezione dei contributi presentati al XIII Convegno Nazionale AISV, tenutosi dal 25 al 27 gennaio 2017 presso la Scuola Normale Superiore di Pisa. Il convegno aveva come tema principale quello della relazione tra i fattori fisici-biologici e i fattori socio-culturali nella determinazione della variazione sincronica e diacronica della forma sonora delle lingue, ma comprendeva anche – come è nella tradizione dei convegni AISV – sessioni incentrate su temi diversi, in funzione delle proposte di comunicazione provenienti dai Soci.

Con la partecipazione di 75 studiosi provenienti da 9 Paesi differenti, due comunicazioni plenarie tenute da John Esling e Susanne Fuchs, una tavola rotonda su balbuzie e teoria fonetica, comprendente gli interventi su invito di Robin Lickley e Pier Paolo Busan, e infine un *panel* sulla Motor Theory curato da Franco Cutugno e Leonardo Badino, il convegno AISV 2017 ha costituito l'avvio di un'intensa fase di approfondimento e confronto tra studiosi di orientamento teorico e/o metodologico anche diverso. Di tutto ciò questo volume vuole essere diretta, anche se parziale, testimonianza.

Il volume si apre con gli articoli dei due relatori delle sessioni plenarie, John Esling (Victoria University) e Susanne Fuchs (ZAS Berlin), raccolti nella sezione "Verso una teoria della variazione fonetica: fattori sociali e biologici a confronto".

Prendendo le mosse da decenni di indagine sperimentale sul ruolo della laringe e delle strutture sovralaringee nella determinazione dei contrasti fonetici e fonologici delle lingue, John Esling illustra nel proprio contributo che, per conoscere e modellizzare le attività fisiologiche coinvolte nella produzione di suoni linguistici, è necessario approfondire il problema delle molteplici interdipendenze tra le attività del cavo orale (e primariamente del suo articolatore principale, la lingua, assistita dalla mandibola) e quelle che si realizzano al livello del piano glottale. Importanti sono non solamente le condizioni di apertura o chiusura della glottide e le possibilità oscillatorie delle corde vocali, ma anche e soprattutto le possibilità di inclinazione verticale delle strutture che le sorreggono, nonché le restrizioni e le vibrazioni degli spazi epilaringei, che contribuiscono in modo rilevante a definire variazioni sia nella qualità vocalica, sia nella realizzazione di foni consonantici con diversi gradi di costrizione, sia nelle modifiche della qualità della voce, che in alcune lingue è fattore fondamentale per l'implementazione di distinzioni fonologiche, mentre in altre lingue è veicolo di informazioni socio-comunicative. A permettere la comunicazione tra il livello 'alto' del cavo orale e il livello 'profondo' dello spazio laringeo è quel complesso di strutture anatomico-funzionali comprese tra il piano glottale e la base della lingua (o, posteriormente, il faringe superiore) che Esling dimostra svolgere la funzione di un vero e proprio articolatore, al pari degli articolatori tradizionalmente descritti in riferimento al cavo orale e alle sue strutture periferiche più vicine. Una simile prospettiva, impossibile all'interno di quella visione 'linguocentrica' più spesso perseguita nelle scienze fonetiche tradizionali e basata sull'assunzione (ormai perlopiù smentita dai fatti) che al piano dell'articolatore orale si contrapponga, in modo del tutto indipendente, il piano della 'sorgente' glottale, apre nuove sfide alla ricerca della comprensione dei meccanismi della produzione fonetica e all'elaborazione di modelli con gradi di libertà molto maggiori di quelli tradizionalmente proposti.

Il contributo di Susanne Fuchs è invece finalizzato a tracciare un sintetico quadro storico di come le scienze del linguaggio hanno trattato il tema della variazione fonetica e fonologica nel tentativo di tematizzare la differenza tra cause fisico-biologiche e radici socio-culturali di tale variazione. Ripercorrendo alcune delle tappe principali di tale dibattito, e sintetizzando i risultati più significativi delle proprie recenti ricerche in tale senso, l'autrice mostra come i livelli di volta in volta presi in esame nello studio della variazione siano cresciuti sia nel numero che nel grado di complessità. Nell'ambito dei fattori sociali, all'analisi dei gruppi interni alla comunità dei parlanti si aggiunge oggi lo studio dell'interazione e delle dinamiche di adattamento o convergenza tra dialoganti; nell'ambito dei fattori biologici, allo studio della morfologia del cavo orale del locutore si affianca la presa in carico di caratteristiche biomeccaniche, cognitive e persino fenotipiche dei parlanti. Numero e complessità dei fattori, dunque, aumentano: il problema fondamentale che ne deriva, secondo Fuchs, è quello della consapevolezza che la relazione tra i fattori è non lineare, e non si presta ad indagini monodimensionali né a facili predizioni sui rapporti di causa-effetto. La risposta che le scienze del linguaggio contemporanee devono fornire è perciò un approccio integrato e multidisciplinare, che permetta a studiosi con specializzazioni differenti di incontrarsi e confrontarsi su problemi comuni, elaborare strategie complessive, giungere a generalizzazioni che vadano al di là del singolo studio. In questa prospettiva è possibile anche comprendere fino in fondo il significato della variazione e porre le domande di ricerca in ottica nuova: non più chiederci, dunque, come sia possibile che gli essere umani si parlino e si capiscano *nonostante* l'enorme tasso di variabilità insito nel segnale, ma piuttosto come accada che il linguaggio umano funzioni *proprio in virtù* e per merito anche della variazione fonetica.

La seconda parte del volume è intitolata "Sociale e biologico nella variazione indotta dalle caratteristiche del parlante, della lingua e del contesto comunicativo" e raccoglie nove contributi che affrontano il tema della variazione in riferimento a fattori segmentali, sovrasegmentali e della qualità della voce. Includendo l'analisi di corpora sociofonetici, multimodali o focalizzati sul rapporto tra produzione e percezione, questi contributi illustrano come anche le innovazioni in termini metodologici giochino un ruolo importante per la futura elaborazione di modelli avanzati della produzione fonetica.

La terza parte del volume è invece dedicata a "Sociale e biologico nella deviazione dalle 'norme': bilinguismo e patologia". Apprendere o perdere progressivamente competenze linguistiche, per cause sociali (ivi compresa la scelta dell'individuo di imparare una lingua straniera) o patologiche, sono fenomeni direttamente influenzati sia dalle caratteristiche biologiche del parlante e del suo apparato fonatorio e percettivo, sia dalle caratteristiche dell'ambiente sociale che lo circonda e della posizione che il parlante vi occupa. Anche le ricadute didattiche o cliniche che i sei contributi qui raccolti direttamente o indirettamente affrontano non possono prescindere dalla comprensione approfondita di quei rapporti non lineari tra le diverse componenti della variazione che fanno l'oggetto degli studi sperimentali.

La quarta parte del volume, "Balbuzie e teoria fonetica", contiene quattro interventi sul tema della balbuzie e del contributo che gli studi in questo ambito possono fornire per la teoria fonetica generale. La balbuzie può essere considerata un prototipo di quel genere di fenomeni complessi del linguaggio che derivano dall'intreccio di fattori socioculturali, psicologici, fisiologici e genetici, e rispetto ai quali è necessario un approccio di studio (e di intervento) integrato. La scienza del domani dovrà fornire risposte in questo senso tanto allo specialista, nell'ottica di un avanzamento della teoria fonetica generale, quanto, in termini di ricadute dirette, alla società in senso lato.

Chiude infine il volume una sezione ("Lavori in corso") contenente brevi articoli di ambito applicativo, che illustrano progetti recenti e innovativi nel quadro nazionale delle tecnologie della voce e del parlato.

La realizzazione di questo volume non sarebbe stata possibile senza la collaborazione attenta e costruttiva di tutti gli Autori coinvolti, ai quali va il nostro più sentito ringraziamento. Un grazie speciale va ai membri del Comitato scientifico del volume, che hanno prestato il loro tempo e la loro professionalità al processo di revisione dei manoscritti, sia in fase di proposta iniziale che rispetto agli interventi di modifica successivamente apportati dagli Autori. È con la partecipazione di tutti costoro che siamo perciò liete di consegnare alle stampe questo volume a un anno dalla celebrazione del Convegno da cui prende le mosse, con l'auspicio di massima circolazione e fortuna sia presso la comunità degli studiosi italiani che all'estero.

PARTE I

VERSO UNA TEORIA DELLA VARIAZIONE FONETICA: FATTORI SOCIALI E BIOLOGICI A CONFRONTO

DOI: 10.17469/O2103AISV000001

JOHN ESLING

The laryngeal articulator's influence on voice quality and vowel quality

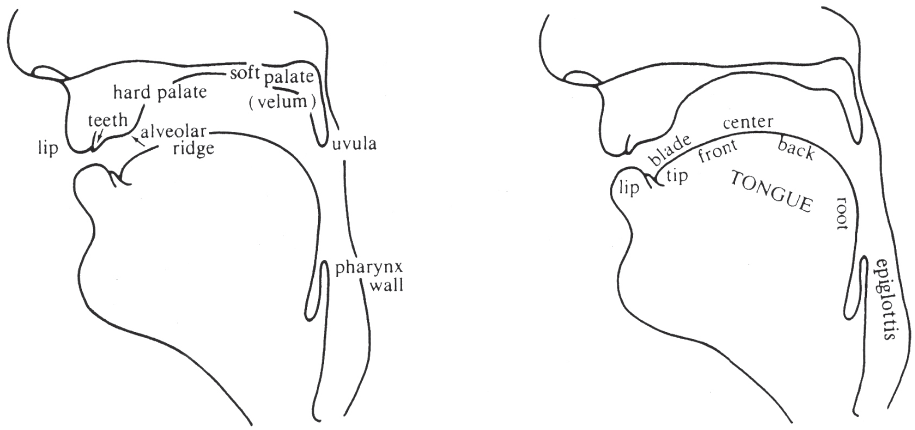
In order to account for sound qualities that emerge from the lower vocal tract, the traditional articulatory model of the vocal tract has to be redrawn. The lower vocal tract incorporates an epilaryngeal tube, aryepiglottic articulators, larynx-height adjustments, and sublingual pharyngeal spaces influenced by epilaryngeal parameters and by lingual-retraction parameters. Laryngoscopic observations reveal how several pharyngeal strictures and laryngeal voice qualities function in a variety of linguistic systems. Parallel cineradiographic, ultrasound, and MRI techniques reveal that laryngeal articulation is critical in shaping tonal register effects and vowel quality. The lower vocal tract is shown to be a significant contributor in shaping the auditory/acoustic output of vowels that are usually defined by uniquely oral parameters. More extreme laryngeal modifications evoke changes to vowel quality that differ by vowel category and by laryngeal effect.

Key words: larynx, epilaryngeal, aryepiglottic, laryngoscopic, vowel quality.

1. *The vocal tract, traditionally*

The objective of this presentation is to redraw the articulatory possibilities of the vocal tract in order to accommodate sound qualities that are produced primarily in the lower vocal tract. The traditional vocal tract model (e.g. Ladefoged, 2001) does not depict a lower vocal tract articulator. Articulations in the lower vocal tract are generally ignored except for the role that the vocal folds play in producing phonation of various types, based largely on the size and shape of the glottis. The traditional model could be called a linguocentric source-filter model. In a source-filter model, glottal aperture is taken to be the deciding factor in the modulation of phonation, and the rest of the vocal tract, or 'supraglottic' vocal tract, is the 'tube' or resonating space where the source is filtered, creating spectral elaborations. In a purely linguocentric model – an articulatory parallel to the acoustic source-filter paradigm, the tongue is assumed to be the primary articulator that produces 'sound shapes' anywhere between the glottis and the lips (with contributions of course from the jaw and from velo-pharyngeal aperture). Figure 1 illustrates the traditional depiction of the laryngeal space, where only the pharynx wall appears as a passive articulator, and only the epiglottis is present as a potential active articulator, presumably associated with the tongue in this regard. There is no depiction of the larynx or glottis at all in this view, but that is presumably because the glottis is assumed to have only one role, dealt with later on in the textbook in a section on 'States of the Glottis,' which also includes descriptions of timing as a laryngeal accompaniment to oral stops (Ladefoged, 2001: 122-132). Phonation types have been elaborately described in articulatory and auditory phonetics, including their acoustic correlates (Catford, 1964, 1977). The modelling of how these various phonation types relate to the structures of the vocal tract and how they are differentiated from each other remains, however, focused on glottal shape in a relatively 2-dimensional conceptualization (Gordon, Ladefoged, 2001). The notion of articulation happening in the larynx is generally discounted. Articulation is commonly thought to be a function of lingual, and mandibular and labial, movements. The larynx is not thought to move in the same ways as the tongue, jaw, or lips, and it has therefore been considered (mistakenly) to be only a housing for the muscles that control glottal aperture.

Figure 1 - *Traditional vocal tract model (Ladefoged, 2001: 3-4). From Ladefoged. Course in Phonetics, 4E. © 2001 South-Western, a part of Cengage, Inc. Reproduced by permission*

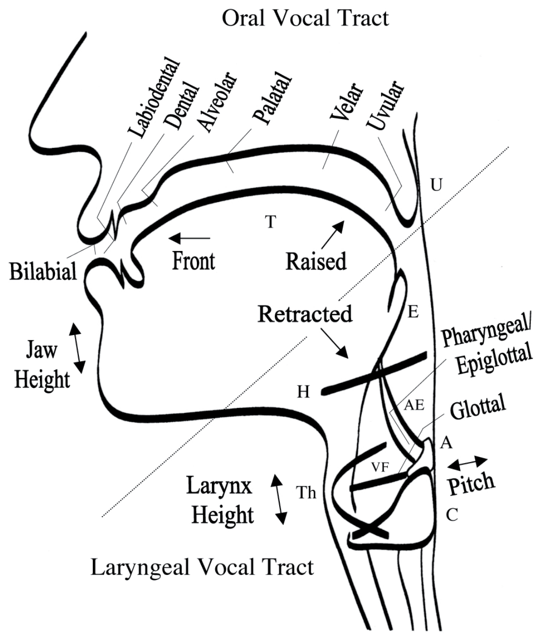


2. *Lower vocal tract articulation, redrawn*

Considerable experimental phonetic research has demonstrated that the lower vocal tract, or laryngeal vocal tract, is more highly elaborated than in lingually-dominated models (Esling, 1996, 2005). The Laryngeal Articulator Model, shown graphically in Figure 2, divides the vocal tract into two parts, a laryngeal vocal tract, beginning at the glottis and continuing upwards to the top of the pharynx below the uvula, and an oral vocal tract, beginning at the uvula and continuing through the lips. This view could be called the 'two-part vocal tract' (Esling, 2010). Actions of the tongue are separated in this model into three primary directions: retracting, raising, and fronting. Retraction of the tongue is allocated to the laryngeal articulator, below the dashed line, because the tongue does not retract, physiologically, without an initiating action occurring in the larynx. This is the principal difference between the Laryngeal Articulator Model and linguocentric models; the tongue is not the primary articulator but rather an accompanying action to a dominantly laryngeal manoeuvre. The actions of the laryngeal articulator reach far beyond just glottal voicing and horizontal-plane alteration in glottal shape or state. The mechanism depicted in Figure 2 is fully elaborated to perform articulations that bear a direct relationship to changes in auditory quality and acoustic resonance. The mechanism also generates various vibratory possibilities. It is therefore a multiple source, giving a more deeply-layered aspect to the notion of phonation types or of glottal state. As we shall see below, there is a substantial vertical aspect to the performance of the laryngeal mechanism, before the airstream reaches the oral section of the vocal tract.

The relationship between the tongue and the laryngeal mechanism (which can also be called the aryepiglottic sphinctering mechanism or the laryngeal constrictor mechanism) has implications for the description of vowel quality. Those vowels in the lower-right corner of the vowel quadrilateral have the potential to be strongly influenced by the action of the laryngeal constrictor, due to the accompanying retraction of the tongue that the constrictor engenders. Vowels in the upper-right corner of the vowel quadrilateral are lingually raised, that is, they are pulled back and upwards, and as such are not as likely to show the effects of laryngealization or pharyngealization (both of which are terms that can be used to describe the general effects of the laryngeal constrictor mechanism, although their phonological usage may refer to more specific phonetic events). Vowels at the left side of the vowel quadrilateral are lingually fronted, and more open vowels exhibit progressively greater degrees of jaw opening, and as such are also not as likely to show the effects of laryngealization or pharyngealization. Thus, vowels such as [ɑ] or [ɔ] can be thought of as having the potential to be more laryngealized or pharyngealized than raised or fronted vowels.

Figure 2 - *The Laryngeal Articulator Model of the vocal tract (Esling, 2005)*



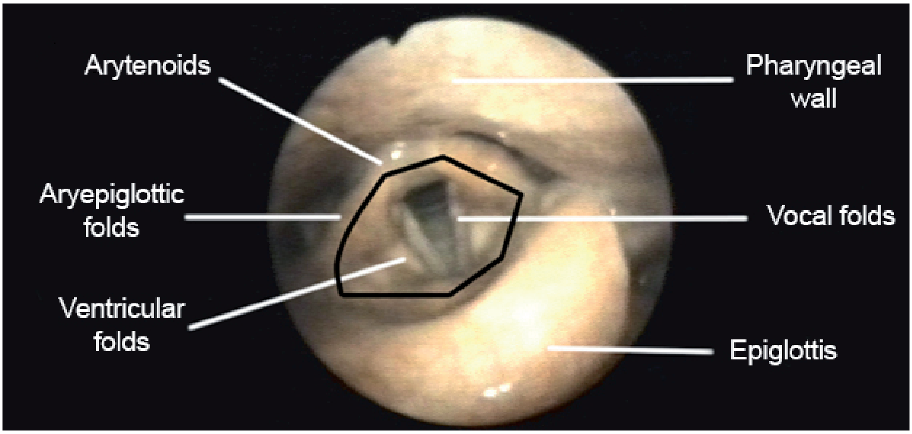
3. *Instrumental approaches to laryngeal articulation*

Our instrumental research over the years has shown that the glottis is not the only source of periodic energy in the larynx and that the ventricular folds and aryepiglottic folds also generate vibrations attested in speech sounds (Moisik, Esling & Crevier-Buchman, 2010). We have also demonstrated that the laryngeal constrictor mechanism controls cavity resonance in pharyngeal articulations, cavity resonance and/or phonation type in 'tonal register' languages, and cavity resonance in 'vowel harmony' languages such as those that have been labelled 'ATR' languages (Edmondson, Esling, 2006). Phonation types have also been realigned with a new perspective on states of the glottis (Esling, Harris, 2005), or more properly 'states of the larynx.' The tools of investigation that we have used in our phonetic research have allowed us to formulate a theory in which the articulatory production of pharyngeal sounds can be explained as being contained within the 'laryngeal articulator' (Esling, 2005). We have examined laryngeal articulation in over two dozen languages from diverse language families across the world to demonstrate the distinctive 'manners of articulation' as well as resonances that the laryngeal articulator can produce.

3.1 Laryngoscopic investigation

Our initial mode of instrumental investigation relied on fibreoptic laryngoscopy, in which endoscopically transmitted images of the lower vocal tract are captured using either a nasally inserted or an orally inserted scope (see Sawashima, Hirose, 1968). Figure 3 demonstrates the view of the articulatory structures in the lower vocal tract, where the distal end of the laryngoscope is just behind and below the uvula. The epiglottis is at the front (bottom), and aryepiglottic folds join the arytenoid cartilages at the back (top, where they hide the oesophageal opening against the posterior pharyngeal wall) to the side borders of the epiglottis, upwards towards the front. The ring designates the aryepiglottic laryngeal 'sphincter,' by which the corniculate tubercles of the arytenoids connect to the cuneiform tubercles of the aryepiglottic folds, the sides of the epiglottis, and the epiglottic tubercle at the base of the epiglottis. This rim forms the closure mechanism that produces the narrowing required for pharyngeal sounds, for what have been called epiglottal sounds, for register adjustments in the larynx accompanying tone, and for resonance modulation in many harmonic systems. The space between the glottis upwards to the aryepiglottic rim is the epilaryngeal tube – a physiologically (and phonetically) critical space in the lower pharynx.

Figure 3 - *A laryngoscopic view of the lower vocal tract, showing anterior structures (bottom), posterior structures (top), and three sets of folds through the length of the epilaryngeal tube*



Full closure of the supraglottic epilaryngeal tube occurs in gagging, coughing, swallowing, and in the phonetic production of epiglottal stop [ʡ]. Far from being a phonetic rarity, epiglottal stop (or epiglottal plosive, sometimes called pharyngeal stop by fieldworkers) is widely attested. At least, it has been widely observed since our description of it established what it is and how it is produced (Esling, 2003; Esling, Fraser & Harris, 2005). Catford did introduce the term 'epiglotto-pharyngeal' to distinguish more extremely retracted sounds from those labelled pharyngeals (Catford, 1968: 326), but although he identified sounds that he would have labelled epiglottopharyngeal fricative, approximant and trill, he was equivocal about the nature of an epiglottopharyngeal stop articulation. Instead of the tongue articulating against the posterior wall of the pharynx, which was the widely-held view at that time, we observe the aryepiglottic folds to advance forwards and upwards from their posterior, open position towards the base of the epiglottis, in a gesture that finishes by completely arresting the passage of air from the lungs. Full aryepiglotto-epiglottal closure results in a pharyngeal/epiglottal stop. Epiglottal stop /ʡ/ is a phoneme in Tigrinya (Semitic), in the Interior Salish languages of British Columbia, and in Nuuchahnulth (Wakashan). It occurs in Iraqi Arabic as a realization of the geminate pharyngeal /ʕʕ/, and it is attested in Berber. It is also known to occur in a number of Caucasian languages.

Tigrinya, Nuuchahnulth, and Iraqi Arabic also have a voiceless pharyngeal fricative /ħ/ in their phonologies, and this sound is homorganic with epiglottal stop. Some Caucasian languages may differentiate a pharyngeal from an epiglottal fricative, but this likely due to a difference in larynx height, where the pharyngeal has a lower larynx position, and the epiglottal has a higher larynx position and therefore smaller resonating spaces and higher-pitched spectral attributes. In languages that have an epiglottal stop phoneme, however, the stop will generally have the same place of articulation (the aryepiglottic folds against the tubercle of the epiglottis) and the same relative larynx height as the fricative. The same applies to languages that also have a voiced pharyngeal approximant /ʕ/ and/or a pharyngeal/epiglottal trill. In this context, it should be noted that the pharyngeal/epiglottal place of articulation is one identical column in the elaborated IPA chart contained in Esling (2010) and in Coey, Esling & Moisik (2014).

The voiced pharyngeal approximant /ʕ/ appears in many languages, with the same articulatory attributes as the stop and fricative, distinguished only by the degree of approximation of the aryepiglottic folds and by glottal vocal fold voicing. As mentioned for Iraqi Arabic, stopping the airflow of [ʕ] by the slightest closure of the aryepiglottic mechanism results in an epiglottal stop [ʡ]. Although the pharyngeal fricative and approximant appear together in the IPA chart, they have demonstrably different manners of articulation, as argued by Laufer (1996). A number of degrees of approximation and of frication are represented for the Pharyngeal/Epiglottal column in the elaborated IPA chart contained in Esling (2010). This same chart is reproduced on the main page of the app, *iPA Phonetics* (Coey, Esling & Moisik, 2014). Iraqi Arabic is a variety in which the pharyngeal approximant /ʕ/ or the pharyngeal fricative /ħ/ can become trilled. This involves the aryepiglottic folds at the upper margin of the epilaryngeal tube, which vibrate, each one against the area around the epiglottic tubercle. When the space of the laryngeal constrictor between the advancing aryepiglottic folds and the epiglottal surface becomes very narrow, as for /ʕ/ or /ħ/, the potential for the aryepiglottic folds to vibrate increases. In Iraqi Arabic, this occurs when /ʕ/ is spoken in a more forceful prosodic context or when /ħ/ is lengthened as a geminate intervocalically. We have seen that when /ʕ/ is lengthened as an intervocalic geminate, it is generally realized as a full stop [ʡ]. Geminate /ħ/ nearly always trills. It could be considered more likely that the voiceless fricative would engender trilling, because of the propensity of the voiceless airstream to induce vibration of adjacent structures. We have used the symbols for an epiglottal fricative to represent these sounds, [ʢ] and [ʜ], since the epiglottal label has been used to imply stronger articulation, and because trilling is a more constricted manner of articulation than either [ʕ] or [ħ]. The other way in which [ʕ] or [ħ] could be enhanced and logically represented by the symbols [ʢ] and [ʜ] would be through a more extreme raising of the larynx. Thus, aryepiglottic trilling and extreme compaction of the epilaryngeal space are complementary gestures that warrant the use of the 'epiglottal' symbols [ʢ] and [ʜ]. This usage is arguably consistent with Catford's (1968) intentions.

In Moisik, Esling & Crevier-Buchman (2010) the vibrations of the aryepiglottic folds are analyzed in detail. The two folds are shown to vibrate in two different phases, which is not unusual for independent structures that do not vibrate against each other but which vibrate against separate parts of the base of the epiglottis. In our research, the left fold vibrates at the same frequency as the vocal folds, but the right fold vibrates at half that frequency, ostensibly due to different angles and pressures with which the cuneiform tubercles of the folds are pressed up against the epiglottis. It may be that many people share this pattern, or at least exhibit a pattern where the two folds are in different phases. Like epiglottal stop, aryepiglottic fold trilling is not unusual physiologically. It occurs during throat clearing and often during loud yelling.

3.2 Cineradiographic evidence

Another method of observing the articulatory functions of the lower vocal tract is by means of cineradiography. Cineradiographic videos of epiglottal stop, voiceless aryepiglottic trill, and voiced aryepiglottic trill were performed at the at Instituto de Ciências Biomédicas, Departamento de Anatomia, Universidade Federal do Rio de Janeiro, with the collaboration of Milton Melciades Barbosa Costa and Leonardo Fuks. Each short video was then analyzed using automated measurements developed in Matlab by Scott Moisik. The epilaryngeal tube and the pharyngeal tube above it were each defined as a region of interest (ROI) in Matlab – a two-dimensional area. In the black-and-white sagittal images of each video clip, approximate areas of the epilaryngeal tube area and the pharyngeal tube can be calculated as the different luminosity thresholds of pixels in each ROI change over time. Each area is measured and tracked over time for each ROI, with most-luminescent pixels indicating a larger area (more open in terms of volume), and least-luminescent pixels indicating a smaller area (more closed in terms of volume). An aryepiglottic-fold-to-vocal-fold distance is also measured, and an epiglottis-vocal-fold angle is calculated by drawing one line from an anterior point on the neck along the vocal folds and another line up along the border of the epiglottis. Measurements were also taken of the height of the cricoid cartilage relative to an arbitrary reference point beneath.

Results for [ɑʡɑ] show that the epilaryngeal tube area decreases to zero during the stop, while the pharyngeal tube area above remains about 75% open. Aryepiglottic-fold-vocal-fold separation also reduces to zero, and the angle between the vocal folds and the epiglottal border above reduces to zero degrees. This is powerful evidence that epiglottal stop is an articulation of full closure and indicates precisely where and how that closure occurs. Results for [ɑʜɑ] show a similar but less extreme pattern. Epilaryngeal tube area decreases to around 40% during stricture, while pharyngeal area above remains about 75% open. This implies that epilaryngeal tube volume is the deciding factor in the articulation of pharyngeals, while (upper) pharyngeal volume is less significant. Aryepiglottic-fold-vocal-fold separation is around 30% of maximum, and the angle between the vocal folds and the epiglottal border above reduces to about 20 degrees. Results for [ɑʢɑ] show tighter constriction than for voiceless [ʜ]. Epilaryngeal tube area decreases to around 10% during stricture, while pharyngeal area above remains about 60% open. Here again, epilaryngeal tube volume is the deciding factor in the articulation of the pharyngeal. Aryepiglottic-fold-vocal-fold separation decreases to around 10% of maximum, and the angle between the vocal folds and the epiglottal border is around 15 degrees. In all three of the videos, the point of maximum stricture is clearly the arytenoid and aryepiglottic structures approximating or closing against the tubercle at the base of the epiglottis. This action pulls the tongue and epiglottis backwards, towards but not touching the posterior pharyngeal wall. Since the pharyngeal space (between the retracted tip of the epiglottis and the arytenoid-aryepiglottic top of the epilaryngeal tube) only reduces partially in area (and presumably volume), we can conclude that these pharyngeal/epiglottal articulations are not primarily a lingual-retraction articulation but rather a primary aryepiglottic articulation (inducing changes to epilaryngeal tube area, viz. volume). These changes are laid out in detail in Moisik (2013).

3.3 Ultrasound of the larynx

Another technique of examining laryngeal function that provides a view of the laryngeal articulator from above as well as from the side in the vertical plane is to combine laryngoscopy with laryngeal ultrasound. The technique, developed by Scott Moisik and known as 'simultaneous laryngoscopy and laryngeal ultrasound' (SLLUS), is illustrated in a study of tonal pitch and register in Mandarin by Moisik, Lin & Esling (2014). Photographs in that article illustrate how the two instruments are used together in the lab. The two methods complement each other; laryngoscopy working well to evaluate larynx state, and ultrasound working well to evaluate larynx height. The mapping of how laryngeal structures correspond to the view obtained from the ultrasound probe placed against the side of the thyroid cartilage is also illustrated in the *iPA Phonetics* app (Coey, Esling & Moisik, 2014). This placement of the probe allows three levels of valves (vocal, ventricular, aryepiglottic) and the epiglottis to be visualized in a side-on representation. Viewing the larynx with ultrasound in the vertical dimension is a particularly effective way of interpreting the variable stricture of the valves of the larynx (Edmondson, Esling, 2006).

Simultaneous laryngoscopy and laryngeal ultrasound of epiglottal stop [iʡi] is evaluated using optical flow analysis of movements in the ultrasound, synchronized with the video of the laryngoscopic image. As the laryngeal articulatory structures rise, flow vectors can be summed and converted to millimetres to track global vertical displacement during laryngeal closure. Elevation of the larynx can be seen in the laryngoscopic video, as the structures rise and compact, approaching the distal end of the scope and reflecting more light. The synchronized ultrasound track allows these movements to be quantified. The onset of closure is accompanied by early approximation of the vocal folds and the ventricular folds, followed by systematic elevation of these coupled sets of folds and of the aryepiglottic folds. As this occurs, the tongue begins to retract. It is interesting to note that closure of the aryepiglottic sphincter appears to be completed prior to full elevation of the larynx. Once maximum elevation is achieved, the vectors tracking the top of the epilaryngeal tube indicate that the aryepiglottic folds are releasing and beginning to drop and that the tongue is beginning to advance, while the glottal-level structures are still maintaining raising. Immediately after this, the glottal level starts to drop, the aryepiglottic folds begin to descend, and the tongue moves forward, leaving the pharyngeal cavity open. In the production of [iʜi], there is an immediate transition from vocalic voicing to laryngeally constricted 'funnelling' of the epilaryngeal tube (into the state of 'whisper') to generate whispered friction as the aryepiglottic folds also begin to vibrate for the voiceless pharyngeal trill. The constrictor (sphincter) continues to narrow as the larynx rises. Voicing for the second vowel begins as the larynx reaches its most elevated point, with the larynx steadily lowering as voicing continues. For [iʕi], because it is fully voiced, it is more difficult to see in the laryngoscopic video where transitions occur. Ultrasound, aligned with the speech waveform, helps us to determine that the larynx continues to rise from the first vowel into the voiced pharyngeal trill. The onset of the second vowel coincides with maximum larynx height, after which the larynx descends to baseline as the mechanism opens at the end of voicing. An interesting comment to add is that many F0-extraction algorithms fail to detect the periodicity of voicing beginning exactly at the point where aryepiglottic trilling starts, since aryepiglottic vibrations can conflict with vocal fold vibrations. However, Scott Moisik's custom F0-extraction algorithm is able to deal with these multiple sources and does a good job of tracing F0 throughout the production of an acoustically challenging sequence such as [iʕi].

These results tell us that pharyngeals/epiglottals are primarily a function of epilaryngeal tube constriction (based on area, but extrapolating to volume). Laryngeal ultrasound is shown to be a suitable technique to evaluate larynx height. It demonstrates that larynx height is a factor in producing pharyngeals/epiglottals and that larynx height is closely coordinated with the action of closing and opening the aryepiglottic sphincter. The combined SLLUS methodology provides us with a means to measure parameters necessary to explain the function of larynx height. We surmise that if larynx height does not correlate with pitch, then the target must be a pharyngeal, as we have seen in these instances. If, on the other hand, larynx height does correlate with pitch, then we must be dealing with a target that is a tonal feature. This allows us to presume that a language such as Iraqi Arabic produces pharyngeal/epiglottal consonants by using the laryngeal constrictor mechanism in coordination with larynx height, whereas a Tibeto-Burman language, where constrictor function and pitch both characterize the 'register' of the syllable, will dedicate larynx height primarily to the production of tone. That is not to say that every dialect of Arabic should favour a raised larynx per se – only that the production of the pharyngeal stricture and larynx height are coordinated. And it is not to say that every Tibeto-Burman language will ignore the relationship between laryngeal constriction and larynx height – only that the tendency is for larynx height to be associated with pitch changes more than with laryngeal constrictor states.

4. *Laryngeal articulation and vowel quality*

Just as laryngeal articulation is related to phonatory quality and to tonal quality (of necessity, because both phonation and pitch changes are sited within the larynx), the fact that the laryngeal space is an articulator proper and not just a vibration generator means that vowel quality cannot be described auditorily or calculated acoustically based only on the parameters of the oral vocal tract. The shapes, volumes, and movements of the laryngeal vocal tract must also be taken into account. As it stands, the study of vowel quality tells us very little about the contribution of lower-vocal-tract resonances to the auditory or acoustic characteristics of vowels.

One way of looking at the configurations of both the oral and laryngeal vocal tracts together is through magnetic resonance imaging (MRI). The view is sagittal, giving an initial impression of area, much like cineradiographic images, but parallel slices can be obtained at successive depths to create data on volume. We have collected preliminary MRI data to test the procedure and to investigate whether it is possible to identify changes in oral articulator posture (and by extension differences in vowel quality) due to contrasting postures of the laryngeal articulator. Data capture is based on earlier pilot work to identify the articulatory nature of glottal stop and of creaky voice using MRI (Moisik, Esling, Crevier-Buchman, Amelot & Halimi, 2015). The MRI tests comprise imaging a series of peripheral vowels and then comparing those same vowel targets when produced with contrasting settings of the laryngeal mechanism, initially in midsagittal section. The MRI images are all static (sustained postures were maintained). The laryngeal conditions include glottal stop (the vowel followed by [ʔ]), epiglottal stop (the vowel followed by [ʡ]), creaky voice (the vowel produced with creaky phonation [̰]), and raised larynx voice (the vowel produced with [ˤ] quality). Canonically, raised larynx voice can be interpreted as the same as pharyngealized voice but higher in the pitch range. In theory, the vowel productions will show predictable positioning of the tongue in the oral cavity, and the laryngeal conditions will evoke not only changes in the 'source' (as in a lingually-based model) but also (because the larynx is also an articulator) changes in the articulatory shape of the lower vocal tract. Presumably then, there should be an interaction between the oral production of the vowel and the laryngeal production of the lower-vocal-tract coarticulation, whereby vowel quality is regulated and normalized, as it were, for the context.

To outline some preliminary observations, the glottal stop context (the vowel followed by [ʔ]) has interesting ramifications for tongue positioning. Generally, the tongue appears to tilt backwards and downwards, as on a fulcrum with the jaw tilting up slightly, when glottal stop closes the vowel. This changes, systematically, the lingual articulation for the vowel. The larynx as a whole does not appear elevated as a function of the glottal stop; in fact, the major effect on the larynx seems to be lowering. This helps us to understand the gradational nature of the action of the laryngeal constrictor mechanism; that is, glottal stop cannot necessarily be predicted to have the same laryngeal behaviour, or at least laryngeal height behaviour, as epiglottal stop, for instance. The epiglottal stop context (the vowel followed by [ʡ]) uniformly involves elevation of the larynx as a whole and simultaneous retraction of the tongue to close off the epilaryngeal tube and significantly narrow the lower pharynx behind the epiglottis. This does not mean, however, that the whole of the tongue is pulled backwards or downwards, as it appears to be in the glottal stop context. Instead, the tongue may have to stretch to attain the upwards approximation necessary to produce sufficiently fronted [i] or sufficiently raised [u]. In these cases, the laryngeal coarticulation forces the tongue to be pulled in two directions at the same time to achieve an adequate vowel quality target. For vowels in the context of creaky voice (the vowel produced with creaky phonation [̰]), the base of the tongue and the jaw appear to swivel downwards (towards the glottis) as the antero-posterior dimension of the larynx shortens. In the raised larynx voice (produced with [ˤ] quality) context, the tongue is pulled backwards and downwards, as it is for glottal stop, but the antero-posterior dimension of the larynx is stretched (lengthened) rather than shortened. This is ostensibly because a raised larynx voice or pharyngealized voice quality requires open air spaces under the tongue/epiglottis to create resonances. The descending tongue body/root is angled back to lie almost parallel above the glottis, which resembles the posture for epiglottal stop, but the need to keep the epilaryngeal tube cavity and the subepiglottal pharyngeal space open means that the cervical spine has to retract posteriorly while the larynx at the front has to advance anteriorly and tilt upwards slightly. By contrast, in the creaky voice context or in the epiglottal stop context, these spaces are effectively obliterated. Similarly to epiglottal stop, the severely retracted tongue of the raised larynx voice context means that the upper part of the tongue must stretch upwards to achieve front vowels ([i ɪ e ɛ æ]) or the raised vowel [u]. Thus, pronounced tongue retraction is common to both epiglottal stop and raised larynx voice due to the severe degree of laryngeal constriction; however, the epilaryngeal space below the flattened epiglottis is spread out anteriorly and posteriorly in raised larynx voice in order to preserve resonance cavity viability. Neither glottal stop nor creaky voice engender the same degree of constriction or retraction.

These preliminary observations suggest that the laryngeal articulator plays a considerable role in the shaping of resonance cavities that contribute to vowel quality. Each laryngeal condition examined here exerts a different effect on the configuration of the lower vocal tract and on the tongue and jaw, especially the lower part of the tongue. Nevertheless, the oral vocal tract preserves the role of distinguishing close vowel quality, even when laryngeal effects are simultaneously applied. The production of fronted [i] or of raised [u] requires the body of the tongue to front or raise even if the mass of the lower tongue is being pulled into the laryngeal space. The jaw closes to assist in these articulations, but the tongue must articulate the elevation movements at the same time as the retraction movements. When this happens, the tongue is pulled in two directions, upwards and downwards. Open vowels behave differently with respect to the relationship between the oral and the laryngeal articulators. The suggestion can also be made that tilt, on a line coursing anteriorly and upwards from the cricoid ring to the front of the jaw, is a function of the type of laryngeal configuration and not just of pitch.

The conclusion can be drawn that models of vowel quality description that consider only oral configurations as the components of articulation that generate the resonances that are heard as auditory quality or seen as spectral magnitudes must be considered deficient. By the same token, to consider that pitch (vocal fold vibratory frequency) is the only laryngeal component that interacts with vowel quality is also insufficient to account for the interactions that occur between the larynx and the oral vocal tract. To complete all aspects of vowel quality description, a triangulation must occur. The geometric relationships among three sets of physiological activities must be reconciled. The first is clearly the positioning of the articulatory structures (primarily the tongue) in the oral vocal tract – an orientation that differs substantially depending on whether the target vowel is close or open. The second is the positioning of the laryngeal articulatory structures in the lower vocal tract – configurations that differ depending on whether the sound being produced is a continuant or a stop, slightly closed or fully closed, or voiced or voiceless. The third is the role of pitch – not only adding vibratory frequency to the speech signal but also requiring that the laryngeal structures change position as glottal length shortens or stretches. The first parameter (oral) is interdependent on the second (laryngeal). Even if the laryngeal cavity is wide open, we must assume that the size and volumes of the laryngeal vocal tract are contributing to the auditory/acoustic characteristics of the ostensibly lingually-shaped vowel. The second parameter (laryngeal) becomes increasingly important as the laryngeal articulator departs from its open position and adopts a more constricted setting. Various constricted settings contribute different aspects of tilt (inclination of laryngeal structures to tongue to jaw), depending on the openness of the epilaryngeal spaces (usually a requirement for the resonance of these spaces), on the kind of phonation type that is being produced (where low-pitched types require a different laryngeal configuration from high-pitched types), and on whether multiple vibrators through to the top of the epilaryngeal tube add more complex requirements to the positioning of structures higher up the chain (because the bottom of the epilaryngeal tube relates differently to superior structures than does the aryepiglottic top of the epilaryngeal tube). The third parameter (pitch) is known to have an intrinsic relationship to vowel quality (Whalen, Gick, Kumada & Honda, 1998), as laryngeal tilt influences the shape of superior structures in their oral performance of the target vowel. However, since laryngeal tilt can also occur as a result of the performance of particular laryngeally constricted targets, it behooves us in our research on vowel quality to consider both the influence of tilt resulting from a manoeuvre adopted to generate a particular pitch as well as the influence of tilt resulting from a manoeuvre adopted to realize a combination of laryngeal postures. Since tilt can occur along different axes, laryngeal tilt arising from a pitch setting alone cannot account for the full chain of articulatory events that bind the lowest (glottal) level of the vocal tract to oral vowel articulation. In this chain of events, laryngeal articulatory configuration is the key intermediary factor – one that plays a more common and influential role than often suspected. The intricacies of these sequential and interconnected relationships throughout the vocal tract are the subject of a paper by Moisik, Esling, Crevier-Buchman & Halimi (submitted).

Acknowledgements to co-researchers

Scott R. Moisik, Nanyang Technological University (NTU) Singapore
Lise Crevier-Buchman, CNRS-UMR7018-LPP, Paris III, ORL-HEGP
Philippe Halimi, Radiologie-HEGP.

Bibliography

Catford, J.C. (1964). Phonation types: The classification of some laryngeal components of speech production. In Abercrombie, D., Fry, D.B., MacCarthy, P.A.D., Scott, N.C. & Trim, J.L.M. (Eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday 12 September 1961*. London: Longmans, 26-37.

Catford, J.C. (1968). The articulatory possibilities of man. In Malmberg, B. (Ed.), *Manual of phonetics*. Amsterdam: North-Holland Publishing Company, 309-333.

Catford, J.C. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Coey, C., Esling, J.H. & Moisik, S.R. (2014). *iPA Phonetics*, Version 1.0 [2014]. Department of Linguistics, University of Victoria.

Edmondson, J.A., Esling, J.H. (2006). The valves of the throat and their functioning in tone, vocal register, and stress: Laryngoscopic case studies. In *Phonology*, 23, 157-191.

Esling, J.H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. In *Journal of the International Phonetic Association*, 26, 65-88.

Esling, J.H. (2003). Glottal and epiglottal stop in Wakashan, Salish and Semitic. In Solé, M.J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, vol. 2. Barcelona, Spain, 1707-1710.

Esling, J.H. (2005). There are no back vowels: The laryngeal articulator model. In *Canadian Journal of Linguistics*, 50, 13-44.

Esling, J.H. (2010). Phonetic notation. In Hardcastle, W.J., Laver, J. & Gibbon, F.E. (Eds.), *The handbook of phonetic sciences* (2nd ed.). Oxford: Wiley-Blackwell, 678-702.

Esling, J.H., Fraser, K.E. & Harris, J.G. (2005). Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). In *Journal of Phonetics*, 33, 383-410.

Esling, J.H., Harris, J.G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In Hardcastle, W.J., Mackenzie Beck, J. (Eds.), *A figure of speech: A Festschrift for John Laver.* Mahwah, NJ: Lawrence Erlbaum, 347-383.

Gordon, M., Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. In *Journal of Phonetics*, 29, 383-406.

Ladefoged, P. (2001). *A course in phonetics* (4th ed.). Orlando: Harcourt College Publishers.

Laufer, A. (1996). The common [ʕ] is an approximant and not a fricative. In *Journal of the International Phonetic Association*, 26, 113-118.

Moisik, S.R. (2013). The epilarynx in speech. PhD Dissertation, University of Victoria.

Moisik, S.R., Esling, J.H. & Crevier-Buchman, L. (2010). A high-speed laryngoscopic investigation of aryepiglottic trilling. In *Journal of the Acoustical Society of America*, 127, 1548-1559.

Moisik, S.R., Esling, J.H., Crevier-Buchman, L., Amelot, A. & Halimi, P. (2015). Multimodal imaging of glottal stop and creaky voice: Evaluating the role of epilaryngeal constriction. In *The Scottish Consortium for ICPhS 2015* (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. Paper number 247.

Moisik, S.R., Esling, J.H., Crevier-Buchman, L. & Halimi, P. (submitted). Putting the larynx in the vowel space: Studying larynx state across vowel quality using magnetic resonance imaging. In *Journal of Phonetics*.

Moisik, S.R., Lin, H. & Esling, J.H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). In *Journal of the International Phonetic Association*, 44, 21-58.

Sawashima, M., Hirose, H. (1968). New laryngoscopic technique by use of fiberoptics. In *Journal of the Acoustical Society of America*, 43, 168-169.

Whalen, D.H., Gick, B., Kumada, M. & Honda, K. (1998). Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. In *Journal of Phonetics*, 27, 125-142.

DOI: 10.17469/O2103AISV000002

SUSANNE FUCHS

Changes and challenges in explaining speech variation: A brief review

In this brief overview I review how our understanding of speech variation has changed over the last decades. I argue that depending on the motivation of the scientific investigations and the theoretical points of view, variation has been discussed with respect to biological, social and communicative factors as well as with respect to the nonlinearities between acoustics, articulation, and perception. Since all of these factors can potentially interact with each other, the core challenges now are to explain the underlying mechanisms, to use appropriate methodologies for the analyses, and to understand how far the underlying mechanisms can be generalized to other speakers or communicative events. I conclude that speech variation is not an obstacle, but rather a rich source that allows us to examine the many facets of language.

Key words: variation, sociophonetics, speaker-specific behaviour, acoustic-articulatory relations.

1. *Changes in explaining speech variation*

The aim of this paper is to provide a general review of how our understanding of speech variation has changed over the last decades, and how speech variation has been explained, depending on the perspective of the researcher. The emphasis lies specifically on acoustic and articulatory variations that have their origin in dialectal, social, communicative and biological factors. Moreover, the nonlinear relations between acoustic, articulation and perception are taken into account. These and other constraints are particularly challenging when discussing speech variation. This review is not intended to provide a summary of studies in this area. Instead, some important papers have been selected which, in my view, mirror how variation has been approached at a certain time.

1.1 Changes in understanding dialectal, social and communicative influences

Finding variation in speech is not novel in itself, it has persisted since the early days of empirical speech research performed using technological equipment. Peterson, Barney (1952) are among the pioneers in the area of acoustics, analysing American English vowel formants of 76 speakers (men, women and children) with different dialectal background and the perception of these vowels by 70 listeners (men, women) with the same dialectal background as the speakers. On the basis of their results they were able to show that speaker productions and listener judgements were influenced by their dialectal background. Furthermore, they provided evidence that variation in the acoustic signal has vowel-specific consistencies and cannot be treated as random noise.

Another famous study focused on acoustic variations that can be attributed to social factors. Labov (1963) investigated the frequencies and distribution of diphthongs on Martha's Vineyard, an island with specific geographical patterns (e.g. up-island: a rural area; down-island: an area with larger towns and tourism) and social structures (four ethnic groups: English descents, Portuguese descents, Indian descents and a mixture of other groups; lowest average income; smallest amount of rich people; highest rate of unemployment in the state with a huge dependency on seasonal work in tourism; highest percentage in the state of married working woman with children). The speech analyses were based on 69 interviews with the islanders. Vowel centralization of the diphthongs was found particularly in inhabitants of the up-island while more open variants occurred in the speech of inhabitants living down-island with a larger influence by the mainland. He also noted that seasonal tourists did not directly affect the islanders' speech production, and differences were grounded in long-term processes of social disparities.

More recently, long-term effects of sound change have also been described in regards to a particularly famous social figure, Queen Elizabeth II of the United Kingdom (Harrington, Palethorpe & Watson, 2000; Harrington, 2006, 2007). These authors used the Queen's Christmas broadcast recordings to analyse changes in her speech between the 1950s and the 1980s. Such long-term analyses are exceptional, because longitudinal data from the same speaker are rarely available. Harrington and colleagues were able to show that from the 1950s to the 1980s, the Queen's vowel productions moved in the direction of English speakers who are younger and lower in the social hierarchy.

Within the last decades, there have also been studies focussing increasingly on speech variation that is due to short-term adaptation of a speaker towards a specific listener and a communicative situation. Such an approach does not exclude the possibility of long-term adaptations. Short-term adaptations have been covered under several terms with different definitions, such as "interactive alignment" (e.g. Pickering, Garrod, 2004), "convergence" (e.g. Manson, Bryant, Gervais & Kline, 2013), "inter-personal coordination" (e.g. Tolston, Shockley, Riley & Richardson, 2014), "speech imitation" (e.g. Garnier, Lamalle & Sato, 2013), "accommodation" (Giles, Coupland & Coupland, 1991), and "entrainment" (e.g. Levitan, Gravano & Hirschberg, 2011). For conversations, it has been shown that interlocutors adapt to each other at various levels. They can adapt in the use of words and their meaning (convergence in lexical and semantic representations, see e.g. Garrod, Anderson, 1987), in the use of syntactic structures (convergence in syntactic representations, see e.g. Branigan, Pickering & Cleland, 2000) and in speech rate or fundamental frequency (phonetic convergence, see e.g. Babel, 2009; Babel, Bulatov, 2012). However, it is still unclear where the origin of this convergence lies and how it emerges between humans, despite their not being physically connected.

One view proposes that convergence in dialogue is the result of shared linguistic representations (Pickering, Garrod, 2004). Pickering and Garrod assume that convergence is grounded in an automatic priming process. Priming refers to a process of increased sensitivity to a certain stimulus due to prior experience, e.g. the prior presentation or production of a word enhances later perceptual identification of this word (Jakoby, 1983). This process is automatic and can even occur in cases in which the task is "not to converge" (Issartel, Marin & Cadopi, 2007).

Such a view is somewhat different from the Communication Accommodation Theory (CAT) by Giles *et al.* (1991). In their framework, the degree of convergence is not automatic, as convergence is seen from a sociolinguistic perspective where interlocutors adjust in the direction of increased mutual similarity to facilitate communication and establish appropriate social distance in the respective communicative situation.

Eckert (2008: 455) proposes, "speaking in the social world involves a continual analysis and interpretation of categories, groups, types, and personae and of the differences in the ways they talk". Social meaning and variation of language is then studied with respect to style, which can flexibly change according to the many factors and functions involved in communicative situations and their interpretation (for further review, see Eckert 2012).

Thus, speech variation has been approached from at least two different perspectives: long-term changes due to, e.g., dialectal background, social status, age, and short-term adaptations to the listener and situation. Long-term adaptations and the respective changes are often discussed in light of sound change phenomena in larger communities and short-term adaptations are seen with respect to the flexibility and individual interpretation of a social agent under the situational circumstances.

1.2 Biological factors explaining speech variation

In his pioneering empirical work, Fant (1966) analysed formants of sustained vowel productions in seven male and seven female speakers of Swedish. On the basis of this speech material, he was able to show that on average, females have formant frequencies 18 percent higher than those of males. This scaling factor is inversely proportional to the vocal tract length of the speaker, i.e. higher formant values go hand in hand with smaller vocal tracts in females in comparison to males. However, Fant also reported vowel-specific effects. Specifically, low back vowels are produced with much higher first formants in females than in males. These production differences among male and female speakers are larger than in any other vowel examined. Fant attributed these findings to the relatively long pharyngeal cavity of male speakers in comparison to female speakers. He also mentioned another potential influence on vowel-specific effect – the smaller laryngeal cavity in females.

A substantial body of work followed Fant's seminal study using the progress in technology and computational power. Fitch, Giedd (1999) used a large sample of 129 people (53 females) with a normal Body-Mass-Index in an age range from 2 to 25 years as part of a larger study on brain development. Body features (weight and height) were taken into account as well as selected measurements based on scans of Magnetic Resonance Images of the vocal tract. Vocal tract length strongly correlated with body features: taller and heavier people had a longer vocal tract. When the effect of body size was removed from the statistical analysis, effects of age and sex remained with specific proportions between the oral and pharyngeal cavity for males and females. Additional analysis revealed that sex-differences in pharyngeal length occur after puberty.

So far, Vorperian, Wang, Schimek, Durtschi, Kent, Gentry & Chung (2011) have analysed the largest sample of which I am currently aware (Magnetic Resonance Imaging scans & computer tomography images of 605 people from birth to 19 years of age), adding more details to the nonlinear development of vocal tract proportions and discussing the implications for speech acoustics. Surprisingly few studies, though, have investigated changes in the upper vocal tract in the elderly, probably because the most significant changes occur from birth to adulthood. However, these changes are crucial particularly in the discussion of sound change when comparing older and younger adults (Reubold, Harrington & Kleber, 2010). Xue, Hao (2003) are an exception, analysing data from 38 younger speakers (20-30 years old) and 38 older speakers (65-87 years old, all healthy). They found a larger oral cavity volume in the elderly, a slightly longer mouth cavity, and no difference in pharyngeal length. Overall vocal tract length did not differ significantly. The larger mouth cavity in the elderly coincided with lowered first formant values. Further empirical work is needed to describe and evaluate vocal tract changes on a continuous scale over the entire lifespan.

From the description of the overall vocal tract shape, more recent studies have focused on specific parts of the vocal tract, e.g. the size and shape of the palate (e.g., Brunner, Fuchs & Perrier, 2009; Fuchs, Toda, 2010; Yunusova, Rosenthal, Rudy, Baljko & Daskalogiannakis, 2012; Lammert, Proctor & Narayanan, 2013; Weirich, Fuchs 2013; Weirich, Fuchs, Simpson, Winkler & Perrier, 2016). For instance, Brunner *et al.* (2009) investigated differences in the coronal plane of the palate and discussed these with respect to speakers' articulatory precision in the production of high vowels and /j/. All these sounds are realized with a considerable amount of tongue-palate contacts. The authors assumed that speakers with a flat palate shape have to limit their articulatory variability in order to keep acoustic variability within a tolerable range, while speakers with a dome-shaped palate are less constrained. These assumptions were generally confirmed by means of electropalatographic data from 32 speakers of different languages. Speakers with a flat palate consistently showed a lower level of articulatory variability than speakers with dome shaped palates (who were either variable or not), while acoustic variability did not differ among the speakers.

The palatal inclination angle has been discussed with respect to the production of the phonemic /s/-/ʃ/ contrast in German (Weirich, Fuchs, 2013). The authors tested their hypotheses on the basis of speech samples of mono- and dizygotic twin pairs as well as a group of unrelated speakers. They were able to provide evidence for speaker-specific articulatory strategies depending on the alveolo-palatal inclination angle. Speakers with a flat angle only retracted the tongue for the postalveolar fricative while speakers with a steep angle additionally elevated the tongue.

The particular shape of the palate has even been discussed as a bias in the development of clicks in Khoisan languages, even if it is clear that speakers can in principle learn every language independently of their vocal tract properties (Moisik, Dediu, 2015). The authors follow up on earlier work that reports a lack of a clear alveolar ridge in speakers of these click languages. Moisik, Dediu suggest that less articulatory effort is needed for the production of clicks when speakers have a smooth palatal profile. They provide evidence for their suggestion by means of simulations with a flat and a steeper alveolar ridge using a 3D biomechanical tongue model.

Different modelling approaches were required for a recent investigation (Weirich *et al.*, 2016) of the degree of jaw opening in males' and females' speech, because it is unclear whether speakers only adapt to their particular vocal tract anatomy or compensate for it. Weirich and colleagues considered the possibility that the longer pharynx in male speakers coincides with a greater distance between the condyle and the gnathion, which in turn also results in a larger jaw displacement. Simulations with prototypical male and female models showed a complete linguo-pharyngeal closure for the male model producing a prominent low vowel, while such a closure was not found in the female model. The authors therefore suggest that in the production of low prominent vowels, males do not open their jaw very widely, because this may cause a linguo-pharyngeal closure if the tongue does not compensate for it. The study is particularly interesting in light of the reported sociophonetic finding that male mumbling comes across as macho (Heffernan, 2010).

Consistent speaker-variation has not only been described at the level of vocal tract morphology, but also in regards to parts which had been considered rather invariant among human subjects for a long time. Golestani and colleagues (Golestani, Price & Scott, 2011) carried out an anthropomorphic analysis of brain structures in speech areas for two different participant groups: expert phoneticians who had received years of training in sound production and perception and a control group with no particular training. Golestani *et al.* were able to find differences in the surface area of the pars opercularis, a structure involved in phonological processing. The size of the area correlated positively within the group of expert phoneticians with years of phonetic training. Thus, training can affect brain plasticity in speech-related areas. The authors also found a greater gyrification in the auditory cortex of the expert phonetician brain, but this did not show any correlation with training; it most likely already develops in utero.

Finally, one might come to the conclusion that even if there can be consistent inter-speaker variation at the level of vocal tract or brain morphology, it might not occur at the genetic level in the non-clinical population. But even at this level, our understanding has changed over the last decades. The popular assumption that genes defining our physical properties are fixed and assigned at birth without any further changes is no longer viable (Dediu, 2015). Furthermore, even if humans are genetically very close to their ancestors, the idea that the mutation of a single gene brought language to the human species (as is often proposed by researchers in generativism) has become nothing more than a fantasy (Dediu, Christiansen, 2016). According to the authors, a certain gene can produce different proteins at different times in different tissues. "The genetic foundations of language and speech are extremely complex and there is no gene 'for' language (Fisher, 2006). Instead, there are many genes interacting in complex regulatory networks tuned to many contextual cues and influencing many aspects of the phenotype. Genes are not monolithic units with simple and clear functions but instead there is pervasive gene regulation at multiple levels and constant interaction with the environment" (Dediu, Christiansen, 2016: 367).

To summarize, consistent biological between-speaker variation can be found at different macroscopic and microscopic levels. They have been discussed in terms of speech motor control, individual differences (Fuchs, Pape, Petrone & Perrier, 2015a), and even language evolution (Dediu, Jannsen & Moisik, 2016). The first robust empirical data focussed on visible and audible differences in vocal tract morphology. More recently, these have been complemented by work looking behind the surface structures and examining, for example, the level of the brain, the level of gene regulation, and the level of biomechanics, adding increasing detail to the complex picture and sometimes challenging theoretical models of speech and language production.

1.3 Relation between different levels

A complicating factor in the description of speech variation comes into play when considering different levels, i.e. acoustics, articulation and perception. In many instances, these levels have a nonlinear relationship and a certain amount of variation at one level does not necessarily coincide with variation at another level.

Different relationships have yielded very influential theoretical concepts in phonetics. In his famous paper on the quantal nature of speech, Stevens (1989) described the non-linear relationship between acoustics and articulation as well as between acoustics and perception in three different regions of the articulatory domain. Regions I and III characterize a relation of relative acoustic stability (with only small changes visible as plateau-like shapes), but substantial articulatory variation. Both regions differ qualitatively with respect to their acoustic values. Region II is a threshold area between regions I and III, characterized by stability in articulation and huge variability in the acoustics. The articulatory-acoustic relations, according to Stevens (see page 5), are quantal in nature and can also be applied to the relation between perception and acoustics. Stevens interprets these relationships in connection with distinctive phonological features. Stable acoustic regions (regions I and III) would be favoured in phonological inventories. Thus Stevens supposes a primacy of invariant or relatively stable acoustic regions over variable articulatory motions for speech production (Blumstein, Stevens, 1979; Stevens 1989).

Further support for this primacy comes from studies on motor equivalence in speech production, following the general motor control principle that several possibilities exist to reach a defined goal. For speech production, trading relations have been shown, e.g., between tongue body raising and lip protrusion during the production of /u/ (among others, Perkell, Matthies, Svirsky & Jordan, 1993; Savariaux, Perrier & Orliaguet, 1995). This leads the authors to infer that goals are defined in the acoustic/auditory domain, because this domain is relatively stable across conditions, while articulatory motions vary for this particular sound.

Another very influential approach discussing the relation between speech production and perception is the theory of hypo- and hyper articulation (H&H by Lindblom, 1990). It rests on empirical evidence that there is no invariance in the speech signal. Lindblom suggests that speech production varies along a continuum between hyper- and hypo articulation. Hyper articulation may be a requirement under certain situational conditions (e.g. the interlocutor is in another room; there is background noise, etc.). The effort manifests on the side of the speaker who must produce a message and increase discriminability for the listener. Hypo-articulation follows the principles of minimal motor effort for the speaker, but goes hand in hand with less perceptual discriminability and increased effort by the listener to understand the message. The H&H theory suggests that flexible adaptations of speakers to reception (social and communicative constraints) and production (physiological and cognitive constraints, see Lindblom 1990: 418) are responsible for the variation between hyper- and hypo-speech.

More recently, Perkell and colleagues (Perkell, Guenther, Lane, Matthies, Stockmann & Tiede, 2004a; Perkell, Matthies, Tiede, Lane, Zandipour, Marrone 2004b) found speaker-specific effects in the relation between speech production and perception. They hypothesized that inter-speaker variation may be attributed to individual perceptual capabilities. Those speakers who are able to accurately discriminate phonemic contrasts are also the ones who could distinctively produce a phonemic contrast. 19 speakers were recorded with electromagnetic articulography and acoustics producing the words *cod*, *cud*, *who'd*, and *hood*. The same speakers underwent a perception experiment with an annotation and a discrimination task of the speech material. Speakers with high perceptual discriminability scores were the ones who produced the vowel contrasts with a larger distance in the acoustic and articulatory space in comparison to the speakers with less perceptual discriminability, providing a reason for speaker variation based on perceptual capacities.

A study by Cangemi, Krüger & Grice (2015) showed that the speaker-specific behaviour in perceptual discriminability and articulatory precision is not a general property of a "universally intelligible speaker" or a "universally proficient listener", but rather that how a speaker behaves with respect to a particular listener depends on the particular dyad, at least for intonational contrasts of focus. Cangemi and colleagues recorded five native speakers of German using electromagnetic articulography with target words under different focus conditions. Lip motions and acoustic pitch accents were analysed. The production experiment was followed by a perception experiment with 20 different listeners judging the productions of the five speakers. The most original finding was that one particular subject can be more intelligible than most others for a particular listener and less intelligible than most other speakers for another listener. Thus, the same subject can be judged on very different ends, depending on the respective dyad.

The complex interplay and the nonlinearities between different levels involved in speech production and perception make an investigation of speech variation challenging, since this variation may be specific to one particular level and may depend on the speech material, the speaker task, and communicative constraints.

In summary, in the last decades researchers dedicated their attempts to the description of nonlinearities between acoustics, articulation, and perception with the primary interest being to find regularities and explain phonemic inventories. More recent studies focus on the flexibility of these nonlinear relations due to speaker-specific or communicative constraints.

2. *Challenges in explaining speech variation*

Researchers trying to explain the variation in speech production and perception face various challenges. I will provide four of them related to the following topics: generalization, intra- and inter-speaker variability, single time point and time series analysis, and multidimensional factors. I will try to offer explanations on the basis of selected examples.

2.1 How far can we generalize?

Experiments are often designed in such a way that a few factors with a few levels are varied and the dependent variables studied. Increasing the number of factors to higher than three has the disadvantage that it is hard to interpret potential interactions among them. If a factor has several levels, additional post-hoc analyses are required to understand the relation between the different levels. In this respect, experimental designs with just a few independent variables and levels are favoured. However, less comprehensive designs may restrict the researcher with regards to the generalizations that can be drawn. In a recent investigation, Koenig, Fuchs (2015) studied the effect of loudness (normal versus loud speech) on German vowel production in different speech tasks. The investigation was motivated by previous efforts which a) focussed to a large extent on low vowels rather than sampling broadly across the vowel space, and b) provided average values for the whole vowel space without being explicit about vowel-, speaker-, or task effects. Moreover, loud speech is often taken as an intervention method for speech therapy (e.g. within the Lee Silverman voice treatment). Understanding how much loud speech affects formant values across several vowels, speakers, and tasks is fundamental to evidence-based treatment.

Figure 1 - *F1 and F2 space (in Hz) consisting of tense German /a, i, u/ vowels, spoken by 11 female speakers in normal (grey) and loud (black) speech conditions; from Koenig & Fuchs (2015)*

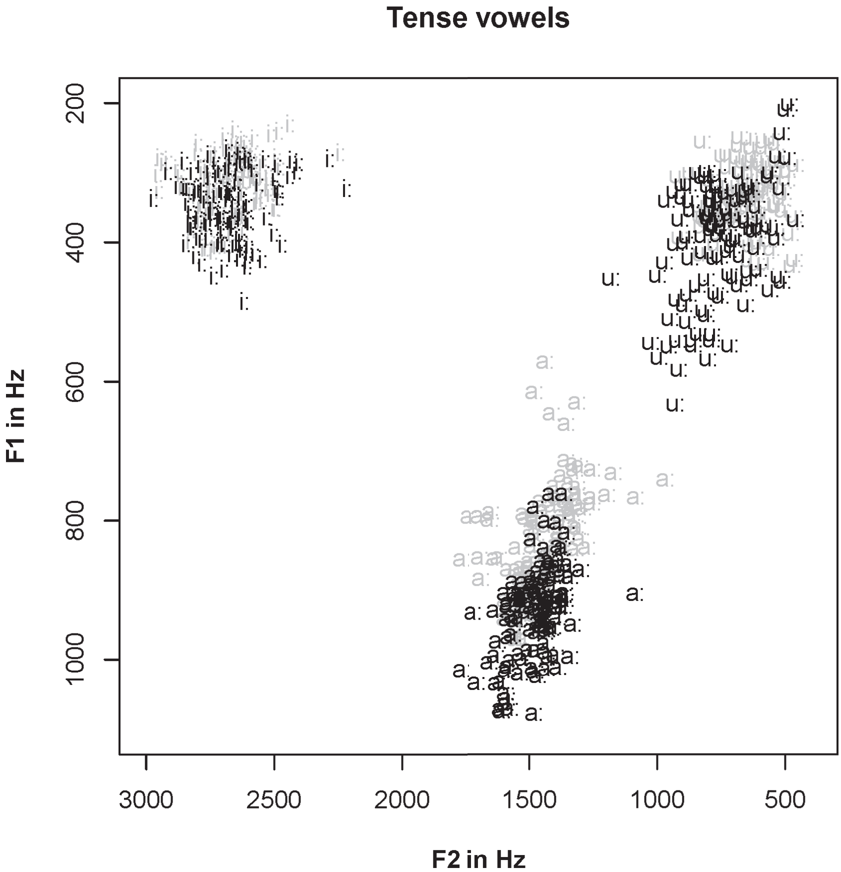
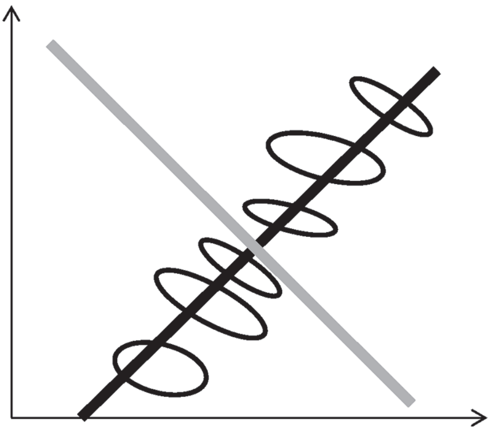


Figure 1 displays the results for tense vowels in a question-answer task. Normal vowel productions are visualized in grey and loud speech in black. Significant differences due to loudness are produced in the first formant of low vowels, while in high vowels these differences are weak or absent. The study is a nice example that investigating only low vowels does not allow generalizations to all other vowels. The often described increase in vowel space (based on the area between the corner vowels /a, i, u/ in loud speech may actually be a result of the low F1 value in /a/ only and not a result of more peripheral vowels in general. Hence, it is advisable to draw conclusions that are based on the specific speech material and task.

2.2 How much are global effects mirrored in intra-speaker variation?

The second example is related to intra- and inter-speaker variation. In most studies, researchers are interested in global behaviour across speakers and a given task, while single speakers' behaviour is treated as random noise. If several data points are recorded from single subjects, the distribution of these must be considered. Figure 2 shows a schematic example for an extreme case. The six black dispersion ellipses correspond to distributions of six different subjects. The main direction of variance in these ellipses clearly shows a negative within-speaker correlation (depicted as the schematic grey regression line). If statistical analysis were to be based solely on average values of all speakers, a positive correlation would be found (depicted as the black regression line). Even though this example may be an extreme case, it makes sense to look at intra- as well as inter-speaker variation, to record several data points for each subject, to graphically explore a potential bias between intra- and inter-speaker effects, and to include speaker-specific slopes in statistical models.

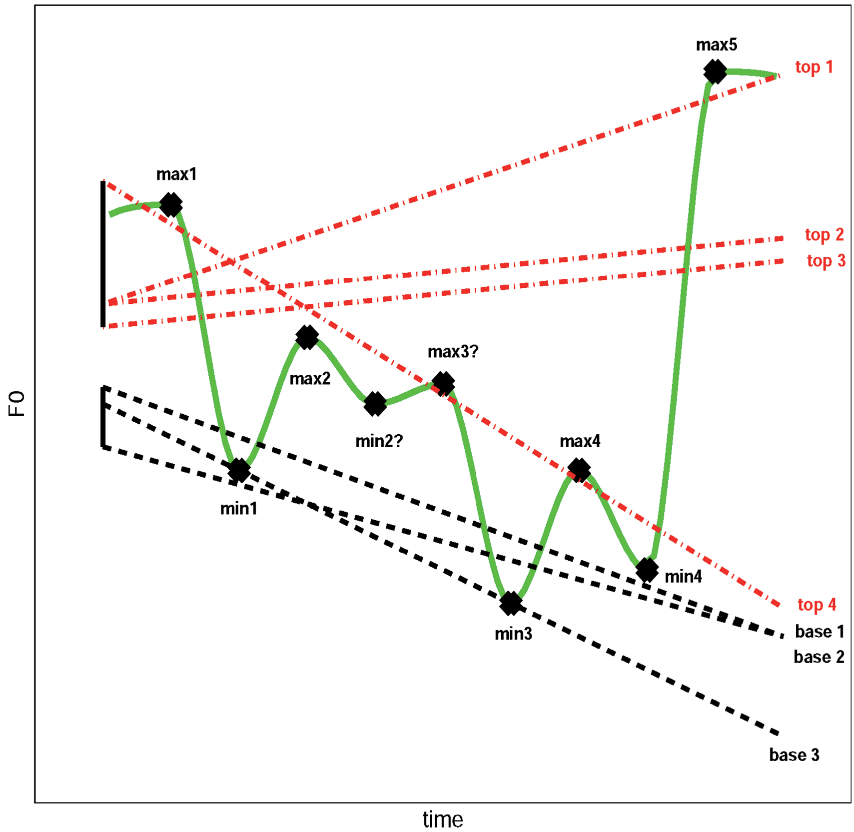
Figure 2 - *Schematic view of intra- and inter-speaker variation of two measured parameters. The dispersion ellipses correspond to variation in six single speakers. The grey regression line schematically displays the variation within speakers while the black regression line to variation between speakers*



2.3 Single time point analyses versus time series analyses

Another important challenge we are currently facing is the choice of either analysing selected temporal landmarks or the entire time series in the temporal window of interest. More and more studies oppose "magic moment measures" (Vatikiotis-Bateson, Barbosa & Best, 2014), because the choice of the selected time points may be primarily driven by certain theoretical concepts. Single time point analysis can be valuable in one case, but unreliable in another, since speech production and perception are complex dynamic processes involving the flow of coordinated articulatory motion and transitory acoustic states. The recent advances in tools able to take the whole time series into account are particularly helpful in this respect. The third example will illustrate this point. The example deals with a phenomenon called f0 declination, reflecting the gradual decrease of the fundamental frequency (see Fuchs, Petrone, Rochet-Capellan, Reichel & Koenig, 2015b, and references therein). F0 declination has been measured in various ways: As the regression line that is fit through all f0 values within a given time interval, as the topline, a regression line going through all high pitch accents, a baseline moving through the low tones, and the midline, a regression line between the top- and the baseline. One of the difficulties for the calculation of top-, base- and midlines is the selection of the local high and low tones, which very much depends on the phonological theory. In Figure 3, it is shown that the slope of the topline and baseline can vary considerably when including or excluding some of the pitch accents. For instance, topline 4 has a negative slope when the boundary tone (max 5) is not included, while all other toplines show a positive slope when the boundary tone is included.

Figure 3 - *Stylistic f0 contour with potential pitch maxima and minima (black dots). Based on the inclusion of certain maxima and minima, different f0 top- and baselines (dashed and dashed-dotted lines) were calculated. This figure was adopted from Reichel, Mády, 2013. Thanks to Uwe Reichel for making it available*



When all data points are taken into account for the calculation of the regression line, the results will be less affected by local peaks and valleys, but may also be subject to noise, due for instance to micro-prosodic perturbations.

2.4 Teasing apart different influences

The largest challenge of all might be to tease apart the different factors (see § 1.1-§ 1.3). Doing so presupposes a researcher or a research team that has a broad background expertise and is aware of all the factors that could potentially have an impact on the observed dependent variables. For example, researchers in sociophonetics should be aware of the biological factors that might also contribute to their findings and know the literature in this area. Reversely, this is also true for researchers who are primarily interested in the biological origin of language. Thus, researchers must critically question at which level they expect variation, how much this expectation is driven by their own theoretical and conceptual thinking, and to what extent they may be "blind" to other areas. Luckily, the existence of vast publication databases allows researchers to carry out comprehensive inter-disciplinary literature reviews reaching beyond their main research area.

Additionally, speaker-specific physically realistic models (e.g. Winkler, Fuchs, Perrier & Tiede, 2011; Stavness, Nazari, Perrier, Demolin & Payan, 2013) are available that help us to better separate articulatory behaviour that is an adaptation to specific vocal tract properties from cases in which speakers compensate for particular properties. It also allows investigation of the complex relations between speaker-specific anatomy, muscle recruitment, articulation and acoustics with the drawback of being time-consuming and computationally expensive.

Finally, open access articles and respectful data sharing among researchers may be a rich source for explaining speech variation from different angles and at the same time cutting costs and reducing effort.

3. *Conclusion*

This review has shown that variability in speech has been found at various levels since the earliest empirical studies on the topic were first conducted. This should not imply that there were no approaches supporting the idea of invariance. However, within the last decades, the body of empirical evidence indicating that variation is everywhere has continued to grow. This is the case not only at the macroscopic level (such as body size and vocal tract length), but also at the microscopic level (such as in brain morphology or genetics). Speakers differ in terms of their body features, language and culture, and social behaviour, and they can flexibly adapt to communicative constraints.

Depending on the focus of their studies, scientists have provided explanations for variation that range from social to communicative to biological in nature. To overcome the boundaries between research disciplines, we should be aware of the limits of our own conceptual thinking when interpreting variation. A comprehensive knowledge of different research perspectives, theoretical plurality, critical thinking, and/or working in interdisciplinary teams are among the factors which could help to allow future work to disentangle potentially co-occurring processes (Fuchs, Lancia, 2016). I believe we now have access to many sources (e.g. multidimensional and multisensory recording techniques, statistics, computational power for processing huge datasets, open access libraries, servers for scientific data sharing), which allow us to proceed in such a direction.

What is the main takeaway from this brief review?

I wish to conclude with something Georg Meyer said at our first summer school "Cognitive and physical models of speech production, perception and perception-production interaction" in Lubmin 2004: "variability is not the enemy, variability is our friend". I would even go a step further and say that variability is a rich source that allows us to examine the many facets of language in detail, something which would be impossible if variation did not exist.

Acknowledgments

This work was supported by the German Bundesministerium für Bildung und Forschung (BMBF) (Grant Nr. 01UG1411) and the Leibniz Society. I would like to thank my colleagues at the Laboratory Phonology Group at ZAS, the editors, Barbara Gili Fivela and my external collaboration partners for the fruitful exchanges, comments and discussions related to these topics. Thanks to Olivia Maky for proofreading.

Bibliography

Babel, M.E. (2009). Phonetic and social selectivity in speech accommodation. PhD Dissertation, University of California, Berkeley.

Babel, M.E., Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. In *Language and Speech*, 55(2), 231-248.

Blumstein, S.E., Stevens, K.N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. In *The Journal of the Acoustical Society of America*, 66(4), 1001-1017.

Branigan, H.P., Pickering, M.J. & Cleland, A.A. (2000). Syntactic coordination in dialogue. In *Cognition*, 75, B13-25.

Brunner, J., Fuchs, S. & Perrier, P. (2009). On the relationship between palate shape and articulatory behavior. In *The Journal of the Acoustical Society of America*, 125(6), 3936-3949.

Cangemi, F., Krüger, M. & Grice, M. (2015). Listener-specific perception of speaker-specific productions in intonation. In Fuchs, S., Pape, D., Petrone, C. & Perrier, P. (Eds.), *Individual differences in speech production and perception.* Frankfurt am Main-Berlin-Bern-Bruxelles-New York-Oxford-Wien: Peter Lang Publisher, 123-145.

Dediu, D. (2015). *An introduction to genetics for language scientists*. Cambridge: Cambridge University Press.

Dediu, D., Christiansen, M.H. (2016). Language evolution: constraints and opportunities from modern genetics. In *Topics in Cognitive Science*, 8(2), 361-370.

Dediu, D., Jannsen, R. & Moisik, S.R. (2017). Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. In *Language and Communication*, 54, 9-20.

Eckert, P. (2008). Variation and the indexical field. In *Journal of Sociolinguistics*, 12(4), 453-476.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. In *Annual Review of Anthropology*, 41, 87-100.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. In *STL-QPSR*, 7(4), 22-30.

Fisher, S.E. (2006). Tangled webs: Tracing the connections between genes and cognition. In *Cognition*, 101(2), 270-297.

Fitch, T., Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. In *The Journal of the Acoustical Society of America*, 106(3), 1511-1522.

Fuchs, S., Lancia, L. (2016). Seeing speech production through the window of complex interactions: Introduction to supplement containing selected papers from the Xth International Seminar on Speech Production (ISSP) in Cologne. In *Journal of Speech, Language, and Hearing Research*, 59, S1555-S1557.

Fuchs, S., Pape, D., Petrone, C. & Perrier, P. (Eds.) (2015). *Individual differences in speech production and perception.* Frankfurt am Main-Berlin-Bern-Bruxelles-New York-Oxford-Wien: Peter Lang Publisher.

Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U.D. & Koenig, L.L. (2015). Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. In *Journal of Phonetics*, 52, 35-45.

Fuchs, S., Toda, M. (2010). Can biological or sociophonetic factors explain the difference in male versus female /s/. In Fuchs, S., Zygis, M. & Toda, M. (Eds.)., *Turbulent sounds. An interdisciplinary guide*. Berlin: Mouton de Gruyter, 281-302.

Garnier, M., Lamalle, L. & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. In *Frontiers in Psychology*, 4, 600.

Garrod, S., Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. In *Cognition*, 27, 181-218.

Giles, H., Coupland, N. & Coupland, J. (1991). Accomodation Theory: Communication, context, and consequence. In Giles, H., Coupland, J. & Coupland, N. (Eds.), *Contexts of Accomodation: Developments in Applied Sociolinguistics*. New York: Cambridge University Press, 1-68.

Golestani, N., Price, C.J. & Scott, S.K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. In *Journal of Neuroscience* 31(11), 4213-4220.

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In *Laboratory phonology*, 9, 125-143.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. In *Journal of Phonetics*, 34(4), 439-457.

Harrington, J., Palethorpe, S. & Watson, C.I. (2000). Does the Queen speak the Queen's English? In *Nature*, 408, 927-928.

Heffernan, K. (2010). Mumbling is macho: Phonetic distinctiveness in the speech of American radio DJs. In *American Speech*, 85, 67-90.

Issartel, J., Marin, L. & Cadopi, M. (2007). Unintended interpersonal co-ordination: Can we march to the beat of our own drum? In *Neuroscience Letters*, 411, 174-179.

Jakoby, L.L. (1983). Perceptual enhancement: Persistent effects of an experience. In *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 21-38.

Koenig, L.L., Fuchs, S. (2015). Acoustic effects of loud speech and interrelationships among measures. In The Scottish Consortium for ICPhS (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK: the University of Glasgow. Paper number 424.

Labov, W. (1963). The social motivation of a sound change. In *Word*, 19(3), 273-309.

Lammert, A., Proctor, M. & Narayanan, S. (2013). Morphological variation in the adult hard palate and posterior pharyngeal wall. In *Journal of Speech, Language, and Hearing Research*, 56(2), 521-530.

Levitan, R., Gravano, A. & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2. Association for Computa tional Linguistics*, 113-117.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W.J. & Marchal, A. (Eds.), *Speech production & speech modeling*. Netherlands: Kluwer Academic Publishers, 403-439.

Manson, J.H., Bryant, G.A., Gervais, M.M. & Kline, M.A. (2013). Convergence of speech rate in conversation predicts cooperation. In *Evolution and Human Behavior*, 34(6), 419-426.

Moisik, S.R., Dediu, D. (2015). Anatomical biasing and clicks: Preliminary biomechanical modelling. In Little, H. (Ed.), *Proceedings of the Satellite event: The evolution of phonetic capabilities: causes, constraints, consequences*, Glasgow, UK, 12 August 2015, 8-13.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E. & Tiede, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. In *Journal of the Acoustical Society of America*, 116, 2338-2344.

Perkell, J.S., Matthies, M.L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E. & Guenther, F.H. (2004). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination & use of an articulatory saturation effect. In *Journal of Speech, Language, and Hearing Research*, 47, 1259-1269.

Perkell, J.S., Matthies, M.L., Svirsky, M.A. & Jordan, M.I. (1993). Trading relations between tongue-body raising & lip rounding in production of the vowel /u/: a pilot motor equivalence study. In *Journal of the Acoustical Society of America*, 93, 2948-2961.

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.

Pickering, M.J., Garrod, S. (2004). The interactive-alignment model: Developments and refinements. In *Behavioral and Brain Sciences*, 27(2), 212-225.

Reichel, U.D., Mády, K. (2013). Parameterization of F0 register and discontinuity to predict prosodic boundary strength in Hungarian spontaneous speech. In Wagner, P. (Ed.), *Elektronische Sprachverarbeitung. Studientexte zur Sprachkommunikation*. Dresden: TUDpress, 223-230.

Reubold, U., Harrington, J. & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: a longitudinal analysis in adult speakers. In *Speech Communication*, 52(7), 638-651.

Savariaux, C., Perrier, P. & Orliaguet, J.P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: a study of the control space in speech production. In *Journal of the Acoustical Society of America*, 98, 2428-2842.

Stavness, I., Nazari, M.A., Perrier, P., Demolin, D. & Payan, Y. (2013). A biomechanical modeling study of the effects of the orbicularis oris muscle and jaw posture on lip shape. In *Journal of Speech, Language, and Hearing Research*, 56(3), 878-890.

Stevens, K.N. (1989). On the quantal nature of speech. In *Journal of Phonetics*, 17, 3-45.

Tolston, M.T., Shockley, K., Riley, M.A. & Richardson, M.J. (2014). Movement constraints on interpersonal coordination and communication. In *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 1891-1902.

Vatikiotis-Bateson, E., Barbosa, A.V. & Best, C.T. (2014). Articulatory coordination of two vocal tracts. In *Journal of Phonetics*, 44, 167-181.

Vorperian, H.K., Wang, S., Schimek, E.M., Durtschi, R.B., Kent, R.D., Gentry, L.R. & Chung, M.K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. In *Journal of Speech, Language, and Hearing Research*, 54, 995-1010.

Weirich, M., Fuchs, S. (2013). Palatal morphology can influence speaker-specific realizations of phonemic contrasts. In *Journal of Speech, Language, and Hearing Research*, 56(6), S1894-S1908.

Weirich, M., Fuchs, S., Simpson, A., Winkler, R. & Perrier, P. (2016). Mumbling: Macho or morphology? In *Journal of Speech, Language, and Hearing Research*, 59(6), S1587-S1595.

Winkler, R., Fuchs, S., Perrier, P. & Tiede, M. (2011). Speaker-specific biomechanical models: From acoustic variability via articulatory variability to the variability of motor commands in selected tongue muscles. In Laprie, Y. (Ed.), *Proceedings of the ISSP Montreal*, 219-226.

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.

Yunusova, Y., Rosenthal, J.S., Rudy, K., Baljko, M. & Daskalogiannakis, J. (2012). Positional targets for lingual consonants defined using electromagnetic articulography. In *The Journal of the Acoustical Society of America*, 132(2), 1027-1038.

PARTE II

SOCIALE E BIOLOGICO NELLA VARIAZIONE INDOTTA DALLE CARATTERISTICHE DEL PARLANTE, DELLA LINGUA E DEL CONTESTO COMUNICATIVO

DOI: 10.17469/O2103AISV000003

DANIELA MEREU

Arretramento di /s/ nel sardo cagliaritano: uno studio sociofonetico

This paper presents a sociophonetic study on a local stereotype of the variety of Sardinian spoken in Cagliari. In particular, the aim is to investigate the use of the realisation of the voiceless and voiced alveolar fricatives /s, z/ as a voiceless postalveolar fricative [ʃ], in pre-consonantal environments (also across word boundaries), e.g. *tostau* [toʃˈtau] 'hard, tough', *disgràtzia* [diʃˈgrattsja] 'bad luck, disaster'. After the acoustic analysis, the paper focuses on the exploration of the patterns of stylistic variation of this variable. The results of the analysis show the activation of the substandard variant [ʃ] related to particular topics.

Key words: Cagliari Sardinian, /s/-retraction, sociophonetic variation, style-shifting.

1. *Introduzione*

Questo studio si propone di presentare un'indagine sociofonetica sull'arretramento di /s/ nel sardo cagliaritano.

La varietà di sardo parlata a Cagliari è stata argomento di poco interesse da parte degli studi di linguistica sarda, in quanto i paradigmi tradizionali della dialettologia si sono focalizzati sempre sulle varietà rurali, ritenute meno aperte agli influssi esterni e quindi maggiormente conservative (Chambers, Trudgill, 1980). Pertanto, negli studi linguistici sul sardo è possibile trovare solamente qualche sporadico riferimento ai tratti fonetici del cagliaritano (Wagner, 1941[1984]; Virdis, 1978; Blasco Ferrer, 1984; Paulis, 1984; Atzori, 1986; Fontana, 1996; Dettori, 2002). Sul versante più propriamente sociolinguistico possiamo invece citare i lavori di Paulis, Pinto & Putzu (2013) e Rattu (2017), in cui vengono forniti dei dati riguardanti il repertorio sociolinguistico di Cagliari.

Il dialetto cagliaritano può essere definito come una varietà a rischio di estinzione (Loporcaro, Putzu, 2013: 205), visto che attualmente i suoi parlanti sono molto pochi sia tra le fasce più giovani della popolazione sia tra quelle più anziane. Al momento non esistono informazioni precise e attendibili sul numero dei parlanti di questa varietà, ma solo qualche dato sui parlanti di sardo in generale. L'indagine condotta da Oppo (2007), che potremmo definire di carattere autovalutativo – visto che non prevedeva dei test linguistici atti a verificare la reale competenza linguistica dei parlanti – rivela che a Cagliari il 57,6% degli intervistati dichiara di parlare sardo. Tuttavia, lo studio più recentemente presentato in Paulis *et al.* (2013) – basato su un questionario che includeva anche degli esercizi volti ad accertare l'effettiva competenza attiva e passiva dei parlanti – riporta che, per quanto riguarda Cagliari, solamente il 31% degli informanti del campione esaminato ha dimostrato di saper parlare il sardo. Occorre aggiungere, tuttavia, che quest'ultima indagine ha registrato anche un 25% di persone intervistate che affermano di parlare il sardo pur dimostrando di non conoscerlo. La somma di queste due percentuali (31% e 25%) si avvicina al dato riportato da Oppo (2007), 57,6%, come è stato messo in rilievo da Pinto (2013). Pertanto, vale la pena notare che, nonostante il numero esiguo di dialettofoni, entrambe le ricerche sembrano evidenziare un atteggiamento positivo nei confronti dei dialetti sardi, poiché anche coloro che non sanno parlare il sardo affermano di conoscerlo. Questo cambiamento ideologico nei confronti dei dialettofoni e l'accresciuto valore dei dialetti risulta essere un fenomeno più generale che si è diffuso in Italia a partire dagli anni Novanta (Berruto, 2002; Dal Negro, Vietti, 2011).

Tenuto conto dunque dello statuto di questo dialetto quale varietà in via di estinzione e delle difficoltà incontrate durante la fase di reperimento dei parlanti, la scelta del contesto nel quale raccogliere la maggior parte dei dati è ricaduta su una confraternita religiosa, l''Arciconfraternita della Solitudine' – situata a Villanova, quartiere storico della città – in quanto rappresenta un punto di ritrovo che vanta una lunga tradizione in città e in cui, pertanto, si riteneva altamente probabile trovare informanti che fossero al contempo originari di Cagliari e dialettofoni.

Con 'Arciconfraternita della Solitudine' si fa riferimento a tre diverse realtà: la confraternita vera e propria, il coro dei cantori di San Giovanni e il circolo.

La confraternita è un sodalizio religioso nato nei primi anni del XVII secolo che si occupa dell'organizzazione di diversi riti religiosi, tra i quali quelli della Settimana Santa sono certamente i più importanti. La particolarità del contesto cagliaritano risiede nel fatto che questi riti sono compiuti contemporaneamente anche da un'altra confraternita, la 'Confraternita del Santissimo Crocifisso' (detta anche 'Confraternita di San Giacomo'), situata anch'essa nel quartiere di Villanova, nella Chiesa di San Giacomo. Entrambe le confraternite svolgono le stesse processioni negli stessi giorni e alla stessa ora, ma seguendo dei percorsi diversi. Questa 'condivisione' di compiti ha fatto sì che tra i due sodalizi si sia creata nel corso degli anni una forte rivalità.

Il coro dei cantori, detto 'Massa di San Giovanni', si occupa dell'esecuzione dei canti durante le soste della processione, mentre il circolo, situato sulla terrazza della chiesa, rappresenta un punto di ritrovo che permette ai confratelli di incontrarsi quotidianamente per scambiare due chiacchiere o bere qualcosa.

Questa confraternita può essere considerata una 'comunità di pratica' (Lave, Wenger, 1991; Wenger, 1998; Eckert, McConnell-Ginet, 1992; Eckert, 2000) perché soddisfa i tre criteri che la definiscono come tale: 1) impegno comune: i confratelli e le consorelle interagiscono regolarmente e prendono parte alle attività confraternali con diversi gradi di coinvolgimento; 2) obiettivo comune: i membri della confraternita hanno il compito di organizzare i riti religiosi; 3) repertorio condiviso: tutti i componenti condividono pratiche sociali, norme linguistiche, credenze e attitudini nei confronti della confraternita.

L'impiego di questo costrutto teorico risulterà molto più proficuo negli sviluppi futuri di questo lavoro, ovvero nel momento in cui verranno messi in relazione i ruoli svolti dai membri della confraternita con i *pattern* di variazione sociofonetica. In questa sede, il modello della comunità di pratica verrà sfruttato solamente durante la fase di analisi stilistica, ovvero in relazione agli argomenti trattati nelle conversazioni e, in particolare, alla correlazione esistente tra questi e l'uso delle diverse varianti sociofonetiche.

Considerato che l'analisi stilistica farà riferimento ai diversi approcci sociolinguistici allo stile del parlato, prima di affrontare lo studio sociofonetico, sembra opportuno illustrarli brevemente.

1.1 Modelli di variazione stilistica

In ordine cronologico, il primo modello proposto per l'analisi della variazione *intraspeaker* è quello laboviano, definito *attention paid to speech*, che è stato descritto e applicato a più riprese (Labov, 1966; 1972; 2001).

Partendo dall'assunto che l'intervista sociolinguistica è un evento in cui il parlato che viene elicitato è di tipo sorvegliato, Labov identifica il *careful speech* come il parlato usato dagli informanti nelle loro risposte alle domande dell'intervistatore, ovvero: «the type of speech which normally occurs when the subject is answering questions which are formally recognized as "part of the interview"» (Labov, 1966: 59).

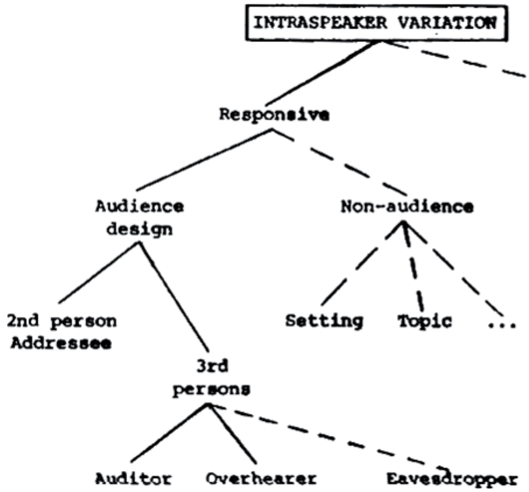
Dato per assodato questo, il sociolinguista americano nota come nell'intervista ci siano però alcune parti in cui è possibile rintracciare il *casual speech* (il parlato spontaneo); si tratta di quelle sezioni che si allontanano dallo schema tipico dell'intervista e che Labov riconduce a: a) il parlato che avviene fuori dall'intervista vera e propria, per esempio, all'inizio o alla fine della registrazione; b) il parlato che avviene con una terza persona; c) il parlato non in diretta risposta a una domanda (le digressioni); d) filastrocche e giochi legati all'infanzia; e) pericolo di morte.

Con lo strumento dell'intervista sociolinguistica, Labov tenta di catturare, complessivamente, cinque tipologie stilistiche, che dispone lungo un *continuum* sull'asse dello stile (*casual speech*; *careful speech*; *reading style*; *minimal pairs*; *word lists*). Secondo il *principio dell'attenzione* gli stili possono essere ordinati lungo un'unica dimensione, misurata dal grado di auto-monitoraggio al parlato. Gli stili più spontanei si posizionano a un'estremità del *continuum*, mentre quelli più sorvegliati all'estremo opposto (Labov, 1972: 112). Quanto maggiore sarà l'attenzione che un parlante rivolge al parlato, tanto più alto sarà il grado di formalità dello stile usato.

Il secondo modello che tenta di spiegare la variazione stilistica è il cosiddetto *audience design*, elaborato da Alan Bell (1984; 2001) e derivato dalla *Speech Accommodation Theory* (Giles, 1973; Giles, Powesland, 1975; Giles, 1980; Giles, Coupland & Coupland, 1991). Secondo Bell, la forza motrice che regola il cambiamento di stile non è l'attenzione al parlato, ma l'interlocutore. «*Audience Design* [...] assumes that persons respond mainly to other persons, that speakers take most account of bearers in designing their talk» (Bell, 1984: 159).

L'*Audience Design* si inserisce in un modello più ampio (Figura 1), all'interno del quale trovano posto anche i cosiddetti *non-audience factors*, come per esempio, *topic* e *setting*.

Figura 1 - *Schema rappresentante il modello Audience Design (Bell, 1984: 162)*



L'ipotesi di Bell è che anche il cambiamento di stile in funzione di questi fattori derivi dall'*audience design*:

> Speakers associate classes of topics *or* settings with classes of persons. They therefore shift style when talking on those topics or in those settings as if they were talking to addressees whom they associate with the topic or setting. [...] The basis of all style shift according to nonpersonal factors lies then in audience-designed shift (Bell, 1984: 181).

Inoltre, in questo modello anche il cambiamento di stile definito *initiative*, che prevede una componente agentiva da parte del parlante, può essere fatto rientrare all'interno del modello dell'*audience design*. Nella formulazione di Bell l'intero complesso dei cambiamenti di stile dipende dall'interlocutore: un parlante che prende l'iniziativa e ridefinisce la situazione attraverso il parlato sta ancora rispondendo all'interlocutore. Pertanto, il cambiamento di stile che nasce da iniziativa personale (*initiative shift*) rappresenta essenzialmente una ridefinizione da parte del parlante della sua relazione con l'interlocutore (Bell, 1984: 185).

L'ultimo modello a cui si accennerà è il cosiddetto *Speaker Design*, elaborato per la prima volta da Arnold *et al.* (1993), costituenti il cosiddetto *California Style Collective*. Prendendo le mosse dal lavoro sull'identità di Le Page, Tabouret-Keller (1985), questa formulazione teorica si fonda sull'idea del parlante visto come agente, ovvero come parte attiva nel processo di costruzione di specifiche identità sociali e non come un mero soggetto passivo che reagisce solamente ai fattori situazionali. In quest'ottica la variazione stilistica rappresenta una risorsa per costruire l'identità del parlante. Negli studi che si fondano sul modello *speaker design*, i *pattern* di variazione linguistica sono visti non come i riflessi di un'identità statica, in quanto definita dalla posizione occupata da un individuo nell'ordine sociale esistente e, dunque, dalla sua appartenenza a uno specifico gruppo sociale, ma rappresentano delle risorse che i parlanti usano per modellare e rimodellare sia le strutture sociali sia la loro posizione rispetto a queste strutture (cfr. Schilling-Estes, 2002). Pertanto, in questo tipo di approccio assume un ruolo fondamentale l'agentività del parlante.

2. *Metodologia della raccolta dati e costituzione del* corpus

Per quanto riguarda le registrazioni, la raccolta dei dati sul campo è stata effettuata impiegando un registratore Zoom H5, con una campionatura a 44,100 Hz e la digitalizzazione a 16-bit.

Lo strumento di elicitazione adottato è quello dell'intervista etnografica semi-strutturata (cfr. Vietti, 2003; Abete, 2012). Le interviste sono state svolte nella confraternita religiosa e si sono focalizzate su argomenti di interesse per i parlanti, come la storia delle confraternite, l'organizzazione dei riti sacri, la mancanza di servizi in città, la lingua sarda, la rivalità tra le confraternite. Il *corpus* di riferimento, che comprende sia interviste individuali sia conversazioni di gruppo, è costituito da circa 6 ore di parlato semi-spontaneo[1]. La semi-spontaneità del parlato dialogico elicitato è dimostrata da una serie di indizi significativi sia sul piano contenutistico sia su quello fonetico e conversazionale. Da una parte, infatti, nelle interviste troviamo la presenza del turpiloquio, insulti tra gli informanti, battibecchi e momenti di grande coinvolgimento emotivo; dall'altra, sono rintracciabili alcuni indizi conversazionali, come le esitazioni, le riformulazioni, le false partenze e le sovrapposizioni (cfr. Berruto, 1993), e altri segnali di tipo fonetico, come l'alta velocità di eloquio e la presenza cospicua di fenomeni di riduzione (per es. elisioni e assimilazioni).

Le interviste analizzate fanno parte di un *corpus* più ampio, di circa 10 ore di parlato, che ha coinvolto complessivamente 13 parlanti[2]. In questa sede verranno presi in considerazione solamente 7 informanti (5 uomini e 2 donne), di età compresa tra i 37 e i 74 anni, che rappresentano i parlanti appartenenti alla confraternita[3].

Le occorrenze sonore esaminate per l'analisi sociofonetica sono 879.

Per poter analizzare i *pattern* di distribuzione delle varianti nei diversi contesti comunicativi interni al macro-evento dell'intervista, e quindi per poter svolgere l'analisi stilistica, le conversazioni sono state pianificate in modo da ottenere risposte omogenee in ogni registrazione. In questo modo, nella fase di analisi è stato possibile prendere in esame gli stessi argomenti e micro-generi e avere a disposizione una stessa gamma di categorie stilistiche per tutti i parlanti[4]. È attraverso questo stratagemma che si è cercato di catturare registri più o meno sorvegliati o, in termini più generali, segmenti di intervista in cui il controllo del parlato lasciava spazio a una maggiore spontaneità (come per esempio avviene nel racconto di aneddoti particolarmente coinvolgenti per il parlante e nel dialogo tra informanti).

In aggiunta a ciò, la tipologia di intervista adottata ha permesso di affrontare lo studio del parlato anche in funzione del diverso interlocutore, in quanto il *corpus*, come è stato già

[1] La predilezione dell'etichetta 'parlato semi-spontaneo' al posto di quella di 'parlato spontaneo' deriva dalla modalità di elicitazione del materiale sonoro, raccolto tramite interviste svolte a microfono palese, che non ci permette di considerare le conversazioni registrate come interazioni naturali a tutti gli effetti.

[2] L'intero *corpus* di parlato è stato raccolto secondo le medesime modalità di escussione.

[3] I restanti soggetti del campione non appartengono al sodalizio confraternale.

[4] La particolare tipologia di dati non ha permesso di avere, sin dal principio, una quantità di parlato uguale per ogni parlante. Pertanto, per rendere il *corpus* omogeneo è stato creato un sottocorpus, per il quale sono stati selezionati in ciascuna intervista, per ogni parlante, tre minuti di parlato per ogni categoria stilistica. Così facendo si è ottenuta una quantità pressoché simile di parlato per tutti gli informanti, bilanciata sulla base del diverso argomento trattato.

anticipato, è costituito sia da interviste singole sia da interviste di gruppo, alle quali hanno preso parte due o più informanti. «Changing the dynamics of the interview away from the one-on-one format can also facilitate the production of casual speech» (Milroy, Gordon, 2003: 66). Nelle interazioni di gruppo l'intervista si è talvolta trasformata in una reale conversazione tra i locutori, in cui il ruolo dell'intervistatrice è stato quasi assente. Questo contesto ha facilitato la produzione del parlato spontaneo, quello che in termini laboviani è definito *casual speech* (Labov, 1966: 59).

3. *Descrizione della variabile sociofonetica oggetto di studio*

Come anticipato, la variabile sociofonetica presa in esame è la realizzazione della fricativa alveolare sorda e sonora /s, z/ come una postalveolare sorda [ʃ], in posizione preconsonantica (anche al confine di parola): es. *is cassonetus* [iʃ kassɔˈnɛtuzu] 'i cassonetti' (Figura 2), *prus mannu* [pruʃ ˈmannu] 'più grande' (Figura 3).

Questo processo fonetico è chiamato anche */s/-retraction* e sembra essere favorito dai contesti preconsonantici, visto che alcuni studi sulla produzione rivelano che /s/ ha un centro di gravità più basso quando precede /t/ rispetto a quando precede una vocale (Baker, Archangeli & Mielke, 2011; Iskarous, Shadle & Proctor, 2011; Stevens, Bukmaier & Harrington, 2015).

Figura 2 - *Spettrogramma di una realizzazione di* is cassonetus *[iʃ kassɔˈnɛtuzu] 'i cassonetti'*

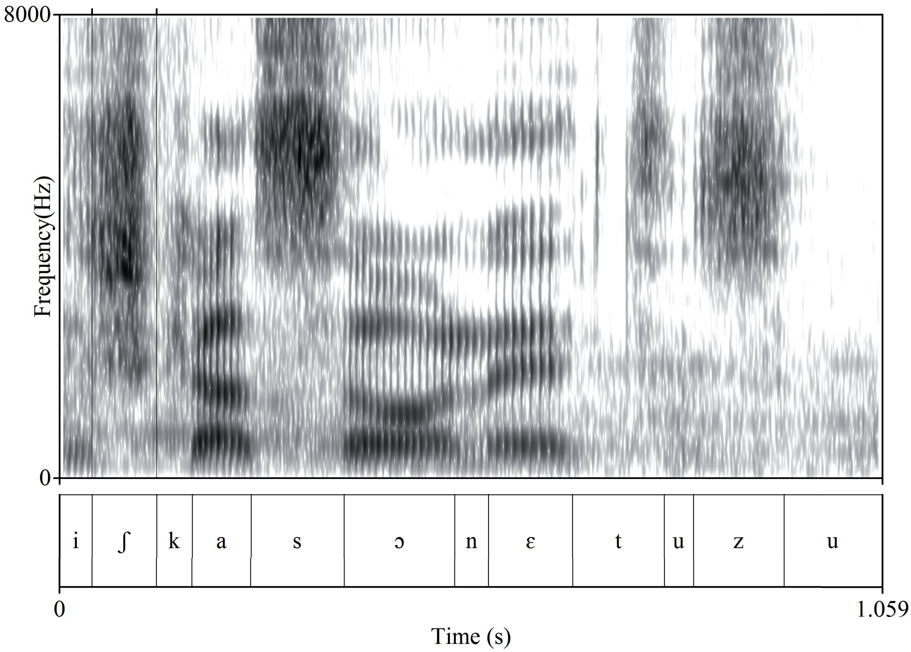
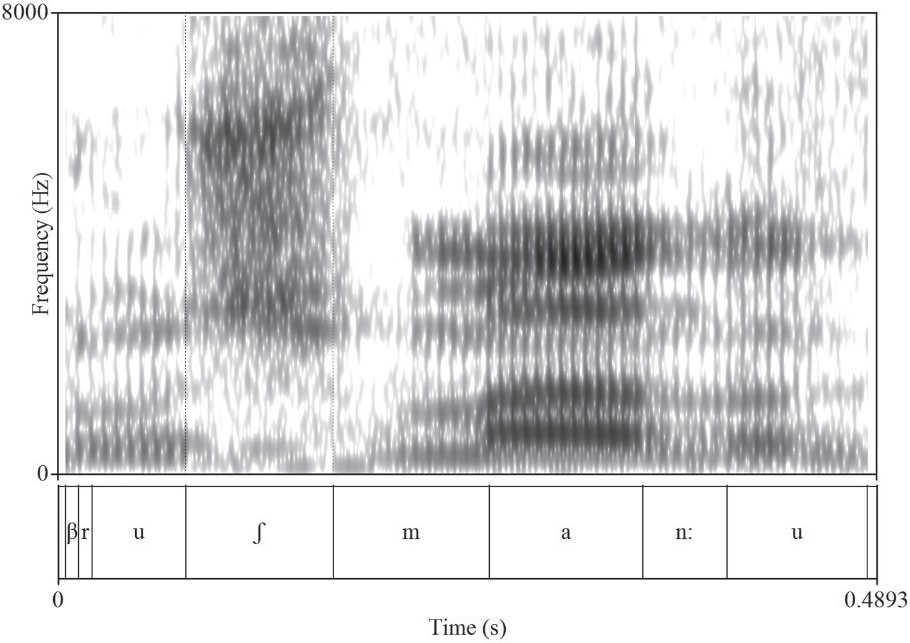


Figura 3 - *Spettrogramma di una realizzazione di* prus mannu *[pruʃˈmannu] 'più grande'*



Per quanto riguarda la variante postalveolare [ʃ], questa risulta essere un suono sordo con un grado variabile di sonorità anche quando essa è seguita da una consonante sonora e perciò ci aspetteremmo al suo posto una fricativa postalveolare sonora [ʒ]. Per verificare la proporzione di sonorità nelle fricative precedenti una consonante sonora è stata utilizzata la funzione *Voice Report* di Praat (Boersma, Weenink, 2017).

Seguendo la classificazione suggerita da Davidson (2015), ogni suono è stato classificato come sonoro se più del 90% del segmento è stato identificato da *Voice Report* come sonoro; sordo, se meno del 10% della costrizione è risultato sonoro; parzialmente sonoro, se la parte sonora della costrizione era compresa tra il 10% e il 90%. Le realizzazioni della fricativa postalveolare che precedono un suono sonoro si sono rivelate tutte parzialmente sonore. Per quanto riguarda l'esempio di *prus mannu* (Figura 3), *Voice Report* ha misurato una percentuale di sordità (*fraction of locally unvoiced frames)* dell'87,5%.

4. *Analisi sociofonetica*

4.1 Segmentazione ed etichettatura

Tutti i file sonori sono stati segmentati usando il *software* Praat. In particolare, la segmentazione delle occorrenze è stata condotta manualmente, tenendo conto sia della forma d'onda sia dello spettrogramma a banda larga, secondo i criteri suggeriti da Jongman, Wayland & Wong (2000) e Nissen & Fox (2005). Nello specifico, l'*on-*

set della fricativa è stato considerato il punto in cui il numero di *zero crossings* cresce rapidamente. Per segnalare l'*offset* del segmento, si è fatto riferimento all'intensità minima immediatamente precedente il silenzio associato alla chiusura dell'occlusiva seguente. Nel caso in cui il suono seguente fosse un'altra fricativa si è cercato di individuare il punto in cui si rintracciava simultaneamente una modificazione della forma d'onda e dell'energia nello spettrogramma.

Nella prima fase dell'analisi, sono state identificate, su base uditivo-percettiva e spettrografica, quattro diverse varianti della variabile studiata: le due varianti alveolari, sorda e sonora, [s] e [z] (varianti standard); una variante intermedia tra [s] e [ʃ], che potremmo definire una sibilante alveolare arretrata [s̠], e la variante postaveolare [ʃ] (variante sub-standard).

4.2 Analisi acustica

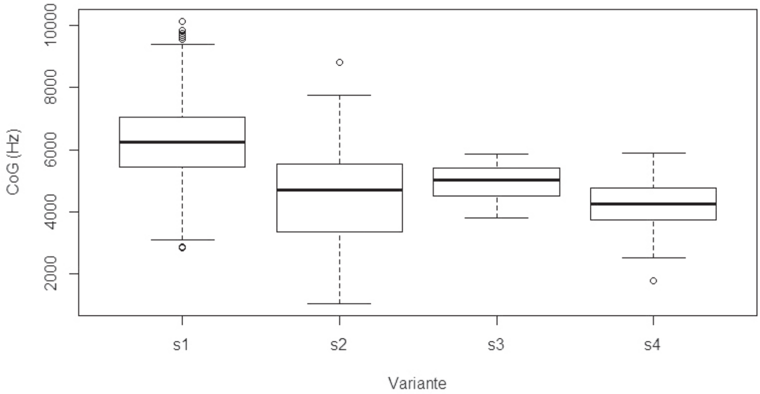
Dopo aver etichettato le occorrenze, è stata condotta l'analisi acustica, il cui fine è stato quello di confermare la precedente categorizzazione delle varianti. Per questa fase della ricerca, si è ricorso all'analisi dei momenti spettrali, uno dei possibili metodi usati per lo studio delle fricative (Forrest *et al.*, 1988; Nittrouer, 1995; Shadle, Mair, 1996; Jongman *et al.*, 2000; Munson, 2001; Munson, 2004; Nissen, Fox, 2005; Stuart-Smith, 2007; Spinu, Lilley, 2016). In particolare, sono stati estratti automaticamente i primi quattro momenti spettrali – centro di gravità (CoG), deviazione standard, obliquità e curtosi – dalle occorrenze segmentate per mezzo dello *script* elaborato da Christian Di Canio nel 2013, al quale sono state apportate alcune modifiche[5]. Tuttavia, considerato che le diverse varianti sono delle sibilanti che differiscono tra di loro per il luogo di articolazione, è stato usato solamente il primo momento spettrale (CoG), che rappresenta un indizio significativo del luogo di articolazione perché fornisce informazioni sulla concentrazione di energia acustica (Baum, McNutt, 1990; Jesus, Shadle, 2002; Munson, 2001; Baker *et al.*, 2011; Stevens *et al.*, 2015). Data la correlazione esistente tra il grado di anteriorità della costrizione e il CoG, il CoG di /s/ dovrebbe essere più alto di quello di /ʃ/.

Prima dell'estrazione dei momenti spettrali, un filtro passa-alto (*high pass filter*) di 500 Hz è stato applicato in modo che fosse possibile confrontare le fricative sorde e sonore, così come suggerito da Munson (2001) e Stuart-Smith (2007).

Il *boxplot* di Figura 4 riporta i risultati dei valori del CoG. È possibile notare che il CoG della fricativa alveolare sorda (s1) è più alto di quello della fricativa postalveolare (s4) (differenza statisticamente significativa al t-test, p-value < 0.001) ed esiste un'altra variante che si posiziona tra la variante standard e quella sub-standard.

[5] Lo script è reperibile all'indirizzo: http://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives_2.0.praat. Data la forte variabilità relativa alla durata delle fricative analizzate e, in particolar modo, la presenza di segmenti molto brevi, i valori acustici sono stati estratti da tre finestre d'analisi (*Window number*: 3) di 10 ms (*Window size*: 0.010).

Figura 4 - *Boxplot raffigurante i valori del CoG estratti dalle produzioni dei parlanti uomini. Legenda: s1: [s]; s2: [z]; s3: [s̠]; s4: [ʃ]*



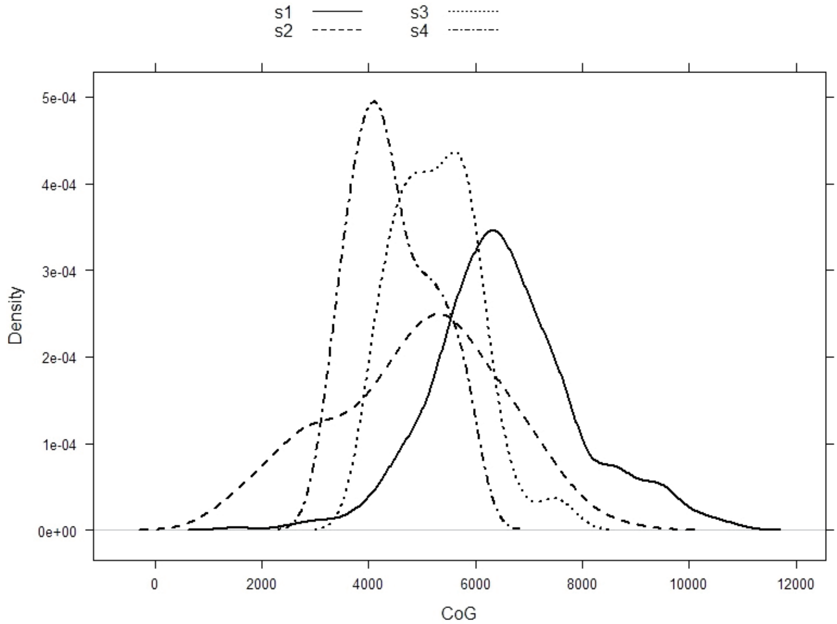
I valori raffigurati nel *boxplot* rappresentano solamente quelli prodotti dai parlanti uomini. Infatti, non avendo normalizzato i dati, si è preferito mantenere separati i valori di CoG relativi alle produzioni maschili dai valori delle produzioni femminili. Inoltre, visto che nel campione di riferimento le donne non producono realizzazioni postalveolari, sono stati riportati esclusivamente i valori maschili.

Riteniamo possa essere utile esplicitare la motivazione che soggiace alla decisione di non normalizzare i dati.

La discussione sulle differenze acustiche esistenti nella produzione delle fricative tra parlanti uomini e donne è al centro di numerosi studi. In particolare, uno degli aspetti più interessanti, dal nostro punto di vista, è il peso che in questa variabilità presentano rispettivamente i fattori di tipo anatomico-biologici e quelli sociofonetici. In estrema sintesi, diremo che le produzioni di /s/ mostrano frequenze più alte nelle donne rispetto agli uomini ed esistono degli studi che dimostrano la validità delle spiegazioni biologiche come base di questa diversità (Schwartz, 1968; Johnson, 1991). D'altra parte, è anche vero che parlanti uomini e donne possono fare ricorso a strategie differenti per articolare /s/ come parte di un processo di rappresentazione dell'identità di genere all'interno di una particolare cultura o società (Stuart-Smith, Timmins & Wrench, 2003; Munson, McDonald, DeBoe & White, 2006; Fuchs, Toda, 2010; Czaplicki, Żygis, Pape & Jesus, 2016). È chiaro quindi che con la normalizzazione dei dati si rischierebbe di perdere anche informazioni sull'identità sociale dei parlanti. Per questi motivi, si è preferito non normalizzare i dati e mantenere distinti i valori acustici per gli uomini e per le donne.

Soffermandoci ancora sul *boxplot* riportato, possiamo osservare come, nonostante sia stato applicato un filtro, i valori del CoG di /z/ siano più bassi rispetto a quelli di /s/. Può essere interessante dunque indagare più da vicino l'andamento dei valori ottenuti dal gruppo di occorrenze riconducibili a /z/.

Figura 5 - *Distribuzione di probabilità del CoG per le produzioni dei parlanti uomini. Legenda: s1: [s]; s2: [z]; s3: [s̠]; s4: [ʃ]*



Questo grafico (Figura 5) ci fornisce una rappresentazione più dettagliata della distribuzione dei valori del CoG delle varianti. Se ci focalizziamo ora solamente sui valori di /z/ (s2), notiamo come sia possibile individuare due diversi picchi, o meglio, una 'gobba', a frequenze più basse, tra i 2000 e i 3000 Hz (a sinistra), e un picco, a frequenze più alte, situato a circa 5500 Hz (a destra).

Il particolare andamento del CoG di /z/, durante la fase di analisi, ha fatto ipotizzare che dietro la concentrazione di valori a basse frequenze potesse celarsi qualche altra variante non prevista. Questa ipotesi è stata rafforzata dal fatto che durante la fase di segmentazione e di etichettatura un certo numero di occorrenze di /z/ sembrava mostrare caratteristiche spettroacustiche riferibili più a delle realizzazioni approssimanti che non a delle fricative. Date queste premesse, si è pensato pertanto che potesse essere utile verificare, per le realizzazioni sonore che erano state etichettate come [z], i valori dell'*Harmonics-to-Noise Ratio* (HNR), un indice che rappresenta la quantità di periodicità del segnale, mediante una funzione di Praat. I risultati dell'analisi hanno evidenziato che effettivamente le produzioni identificate inizialmente come [z] e che mostravano un CoG molto basso si caratterizzavano anche per un HNR molto alto, ovvero una grande quantità di periodicità. Sembra pertanto plausibile considerare queste occorrenze delle approssimanti e non delle fricative. La correlazione tra CoG basso e HNR alto ha permesso di spiegare i valori così bassi del gruppo di occorrenze i cui valori di CoG sono riferibili alla 'gobba' raffigurata nel grafico. La presenza di realizzazioni prodotte con approssimazione invece che con frizione può essere ricondotta alla natura dei dati analizzati, un parlato connesso semi-spontaneo.

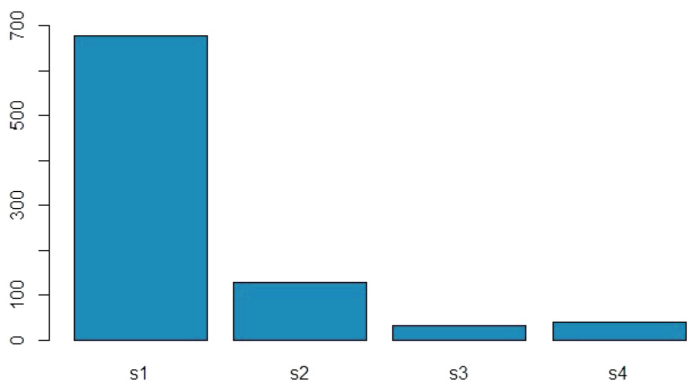
Tuttavia, dato l'obiettivo di tipo sociofonetico di questo contributo, si preferisce rimandare un'analisi più approfondita della questione a un'altra sede.

Proseguendo con l'illustrazione dell'andamento dei valori del CoG per le diverse varianti, appare invece più opportuno giustificare la necessità dell'inserimento di una variante intermedia (s3). La decisione di mantenere questa variante e di non ridistribuire le sue occorrenze nelle altre due varianti (alveolare e postalveolare) è da ricondurre a due motivi principali, uno di tipo metodologico e l'altro più propriamente analitico. Per quanto riguarda la ragione metodologica, visto che si tratta di una variante individuata durante la fase di etichettatura, avvenuta per mezzo dello strumento uditivo-percettivo e di quello spettroacustico, questa variante intermedia rappresenta un indizio delle difficoltà che si sono incontrate nella fase di classificazione delle diverse varianti e, come tale, dimostra la forte variabilità presente. Inoltre, e qui passiamo alla motivazione di tipo analitico, i valori del CoG sembrano confermare questa scelta. Infatti, l'andamento del CoG mostra due picchi diversi e, se è vero che quello a frequenze più basse (a sinistra) potrebbe essere incluso nella curva della variante postalveolare, è altrettanto vero che il picco a frequenze un po' più alte (a destra) risulta essere proprio una via intermedia tra le due varianti e, in quanto tale, giustifica la scelta di mantenere la variante s3.

4.3 Analisi sociolinguistica

Il ridotto numero di occorrenze attestate per [ʃ] non consente al momento di proporre delle ipotesi valide sui vincoli linguistici che regolano il comportamento di questa variabile; tuttavia, possiamo invece fornire qualche spiegazione proprio sulla quantità esigua di queste produzioni rispetto a quelle delle varianti standard.

Figura 6 - *Diagramma a barre raffigurante la distribuzione numerica delle diverse varianti. Legenda: s1: [s]; s2: [z]; s3: [s̠]; s4: [ʃ]*

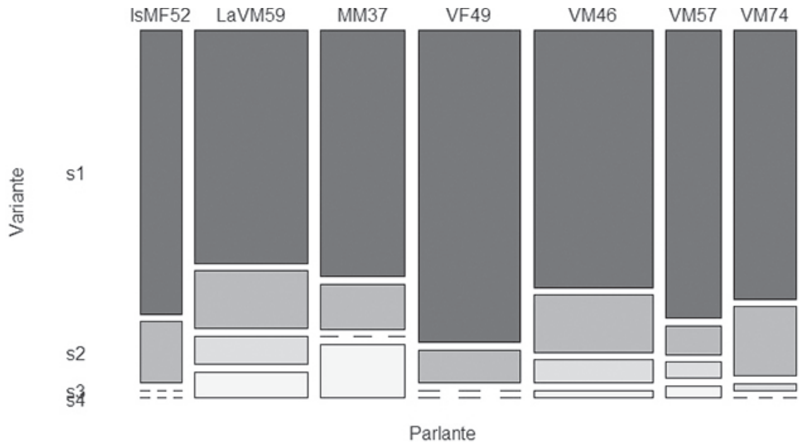


Come è possibile osservare dal grafico in Figura 6, che rappresenta la distribuzione numerica delle diverse varianti, l'arretramento di /s/ è un fenomeno molto marginale (s1: 678; s2: 128; s3: 33; s4: 40).

Si tratta di un tratto altamente stigmatizzato, socialmente e stilisticamente marcato verso il basso, associato a parlanti appartenenti a un gruppo socio-culturale basso e a registri molto informali. Dato che questo fenomeno è oggetto di commenti aperti da parte dei parlanti e che di frequente sui *social network* viene reso graficamente con il grafema <sh> o <sch>, può essere considerato uno stereotipo in senso laboviano (Labov, 1994: 78).

Per maggiore chiarezza, sarà utile mostrare anche come le diverse varianti sono distribuite tra i parlanti (Figura 7).

Figura 7 - *Diagramma a mosaico raffigurante la distribuzione delle diverse varianti tra i parlanti. Legenda: s1: [s]; s2: [z]; s3: [s̠]; s4: [ʃ]*



Un primo dato che salta subito agli occhi è la totale assenza di varianti sub-standard nelle due donne (IsMF52 e VF49)[6].

Nel gruppo di informanti donne dell'intero campione di parlanti intervistati (in totale 4) non si hanno realizzazioni di questo tipo, nemmeno nei contesti comunicativi più informali. Sebbene il numero dei locutori non sia sufficientemente elevato, soprattutto il numero delle informanti donne[7], e il campione non sia bilanciato, questo dato merita comunque di essere approfondito.

Dai pochi dati che si hanno a disposizione il *pattern* di variazione emerso in relazione al genere sembrerebbe rispecchiare la tendenza registrata in molti lavori di sociolinguistica (soprattutto di stampo anglosassone) secondo cui le donne mostrano un atteggiamento più normativo degli uomini nei confronti della lingua.

[6] I nomi dei parlanti sono composti dalla sigla del quartiere di origine (IsM: Is Mirrionis, LaV: La Vega; M: Marina; V: Villanova), seguita dal genere e dall'età del parlante.

[7] Il ristretto numero di donne presenti nel campione è frutto delle difficoltà incontrate durante la raccolta dati nel reperimento di parlanti dialettofoni.

Riassumendo i risultati delle indagini che hanno riguardato la variazione di genere, Labov (1990) riconduce l'intero quadro a due principi generali. Il primo afferma che per variabili sociolinguistiche stabili, cioè che durano nel tempo, gli uomini usano con maggiore frequenza rispetto alle donne le varianti non standard. Pertanto, le donne risultano essere più sensibili degli uomini alle forme linguistiche di prestigio.

Un corollario del primo principio enuncia che nei cambiamenti dall'alto, ovvero quando i mutamenti avvengono al di sopra del livello di consapevolezza (e talvolta mostrano come stabili variabili sociolinguistiche degli stereotipi) le donne favoriscono più degli uomini le forme di maggiore prestigio.

Infatti, poiché i cambiamenti dall'alto condividono molte delle proprietà delle variabili sociolinguistiche stabili, non sorprende che il ruolo dei sessi sia simile e che le donne guidino sia l'acquisizione di nuovi *pattern* di prestigio sia l'eliminazione delle forme stigmatizzate (Labov, 1990: 213).

Il secondo principio formulato da Labov afferma invece che nei cambiamenti dal basso le donne rappresentano spesso le innovatrici.

Come conseguenza di questi principi viene a crearsi il cosiddetto 'paradosso del genere' (*gender paradox*) che consiste in una maggiore adesione alla norma linguistica e in un uso minore di deviazioni da parte delle donne, quando queste deviazioni sono apertamente censurate, mentre mostrano un maggior numero di deviazioni rispetto agli uomini quando queste non sono censurate (Labov, 2000: 40).

I risultati dello studio qui presentato si conciliano con le assunzioni proposte da Labov: anche in questo caso le donne rifiutano le forme sub-standard perché altamente stigmatizzate e usano le varianti standard.

A livello macro-sociolinguistico l'utilizzo delle varianti più prestigiose da parte delle donne si riflette nella scelta della varietà standard, l'italiano, a scapito di quella non standard, il sardo. La predilezione da parte delle donne dell'italiano come principale codice di comunicazione è testimoniata dalla difficoltà incontrata nel reperire parlanti donne che potessero far parte del campione.

4.3.1 Analisi stilistica

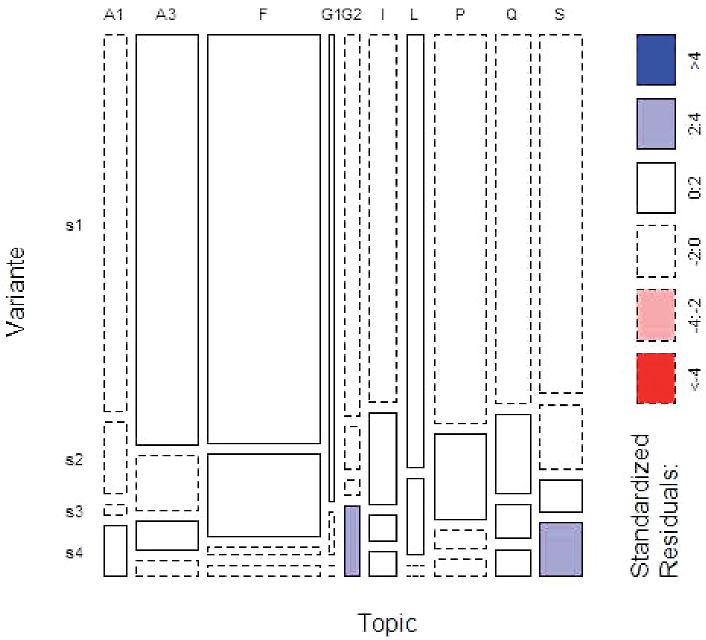
La pianificazione delle interviste secondo i criteri esposti al par. 2 ha permesso di affrontare lo studio della variazione stilistica e di esplorare come i parlanti usano le diverse varianti sociofonetiche nel loro parlato, in funzione sia del diverso argomento sia del diverso interlocutore.

I risultati dell'analisi sono illustrati nel diagramma a mosaico con l'indicazione dei valori residuali (Figura 8). Si tratta di un grafico creato sulla base di un test chi quadrato e i residui standardizzati indicano quanto distano fra loro i valori attesi e quelli osservati per ogni cella, nel particolare modello che si è assunto. Pertanto, i residui ci segnalano quali sono le celle che guidano la mancanza di adattamento[8] al modello di indipendenza. In questo caso, i valori residuali indicano come celle in cui gli scarti tra le frequenze attese e quelle osservate sono significativamente diversi da

[8] In particolare, le celle con i colori verso il blu e i confini continui indicano un'associazione positiva, mentre le celle con i colori verso il rosso e i confini tratteggiati segnalano un'associazione negativa.

zero quelle relative alle categorie 'Dialogo tra informanti' (G2) e 'Rivalità tra confraternite' (S). In altre parole, la variante sub-standard (s4) mostra una frequenza d'uso maggiore in questi due segmenti di parlato.

Figura 8 - *Diagramma a mosaico con l'indicazione dei valori residuali raffigurante la distribuzione delle varianti (s1: [s]; s2: [z]; s3: [s̠]; s4: [ʃ]) per categoria stilistica. Legenda: A1: aneddoto divertente; A3: aneddoto personale; F: descrizione; G1: dialogo con l'intervistatore; G2: dialogo tra informanti; I: lamentela; L: lingua sarda; P: problemi del quartiere; Q: quartiere; S: rivalità tra confraternite*



Proviamo a capire quali possono essere le motivazioni di questa distribuzione, partendo dalla categoria 'Rivalità tra confraternite' (S).

L'osservazione etnografica partecipante, approccio metodologico che ha fatto da guida durante la raccolta dati, ha permesso di capire che questo argomento è di particolare interesse per i componenti della confraternita.

Durante la ricerca sul campo è stata scoperta infatti l'esistenza dell'altra confraternita presente nello stesso quartiere che assolve gli stessi compiti dell''Arciconfraternita della Solitudine'. Entrambe le confraternite accompagnano in cattedrale (secondo le stesse modalità e negli stessi giorni) i simulacri religiosi ed entrambe le associazioni hanno al loro interno un gruppo di cantori che si occupa dei canti religiosi. Pertanto, i fedeli devono scegliere quale delle due processioni seguire. Per questo motivo tra le due confraternite si è venuta a creare una forte rivalità, che dà vita a discussioni su quale delle due sia la più antica e quindi anche quella che detiene il diritto di portare avanti le tradizioni religiose della città.

Tale argomento suscita negli intervistati un grande coinvolgimento emotivo ed è proprio nei segmenti di parlato incentrati su questo tema che viene prodotta una maggiore quantità di varianti postalveolari.

Altro contesto comunicativo che mostra una produzione consistente delle varianti marcate è il 'Dialogo tra informanti', ovvero la conversazione che si svolge tra i partecipanti all'interazione, in quei momenti durante i quali la struttura dell'intervista viene scardinata e diventa un vero e proprio dialogo tra i partecipanti, in cui il ruolo dell'intervistatrice è pressoché assente.

Mentre l'argomento 'Rivalità tra confraternite' si inserisce nelle sezioni canoniche dell'intervista, ovvero nelle risposte che gli informanti danno alle domande poste dalla ricercatrice, il 'Dialogo tra informanti' rientra in quelle parti non riconducibili allo schema domanda-risposta.

Se nel caso della rivalità tra confraternite possiamo parlare di variazione stilistica in funzione dell'argomento, in questo caso è bene parlare di cambio di stile in funzione dell'interlocutore. Nel primo caso, infatti, quando si analizza la variazione in funzione dell'argomento, l'interlocutore è fisso e ciò che cambia è l'argomento. Nel secondo caso, invece, è il cambio di interlocutore ad acquistare rilevanza: gli informanti tendono a utilizzare maggiormente la variante sub-standard quando conversano con gli altri partecipanti all'intervista rispetto a quando parlano con l'intervistatrice.

4.3.2 Discussione dei dati

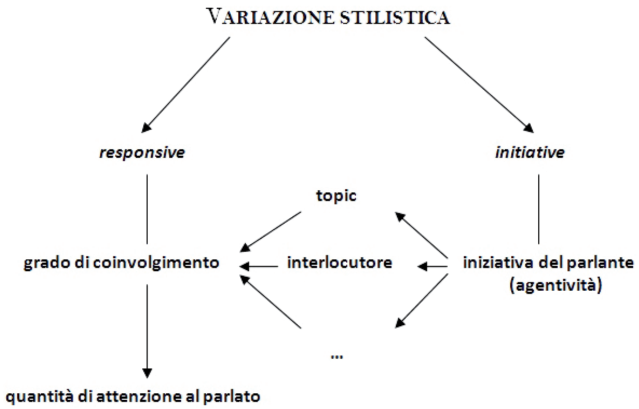
Dopo aver presentato i risultati ottenuti, si mostrerà ora come questi dati possono trovare spazio all'interno delle teorie descritte al par. 1.1. Provando ad applicare i tre modelli ai risultati di questa analisi, è stato elaborato un nuovo schema rappresentante la variazione stilistica. Preme sottolineare che il modello che si proporrà non ha alcuna pretesa di generalizzazione, né intende rendere conto dei meccanismi di variazione stilistica *tout court*, ma risponde solamente a esigenze analitiche ed è stato formulato sulla base dei dati risultanti da questa indagine.

In linea generale, i risultati ottenuti sembrano potersi inserire nello schema laboviano dell'*attention paid to speech* perché l'andamento dei *pattern* di variazione delle varianti sembra disporsi sull'asse della quantità di attenzione al parlato. In generale, quanto più un informante presta attenzione al parlato tanto più risulta essere sorvegliato lo stile, mentre se l'attenzione è rivolta ai contenuti affrontati, il locutore tende a dedicare meno attenzione al modo di parlare. Nei parlanti intervistati gli argomenti in cui maggiormente vengono usate le varianti sub-standard sono quelli nei quali è presente un alto grado di attenzione nei confronti del contenuto, a scapito del parlato stesso. Facendo un passo indietro, possiamo chiederci a cosa sia dovuto il maggiore o minore grado di auto-monitoraggio. Sembra chiaro che maggiore è il coinvolgimento degli informanti nei confronti di un argomento, minore è l'attenzione che si ha quando se ne parla. Così, nell'argomento 'Rivalità tra confraternite', il coinvolgimento è molto alto e il grado di attenzione al modo di parlare è minore, perché sono altre le componenti che in quel momento governano il discorso (es. componenti di tipo emotivo, quali emozione, rabbia, commozione).

Anche il dialogo tra informanti sembra potersi inserire in questo modello perché il grado di attenzione che si ha con persone con cui si è in confidenza è certamente più basso rispetto a quello richiesto in una conversazione con un estraneo.

Un dato che rafforza la presupposizione che a governare la variazione stilistica sia la quantità di attenzione al parlato è la presenza delle varianti marcate soprattutto nelle parti conclusive dell'intervista. In più, il *corpus* analizzato ha registrato alcune occorrenze di un'altra variabile altamente stigmatizzata, che qui non verrà presa in esame (la palatalizzazione dell'occlusiva velare /k, ɡ/ di fronte ad /a/, e.g. *cani* [ˈkʲani] 'cane', *gatu* [ˈɡʲatu] 'gatto'), anche nel dialogo con l'intervistatrice, ma esclusivamente nelle parti finali dell'intervista, dopo oltre un'ora di conversazione, ovvero quando l'attenzione non è più alta come nelle fasi iniziali. L'analisi dei dati, supportata da riflessioni teoriche, ci ha condotto a elaborare un altro modello, a partire da quello elaborato da Bell (1984) e diverso da quelli finora proposti, all'interno del quale i *pattern* di variazione riscontrati sembrano trovare una spiegazione più soddisfacente (Figura 9).

Figura 9 - *Schema proposto per spiegare la variazione stilistica della variabile sociofonetica studiata nel corpus di riferimento*



A differenza dello schema di Bell, questo modello assume un diverso punto di vista nei confronti dei fattori che regolano la variazione: l'interlocutore è posizionato sullo stesso piano del *topic* e costituisce *una* delle variabili che influenzano la variazione stilistica. Interlocutore e *topic* suscitano nel parlante un certo grado di coinvolgimento e questo a sua volta determina un maggiore o minore grado di attenzione nei confronti del parlato.

L'influenza del *topic* sull'uso di specifiche varianti sociofonetiche è stato inoltre dimostrato negli ultimi anni da diversi studi (Becker, 2009; Lawson, 2009; Love, Walker, 2012; Hay, Foulkes, 2016).

Oltre a questa componente 'di risposta', un modello di variazione stilistica dovrebbe prevederne anche una più attiva (*initiative*): la variazione stilistica non è da interpretare solo come il prodotto di una reazione del parlante ai diversi stimoli ai quali è sottoposto da parte del ricercatore o da parte delle situazioni circostanti, ma può essere anche frutto di una presa di iniziativa del parlante, di una componente decisionale, agentiva (cfr. Coupland, 1980; Arnold *et al.*, 1993). Assumiamo quindi che il parlante, in quanto agente linguistico, possa sfruttare le proprie risorse linguistiche non solo in risposta a qualcosa ma anche come mezzo per dare inizio a qualcosa, per esempio per costruire identità sociali, così come previsto dal modello *Speaker Design* (Arnold *et al.*, 1993), o semplicemente per enfatizzare un concetto che si vuole esprimere.

Come è evidente, lo schema proposto condivide con quello di Bell (1984), precedentemente presentato (Figura 1), le due componenti *responsive* e *initiative*, ma sostituisce alla linea tratteggiata la linea continua, in quanto il cambiamento di stile *initiative* è ritenuto autonomo e non dipendente dall'*audience design*.

5. *Conclusioni*

Benché lo studio presentato costituisca un'analisi preliminare, dai risultati emersi si può provare a fare qualche osservazione generale sia sulla bontà del metodo etnografico adottato sia su aspetti riguardanti il comportamento della variabile analizzata.

L'analisi sociofonetica ha dimostrato come il lavoro etnografico sul campo consenta di individuare le categorie locali che meritano di essere tenute in considerazione durante l'analisi (cfr. Eckert, 2000: 69). Nel caso specifico analizzato, l'osservazione etnografica partecipante ha permesso di spiegare perché alcuni argomenti risultino essere i favoriti per l'uso di una particolare variante (come nel caso dell'argomento 'Rivalità tra confraternite').

Infatti, i risultati dell'analisi stilistica hanno sottolineato l'attivazione di una particolare variante, [ʃ], in relazione a specifici segmenti di parlato. Partendo da questi dati e attingendo ai principali modelli teorici di variazione stilistica, è stata infine proposta una nuova direzione interpretativa, finalizzata a dare conto dei meccanismi soggiacenti ai *pattern* di variazione stilistica della variabile sociofonetica oggetto di studio, riscontrati all'interno del *corpus* analizzato.

Ringraziamenti

Questa ricerca non sarebbe stata possibile senza la partecipazione degli informanti, che ringrazio per il tempo che mi hanno dedicato e per la loro disponibilità.

Desidero inoltre ringraziare Alessandro Vietti per gli utili suggerimenti e i preziosi commenti a questo lavoro.

Riferimenti bibliografici

Abete, G. (2012). Aspetti metodologici per lo studio della variazione fonetica nel parlato dialettale. In Bianchi, P., De Blasi, N., De Caprio, C. & Montuori, F. (Eds.), *La variazione nell'italiano e nella sua storia. Varietà e varianti linguistiche testuali.* Firenze: Franco Cesati Editore.

Arnold, J., Blake, R., Eckert, P., Iwai, L., Mendoza-Denton, N., Morgan, C., Polanyi, L., Soloman, J. & Veatch, T. (1993). Variation and Personal/Group Style. Paper presented at New Ways of Analysing Variation 22. Ottowa, Ontario.

Atzori, M.T. (1986). Cagliari e il suo dialetto. In AA. VV. *Dialettologia urbana: problemi e ricerche.* Pisa: Pacini, 1-4.

Baker, A., Archangeli, D. & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. In *Language Variation and Change,* 23, 347-374.

Baum, S.R., McNutt, J.C. (1990). An acoustic analysis of frontal misarticulation of /s/ in children. In *Journal of Phonetics*, 18, 51-63.

Becker, K. (2009). /r/ and the construction of place identity on New York City's Lower East Side. In *Journal of Sociolinguistics*, 13(5), 634-658.

Bell, A. (1984). Language Style as Audience Design. In *Language in Society I*, 13, 145-204.

Bell, A. (2001). Back in style: reworking audience design. In Eckert, P., Rickford, J.R. (Eds.), *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, 139-169.

Berruto, G. (1993). Varietà diamesiche, diastratiche, diafasiche. In Sobrero, A.A. (Ed.), *Introduzione all'italiano contemporaneo. La variazione e gli usi. Vol. 2.* Roma-Bari: Laterza, 37-92.

Berruto, G. (2002). Parlare dialetto in Italia alle soglie del Duemila. In Beccaria, G.L., Marello, C. (Eds.), *La parola al testo. Scritti per Bice Mortara Garavelli*. Alessandria: Edizioni dell'Orso, 33-49.

Blasco Ferrer, E. (1984). *Storia linguistica della Sardegna*. Tübingen: Niemeyer.

Boersma, P., Weenink, D. (2017). *Praat: Doing phonetics by computer*. http://www.praat.org/.

Chambers, J.K., Trudgill, P. (1980). *Dialectology.* Cambridge: Cambridge University Press.

Coupland, N. (1980). Style-shifting in a Cardiff Work-setting. In *Language in Society*, 9, 1-12.

Czaplicki, B., Żygis, C.B.M., Pape, D. & Jesus, L.M.T. (2016). Acoustic evidence of new sibilants in the pronunciation of young Polish women. In *Poznań Studies in Contemporary Linguisitic*, 52(1), 1-42.

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

Davidson, L. (2015). Patterns of voicing in American English voiced obstruents in connected speech. In The Scottish Consortium for ICPhS (Eds.), *Proceedings of the 18th International Congress of the Phonetic Sciences*. Glasgow, UK: The University of Glasgow. Paper number 84.

Dettori, A. (2002). La Sardegna. In Cortelazzo, M., Marcato, C., De Blasi, N. & Clivio, G.P. (Eds.), *I dialetti italiani*. *Storia*, *struttura*, *uso*. Torino: UTET, 898-958.

Eckert, P. (2000). *Linguistic Variation as Social Practice. The Linguistic Construction of Identity in Belten High*. Malden-Oxford: Blackwell.

Eckert, P., McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. In *Annual review of anthropology*, 21, 461-90.

Fontana, P. (1996). Tendenze evolutive foniche nel sardo cagliaritano. In *La grotta della vipera*, 74, 32-36.

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, 115-124.

Fuchs, S., Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In Fuchs, S., Toda, V. &. Żygis, M. (Eds.), *Turbulent sounds. An interdisciplinary guide*. Berlin: Mouton de Gruyter, 281-302.

Giles, H. (1973). Accent mobility: A model and some data. In *Anthropological Linguistics*, 15, 87-109.

Giles, H. (1980). Accommodation theory: Some new directions. In *York Papers in Linguistics*, 9, 105-136.

Giles, H., Coupland, J. & Coupland, N. (1991). *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

Giles, H., Powesland, P.F. (1975). *Speech Style and Social Evaluation*. London: Academic Press.

Hay, J., Foulkes, P. (2016). The evolution of medial (-t-) in real and remembered time. In *Language*, 92(2), 298-330.

Iskarous, K., Shadle, C. & Proctor, M.I. (2011). Articulatory-acoustic kinematics: The production of American English /s/. In *Journal of the Acoustical Society of America*, 129(2), 944-954.

Jesus, L.M.T., Shadle, C.H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. In *Journal of Phonetics*, 30, 437-464.

Johnson, K. (1991). Differential effects of speaker and vowel variability on fricative Perception. In *Language and Speech*, 34(3), 265-279.

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

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). Some principles of linguistic methodology. In *Language in Society*, 1, 97-120.

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

Labov, W. (1994). *Principles of Linguistic Change. Vol. 1: Internal Factors*. Oxford: Blackwell.

Labov, W. (2000). Resolving the gender paradox in the study of linguistic change. In Cipriano, P., d'Avino, R. & Di Giovine, P. (Eds.), *Linguistica storica e sociolinguistica*. Roma: Il Calamo, 35-43.

Labov, W. (2001). The anatomy of style-shifting. In Eckert, P., Rickford, J.R. (Eds.), *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, 85-108.

Lave, J., Wenger, E. (1991). *Situated Learning. Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lawson, R. (2009). Sociolinguistic constructions of identity among adolescent males in Glasgow. PhD Dissertation, University of Glasgow.

Le Page, R.B., Tabouret-Keller, A. (1985). *Acts of Identity: Creolebased Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.

Loporcaro, M., Putzu, I.E. (2013). Variation in auxiliary selection, syntactic change, and the internal classification of Campidanese Sardinian. In Paulis, G., Pinto, I. & Putzu, I.E. (Eds.), *Repertorio plurilingue e variazione linguistica a Cagliari*. Milano: Franco Angeli, 200-244.

Love, J., Walker, A. (2012). Football versus football: Effect of topic on /r/ realization in American and English sports fans. In *Language and Speech*, 56(4), 443-460.

Milroy, L., Gordon, M. (2003). *Sociolinguistics. Method and Interpretation.* London: Blackwell Publishing.

Munson, B. (2001). A method for studying variability in fricatives using dynamic measures of spectral mean. In *Journal of the Acoustical Society of America*, 110(2), 1203-1206.

Munson, B. (2004). Variability in /s/ production in children and adults: Evidence from dynamic measures of spectral mean. In *Journal of Speech, Language, and Hearing Research*, 47(1), 58-69.

Munson, B., McDonald, E.C., DeBoe, N.L. & White, A.R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. In *Journal of Phonetics*, 43(2), 202-240.

Nissen, S.L., Fox, R.A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. In *Journal of the Acoustical Society of America*, 118, 2570-2578.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. In *Journal of the Acoustical Society of America*, 97, 520-530.

Oppo, A. (Ed.) (2007). *Le lingue dei sardi. Una ricerca sociolinguistica. Rapporto finale*. Cagliari: Regione Autonoma della Sardegna.

Paulis, G. (1984). *Introduzione e Appendice alla Fonetica Storica del Sardo di Max Leopold Wagner*. Cagliari: Gianni Trois Editore.

Paulis, G., Pinto, I. & Putzu, I. (Eds.) (2013). *Repertorio plurilingue e variazione linguistica a Cagliari.* Milano: Franco Angeli.

Pinto, I. (2013). Riflessioni sul metodo e primi risultati. In Paulis, G., Pinto, I. & Putzu, I. (Eds.), *Repertorio plurilingue e variazione linguistica a Cagliari*. Milano: Franco Angeli, 131-145.

Rattu, R. (2017). Repertorio plurilingue e variazione sociolinguistica a Cagliari: i quartieri di Castello, Marina, Villanova, Stampace, Bonaria e Monte Urpinu. Tesi di Dottorato, Università di Cagliari.

Schilling-Estes, N. (2002). Investigating Stylistic Variation. In Chambers, J.K., Trudgill, P., Schilling-Estes, N. (Eds.), *The Handbook of Language Variation and Change*. Malden-Oxford: Blackwell, 375-401.

Schwartz, M.F. (1968). Identification of speakers' sex from voiceless, isolated fricatives. In *Journal of the Acoustical Society of America*, 43(5), 1178-1179.

Shadle, C.H., Mair, S.J. (1996). Quantifying spectral characteristics of fricatives. In *Proceedings of the International Conference of Spoken Language Processing*. Philadelphia, PA: ICSLP, 1521-1524.

Spinu, L., Lilley, J. (2016). A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives. In *Journal of Phonetics*, 57, 40-58.

Stevens, M., Bukmaier, V. & Harrington, J. (2015). Pre-consonantal /s/-retraction. In The Scottish Consortium for ICPhS (Eds.), *Proceedings of the 18th International Congress of the Phonetic Sciences*. Glasgow, UK, 10-14 August 2015.

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. In Cole, J., Hualde, J.I. (Eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 65-86.

Stuart-Smith, J., Timmins, C. & Wrench, A. (2003). Sex and gender differences in Glaswegian /s/. In Solé, M.J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the 15th International Congress of the Phonetic Sciences.* Barcelona, Spain, 3-9 August.

Virdis, M. (1978). *Fonetica del dialetto sardo campidanese.* Cagliari: Edizioni della Torre.

Wagner, M.L. (1984). *Fonetica storica del sardo. Introduzione traduzione e appendice di Giulio Paulis*. Cagliari: Gianni Trois Editore.

Wenger, E. (1998). *Communities of Practice. Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

DOI: 10.17469/O2103AISV000004

NICHOLAS NESE, CHIARA MELUZZI

Accomodamento ed emergenza di varianti fonetiche: le affricate dentali intermedie a Pavia e Bolzano

The study explores two situations of language contact between Standard Italian, Italo-Romance dialects and Italian regional varieties in two sociolinguistic settings: a university college in Pavia and a group of second or third generation of Italian immigrants in Bolzano. All speakers were aged between 19 and 35 years. The study offers an acoustic analysis of the realization of dental affricates in word-list reading. The emergence of a new degree of sonority, the so-called intermediate, is highlighted in both group. The interpretation of this phenomenon relies on the Accomodation Theory[1].

Key words: sociophonetics, dental affricates, accomodation theory, language variation, language contact.

1. *Introduzione*

Il presente contributo prende in considerazione la pronuncia delle affricate dentali prodotte da due gruppi di giovani pavesi e bolzanini. Entrambi i contesti di elicitazione dei dati riguardano due comunità linguistiche caratterizzate da un background migratorio, seppur di diversa portata e finalità. I giovani pavesi registrati (Nese, 2016) sono infatti tutti studenti universitari del collegio Giasone del Maino, provenienti da diverse regioni d'Italia. Al contrario, i giovani bolzanini presi in esame già da Meluzzi (2014) rappresentano una seconda generazione d'immigrazione nel capoluogo altoatesino, ma condividono con i coetanei pavesi un background linguistico eterogeneo. Lo scopo di questo contributo è osservare la presenza di due foni affricati intermedi per grado di sonorità nella pronuncia di parole in isolamento di entrambi i gruppi di giovani, argomentando che l'emergenza di tale pronuncia possa essere intesa in ottica sociolinguistica come fenomeno di accomodamento linguistico legato, in ultima istanze, alle dinamiche migratorie che hanno caratterizzato le due comunità di parlanti.

Il contributo è organizzato come segue: nel § 2 verranno presentati brevemente i principi teorici alla base del nostro lavoro, con particolare risalto per il concetto sociolinguistico di accomodamento, nonché delle caratteristiche fonetico-fonologiche della variabile qui esaminata, ossia le affricate dentali; nel § 3 presenteremo i due corpora pavese e bolzanino su cui basiamo l'analisi, che sarà sviluppata nel § 4 del lavoro. Nel § 5 discuteremo, infine, i risultati emersi in ottica sociofonetica, mettendo in relazione la nozione teorica di acco-

[1] Il contributo è stato concepito ed elaborato congiuntamente dai due autori; ai soli fini accademici si attribuiranno i § 3, 4.1 e 6 a Nicholas Nese, laddove i § 1, 2 (e relativi sottoparagrafi), § 4.2 e 5 saranno da attribuirsi a Chiara Meluzzi.

modamento con quanto evidenziato dai nostri dati, per concludere, nella sesta e ultima sezione, con alcuni cenni ai futuri sviluppi del lavoro.

2. *Migrazione e accomodamento linguistico*

I contesti analizzati in questo studio riguardano due comunità di parlanti giovani, di età compresa tra i 18 e i 30 anni, entrambe risultato di un processo migratorio diverso per tipo, durata e prospettiva. Si sta ovviamente parlando di migrazioni interne ossia concernenti parlanti sempre di L1 italiana, ma con diverso background dialettale e sociale. Ciò appare particolarmente evidente nel caso della comunità italofona di Bolzano (Meluzzi, 2014): in questi casi si trattava prevalentemente di una migrazione di gruppo, con la formazione di quartieri appositi per i nuovi immigrati su un territorio a maggioranza germanofona, in cui l'intenzione conservativa del mantenimento della propria L1 era molto forte (Esser, 1980)[2]. Simile, per certi versi, è anche il caso degli studenti pavesi del nostro studio: anche loro provengono da diverse parti d'Italia e si ritrovano tutti stanziati in un unico collegio esaminato. Le differenze sociologiche maggiori rispetto al contesto altoatesino risultano il contesto linguistico di immersione (italofono a Pavia, tedescofono a Bolzano), nonché soprattutto la motivazione dietro la "migrazione" verso un'altra città d'Italia: nel caso bolzanino infatti la migrazione era intesa come stabile e, spesso, riguardava gruppi interi di individui della stessa regione, mentre nel caso pavese riguarda più spesso singoli studenti con la prospettiva di tornare, anche ad intervalli regolari, alla propria città d'origine.

Entrambi i casi presentano punti interessanti di riflessione per quanto riguarda la sociolinguistica del contatto, che spesso (per non dire sempre) si è occupata di contatto tra lingue diverse, mentre nei casi qui analizzati il contatto è avvenuto tra varietà della stessa lingua, l'italiano, con al più l'influenza di altre lingue quali i dialetti Italo-Romanzi dei parlanti. Si è però osservato come la conoscenza e l'uso di dialetti Italo-Romanzi non fosse uniformemente distribuiti tra i giovani analizzati; tale variazione di conoscenza e uso non era neppure correlabile con le caratteristiche sociolinguistiche del parlante (es. provenienza geografica, sesso). A partire da questa situazione sociolinguistica, dunque, lo scopo di questa indagine è di andare a verificare se possa essere in atto, nelle due comunità analizzate, un identico processo di accomodamento linguistico (§ 2.1) in relazione in particolare alla pronuncia di un fono particolarmente instabile nel sistema fonetico dell'italiano, le affricate dentali (§ 2.2), e dunque buon marcatore di eventuali mutamenti linguistici in atto. Per riassumere, la nostra ipotesi è dunque che, benché partendo da premesse leggermente diverse in termini di sociolinguistica della migrazione, nelle due comunità di giovani parlanti bolzanini e pavesi si stiano attuando dinamiche simili di accomodamento linguistico per quanto riguarda il grado di sonorità delle affricate dentali, visibile tramite l'emergenza di una affricata intermedia per sonorità (§ 3).

[2] La specificità della migrazione che ha caratterizzato l'italiano di Bolzano si riflette fortemente sugli atteggiamenti linguistici della comunità italofona, da Meluzzi (2015).

2.1 La teoria dell'accomodamento

La Speech Accomodation Theory (SAT) fu sviluppata inizialmente da Giles (1973) e in seguito ripresa, tra gli anni '70 e '80, da molti studiosi all'interno del paradigma sociolinguistico (es. Coupland, 1984), spesso in contrasto con il paradigma più strettamente laboviano (Labov, 1966). Per questi studiosi risultava particolarmente problematica la nozione di variabilità diafasica introdotta da Labov nei termini di "attention paid to speech" (Labov, 1994): secondo Giles (1973) e poi Bell (1984), infatti, la variabilità diafasica sarebbe da interpretarsi più come un risultato dell'influenza interpersonale tra intervistato e intervistatore a seconda altresì dei differenti contesti socio-culturali, delle ideologie, stereotipi e attitudini linguistiche associati alle diverse lingue, dialetti e varietà in gioco. Per questo motivo, Giles, Coupland & Coupland (1991: 14) giudicavano la SAT come la base per una migliore interpretazione sociolinguistica della variabilità del linguaggio nella conversazione, più che una vera e propria teoria.

Dalla SAT è poi derivata una teoria più specifica riguardante l'accomodamento linguistico (Communication Accomodation Theory – CAT, Giles *et al.*, 1987) che distingueva fondamentalmente due fenomeni in cui poteva realizzarsi l'accomodamento linguistico:

1) convergenza, in cui tra due attanti di una conversazione si realizza un accomodamento reciproco sul piano linguistico, prosodico e/o non-verbale (l'esempio più noto di un accomodamento convergente a livello fonetico è offerto da Coupland, 1984, nonché dai più recenti studi di Pardo, 2006; Pardo, Gibbons, Suppes & Krauss, 2012, tra gli altri);
2) divergenza (Bourhis, Giles, 1977) o "disaccomodation" (Scotton, 1985), che si istanzia in una marcata esasperazione dei tratti linguistici e/o non-verbali che differenziano i due attanti, specialmente in casi di comunicazione inter-linguistica e inter-etnica e nel caso in cui uno dei due partecipanti senta minacciata o sminuita la propria identità etnico-linguistica (Giles, Coupland & Coupland, 1991: 8).

Si è soliti inoltre distinguere tra tre dimensioni principali in cui si può istanziare l'accomodamento (Giles, Coupland & Coupland, 1991: 12), dal momento che questo può essere

a) simmetrico o asimmetrico
b) totale o locale
c) verso l'alto o verso il basso ("upward" vs. "downward")

Il primo punto tiene conto dell'eventuale presenza di stereotipi associati a uno o più elementi linguistici, in relazione anche alla percezione dei ruoli sociali dei due attanti. Il secondo punto specifica che l'accomodamento, in senso sia divergente sia convergente, può riguardare anche solo un elemento del sistema linguistico e non altri, o riguardarne alcuni in senso convergente e altri in senso divergente, o ancora con diversi gradi di convergenza e divergenza tra i diversi tratti. L'aspetto più interessante della CAT concerne tuttavia la direzionalità dell'accomodamento: in un caso l'accomodamento si istanzia verso la varietà, o le variabili, generalmente ritenute come maggiormente prestigiose ("upward accomodation"), mentre al contrario l'accomodamento può anche avvenire verso le varietà che definiremmo più sub-standard, più localmente marcate o avvertite come meno prestigiose ("downward accomodation"). Quest'ultimo caso si avvicina molto a quello che nella sociolinguistica laboviana è stato definito come prestigio coperto ("covert-prestige", Milroy, Milroy, 1985).

Sia la SAT, sia soprattutto la CAT hanno riscosso grande successo tra gli studiosi che lavoravano sull'acquisizione linguistica e la sociolinguistica percettiva. Recentemente queste teorie hanno iniziato ad avere una applicazione sistematica anche ai fenomeni di contatto linguistico derivanti da dinamiche migratorie (es. Otheguy, Zentella & Livert, 2007) o da particolari situazioni di contatto come il caso dei collegi universitari (Pardo *et al.*, 2012). Il presente contributo si inserisce proprio in questa linea di ricerca sull'accomodamento in due diverse situazioni di contatto linguistico.

2.2 Analizzare l'accomodamento: le affricate dentali

Le affricate dentali rappresentano due fonemi rari e marcati a livello sia tipologico (Maddieson 1984) sia acquisizionale (Costamagna 2003; Celata, 2004: 78). In particolare, anche l'italiano, unica lingua romanza a preservare sia /ts/ che /dz/ nel proprio inventario fonologico, non presenta un alto rendimento funzionale dell'opposizione tra sorda e sonora (De Dominicis, 1999). Gli studi condotti soprattutto da Canepari (es. Canepari, 1979) e da altri dialettologi italiani (Telmon, 2003; Loporcaro, 2009; Foresti, 2010; Rizzi, 1989) hanno evidenziato la grande variabilità diatopica di questi foni sia per quanto concerne il luogo di articolazione, sia per la distribuzione delle varianti sorda e sonora, specialmente in contesto post-nasale e post-liquida. In particolare dopo /l/ si registra la differenza maggiore tra varietà settentrionali e varietà meridionali di italiano: le prime tendono infatti a una resa prevalentemente sorda, laddove per le seconde è attestata maggiormente la sonora, soprattutto in Calabria, Campania e, secondo Canepari (1979: 79), anche negli accenti siciliani più marcati.

Stante questo quadro molto diversificato, vi si aggiunge una non chiara conoscenza della norma linguistica, almeno per quanto concerne la pronuncia di questi foni. All'interno di un paradigma di ricerca di tipo laboviano, inoltre, l'attribuzione di valori di prestigio all'una o all'altra pronuncia sono da assumersi più come ipotesi di ricerca, soprattutto per l'alto valore identitario dei dialetti e delle varietà regionali, oltre alla già menzionata mancanza di uno standard preciso di riferimento.

In questo quadro, le affricate dentali diventano una buona "cartina al tornasole" (Meluzzi, 2014: 50) per evidenziare mutamenti fonetici in atto, specialmente in ottica di comunità migratorie o in situazione di contatto linguistico prolungato tra varietà diverse di italiano.

3. *I dati*

I dati utilizzati nel presente lavoro sono stati raccolti a Pavia e Bolzano tra il 2011 e il 2015. La raccolta dati è stata condotta dai due autori in momenti diversi della propria attività di ricerca (Meluzzi, 2014; Nese, 2016). In entrambi i casi, trattandosi di una ricerca di tipo sociofonetico, è stato necessario un certo grado di integrazione preliminare da parte dei ricercatori all'interno delle comunità studiate. In tal senso, per quanto riguarda il caso di

Pavia, l'appartenenza del ricercatore alla comunità collegiale è stato un fattore determinante al fine di poter osservare i parlanti in un contesto di parlato spontaneo.

Nel caso di Bolzano invece si è cercato comunque di inserirsi all'interno delle reti sociali della comunità italofona adottando un approccio tradizionale, basato sulla definizione di *rete sociale* (Milroy, Milroy, 1985), ovvero attraverso la frequentazione di luoghi pubblici ed eventi culturali presso i quartieri italofoni, e curando inoltre i rapporti personali in modo da ampliare la propria rete di conoscenze. Allo stesso tempo si è cercato di sfruttare le potenzialità dei social network, in particolare Facebook, al fine di poter raggiungere più rapidamente un numero maggiore di soggetti.

La raccolta dati si è svolta per mezzo di interviste frontali semi-strutturate mirate a elicitare i diversi livelli diafasici del parlato. L'intervista è stata suddivisa in più parti: in questo modo si sono potuti osservare sia il parlato spontaneo e semi-spontaneo, attraverso domande generiche per permettere all'informante di parlare il più liberamente possibile, sia il parlato controllato, tramite una serie di compiti linguistici, tra cui la lettura di una lista di parole in isolamento e di alcuni scioglilingua. Occorre precisare che le parole analizzate nei due corpora non corrispondono, ma sono accomunate dagli stessi contesti fonologici.

Per questa seconda parte di intervista che prevedeva la lettura di parole si è deciso di utilizzare esclusivamente termini esistenti; tale scelta è stata dettata dal voler analizzare lo stesso *item* in diafasia qualora fosse stato realizzato anche nella prima parte di intervista. In secondo luogo la scelta delle parole concerne i diversi contesti fonologici in cui si possono realizzare le affricate dentali, ovvero iniziale assoluto (#C_), intervocalico scempio (VCV), intervocalico geminato (VCCV) e post-sonorantico (SCV), intendendo con il termine sonoranti la consonante nasale /n/, la liquida /l/ e la rotica /r/ (solo nel corpus di Bolzano). I *tokens* contenenti un'affricata dentale sono in totale 213 per il corpus pavese e 530 per quello bolzanino. Data la grande disparità tra i due corpora, non è possibile effettuare un confronto diretto; pertanto in questo contributo i due corpora verranno utilizzati come due casi di studio e confrontati indirettamente per osservare l'emergenza di fenomeni sociofonetici comuni.

Tabella 1 - *Dettagli dei due corpora analizzati in questo contributo: il corpus pavese fa riferimento alla tesi di Nese (2016); i dati bolzanini rientrano nel corpus CItaBol, legato ai lavori di Meluzzi (2014, 2016)*

	Corpus pavese	*Corpus bolzanino*
Data	Gen.-Mar. 2015 + Dic. 2015	Ott. 2011-Mag. 2012
Parlanti	12 (7 M, 5 F)	8 (4 M, 4 F)
Ore di registrazione	2h 57'	2h 40'
Tokens /ts dz/ (lettura di parole)	213	530

La Tabella 1 riassume le caratteristiche principali dei due corpora messi a confronto. Gli informanti del corpus di Pavia presi in considerazione per questo lavoro sono 12 studenti universitari, 6 matricole e 6 laureandi. Il campione è stato inoltre bilanciato tenendo conto del genere (7 maschi e 5 femmine) e della provenienza (6 del nord e 6 del sud). La selezione è stata fatta in modo tale che le 6 matricole provenissero dalle stesse province dei

6 studenti laureandi, in modo da poter effettuare un'analisi in diacronia apparente; le 6 matricole hanno ripetuto l'intervista a distanza di 11 mesi e questo ha permesso di svolgere un'analisi in diacronia reale. A livello di strumentazione, le interviste pavesi sono state realizzate all'interno di una stanza del Collegio Universitario Giasone del Maino con un registratore digitale Agptek e un microfono con clip della stessa marca, con un *sampling rate* di 48 kHz. Per quanto riguarda il corpus di Bolzano, il corpus CItaBol è strutturato in diacronia apparente, con un totale di 42 informanti registrati, di cui 8 tra i 17 e i 35 anni (4 maschi e 4 femmine), 23 tra i 35 e 60 anni (14 maschi e 9 femmine) e 11 oltre i 60 anni (4 maschi e 7 femmine). Ai fini della nostra ricerca sono stati considerati esclusivamente i dati degli informanti di età compresa tra i 17 e 35 anni che hanno prodotto un totale di 2h 40' di registrazione, mentre il corpus pavese consta di 2h 57' di registrazione[3]. Le interviste si sono svolte in un'aula appositamente riservata all'interno della Biblioteca della Libera Università di Bolzano, con un registratore ZOOM H2 munito di microfono esterno Sony, con un *sampling rate* di 48 kHz.

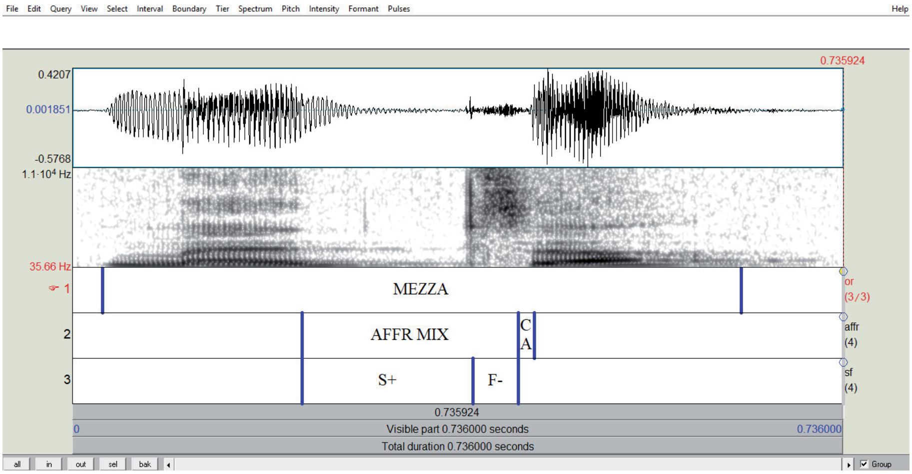
Tutte le registrazioni sono in formato WAV; in entrambi i contesti, inoltre, tutti i compiti linguistici che richiedevano un supporto visivo (lettura di parole, etc.) si sono svolti con l'ausilio di un PC portatile, posto a debita distanza dal microfono per evitare interferenze. Dall'analisi qualitativa degli spettrogrammi e delle forme d'onda, infatti, non sono state rilevate interferenze imputabili a cariche elettrostatiche derivanti dal PC portatile.

In un primo momento i file audio sono stati annotati ortograficamente per mezzo del software ELAN 4.8.1, dopodiché è stata effettuata l'annotazione acustica con l'ausilio del software PRAAT 5.4 esclusivamente sull'elenco di parole in isolamento contenenti le affricate dentali.

Per l'annotazione è stato adottato lo stesso protocollo di ricerca elaborato e utilizzato da Meluzzi (2014: 65-75) strutturato su tre livelli, come mostrato in Figura 1. La prima riga di annotazione, definita *orthographical*, prevede solamente la trascrizione del token, in maniera appunto ortografica (cfr. Fig 1). La seconda riga di annotazione, definita *affricate*, individua esclusivamente l'affricata; l'annotazione "AFFR" è accompagnata dal segno "+" qualora l'esito sia sonoro, "-" in caso di produzione sorda. L'analisi acustica condotta da Meluzzi (2014) aveva permesso di individuare un grado intermedio di sonorità, annotato "MIX", in cui alla prima fase occlusiva sonora seguiva un momento fricativo sordo; la presenza di affricate intermedie è stata individuata anche nei dati pavesi di Nese (2016). Si è reso necessario introdurre l'etichetta "CA" (*coarticulation)*, in quanto in diversi tokens era apprezzabile uno stacco tra la fine della frizione e l'inizio della vocale successiva. Infine la terza riga di annotazione, denominata *stop&fric*, aveva l'obiettivo di dividere i diversi momenti dell'affricata, in particolar modo quello occlusivo e fricativo. A livello di segmentazione, la fase di occlusione veniva fatta terminare subito dopo il *burst* ed è stata annotata come "S" (*stop*); qualora il *burst* fosse stato assente si è preso come punto di riferimento l'inizio della fase fricativa, che è stata indicata come "F" (*friction*).

[3] In Meluzzi (2014) i parlanti giovani avevano effettuato due ripetizioni della lista di parole, per un totale di 1060 tokens. Dal momento che non si erano evidenziate differenze statisticamente significative tra le due ripetizioni, in questa sede si è deciso di prendere in considerazione solo la prima ripetizione della lettura della lista di parole. Ciò ha inoltre consentito di non aumentare la disparità di tokens tra i due corpora.

Figura 1 - *Il protocollo di annotazione ripreso da Meluzzi (2014: 65-75) su tre livelli (nome della parola target, individuazione dell'affricata dentale, divisione tra elemento occlusivo e fricativo dell'affricata dentale). In questo esempio si osserva la presenza di una affricata intermedia per grado di sonorità (dal corpus di Nese, 2016)*



Occorre precisare che l'annotazione così implementata e, di conseguenza, l'individuazione delle diverse varianti dell'affricata dentale, è stata condotta sulla base dell'ispezione visiva dello spettrogramma e della forma d'onda, unitamente a una attenta valutazione acustica, specialmente per quanto riguarda la parte fricativa nel caso di affricate con grado di sonorità intermedio. Inoltre, come osservato da Meluzzi (2014: 73), il cambiamento nella curva di intensità, automaticamente visualizzato su PRAAT, ha spesso servito come conferma delle annotazioni effettuate. Un'indagine specifica di questo e altri parametri acustici, più specificamente legati all'analisi della sonorità (cfr. Hawkins, Nguyen, 2004), rappresenta senz'altro uno sviluppo futuro della ricerca sui dati finora raccolti da entrambi gli autori di questo contributo.

4. *Analisi*

Il lavoro di analisi dei dati è stato condotto principalmente sul parametro di sonorità. Per poter impostare una matrice di dati sia sociali che linguistici e calcolare la significatività statistica delle correlazioni sociolinguistiche ci siamo serviti del software IBM SPSS 20.

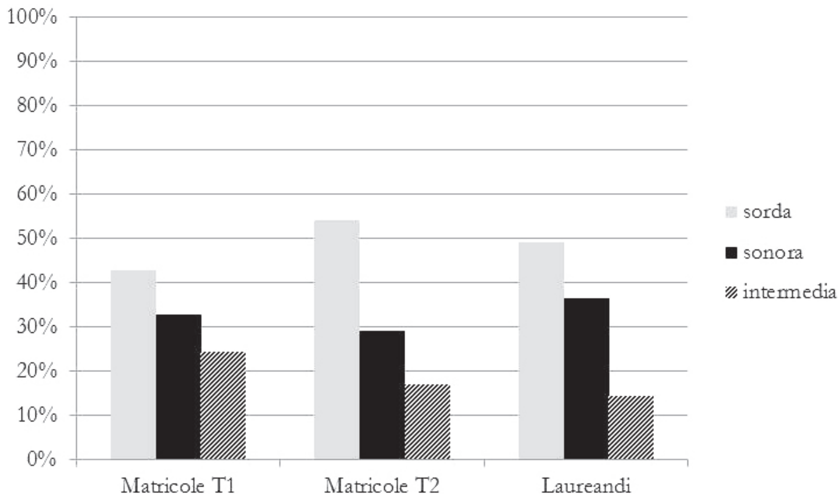
Un primo confronto tra i due sotto-corpora, ha indicato una significativa presenza di affricate dentali intermedie sia nelle realizzazione dei giovani bolzanini sia in quella dei giovani pavesi: per Bolzano (530 tokens), le realizzazioni intermedie erano pari al 20,8%, a fronte di un 38,3% di esiti sonori e di un 40,2% di esiti sordi; il corpus pavese (213 tokens) rivela una prevalenza di esiti sordi (48,8%), laddove per gli esiti sonori e intermedi si registra rispettivamente il 32,9% e il 18,3% dei casi.

Si è quindi analizzato la distribuzione del grado di sonorità rispetto alle diverse variabili sociali nei due sotto-corpora.

4.1 Il corpus pavese

Per il corpus pavese, sono state considerate tre variabili sociali, accanto alla *sigla dell'informante*: *genere*, *provenienza*, *gruppo*. La variabile *gruppo* prevede tre valori, matricole T1 (studenti iscritti al primo anno di un corso di laurea triennale o magistrale), matricole T2 (stesso gruppo di informanti intervistato dopo 11 mesi) e laureandi. Specifichiamo che ai fini del presente lavoro prenderemo in considerazione solo i risultati relativi all'analisi in diacronia reale.

Figura 2 - *Distribuzione del grado di sonorità dell'affricata dentale realizzata come sorda, sonora o intermedia dai giovani del gruppo pavese (213 tokens) divisi secondo il gruppo di appartenenza ossia matricole appena arrivate (T1), matricole registrate dopo un anno di permanenza a Pavia (T2) e laureandi registrati in contemporanea al primo gruppo.* $\chi^2(4) = 3{,}703$, $p > 0{,}05$



Nel corpus pavese, la variabile *gruppo* ha mostrato (Figura 2) una distribuzione omogenea della sonorità nei due sottogruppi e si riscontra inoltre un incremento degli esiti sordi nel sottogruppo *matricole T2*, parzialmente bilanciato da un calo degli esiti intermedi. È stato eseguito il test del chi-quadrato, che non ha tuttavia confermato la validità statistica dei risultati; questo ci ha suggerito, data l'evidente variazione diacronica, di utilizzare i dati in Figura 1 come prima rappresentazione, da intendersi in modo qualitativo, dei nostri dati, andando poi a raffinare l'analisi quantitativa su di essi indagando l'apporto esplicativo di altre variabili. Si è deciso di introdurre la variabile *provenienza*, dividendo tra collegiali provenienti dal nord e provenienti dal sud[4]. Così facendo, tramite un'analisi trivariata, sonorità*gruppo*provenienza, si è ottenuto dei risultati statisticamente significativi al test

[4] Data la limitatezza del corpus, non è stato infatti possibili introdurre distinzioni di grana più fine relativamente alla provenienza.

del chi-quadrato ($\chi^2(2) = 8{,}815$, $p = 0{,}012$) nel complesso confermando l'ipotesi di partenza, ovvero che è possibile accertare e quantificare una variazione linguistica in diacronia.

Figura 3 - *Distribuzione del grado di sonorità delle affricate dentali (sorda, sonora, intermedia) nel gruppo delle matricole T1 (70 tokens) a seconda della città di provenienza, con una macro-distinzione areale tra nord e sud d'Italia.* $\chi^2(2) = 8{,}815$, $p = 0{,}012$

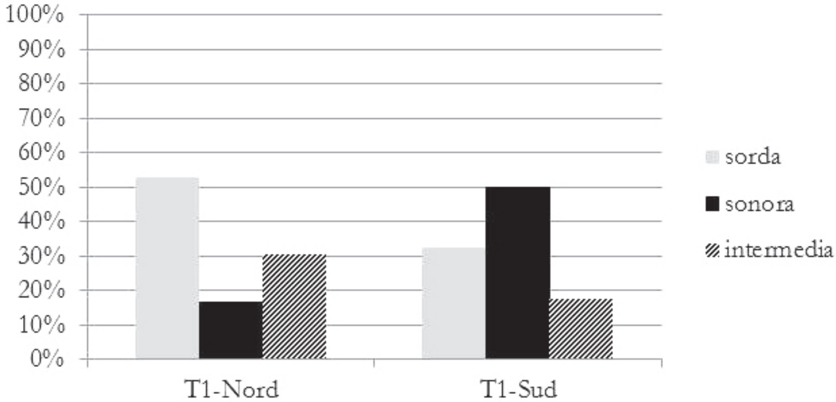
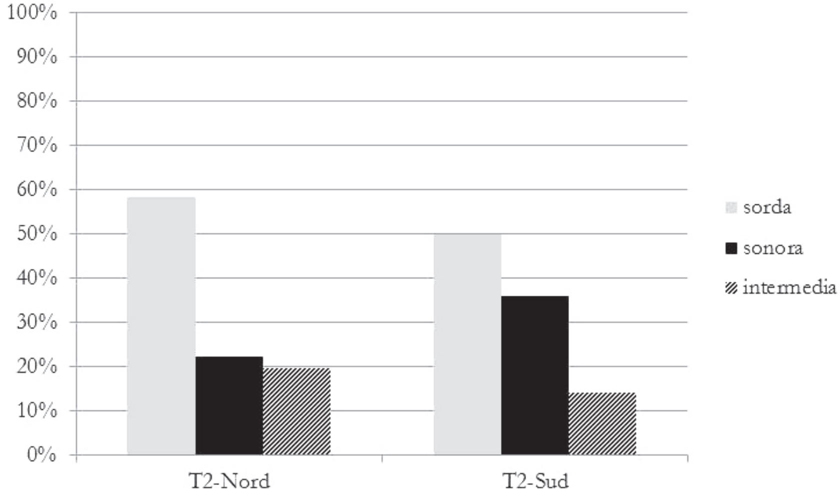


Figura 4 - *Distribuzione del grado di sonorità delle affricate dentali (sorda, sonora, intermedia) nel gruppo delle matricole T2 (70 tokens) a seconda della città di provenienza, con una macro-distinzione areale tra nord e sud d'Italia.* $\chi^2(2) = 1{,}715$, $p > 0{,}05$



Dai grafici delle Figura 3 e 4, si può infatti notare, per quanto riguarda gli studenti del nord, un aumento degli esiti sordi, +5,5% per gli studenti del nord e +17,6% per quelli del sud, parallelamente ad una diminuzione degli esiti intermedi, rispettivamente -10,6% e -3,7%. Nel gruppo degli studenti del sud, invece, questo aumento delle sorde e diminuzione delle intermedie è accompagnato da un significativo calo delle sonore. Come detto in precedenza, questi dati sono solo in parte significativi statisticamente; più precisamente il chi-quadrato ha un valore < 0,05 solo per il gruppo *matricole T1*. Questo risultato ci porta a formulare una prima interpretazione: nel gruppo *matricole T1* la variabile sociale *provenienza* appare essere distintiva, ossia portatrice di significato sociolinguistico, dato che influenza la variazione del parametro di sonorità dell'affricata dentale; nel gruppo *matricole T2* invece ciò non accade, ossia il parametro di *provenienza* non è più determinante per questa variabile linguistica.

Tenendo conto del fatto che gli studenti registrati nel gruppo *matricole T1* e *matricole T2* sono sempre gli stessi a distanza di tempo, si può quindi affermare che la variabile di *provenienza* abbia perso significatività a fronte di una permanenza prolungata all'interno del collegio "Giasone del Maino" di Pavia. Ciò può essere dunque interpretato come effetto di un accomodamento, in particolare da parte degli studenti del sud, verso una pronuncia più settentrionale, almeno per quanto riguarda il parametro qui esaminato, ossia il grado di sonorità delle affricate dentali.

Abbiamo provato a valutare l'interazione tra la variabile di *genere* e quella di *provenienza* ed è emerso un risultato statisticamente significativo solo per il sottogruppo *matricole T1* femmine, con un valore p = 0,03. Questo risultato, associato a quello delle *matricole T2* femmine, che ha valore p > 0,05, è in linea con l'ipotesi avanzata in precedenza, secondo cui si sia verificato un cambiamento, in favore di un accomodamento verso uno standard più settentrionale, nel gruppo *matricole T2*. Grazie a questa ulteriore analisi possiamo anche spingerci ad ipotizzare che sia il gruppo delle femmine quello maggiormente variabile in senso innovatore.

Per concludere la nostra analisi si è deciso di andare ad indagare possibili variazioni anche all'interno dei diversi contesti fonologici in cui si realizza l'affricata dentale, dato che la sonorità di questi foni è molto variabile diatopicamente, in particolare nel contesto post-nasale e intervocalico geminato.

Figura 5 - *Realizzazione delle affricate dentali come sorde, sonore o intermedie nelle matricole T1 (70 tokens) in quattro diversi contesti fonologici (iniziale assoluto, post-sonorantico, intervocalico geminato, intervocalico scempio).* $\chi^2(6)= 29{,}792,\ p\leq 0{,}001$

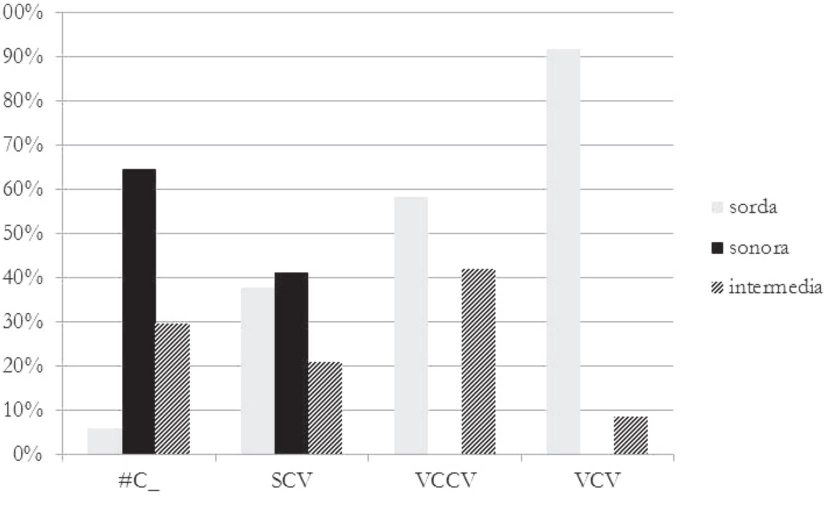
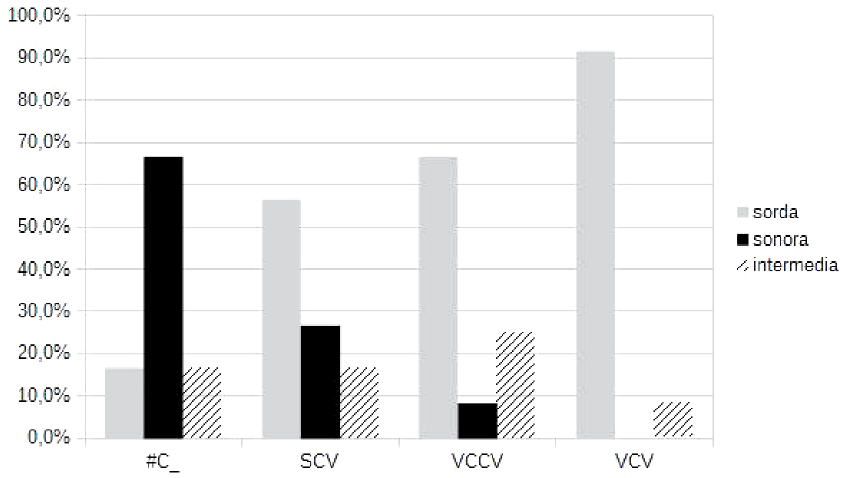


Figura 6 - *Realizzazione delle affricate dentali come sorde, sonore o intermedie nelle matricole T2 (70 tokens) in quattro diversi contesti fonologici (iniziale assoluto, post-sonorantico, intervocalico geminato, intervocalico scempio).* $\chi^2(6) = 23{,}198$, $p = 0{,}001$



Abbiamo pertanto considerato la variabile *gruppo* per osservare un'eventuale variazione diacronica. Dal confronto tra i grafici nelle Figure 5 e 6 si può notare come in tutti i contesti fonologici gli esiti sordi siano in progressivo aumento rispetto agli altri gradi di sonorità e soprattutto rispetto alle realizzazioni intermedie, che rivelano al contrario una diminuzione in tutti i contesti; per quanto riguarda il contesto #C_ si osserva invece un aumento delle occorrenze sonore: specifichiamo che tale risultato è conforme allo "standard" settentrionale.

4.2 Il corpus bolzanino

Per il corpus bolzanino, i dati relativi ai 7 parlanti più giovani presenti in CItaBol (§ 3) sono stati considerati in relazione alla variabile di *genere* e quindi confrontati con le realizzazioni dei parlanti più anziani, in un'ottica di diacronia apparente (Labov, 1994).

Una prima esplorazione dei dati di questo sotto-corpus ha consentito di confermare come la variazione nel grado di sonorità delle affricate dentali sia legata anche al contesto fonologico. La Figura 7 mostra infatti come i giovani bolzanini tendano a pronunce quasi completamente sonore a inizio di parola (es. *zappa*) con percentuali del 72,7%, mentre una resa prevalentemente sorda nel contesto intervocalico sia scempio (53,8%) che geminato (53,1%). In questi contesti intervocalici, così come soprattutto nel contesto post-sonorantico (ossia dopo /r/, /l/ e /n/) è percentualmente significativa la presenza di realizzazione intermedie per grado di sonorità (28,9%).

Figura 7 - *Realizzazione delle affricate dentali come sorde, sonore o intermedie nel corpus bolzanino (530 tokens) in quattro diversi contesti fonologici (iniziale assoluto, post-sonorantico, intervocalico geminato, intervocalico scempio).* $\chi^2(8)= 108,153,\ p\leq 0,001$

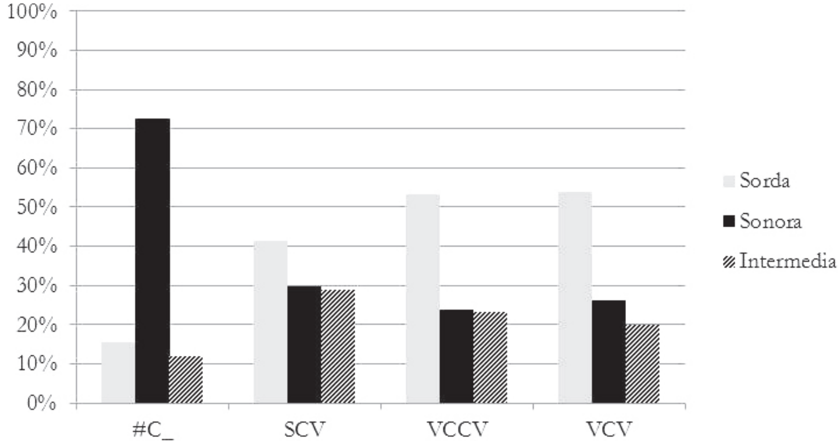
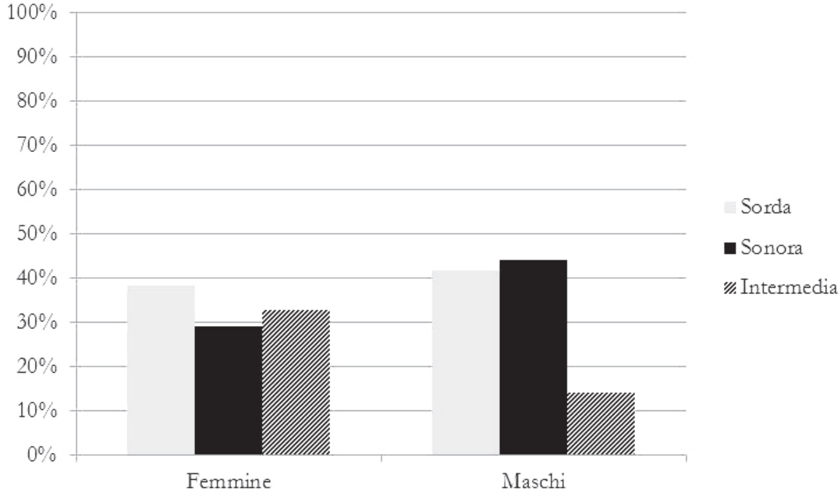
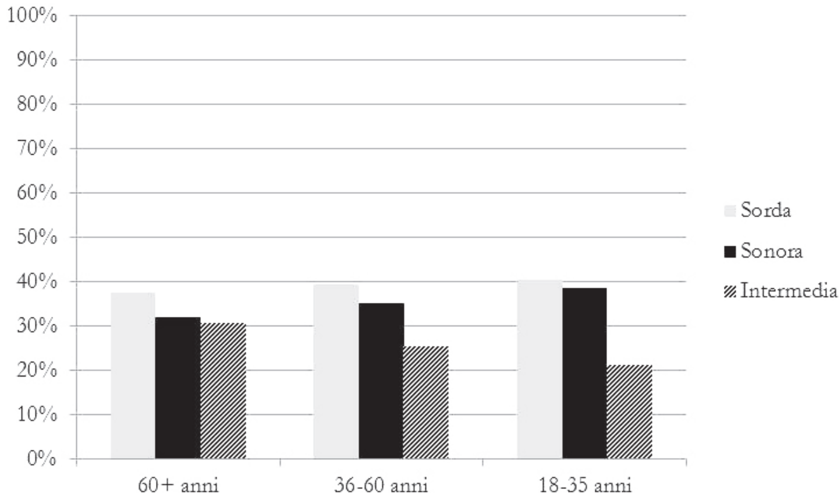


Figura 8 - *Realizzazione delle affricate dentali come sorde, sonore o intermedie nel corpus bolzanino (530 tokens) a seconda del genere del parlante.* $\chi^2(2)= 28,299,\ p\leq 0,001$



Rispetto alla variabile di genere, anche nel corpus bolzanino è possibile osservare una maggiore presenza di affricate dentali intermedie nelle realizzazioni femminili (32,7%) rispetto al gruppo degli uomini (13,9%). Inoltre, se la percentuale di realizzazione di sorde tra uomini e donne sono abbastanza simili, diminuisce nel gruppo femminile la presenza di affricate sonore: il gruppo maschile realizza affricate sonore nel 44,2% dei casi, laddove il gruppo femminile mostra percentuali del 29,1%. Si può quindi dedurre che questa diminuzione di rese sonore nel gruppo femminile si rispecchi in un aumento delle produzioni intermedie all'interno di questo stesso gruppo.

Figura 9 - *Realizzazione delle affricate dentali come sorde, sonore o intermedie nel corpus bolzanino (530 tokens) a seconda dell'età del parlante distinguendo tra tre fasce (over 60, 36-60 anni, 18-35 anni)*[5]. *$\chi^2(2)$= 28,299, p≤ 0,001*



Non avendo a disposizione per il corpus bolzanino dei dati in diacronia reale, si è quindi deciso di operare un confronto in diacronia apparente sulla base dei dati raccolti nel corpus CItaBol (Meluzzi, 2016). Dai dati in Figura 9 emerge come la maggior parte delle realizzazioni intermedie si ritrovano nel parlato dei più anziani (30,5%) che corrispondono anche alla prima generazione d'immigrazione a Bolzano. Si delinea inoltre una chiara tendenza nel gruppo dei giovani a diminuire l'uso di realizzazioni intermedie, almeno in un compito molto formale come la lettura di parole, a favore di realizzazioni o completamente sonore o completamente sorde. Tuttavia anche nel gruppo dei giovani, come si è potuto appurare dai grafici precedenti, le affricate intermedie continuano a essere utilizzate con percentuali apprezzabili (20,9%), confermando il fatto che questa variante sia comunque presente nei dati e non possa essere attribuibile a meri processi di ipo- o iper-articolazione.

5. *Discussione*

L'analisi della realizzazione delle affricate dentali condotta in questo studio ha mostrato come i foni che in Meluzzi (2014) erano stati definiti come "intermedi" per grado di sonorità si ritrovano tanto nei giovani (e non) bolzanini, quanto nei giovani pavesi del collegio Giasone del Maino (Nese, 2016). Il confronto indiretto tra i due sotto-corpora sembra confermare l'ipotesi che questi foni tendano a comparire come risultato di un processo di contatto linguistico tra lingue (italiano standard e dialetti Italo-Romanzi) e varietà di regionali di italiano. L'emergenza di una "terza via" nella

[5] Per le considerazioni metodologiche in merito ai criteri per l'individuazione di queste tre fasce d'età si veda Meluzzi (2014: 76).

realizzazione di un fono chiaramente marcato diatopicamente era già stata evidenziata da Nocchi, Filipponio (2012) per le sibilanti dell'italiano di Livorno. La situazione delineata in questo studio ci porta a condividere con gli autori precedenti la visione dell'emergenza di un terzo tipo di realizzazione dell'affricata dentale, intermedia appunto per grado di sonorità. Nell'ottica della CAT precedentemente esposta (§ 2.1) possiamo dire che sia nel caso bolzanino sia in quello pavese ci troviamo di fronte a un accomodamento verso l'alto, ossia verso una varietà sentita come maggiormente prestigiosa. Nei nostri casi tale varietà è rappresentata da un tipo di italiano sentito genericamente come "settentrionale". Ciò è evidente nel subcorpus pavese nel momento in cui si è osservato come le matricole del Nord modifichino sostanzialmente meno dei coetanei del Sud la propria pronuncia delle affricate dentali dopo un anno di permanenza a Pavia. Questo, inoltre, risulta in linea con quanto evidenziato anche da studi precedenti su fenomeni prosodici più generali sull'accomodamento linguistico nei collegi universitari (Pardo, 2006; Pardo *et al.*, 2012).

6. *Conclusioni e prospettive future*

L'emergenza delle affricate dentali intermedie tanto a Pavia quanto a Bolzano conferma quindi l'ipotesi, già avanzata in Meluzzi (2014, 2016), che questi foni rappresentino una variante di accomodamento in situazioni di contatto linguistico, temporaneo o a lungo termine. Da questi dati, tuttavia, non è ancora chiaro se si tratti di un accomodamento totale del sistema o se invece riguardi solo foni, come appunto le affricate dentali, in cui l'alternanza tra realizzazione sorda e sonora è diatopicamente marcata. Lo studio di altre variabili sui dati utilizzati per questo studio potrà sicuramente offrire una migliore chiave di lettura in questo senso, andando ad evidenziare quegli elementi del sistema linguistico oggetto di accomodamento divergente o convergente a seconda dei casi. Inoltre, l'analisi andrà ampliata prendendo in considerazione anche i dati delle interviste, anche in questo caso in entrambi i sub-corpora.

Inoltre, questo tipo di analisi ha permesso di sottolineare l'importanza della connessione tra la componente sociale e quella fonetica, all'interno di un paradigma di ricerca che appunto vuole essere di tipo sociofonetico, riferendoci in particolar modo alla nozione di *accomodamento linguistico* intesa come un potente strumento per comprendere la variazione linguistica a livello fonetico a grana fine specialmente nei contesti migratori, sia stabili che stagionali, che hanno una notevole influenza sull'organizzazione fonetica del parlato.

Riferimenti bibliografici

Bell, A. (1984). Language style as audience design. In *Language in Society*, 13, 145-204.

Bourhis, R.Y., Giles, H. (1977). The language of intergroup distinctiveness. In Giles, H. (Ed.), *Language, Ethnicity and Intergroup Relations*. London: Academic Press, 119-135.

Canepari, L. (1979). *Italiano standard e pronunce regionali*. Padova: Cluep.

Celata, C. (2004). *Acquisizione e mutamento di categorie fonologiche. Le affricate in italiano*. Pavia: Franco Angeli.

Costamagna, L. (2003). Affricates in Italian as L2: the role of psycho-attitudinal parameters. In Costamagna, L., Giannini, S. (Eds.), *La fonologia dell'interlingua. Principi e metodi di analisi*. Milano: Franco Angeli, 95-129.

Coupland, N. (1984). Accomodation at work: Some phonological data and their implications. In *International Journal of the Sociology of Language*, 46, 49-70.

De Dominicis, A. (1999). *Fonologia comparata delle principali lingue europee moderna*. Bologna: Clueb.

Esser, H. (1980). *Aspekte der Wanderungssoziologie*. Darmstadt and Neuwield: Luchterthand.

Foresti, F. (2010). *Profilo linguistico dell'Emilia Romagna*. Roma-Bari: Laterza.

Giles, H. (1973). Accent mobility: a model and some data. In *Anthropological Linguistics*, 15, 87-105.

Giles, H., Coupland, J. & Coupland, N. (1991). *Contexts of Accomodation. Develop ments in applied sociolinguistics*. Cambridge: Cambridge University Press.

Giles, H., Mulac, A., Bradac, J.J. & Johnson, P. (1987). Speech accomodation theory: the next decade and beyond. In McLaughlin, M. (Ed.), *Communication Yearbook 10*. Newbury Park, CA: Sage, 13-48.

Hawkins, S., Nguyen, N. (2004). Effects on word recognition of syllabic-onset cues to syllable-coda voicing. In Local, J., Ogden, R. & Temple, R. (Eds.), *Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 38-57.

Labov, W. (1966). *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

Labov, W. (1994). *Principles of Linguistic Change, vol. 1: Internal Factors.* London: Blackwell.

Loporcaro, M. (2009). *Profilo linguistico dei dialetti italiani*. Roma-Bari: Laterza.

Maddieson, J. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Meluzzi, C. (2014). Le affricate dentali nell'italiano di Bolzano. Un approccio sociofonetico. Tesi di Dottorato, Università di Pavia/Bolzano.

Meluzzi, C. (2015). Dialects and linguistic identity of Italian speakers in Bozen. In *Globe: A Journal of Language, Culture and Communication*, 1, 1-16.

Meluzzi, C. (2016). A New Sonority Degree in the Realization of Dental Affricates / ts dz/ in Italian. In Ball, M.J., Müller, N. (Eds.), *Challenging Sonority. Cross-linguistic Evidence*. Sheffield: Equinox, 252-275.

Milroy, J., Milroy, L. (1985). *Authority in Language. Investigating Language Prescription and Standardization*. London: Routledge.

Nese, N. (2016). Le affricate dentali nel collegio universitario 'Giasone del Maino'. Un approccio sociofonetico. Tesi di Laurea triennale, Università di Pavia.

Nocchi, N., Filipponio, L. (2012). Lo vuoi co[z]ì o co[s]ì? A sociophonetic study on sibilants in the regional Italian of Livorno (Tuscany). In Calamai, S., Celata, C. & Ciucci, L. (Eds.), *Sociophonetics, at the crossroads of speech variation, processing and communication*. Pisa: Edizioni della Scuola Normale Superiore, 53-56.

Otheguy, R., Zentella, A.C. & Livert, D. (2007). Language and dialect contact in Spanish in New York: toward the formation of a speech community. In *Language*, 83, 770-802.

Pardo, J.S. (2006). On phonetic convergence during conversational interaction. In *Journal of the Acoustical Society of America*, 85, 2088-2392.

Pardo, J.S., Gibbons, R., Suppes, A. & Krauss, R.M. (2012). Phonetic convergence in college roommates. In *Journal of Phonetics*, 40, 190-197.

Rizzi, E. (1989). *Italiano regionale e variazione sociale: il caso di Bologna*. Bologna: Clueb.

Scotton, C.M. (1985). What the heck, sir: Style shifting and lexical colouring as features of powerful languages. In Street, R.L., Cappella, J.N. (Eds.), *Sequence and Pattern in Communicative Behaviour*. London: Edward Arnold, 103-119.

Telmon, T. (2003). Varietà regionali. In Sobrero, A.A. (Ed.), *Introduzione all'italiano contemporaneo. Vol. II: La variazione e gli usi*. Roma-Bari: Laterza, 93-149.

DOI: 10.17469/O2103AISV000005

DUCCIO PICCARDI

Sociophonetic factors of speakers' sex differences in Voice Onset Time: A Florentine case study

The paper shows the results of a sociophonetic analysis of the so-called *gorgia enfatica* (Castellani, 1960), i.e. the allophonic presence of voiceless aspirated plosives in strong positions attested in the vernacular variety of the city of Florence. 24 native speakers were involved in a production test (read speech) followed by a perceptional counterpart (*matched-guise* and open-ended interview). A statistically significant relationship between male speakers and longer Voice Onset Times emerged from our quantitative analysis of production. This distribution was recognized by the speakers, that evaluated the trait using comments fit for a tentative reconstruction of an *indexical field* (Eckert, 2008). Our data corroborate the need for a sociophonetic shift in the research methods concerning the relation between speakers' sex and VOT production (Oh, 2011).

Key words: Gorgia enfatica, Voice Onset Time, indexical field, gender, sociophonetics.

1. *State of the art*

1.1 Voice Onset Time – speakers' sex patterns

Since the first definition of *Voice Onset Time* (VOT) in Lisker, Abramson (1964), this consonantal subsegmental feature has been studied in relation to many linguistic (voicing, speech rate, pitch, place of articulation, following vowel quality etc.) and social variables. In particular, the history of the studies about speakers' sex differences in VOT shows a recent interesting shift towards research paradigms focused on the observation of the trait as it is structured and motivated in a single local speech community, gradually abandoning explanations based on universal or physiological properties. Swartz (1992) was the first research dedicated to the analysis of VOT-sex patterns. The scholar noticed that female (F) English speakers produce voiceless plosives with significantly longer VOTs than their male (M) counterparts[1]. This pattern was later confirmed by Ryalls, Zipprer & Baldauff (1997), Whiteside, Irving (1997; 1998), Koenig (2000), Whiteside, Henry & Dobbin (2004), Robb, Gilbert & Lerman (2005) and Morris, McCrea & Herring (2008)[2]. It should be noted that not all the mentioned studies are comparable in terms of research

[1] The test consonants of Swarz (1992) were [t] and [d]. Data on this topic were also previously reported in Smith (1978) and Sweeting, Baken (1982). In addition to that, as Swarz himself mentions (Swarz 1992: 984), Molfese, Hess (1978) and Molfese, Molfese (1988) can be seen as examples of the beginning of a structured scientific tradition about the correlations between VOT and speakers' sex, focused in these first attempts on perception and not on production.

[2] A quantitative overview of the results presented in these studies can be found in Oh (2011: 60).

methods, neither they have a uniform evaluation of stylistic phenomena: for example, the status of speech rate remains at least controversial if we compare its non-significant effect in Swartz (1992) with the results presented in Morris *et al.* (2008), that, analyzing monosyllables, loses the statistical significance of the sex opposition. The apparent regularity of this VOT-sex pattern led to conclusions inferred from anatomical differences in speakers' phonatory apparatus, such as men's wider supraglottic space and women's shorter and stiffer vocal folds[3]. Observing the significant distance between female VOT values of voiceless and voiced plosives, Whiteside, Irving (1998) try to further add a broad stylistic and sociolinguistic argument: the authors hypothesize that the pattern is a consequence of an optimization of the voicing contrast in female speech, strictly correlated to a slower and clearer style[4]. This first explanatory tendency, that we may call the *physiological/universal line*, has proved insufficient as a result of a gradual diversification of the linguistic contexts taken into account. The first turning point was the observation of the evolution of VOT-sex patterns during the developmental age. Whiteside, Marshall (2001) found that male English speakers around the age of 9 have a dramatic increase in VOT lengths, not shown by their female counterparts and related to sex-determined physical growth. VOT values become adultlike around the age of 11: from the English male speakers' point of view, this is an inhibition of a physical potential, determined by sociophonetic factors[5]. Nissen, Fox (2009) focused on pre-pubescent children, pointing out that female greater values are established around the age of 5, before the growth mechanism explained by Whiteside, Marshall (2001). Since there is no appreciable sexual dimorphism in the vocal tract at the age of 5, the authors conclude that their results «may be based on male-female archetypes present in adult production patterns» (Nissen, Fox, 2009: 1376). The other fundamental turning point concerns the analysis of non-English-speaking communities. Oh (2011) was the first to find a statistically significant relation between male speakers and longer VOTs, through the observation of the aspirated voiceless plosives of Seoul Korean. The author links his findings to an ongoing change led by female speakers, that are substituting VOT length with f0 values in the phonological cues of the voiceless aspirated class[6]. Oh (2011) explicitly tries to undermine the *physiological/universal line*, promoting the birth of a new *sociophonetic line*: «[...] gender differences in VOT values index sociophonetic gender variations, and [...] sociophonetic patterns can differ across languages or dialectal group» (Oh, 2011:

[3] Both of these features grant male speakers a facilitation in «the formation of sub and supraglottic pressure differences and therefore vocal fold vibration» (Oh, 2011: 60). See Smith (1978), Pickett (1980), Swartz (1992), Titze (1994), Whiteside, Irving (1997; 1998), Koenig (2000) for further discussion on this point.

[4] This claim seems to reflect the results of Scharf, Masur (2002) based on German speakers and those of Peng, Chen & Lee (2014) (Sixian Hakka and Taiwanese Mandarin); however, caution should be taken in the indiscriminate application of this criterion to other cases of VOT differences as cues of phonemic distinction. Alzoubi (2016) shows for example greater emphatic-plain contrast in male Jordanian Arabic speakers (involving VOT length) than in their female counterparts.

[5] A later similar study (Whiteside *et al.*, 2004), does not show the same tendency for male speakers. The idea has been recently rediscussed in Yu, De Nil & Pang (2015), with particular attention to its sociophonetic implications.

[6] This change has been confirmed by Kang (2014) and further extended to the affricates by Lee (2016).

65)[7]. In other words, researchers should evaluate the sociophonetic potential of VOT in the examined speech communities before interpreting distributional data as an outcome of the sexual dimorphism of the phonatory apparatus or common behavioral patterns. In recent years, many scholars added elements to our diatopic knowledge about this correlation. Abudalbuh (2011) found no statistical significance for his F > M VOT results regarding the emphatic plosives of Jordanian Arabic. Swedish does not seem to have any significant VOT-sex pattern, as shown in Lundeborg Hammarström, Larsson, Wiman & McAllister (2012). Li (2013) (Mandarin Chinese of Songyuan) loses the statistical significance of his F > M long-lag VOT tendency after controlling the effect of speech rate, but maintains under the same conditions a strong M > F pattern for short-lag VOTs. Cheng (2013) reports very slight sex-related differences for VOTs in Sixian Hakka, showing non-significant F > M for the aspirated series and M > F for the unaspirated one. This tendencies for Sixian Hakka acquire statistical significance in Peng *et al.* (2014), comparable to a similar trend found for the Mandarin of Taiwan. Madhu, Mahendra & Sreedevi (2014) add to this picture some values taken from Telugu plosives. In this Dravidian language, the authors find a general M > F pattern, even if significant only for few plosive types. In Brinca, Araújo, Nogueira & Gil (2016), we find data about a Romance language, Portuguese, in which pre-adolescent children do not show sex-related differences in VOT length. Coming to the Italian dialectal area, only some Calabrese varieties[8] are covered in relation to our topic. The partial results of Romito and Ciardullo research project in San Giovanni in Fiore (Romito, Ciardullo & Tarasi, 2015) notice a slight, non-significant F > M tendency. A non-significant M > F trend was found in Lamezia Terme by Nodari, 2015, even if a strong interaction between the sex variable and the selected class variable (*orientamento verso la scuola*, "scholastic proficiency") reconstructs a suggestive example of sex-prestige pattern, as the lowest-class male speakers produce the longest VOT lengths, interpreted as a sub-standard trait. As shown by this overview, our knowledge of VOT-sex patterns as they are structured in various languages of the world dramatically increased since Oh (2011), that remains *de facto* the only study showing and motivating a strong M > F tendency for voiceless plosives. Unfortunately, few researchers until now decided to embrace Oh's suggestion about following a *sociophonetic line*, reporting instead unexplained distributional data[9]. In order to tackle one of the biggest challenge of sociophonetics, the distinction between socially meaningful traits and biologically determined ones (see Calamai, 2015: 71), we should

[7] Before the fundamental Oh (2011) few other authors tried to exit the anglophone area without finding statistically significant results (e.g. Bijankhan, Nourbakhsh, 2009, for Persian). Interestingly enough, Oh (2011) does not quote the results of Han (2010), that show English-like VOT-sex patterns in the southeastern Korean city of Daegu. We can consider the comparison between the Daegu and the Seoul distribution as one of the few existing proofs of Oh's claim about the possibility of sociophonetic interdialectal differences in this correlation.

[8] In Calabrese dialects aspirated voiceless plosives are a remarkable local trait (e.g., Sorianello, 1996).

[9] Exceptions can be found in Abudalbuh (2011) and Nodari (2015). However, these authors make sociophonetic inferences from their distributional results, without anchoring production to speakers' perception and evaluation. Cheng (2013: 215) weights up the importance of the observed patterns in explaining some acquisitional trends of foreign brides coming to Hakka-speaking communities, but the topic is no further discussed.

try to answer questions like: what type of indexicality VOT values hold? What is speakers' evaluative perception of the trait? Is there a structurally motivated reason to the selection of a specific VOT-sex pattern? We kept in mind this goal during the analysis of an Italian dialectal trait, the so-called Florentine *gorgia enfatica*.

1.2 The *gorgia enfatica*

Castellani (1960) named *gorgia enfatica* the allophonic presence of voiceless aspirated stops ([k^h, t^h, p^h]) in strong position in the vernacular variety spoken in Florence. The trait was firstly noticed during Scheuermeier's fieldwork and then reported (with a dated and now ambiguous transcription system) in the *Linguistic and Ethnographic Atlas of Italy and Southern Switzerland* (AIS). Giacomelli (1934) tried to further investigate the Tuscan chief-town in order to find confirmation of the allophones. The researcher not only recognized the phenomenon, but also collected an interesting metalinguistic comment about it interviewing a group of native speakers (Giacomelli, 1934: 195-200). Moving to the exegetical hypotheses, two divergent theories about the *gorgia enfatica* have been formulated during the second half of the last century. In Arrigo Castellani's (1952; 1960) substratist view, the *gorgia enfatica* can be found only in Florence and in some areas of the Mugello, and represents the last observable remnant of the Etruscan phonological series of voiceless aspirated stops, reduced in weak positions to fricative allophones (the *gorgia*). However, during the following years, similar consonantal strengthening processes were found not just in other Tuscan areas, but also in the entire Italian linguistic domain (see Izzo, 1972). This led to a series of interpretations that saw the *gorgia enfatica* as not significantly linked to the dialects of Tuscany or to Florence, but generated by prosody (i.e., hyperarticulation in Izzo, 1972; emotional speech in Giannelli, 1976) or universal articulatory gestures (Giannelli, Savoia, 1978; 1979-1980, Giannelli, 1983). These two lines are supported without distributional experimental studies about the phenomenon; moreover, following the anti-substratist exegetical path does not explain the peculiar data registered in the AIS, nor Giacomelli's verification of the trait. This clash between local significance and general interpretations[10], somehow reflecting what we said about VOT-sex patterns, seemed to us the ideal research field to test the exegetical potential of sociophonetic methodologies.

2. *Methods of the sociophonetic investigation*[11]

2.1 Participants, setting and structure of the protocol

24 speakers, all born and living in Florence, were recruited for this research. Most of the participants knew us from a long time; if additional individuals belonging to specific so-

[10] From the point of view of traditional Italian dialectology, a substratum hypothesis loses reliability if the trait under inquiry can also be found outside the geographical area of its supposed linguistic substratum (i.e. the so-called *chorographic* requirement).

[11] What we present here is a thematic excerpt and in-depth analysis of the results reported in Piccardi, (in press). In the op. cit. the reader will find reflections about the historical depth of the trait with an attempt of disambiguation of the problematic use of the term "emphasis". In this sense, a prosodic

cial classes were required, we asked the already agreeing participants to search for plausible candidates[12]. All the speakers involved were aware that they were contributing to a dialectological research about the variety of their city; however, we did not provide further information until the final debriefing (see below). Following the methods of one of the few available sociolinguistic studies on the normal *gorgia*, Cravens, Giannelli (1995), we selected three social variables, M vs F (sex variable), I vs II vs III (age variable, adjusted to reflect a generational criterion[13]) and G vs NG (Graduates vs NonGraduates, class variable[14]). We tested two subjects per each possible combination. The survey was conducted contacting two speakers per session. These dyads were composed by socially homogeneous individuals, already acquainted with each other. In order to minimize auditor effects[15], the only other person on the scene of the interviews was the researcher. The tests were performed in quiet rooms of private habitations. We built for our purposes a three-step mixed methods protocol based on the integration of quantitative analysis of production with qualitative evaluation of speakers' perception. The three phases of the protocol were presented to our participants during a single session, following a fixed order. The duration of each session was approx. 45 mins. The general structure of the protocol was devised with the aim of triangulating our data through *significance enhancement* (Collins *et al.*, 2006: 83-87), i.e., the potential mutual optimization of the understanding of quantitative and qualitative results, and the discovery of causal relationships through the attribution of meaning to quantities.

2.2 Production task

Materials. We prepared a session of read speech consisting in a story-telling task and 60 sentences built around a selection of target words containing /k, t, p/ in the following "strong" contexts: initial (*C-*), syntactic doubling (RF, *Raddoppiamento Fonosintattico*[16]),

variable of emphasis was tested explicitly suggesting an emotive utterance of two of the four frames presented (see below). This aspect of the structure of our research should be taken into account in the comparison between our production data and other sources. VOT lengths discriminated by the prosodic variable can be found in our mentioned paper.

[12] This procedure is called the *friend of a friend method* (Milroy, 1987: 52-56), and is usually considered useful to create a privileged position for the observation of naturalistic behaviors.

[13] The cut-off ages selected in Cravens, Giannelli (1995) were 15-25, 35-45 and above 65. Means and standard deviations of our revised categories are as follows: I (25; 1); II (47,5; 5,17); III (73,5; 7,34).

[14] About the class variable, we decided to reformulate the binary partition chosen in the op. cit. (blue collars vs white collars) using Hudson's idea of "sophistication" (Hudson, 1996: 198-199). Hudson proposed a macrosocial division based on the stereotypes of "rough" and "sophisticated" people. This division have predictable consequences on the sex-prestige pattern, and is appreciably practical since it excludes census, focusing on cultural and occupational traits of the speakers. We decided to not introduce other layers of social distinction to maintain a degree of comparability with the previous (AIS etc.) twofold distributions. In our research, a speaker was considered "sophisticated" if a) holding a university degree, and b) having an occupation that requires the aforementioned degree. Conversely, a "rough" speaker was selected if a) not holding a university degree, and b) having an occupation that does not require any kind of degree. Hybrid cases were excluded from this bipartition.

[15] This terminology was introduced by Bell (1984) to describe the effects on style-shifting of the presence on the scene of a conversation of other interlocutors besides the addressee.

[16] See, e.g. Loporcaro (1997: 41-51).

gemination (-*C*:-), postconsonantal (-C*C*-) and postsibilant (-s*C*)[17]. Each target-word was repeated in 4 different frames, presented to the speakers written in a morphosyntactically marked variety of Regional Tuscan Italian. For this first experimental verification of the *gorgia enfatica*, the words were chosen from the available bibliography (AIS; Castellani, 1960) reporting impressionistic data. This study is thus not concerned with potentially conditioning factors, such as word length, frequency, prosodic structure and the quality of the vowels following the plosives[18]. Anyways, the words were not placed in prepausal positions, so as to rule out the effects of contextual lengthening.

Procedure. The dyads were first invited to have an informal conversation in the room of the interview (approx. 20 mins.). We tried to limit our direct verbal interactions in the conversations: in fact, this phase was devised with the aim of mitigating the participants' *attention to speech* (see, e.g. Labov, 1984), accustoming them to have spontaneous discussions while we were observing the scene. Then, a sheet of paper with the texts of the short story and the 60 sentences was given to each participant. The speakers were asked to read the written contents addressing the other participant, using an informal tone. They were allowed to read silently the proposed excerpts before pronouncing them. One single production was digitally recorded for each speaker, using a SONY ICD-UX533.

Analysis. In order to increase the number of our tokens, all voiceless plosives in strong positions elicited during the production phase were analyzed in PRAAT 6.0.14 through manual annotation of spectrograms. Our corpus consisted of 1041 /k/, 2287 /t/ and 738 /p/. Only VOTs preceded by a release burst were considered valid (Marotta 2004)[19]. Measurements were taken from the release burst to the onset of periodicity in the acoustic waveform (Francis, Ciocca & Yu, 2003). The reported results will not consider potential effects of speech rate on VOT lengths[20]. A series of one-way ANOVA statistical tests were performed in Matlab R2016a[21], with the aim of observing the effects of sex ($df = 1$), class ($df = 1$) and age ($df = 2$) comparing the VOT means for each phoneme in each linguistic context. We controlled the family-wise inflation of type I error adjusting the α significance level with a Bonferroni correction, dividing it by the number of performed tests (n), i.e. $\alpha/n = 0{,}05/45 = 0{,}001$.

[17] This last context was discriminated from the generic postconsonantal one to better evaluate the potential inhibition on the production of long-lag VOTs triggered by assimilations of manner and laryngeal features (Izzo, 1972: 107).

[18] For this last aspect, some observations about the lengthening effect of the following high vowels have been presented in Piccardi (in press). The tested words are listed below; <ʰ> stands for the position of the aspirations discussed in our sources: *chʰe* ("that"), *cʰane* ([iˈk:ʰane], "the dog"), *tabaccʰo* ("tobacco"), *mercʰato* ("market"), *nascʰondino* ("hide-and-seek"), *tʰu* ("you"), *sottʰerrare* ("to bury"), *tʰabernacolo* ([i t:ʰaberˈnaholo], "the aedicule"), *soltʰanto* ("only"), *stʰazione* ("station"), *pʰettinare* ("to comb"), *pʰaese* ([i p:ʰaˈeze], "the village"), *ceppʰo* ("stump"), *un pʰo'* ("a bit"), *aspʰettare* ("to wait").

[19] A small subset of plosives showed indeed a short release phase without an audible release burst. These were classified as belonging to a different phonetic category (unreleased plosives; see e.g. Mioni, 2001: 39) and thus excluded from our computation.

[20] We thank an anonymous reviewer for providing us with the reference to a recent paper discussing the need for speech rate normalization in VOT measurements (Nakai, Scobbie, 2016). We intend to take part to this ongoing debate in a separate publication.

[21] We deeply thank Dr. Federico Becattini (University of Florence) for the assistance in this part of the research.

2.3 Perception tasks[22]

Materials. Three pairs of short acoustic stimuli (approx. 5 secs. each), one per plosive point of articulation, were prepared for an informal matched-guise session (Lambert, 1967), recording the voice of the researcher (native speaker of the Florentine dialect) with the same, above stated equipment. The variables were situated in new (i.e., not previously presented) lexical items, with the aim of dismissing the possibility for eventual production-perception correlations of being constrained at the word level. The phrases had a content typologically homogeneous to the tone of the excerpts used for the production phase, i.e. light-hearted and conversational. First, three sentences were spontaneously read by the researcher while not producing the *gorgia enfatica*. These three excerpts were labelled as non-aspirating guise, also serving as a VOT durational baseline ([k]: 28 ms., [t]: 20 ms., [p]: 17 ms.) for the production of aspirated counterparts. The aspirating guise was enacted trying to replicate the sentences using the *gorgia enfatica*, aiming to add approx. 30 ms.[23] to the VOT baselines while maintaining a comparable intonational pattern. The best VOT candidates ([k^h]: 62 ms., [t^h]: 51 ms., [p^h]: 52 ms.) were selected checking the spectrograms in PRAAT. The last part of our three-step protocol consisted in an open-ended interview (Marcato, Ursini & Politi, 1974: 28-30) about the metalinguistic competence of our participants. A topic guide was previously written and printed on a sheet of paper, readable only by the researcher. The guide was structured so as to progressively narrow the focus of the discussion, starting from topics of very broad linguistic interests and ending with a metalinguistic commentary on the trait under scrutiny, not without a direct confrontation with our research questions[24]. Six main topics of discussion were tackled during the interviews: 1) local terminology for non-standard speech varieties; 2) situations considered suitable for the above mentioned varieties; 3) non-standard traits emerging in the above mentioned situations; 4) local terminology for the most salient of the above mentioned traits, i.e. the *gorgia*; 5) linguistic contexts of the above mentioned trait; 6) [contrastively] metalinguistic discussion about the phenomena occurring in the specular contexts, i.e. the *gorgia enfatica*.

Procedure. The recorded excerpts were reproduced in a media player through the integrated speakers of a laptop (Asus SonicMaster). The six stimuli were presented as pairs of different guises; however, no internal order of reproduction was fixed. The rhythm of the test was dictated by the demands of the hearers, and multiple reproductions of the materials were permitted. Two open-ended questions were asked to the participants: "Are there any differences between the first and the second recording, and, if yes, what do they

---

[22] This phase of the inquiry has been inspired by the qualitative component of Campbell-Kibler's (e.g. 2007: 33-39) classic matched-guise protocol, reorganizing her open-ended questions, explicit identification of the changing variables and guise opening.

[23] We choose this value in the 10-40 ms. of the *Just Noticeable Differences* (see Calamai, Ricci, 2005: 73-74) in terms of length considering that «the subjects are much more sensitive to changes in vowel duration than to changes in consonant duration» (Huggins, 1972: 1270).

[24] This top-down procedure is usually called *funneling*, «an indirect way of approaching the heart of the matter, and may be used as a way of manipulating the respondents' degree of awareness of the object of inquiry» (Kristiansen, 2010: 538), enhancing it.

consist in?"[25]; "What can you tell me about speaker X and speaker Y?". Participants' answers were recorded and summarized taking written notes. The same was done during the open-ended interview. The last, more important point of the discussion was presented in three consecutive phases. First, participants were invited to reproduce the *gorgia enfatica*, instructing them to emit a puff of air after the normal consonants. Then, the previously heard guises were opened, i.e. the participants were made aware that they were evaluating the same person enacting two different personae[26]. Finally, they were asked to directly comment on our research questions. More information on the research were provided to those who manifested interest.

Table 1 - *VOT values of the Male (M) and Female (F) classes: Means (μ) and Standard Deviations (σ) calculated in ms. Statistical significant values after the Bonferroni correction can be found in the black frames*

	μ M	*σ M*	*μ F*	*σ F*
k-	39,1	10,1	36,1	13,6
RFk	39,2	12,6	37,3	18,8
k:	38	13,4	39,4	14,8
Ck	34,4	10,3	30,3	10,1
Sk	29,6	10,6	27,7	9,8
t-	29,2	9,7	23,9	9,5
RFt	25,2	10,4	20,4	8,8
t:	27,3	10,5	23,3	12,4
Ct	23	8,3	19,5	7,8
St	19,7	6,6	17,9	5,6
p-	14,9	5,8	12,1	6,1
RFp	16,5	7,5	15,6	11,3
p:	19	7,3	15,4	7,7
Cp	18,2	7,6	19	9,4
Sp	13,4	5,5	10,4	5,4

[25] This short AX discrimination task could be seen as based on a rhetorical question, since the stimuli were created *ad hoc* to be discriminated. The discrimination was instrumental to a) rule out the presence of ideological processes inhibiting participants' will to point out differences (i.e. *erasures*; Irvine, Gal, 2000); b) focus participants' attention on the analyzed variable.

[26] The aim of guise opening is to retrieve impressions on the variable as an individual stylistic resource (Soukup, 2013).

3. *Results*

3.1 The quantitative analysis of production

A general M > F tendency is observable throughout our considered contexts. Some of them hold statistical significance even after the Bonferroni correction (see Tables 1-3): -C*k*- (F = 11,84, p = 0,0006); *t*- (F = 16,52, p = 0,00007), -RF*t*- (F = 17,38, p = 0,00004), -*t*:- (F = 17,61, p = 0,00003), -C*t*- (F = 41,64, p = $1{,}97^{-10}$); -s*p*- (F = 11, p = 0,001)[27].

Table 2 - *VOT values of the NonGraduated (NG) and Graduated (G) classes: Means (μ) and Standard Deviations (σ) calculated in ms. Statistical significant values after the Bonferroni correction can be found in the black frames*

	μ NG	*σ NG*	*μ G*	*σ G*
k-	40,1	11,6	35,1	12,1
RFk	41,3	17,2	35,1	13,4
k:	42,5	14,2	35,6	13,3
Ck	34,3	10,7	30,5	9,8
Sk	30,6	10,8	26,1	9,1
t-	30,3	10,5	23	7,9
RFt	25,1	10,9	20	7,6
t:	27,2	12,7	23	10
Ct	23,6	8,7	19,3	7,4
St	19	6,2	18,5	6,1
p-	14,2	6,5	12,7	5,6
RFp	16,6	10,4	15,5	8,4
p:	16	8,9	18,4	6,1
Cp	19	8,4	18,3	8,8
Sp	12,9	6,7	10,9	4,1

Our data confirm the "vernacular" distribution suggested by the AIS, Giacomelli and Castellani. A constant pattern NG > G is evident in our corpus, with widespread statistical significance: *k*- (F = 11,58, p = 0,0007), -*k*:- (F = 11,81, p = 0,0007), -C*k*- (F = 9,95, p = 0,001); *t*- (F = 30,48, p = $1{,}19^{-7}$), -RF*t*- (F = 16,68, p = $5{,}76^{-5}$),

[27] *k*- (F = 3,99, p = 0,04), -s*t*- (F = 8,69, p = 0,003), *p*- (F = 6,24, p = 0,01) and -*p*:- (F = 8,91, p = 0,003) lost their significance after the adjustment.

-t:- (F = 26,49, p = $3{,}61^{-7}$), -C*t*- (F = 49,26, p = $5{,}05^{-12}$)[28]. Few contexts proved significant for the age variable, none after setting $\alpha = 0{,}001$; moreover, no general tendencies can be observed[29]. This led us to search for particularly relevant oppositions between the productions of subsets collocated in single age categories, developing differences on the sex and class axes. This seemed to be the case: three radical oppositions emerge from one-way ANOVA tests comparing the means of all VOT values between specific social subsets, i.e. M I G > F I G (F = 94,92, p = $5{,}01^{-21}$), M II NG > M II G (F = 94,79, p = $6{,}03^{-21}$) and M III NG > F III NG (F = 67,88, p = $9{,}59^{-16}$).

3.2 The qualitative analysis of perception[30]

3.2.1 The matched-guise test

Every speaker was able to perfectly recognize the segmental place of variation and the modalities of the variation itself. Unfortunately, our 30 ms. VOT distinctions were often considered an unnatural emphasization of "florentinity" or a clumsy imitation of the Florentine dialect made by a foreigner (see below). Nevertheless, 13/24 speakers recognized [k^h] as "more Florentine"; similar comments were elicited by 10/24 speakers for [t^h] and by 2/24 speakers for [p^h]. This scale of proneness to identify consonantal allophones reflects the metalinguistic consciousness about the targets of the weakening process of the normal *gorgia*[31]. Moreover, as it is shown by the reported means, bilabial allophones with a perceivable aspirated release phase are indeed rare. The social coordinates attributed to the aspirated allophones, if defined as "more Florentine", were uniform and consistent with the distribution of production values. The *gorgia enfatica* was considered manlier and more appropriate in informal situations. Interestingly enough, only male speakers specified those situations as funny conversations among friends. Three speakers evoked very specific images related to the utterance of aspirated allophones. F III NG 2 remembered the face of his grandfather, slightly moved. She reproduced his voice replicating the aspiration process on a new lexical item, [ink^{h}anˈtina] ("in the cellar"). F II NG 2 thought about an old husband tiredly speaking to his wife. She also described the setting of her image, a typical Tuscan living room. M III G 2 identified the aspirat-

---

[28] -RF*k*- (F = 8,49, p = 0,003), -s*k*- (F = 5,70, p = 0,01) and -s*p*- (F = 4,62, p = 0,03) lost their significance after the adjustment. /t/ aspirated allophones seem to hold a considerable degree of potential social distinctiveness (Hudson, 1996:249). Along this line, we noticed some reconsiderations in the distinctiveness view of aspirated allophones after hearing the /t/ ones.

[29] Singh, Keshet, Gencaga & Ray (2016) come to a similar conclusion with the analysis of a large corpus of 630 American English speakers. The absence of biologically determined VOT-age pattern does not imply the impossibility for VOT values to index a social category linked to age criteria.

[30] In the next two sections, we will briefly comment only the perceptual results instrumental in explaining the retrieved VOT-sex pattern. For further details, see Piccardi (in press).

[31] The exact quantities of the reactions are reported here for the sole purpose of exemplifying the different impacts on the metalinguistic consciousness of the participants caused by the three plosives. The proposed qualitative data should be interpreted as a hint towards the coexistence of exaggeration and dialectality in a single trait.

ing voice as a father speaking in an authoritarian way to his son. Upon our request to mentally try to switch the sex of the imagined figures (i.e., grandmother, wife, mother to daughter), the speakers firmly denied the possibility: in their sensibilities, the aspirated allophones were linked to an emotional universe connoted as virile. Finally, older speakers were more inclined to identify the aspirated allophones as part of the Florentine diasystem, followed by the young participants and the members of the second age group (III > I > II).

Table 3 - *VOT values of the First (I), Second (II) and Third (III) age classes: Means (μ) and Standard Deviations (σ) calculated in ms*

	μ I	*σ I*	*μ II*	*σ II*	*μ III*	*σ III*
k-	35,2	12	39,3	13	38,3	11,1
RF k	41,9	19,7	37,8	11,7	34,3	12,4
k:	40,1	14,8	40,3	13,5	36,6	13,8
Ck	33	11,5	32,4	9,9	32	10
Sk	29	11,3	26,8	9,6	29	9,7
t-	27,1	11,3	28,1	10,9	24,2	7,1
RF t	22,1	9,9	24	11,2	22,3	8,2
t:	26,1	12,8	26,4	11,9	23,4	9,7
Ct	21,7	8,4	20,8	8,5	21,1	7,9
St	18,2	6,9	19,9	6	18,1	5,3
p-	14	5,7	14,4	7,6	12	4,5
RF p	16,7	11,3	15	7,9	12,7	7,9
p:	18	7,7	17,5	7,7	15,8	7,6
Cp	20,9	8,9	17,1	7,1	18,1	9,2
Sp	12,6	6,7	11,8	5,3	11,2	4,6

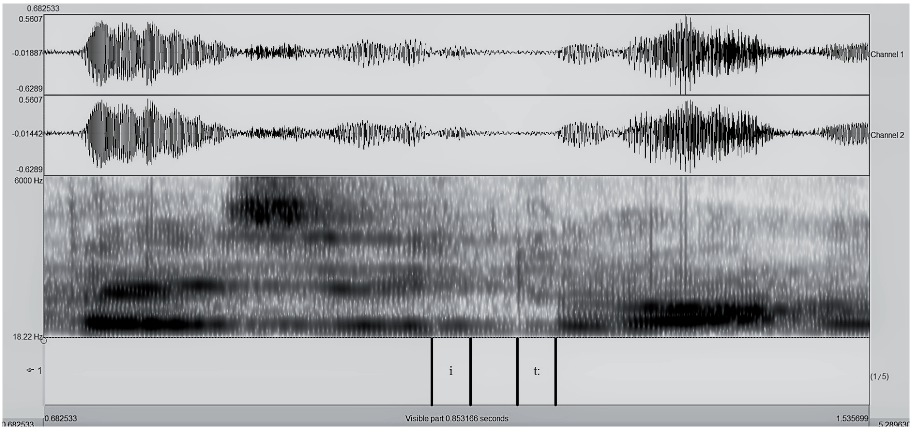
3.2.2 The open-ended interview

This order was not reflected by the metalinguistic competence emerged from our open-ended interviews. In our results, metalinguistic competence increased with the level of speakers' education (G > NG) and decreased with age (I > II > III)[32]. Probably the most interesting recurrent element emerged from our interviews is the referential ambiguity in the popular use of the word *aspirazione* ("aspiration"). *Aspirazione* has been used to refer to the Tuscan spirantization phenomenon since the XVI century, and is still today much more popularly widespread than the term *gorgia*. As Izzo, 1972 pointed out, this use of the term can no longer be considered acceptable in scientific contexts, for its potentially confusing ambiguity with proper aspiration processes (i.e., IPA [ʰ]). Following the informal nature of the open-ended interview model, we used with our speakers *aspirazione* as a synonym of *gorgia*.

[32] These impressions were also anchored to a quantification of the traits suggested during our third point "What are the elements of your speech that you use in the situations you previously considered fitting for the dialect?". No significant differences were observed between M and F speakers.

However, we soon noticed that the speakers spontaneously extended the referential field of the word to the observation of real aspirates in strong positions, i.e. the *gorgia enfatica*. These unexpected extensions happened 11 times with different speakers, that introduced new lexical items in the conversation as examples of what they meant with *aspirazione* (*gorgia enfatica*; see Piccardi, in press, for further details). We had to infer that for our speakers *aspirazione* did not refer to a precise phonetic identity, but to an acoustical impression of a continuous sound that partially or totally places itself in the slot of a St. Italian plosive[33]. Coming to the last point of our interview, the explicit metalinguistic reflection about *the gorgia enfatica* with a guided pronunciation of the phenomenon, all the members of I were instantly capable to comment and identify the trait[34]. As we already mentioned, this ability decreased with age (II: 5/8; III: 3/8).

Figure 1 - Paese di tuo nonno *("village of your grandfather") pronounced with RF by M I NG 1. VOT length of [t^h]: 41 ms*

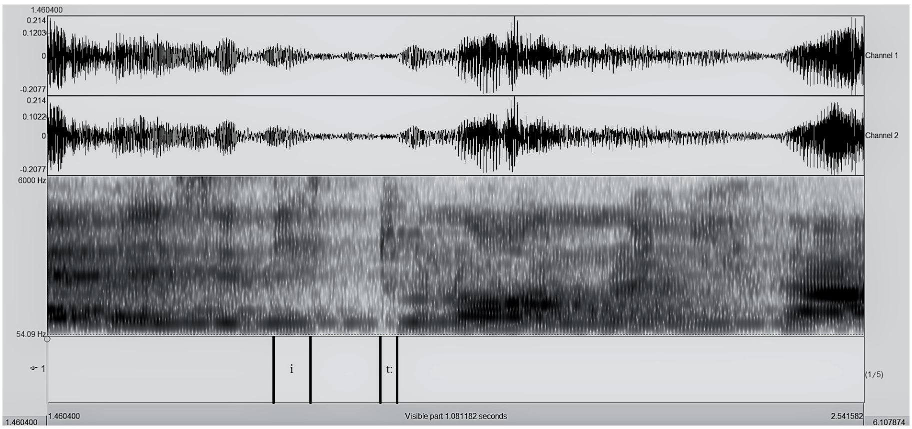


The main social coordinate attributed to the trait was "typical of male speakers", emerged in 12 interviews. Among these, 5 female speakers further specified that they heard the trait uttered by rustic, rude men (NG). Six interviews also presented a secondary attribution to III, class considered the depositary of archaic, authentic and rustic linguistic traits[35]; attributions to I were implicit in some diaphasic con-

[33] The equivalence is plausible from an acoustic point of view, from the moment that [h] (the most typical and recognized realization of the Florentine *gorgia*) and [^{h}] share the same properties (Stevens, 1999: 451).
[34] All the data and impressions presented below were gathered during the phase before the last overt discussion on our research questions.
[35] An anonymous referee advanced the intriguing hypothesis that speech rate could be a determining factor in this association. From the moment that a) speech rate reportedly decreases with aging (Yuan, Liberman & Cieri, 2006); b) tempo has a strong effect on the perception of speakers' age (Skoog Waller, Eriksson & Sörqvist, 2015); c) VOT length can increase as speech rate decreases (Summerfield, 1975); hearers could possibly link long VOTs with their occurrences in slow speech styles, and thus attribute them to old speakers. At this point of the research, we have no concrete elements to evaluate this lead. During the open-ended interview, our participants were reflecting on their own guided pro-

siderations. *Gorgia enfatica* was in fact perceived as a trait used in informal situations among friends in 8 interviews. At this point, any references to members of the family disappeared. Of those 8 interviews, 6 were with male speakers and the other 2 (young women) explicitly recognized the trait as used in groups of male-only friends. Finally, we found a M I tendency to celebrate the trait ("a sign of Florentine pride": M I NG 1; "a way to express our peculiar way of being Florentine": M I G 1) opposed to a F I one to firmly reject it ("yuk! It's only for boys": F I NG 2; "it's just for rude old countrymen!": F I G 2).

Figure 2 - *Paese di tuo nonno ("village of your grandfather") pronounced with RF by F I G 1. VOT length of [t]: 22 ms*



4. *From sex to gender*

In our research, we saw a degree of consistency between the distribution of long-lag VOTs revealed by production data and the evaluation suggested by our speakers. We can now try to move to an exegetical section and see if there are any socially structured reasons for this pattern, overriding supposed biological factors, and if aspiration can be invested by a peculiar set of social meanings. In order to dwell on these aspects, we should interpret Oh's reference to the speaker *gender* not just as an unmarked synonym of biologically determined *sex*, but in the full sense of «social elaboration of biological sex», «not something we are born with, and not something we have, but something we do» (Eckert, McConnell-Ginet, 2003: 10). We will therefore shift from now on the meaning of our M and F from predetermined tags of social categories to social constructs composed by variable bundles of meanings (Kiesling, 2013: 461).

duction of long VOTs; moreover, the connection with speakers' age was never elicited independently from the ones with speakers' sex and conversational context. In our first approximation, we feel more confident in considering the reference to old speakers as a function of the general sex-prestige pattern.

4.1 Explaining the roleplay

Somehow replicating the experience described in Giacomelli (1934), a key element to stimulate metalinguistic reflection about our analyzed trait was the phase of guided production of it. The importance of production for evaluation is one of the most explicit prerequisites (Iannàccaro, 2002: 137-138) for the insertion of a linguistic trait in the Labovian class of *markers*, composed by elements «not at the same high level of social awareness [of stereotypes], but show[ing] consistent stylistic and social stratification» (Labov, 1994: 78). A median position in the three-step Labovian scale of awareness does not fully explain the articulate social roleplay emerged from our interviews. Starting from the elder speakers, ANOVA revealed a radical opposition in VOT lengths between the subclasses M III NG and F III NG. What we call here F III NG is a social category often cited in perceptual dialectology[36] to typically hold a grudge against dialect and substandard linguistic traits, considered potentially humiliating and detrimental to the social promotion of the family. Our interviews delineated a similar picture: F III NG 2 admits being embarrassed by her inability to speak in St. Italian in shops and other everyday situations; F III NG 1 does not understand our interest in dialect and wishes it could just disappear. Directly connected is the scarce occurrence of the dialectal art. m. sg. def. form [i] in the speech of these two speakers: [i] is one of the few perceived trigger of RF, commonly evaluated as local and informal (Agostiniani, 1992). Our data thus suggest that M III NG > F III NG could be a simple case of sex-prestige pattern, characterized by the female avoidance of long-lag VOTs recognized as manly in the sense of substandard[37] and rustic. The absence of further specified ideological reinterpretation of the trait cannot be extended to the opposition observed in I, M I G > F I G. Our interviewed young Florentines showed a deep knowledge about the distribution and plausible contextual use of the trait, i.e. fit for informal conversation in a group of male-only friends. We can consider this kind of competence an example of what Iannàccaro (2002) calls *consapevolezza linguistica* ("linguistic awareness"), as it is opposed to *coscienza linguistica* ("linguistic consciousness"). Linguistic awareness about a trait implies a certain degree of metalinguistic knowledge about it, whereas linguistic consciousness does not: this is the case of the propensity shown by elder speakers to evaluate the trait during the matched-guise session, but not during the open-ended interviews[38]. It seems that

[36] For further references, see Iannàccaro (2002: 116-117).

[37] This claim is reinforced by the data reported by Sorianello (1996: 137). VOT means of voiceless plosives in strong positions uttered by Florentine speakers of St. Italian are far shorter ([pː]: 9,25 ms.; [tː]: 14,25 ms.; [kː]: 27 ms.; [rp]: 10,5 ms.; [rt]: 15,50 ms.; [rk]: 26,25 ms.) than those of our Florentine dialect.

[38] It should be noted that we are interpreting here both *consapevolezza* and *coscienza linguistica* as two different levels of the same category of linguistic markers. The *gorgia enfatica* was never explicitly mentioned in the higher, more general and less induced layers of the open-ended interview; this leads us to think that the trait is not commonly available for overt commentary, not even for the young Florentine speakers (i.e. it is not a stereotype in the Labovian acceptation of the term).

young male Florentines intentionally use long-lag VOTs to index the belonging to a group of friends; in other words, the *gorgia enfatica* could represent an element of male *youth language* in Florence[39]. The use of substandard traits in male *youth language* to convey a sense of aggressiveness, rusticity and idealized manliness is a solid acquisition of dialectology (see, e.g. Cortelazzo, 1995: 585). The enthusiastic comments that we retrieved about the *gorgia enfatica* are a good proof of the *covert prestige* (Trudgill, 1972; Labov, 1990) that the trait holds among young men. From this point of view, the firm rejection of long lag VOTs by F I G speakers is not surprising. F I G is a subcategory of speakers typically associated with *overt prestige* traits (Cravens, Giannelli, 1995), considered the key of self-promotion and social acceptance. Nielsen (2011), developing the idea of finding traces of episodic memory in speech production, proved that VOT lengths can be a target of the so-called *imitation principle*: consistent exposure to lengthened VOTs induces the speakers to replicate the subsegmental trait in the production of new lexical items involving new types of plosives. In this sense, young people can easily take long-lag VOTs from a sensible distribution among the elders, and build an identity feature on them. The step from the evaluation of the trait among the elders to that of the young, idealizing manliness in the search for self-identity, is describable in terms of indexicality (Silverstein, 2003) by an opposition between a N^{th} and a $N^{th}+1$ order[40]. The second age class remains with an indefinite status, conditioned by the personal biography of the single speaker and by the typology of *social networks* (Milroy, Milroy, 1978) linked to him. However, in what seems to be an *age-grading*[41] distribution hinting to a linguistic evolution during the lifespan of the speaker without stable age classes in the corresponding society, our third radical opposition, M II NG > M II G, could be interpreted as the snapshot of a *retrenchment* phenomenon, «a retreat from the non-standard variants used in youth by stabilization» (Evans Wagner, 2012: 375)[42]. We can conclude this section noticing how the potentiality of aspiration to index subsets of the primary less-marked ideological reference[43] could be considered an example of what Irvine, Gal (2000: 38) calls *fractal recursivity*, "the projection of an opposition, salient at some level of relationship, onto some other level".

[39] Interestingly enough, Giacomelli (1934) describes a spontaneous conversation in a group of male young friends from Florence and Vicchio in which he heard frequent productions of voiceless aspirates.
[40] Even if the relation between the two orders is clear, we are not able to determine the exact indexical orders without further analysis about the linguistic ideologies associated to the trait. For some suggestions, see below.
[41] «Apparent time studies in which men are the most frequent users of a variant might be more indicative of age grading than change» (Evans Wagner, 2012: 374).
[42] Of course, the only way to confirm linguistic retrenchments would require a large-scale real-time research observing the behavior of a group of individuals during the passage from adolescence to adulthood (Chambers, 2003: 199). Our terminology should be intended as resulting from the interpretation of the distributional results of a cross-sectional study.
[43] In our case, the general idea of virility can further index the belonging to social subgroups of men not involved in "sophisticated" activities, and possibly the identification of boys in adolescent peer groups.

4.2 Exaggeration as a pathway to manliness

Eckert, McConnell-Ginet (2003) gives great emphasis to the importance of exaggeration of biological differences in the construction of both masculine and feminine genders (see, e.g., op. cit.: 10, 13, 20). In particular, social psychology identifies a fundamental role of exaggeration in the expression of manliness. Desirable and acceptable masculine traits have a narrower definition than their feminine counterparts; moreover, masculinity tends to be more respected than femininity[44]. These facts lead manliness to an overexposure to social threats, and ultimately to defense mechanisms based on the so-called *masculine overcompensation thesis*. We saw how the use of dialectal traits is considered by young male speakers a viable metalinguistic stratagem to convey a sense of rudeness and idealized virility. However, in Florence the most marked local trait, the spirantizing *gorgia*, is far from being stigmatized. Recent studies (Marotta, 2014: 159; Biliotti, Calamai, 2012) suggest that a tendency to consider the *gorgia* an *overtly prestigious* trait is spreading throughout Tuscany. In this linguistic context, male *youth language* clearly needs to select, or create, something else. Italian linguistics reports examples of two different types of what we may call *linguistic exaggeration*. Franceschi (1969) analyses the historical consequences of *interdialectal* exaggeration: an observable overuse of a linguistic feature, targeting historically unmotivated contexts, could point to the late arrival of the feature in the diasystem as an erroneously replicated linguistic trend. In Franceschi's examples, we also find an ironical exaggerated imitation of the stereotypical *gorgia* by foreigners, [tosˈk^{h}ana] ("Tuscany"; Franceschi, 1969: 53). This seems to reflect the comments often generated by our unrealistically lengthened VOTs used in the matched-guise phase, hinting to the possibility of a folk reinterpretation of spirantized allophones in weak positions as aspirated plosives in strong positions. The other typology is *intradialectal* exaggeration. In 1979, Craffonara noticed some historically unexplainable palatalized realization of /ka, ga/ in Dolomitic Ladin. The author argues that those individual and irregular occurrences could be an attempt to convey a sense of local identity through overextension of the most recognizable dialectal feature, i.e. palatalization (Craffonara, 1979: 71). Even more pregnant for our exegetical attempt, Del Puente (1995) shows a clear decline of the metaphony in the dialect of Naples. However, few social categories prove the existence of a countertrend that even lead to an overextension of the metaphonized variants to etymologically unmotivated contexts. These categories consist in young low educated male speakers of the peripheral neighborhood and old male speakers of the same extraction. The trait holds *covert prestige* and is used to convey a sense of being an authentic Neapolitan man. These exaggerations are explicitly recognized and commented by the female counterparts; moreover, a *retrenchment* phenomenon is observed (Del Puente, 1995: 55-56), completing a picture strikingly similar to ours. Considering the formal implications of the example found in Franceschi (1969),

[44] See Willer, Rogalin, Conlon & Wojnowicz (2013: 984-985); Cheryan, Schwartz Cameron, Katagiri & Monin (2015).

and the referential ambivalence of the word *aspirazione*, that hints to a unified folk view of all the types of plosive allophones, *gorgia enfatica* could be born as a *covertly prestigious intradialectal* exaggeration of the overtly prestigious spirantization, the *gorgia*. In addition to that, aspiration seems to hold some ideological implications that make the trait an optimal candidate for the described purposes.

4.3 Why aspiration? Plausible *addenda* to the indexical field

In 2012, Penelope Eckert formalized the birth of a third wave of studies on social meanings attributed to linguistic variation. The third wave does not link evoked social meanings to general sociolinguistic primary categories (first wave), nor to local categories (second wave), but believes that

> variables cannot be consensual markers of fixed meanings; on the contrary, their central property must be indexical mutability. This mutability is achieved in stylistic practice, as speakers make social-semiotic moves, reinterpreting variables and combining and recombining them in a continual process of bricolage [...] (Eckert, 2012: 94).

This multilinear view of indexicality finds its main tool of formal representation in the *indexical field* model (Eckert, 2008), "a constellation of ideologically linked meanings, any region of which can be invoked in context" (Eckert, 2012: 94). Fully released, aspirated /t/ allophones are the core of one of the finest examples of the exegetical power of the *indexical field* (see, e.g., Eckert, 2008: 469; Drummond, Schleef, 2016: 57). These allophones in fact have been described in a good number of different social contexts as peculiar indexes of social meaning. Bucholtz (1996: 125) finds /t/ releases indexing intelligence and resistance to trendy speech styles in a group of high school "nerd" girls (and boys) in San Francisco. Benor (2001) attributes to the high frequency of aspirated final /t/ in the speech of boys in an Orthodox Jewish school of North Carolina a desired trait of erudition: aspiration is in fact not only distributed by gender, but also by level of instruction, being more frequent in students of the Yeshiva. In Podesva's studies on gay speech (see, e.g., Podesva, 2007), the author observes the linguistic behavior of a gay doctor. This doctor produced more aspirated /t/s in the clinic than in informal situation, indexing precision and professionality. However, during a barbecue with friends, the doctor uttered less fully released /t/s, but with much longer VOTs: he was using a parodic exaggeration of his professional persona, indexing exasperation fit to represent a "gay diva". Eckert (2012: 97) expands on these findings suggesting that aspiration, as a fortition phenomenon, «can index emphasis or force, hence focus, power, or even anger». We found a similar phenomenon of *iconization* (Irvine, Gal, 2000) in some of our interviews. In an attempt to disambiguate the two differently perceived types of *aspirazione*, F II NG 2 defines [t^h] as "aspirated... in the *strong* sense"; more significantly, evaluating the *gorgia enfatica*, M II G 2 says that "it's a manly pronunciation... women prefer *weakened* forms". Drummond, Schleef (2016) adds a proof to this argument with the indexical field reconstructed for the intervocalic t-glottalling, that shows a central branch tagged as "more laidback". The authors also try to unify the distributions of aspiration reported in the quoted studies noticing that "the meaning of the feature must relate to something shared by all these [groups or identities]; they all exploit indexical values linked to hyperarticulation (i.e. extremely articulate and clear speech)" (Drummond, Schleef, 2016: 53). We saw how hyperarticulation is often interpreted as a path to express intelligence and precision: from this point of view, the distribution and evaluation of the *gorgia enfatica* as rustic and informal seems at least controversial. However, as Podesva (2013: 428) argues about non-modal phonation types, "social meanings [...] are culturally specific and should not be reduced to purely iconic or unanalyzed associations [...]". Hyperarticulation is also a correlate of *stancetaking* (Freeman, 2014), intended here as "modality, evaluation, subjectivity, epistemicity, footing, alignment, assessment, agreement, and so on, [used] to refer to a speaker's or writer's attitude, displays of emotions and desires, expressions of beliefs and certainty toward given issues, people, and the speakers' coparticipants" (Haddington, 2004: 103). The comments of our speakers suggest some level of *stance accretion*[45] based on this correlation. M III G 2 thinks that the *gorgia enfatica* is a virile communicative strategy to stress an opinion, enriched by a feeling of irritation. M I NG 2 is surprised by the perceived coincidence between a trait used to clarify concepts in professional environments and an acoustic memory of aspirations used in friendly speech during the adolescence. M II G 2 is confused by a trait that conveys at the same time dialectality and firmness, precision. We can thus assume that Florentine men may find in long-lag VOTs a way to express an identity of firm stance-takers, decision-makers, ideologically dynamic people.

5. *Conclusions*

Shifting from production to perception, from sex to gender, we noticed how an instrumentally retrieved, statistically significant distribution of an allophone could potentially be the core of an ideological constellation, imbued with social meanings at various levels of analysis. From this point view, Oh's suggestion to link distributional studies on VOT-sex patterns to in-depth sociophonetic descriptions of the chosen speech community is still actual and exegetically promising. Much work is yet to do on both the production and the perception of the *gorgia enfatica*; however, the integration of these two aspects of the research in a unified protocol proved itself useful for a first level of description of the social factors conditioning the observable distribution of aspirated voiceless plosives in the Tuscan chief-town. "People (perhaps especially men) perform gender" (Johnson, 2006: 487), and sociophonetics has the potential to explain these performances.

[45] «The way in which stances accumulate in more durable structures of identity» (Bucholtz, Hall, 2005: 596). In other words, it is for example the process through which people frequently showing joyful behaviors can be seen *tout court* as joyful.

Bibliography

ABUDALBUH, M. (2011). Effects of gender on the production of emphasis in Jordanian Arabic: A sociophonetic study. In *Kansas Working Papers in Linguistics*, 32, 20-47.

AGOSTINIANI, L. (1992). Su alcuni aspetti del "rafforzamento sintattico" in Toscana e sulla loro importanza per la qualificazione del fenomeno in generale. In *Quaderni di Linguistica dell'Università di Firenze*, 3, 1-28.

ALZOUBI, A. (2016). A sociophonetic study of the effect of gender on emphatic-plain contrast in Jordanian Arabic. In *Journal of the Acoustical Society of America*, 140, 3113.

BELL, A. (1984). Language style as audience design. In *Language in Society*, 13, 145-204.

BENOR, S. (2001). Sounding learned: the gendered use of /t/ in Orthodox Jewish English. In JOHNSON, D.E., SANCHEZ, T. (Eds.), *Penn working papers in linguistics: selected papers from NWAV 29*. Philadelphia: University of Pennsylvania, 1-16.

BIJANKHAN, M., NOURBAKHSH, M. (2009). Voice onset time in Persian initial and intervocalic stop production. In *Journal of the International Phonetic Association*, 39(3), 335-364.

BILIOTTI, F., CALAMAI, S. (2012). Linguistic opinions and attitudes in Tuscany: verbal guise experiments on the varieties of Arezzo and Florence In CALAMAI, S., CELATA, C. & CIUCCI, L. (Eds.), *Sociophonetics, at the crossroads of speech variation, processing and communication.* Pisa: Edizioni della Scuola Normale Superiore, 1-4.

BRINCA, L., ARAÚJO, L., NOGUEIRA P. & GIL, C. (2016). Voice onset time characteristics of voiceless stops produced by children with European Portuguese as mother tongue. In *Ampersand*, 3, 137-142.

BUCHOLTZ, M. (1996). Geek the girl: language, femininity and female nerds. In WARNER, N., AHLERS, J., BILMES, L., OLIVER, M., WERTHEIM, S. & CHEN, M. (Eds.), *Gender and belief systems.* Berkeley: Berkeley Women and Language Group, 119-131.

BUCHOLTZ, M., HALL, K. (2005). Identity and interaction: a socio-cultural linguistic approach. In *Discourse Studies*, 7, 585-614.

CAMPBELL-KIBLER, K. (2007). Accent, (ING), and the social logic of listener perceptions. In *American Speech*, 82, 32-64.

CALAMAI, S. (2015). *Introduzione alla sociofonetica*. Roma: Carocci.

CALAMAI, S., RICCI, I. (2005). Un esperimento di matched-guise in Toscana. In *Studi Linguistici e Filologici on Line*, 3(1), 63-105.

CASTELLANI, A. (1952). *Nuovi testi fiorentini del Dugento*. Firenze: Sansoni.

CASTELLANI, A. (1960). Precisazioni sulla gorgia toscana. In *Boletim de Filología*, XIX, 242-262.

CHAMBERS, J.K. (2003), *Sociolinguistic Theory*. 2nd edition. Oxford: Blackwell.

CHENG, M. (2013). Voice Onset Time of Syllable-Initial Stops in Sixian Hakka: Isolated Syllables. In *Journal of National Taiwan Normal University*, 58(2), 193-227.

CHERYAN, S., SCHWARTZ CAMERON, J., KATAGIRI, Z. & MONIN, B. (2015). Manning Up. Threatened Men Compensate by Disavowing Feminine Preferences and Embracing Masculine Attributes. In *Social Psychology*, 46, 218-227.

Collins, K.M.T., Onwuegbuzie, A.J. & Sutton, I.L. (2006). A Model Incorporating the Rationale and Purpose for Conducting Mixed-Methods Research in Special Education and Beyond. In *Learning Disabilities: A Contemporary Journal*, 4(1), 67-100.

Cortelazzo, M.A. (1995). La componente dialettale nella lingua delle giovani e dei giovani. In Marcato, G. (Ed.), *Donna & linguaggio*. Padova: CLEUP, 581-586.

Craffonara, L. (1979). Zur Palatalisierung von CA und GA in den Sellatälern. In *Ladinia*, III, 69-93.

Cravens, T.D., Giannelli, L. (1995). Relative salience of gender and class in a situation of multiple competing norms. In *Language Variation and Change*, 7, 261-285.

Del Puente, P. (1995). La metafonia napoletana: Un tentativo di analisi sociolinguistica. In *L'Italia dialettale*, 58, 49-68.

Drummond, R., Schleef, E. (2016). Identity in variationist sociolinguistics. In Preece, S. (Ed.), *The Routledge Handbook of Language and Identity*. London: Routledge, 50-65.

Eckert, P. (2008). Variation and the indexical field. In *Journal of Sociolinguistics*, 12(4), 453-476.

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. In *Annual Review of Anthropology*, 41, 87-100.

Eckert, P., McConnell-Ginet, S. (2003). *Language and Gender*. Cambridge: Cambridge University Press.

Evans Wagner, S. (2012). Age Grading in Sociolinguistic Theory. In *Language and Linguistic Compass*, 6(6), 371-382.

Franceschi, T. (1969). Il principio dell'esagerazione come criterio di ricerca linguistica. In *Archivio Glottologico Italiano*, LIV, 49-105.

Francis, A.L., Ciocca, V. & Yu, J.M.C. (2003). Accuracy and variability of acoustic measures of voicing onset. In *Journal of the Acoustical Society of America*, 113(2), 1025-1032.

Freeman, V. (2014). Hyperarticulation as a signal of stance. In *Journal of Phonetics*, 45, 1-11.

Giacomelli, R. (1934). Controllo fonetico per diciassette punti dell'AIS nell'Emilia, nelle Marche, in Toscana, nell'Umbria e nel Lazio. In *Archivum Romanicum*, XVIII, 153-211.

Giannelli, L., (1976). *Toscana*. Pisa: Pacini.

Giannelli, L., (1983). Aspirate etrusche e gorgia toscana: valenza delle condizioni fonetiche dell'area toscana. In Agostiniani, L., Giannelli, L. (Eds.), *Fonologia etrusca, fonetica toscana: il problema del sostrato*. Firenze: Olschki, 61-102.

Giannelli, L., Savoia, L.M. (1978). L'indebolimento consonantico in Toscana. In *Rivista Italiana di Dialettologia*, II, 23-58.

Giannelli, L., Savoia, L.M. (1979-1980). L'indebolimento consonantico in Toscana. In *Rivista Italiana di Dialettologia*, IV, 38-101.

Haddington, P. (2004). Stance Taking in News Interviews. In *SKY Journal of Linguistics*, 17, 101-142.

Han, K. (2010). The Interaction between Phonetic Environment and Speaker Sex on VOT Differences. In *The Journal of Linguistic Science*, 54, 295-308.

Hudson, R.A. (1996). *Sociolinguistics*. 2nd edition. Cambridge: Cambridge University Press.

Huggins, A.W.F. (1972). Just Noticeable Differences for Segment Duration in Natural Speech. In *The Journal of the Acoustical Society of America*, 51(4), 1270-1278.

Iannàccaro, G. (2002). *Il Dialetto Percepito. Sulla reazione di parlanti di fronte al cambio linguistico*. Alessandria: Edizioni dell'Orso.

Irvine, J.T., Gal, S. (2000). Language Ideology and Linguistic Differentiation. In Kroskrity, P.V. (Ed.), *Regimes of Language*. Santa Fe, NM: Sch. Am. Res. Press, 35-83.

Izzo, H.J. (1972). *Tuscan & Etruscan*. Toronto: University of Toronto Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. In *Journal of Phonetics*, 34, 485-499.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. In *Journal of Phonetics*, 45, 76-90.

Kiesling, S.F. (2013). Constructing identity. In Chambers, J.K., Schilling, N. (Eds.), *The handbook of language variation and change*. 2nd ed. Malden: John Wiley and Sons, 448-467.

Koenig, L.L. (2000). Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds. In *Journal of Speech, Language, and Hearing Research*, 43, 1211-1228.

Kristiansen, T. (2010). Investigating language in space: Experimental techniques. In Auer, P., Schmidt, J.E. (Eds.), *Language and Space: An International Handbook of Linguistic Variation. Volume 1: Theories and Methods.* Berlin: Walter de Gruyter, 528-549.

Labov, W. (1984). Field methods of the Project on Linguistic Change and Variation. In Baugh, J., Sherzer, J. (Eds.), *Language in Use*. Englewood Cliffs: Prentice-Hall, 28-53.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. In *Language Variation and Change*, 2, 205-251.

Labov, W. (1994). *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford: Blackwell.

Lambert, W.E. (1967). The social psychology of bilingualism. In *Journal of Social Issues*, 23, 91-109.

Lee, H. (2016). Consistent sound change between stops and affricates in Seoul Korean within and across individuals: A diachronic investigation. In *Journal of the Acoustical Society of America*, 140, 491-496.

Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. In *The Journal of the Acoustical Society of America*, 133, 142-147.

Lisker, L., Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. In *Word*, 20(3), 527-565.

Loporcaro, M. (1997). Lengthening and raddoppiamento fonosintattico. In Maiden, M., Parry, M. (Eds.), *The Dialects of Italy*. London and New York: Routledge, 41-51.

Lundeborg Hammarström, I., Larsson, M., Wiman, S. & McAllister, A. (2012). Voice onset time in Swedish children and adults. In *Logopedics, Phoniatrics, Vocology*, 37(3), 117-122.

Madhu, S.R.B., Mahendra, K.N. & Sreedevi, N. (2014). Voice Onset Time across Gender and Different Vowel Contexts in Telugu. In *Language in India*, 14(12), 252-263.

Marcato, G., Ursini, F. & Politi, A. (1974). *Dialetto e italiano: status socioeconomico e percezione sociale del fenomeno linguistico*. Pisa: Pacini.

Marotta, G. (2004). Non solo spiranti. La gorgia toscana nel parlato di Pisa. In *L'Italia Dialettale*, 62, 27-60.

Marotta, G. (2014). New parameters for the sociophonetic indexes. Evidence from the Tuscan varieties of Italian. In Celata, C., Calamai, S. (Eds.), *Advances in Sociophonetics*. Amsterdam-Philadelphia: John Benjamins, 137-168.

Milroy, J., Milroy, L. (1978). Belfast: Change and Variation in an Urban Vernacular. In Trudgill, P. (Ed.), *Sociolinguistic Patterns in British English*. Baltimore: University Park Press, 19-36.

Milroy, L. (1987). *Language and Social Networks*. Oxford: Blackwell.

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

Molfese, D.L., Hess, T.M. (1978). Hemispheric Specialization for VOT Perception in the Preschool Child. In *Journal of Experimental Child Psychology*, 26, 71-84.

Molfese, D.L., Molfese, V.J. (1988). Right hemisphere responses from preschool children to temporal cue to speech and nonspeech materials: electrophysiological correlates. In *Brain and Language*, 33, 245-259.

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.

Nakai, S., Scobbie, J.M. (2016). The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. In *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1-31.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. In *Journal of Phonetics*, 39, 132-142.

Nissen, S.L., Fox, R.A. (2009). Acoustic and spectral patterns in young children's stop consonant productions. In *Journal of the Acoustical Society of America*, 126, 1369-1378.

Nodari, R. (2015). Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio. Studi AISV 1*. Milano: Officinaventuno, 139-153.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. In *Journal of Phonetics*, 39, 59-67.

Peng, J., Chen, L. & Lee, C., (2014). Voice Onset Time of Initial Stops in Mandarin and Hakka: Effect of Gender. In *Taiwan Journal of Linguistics*, 21(1), 63-80.

Piccardi, D. (in press). La cosiddetta "gorgia enfatica" fiorentina: una ridefinizione in chiave sociofonetica. In *L'Italia Dialettale*, 79.

Pickett, J.M. (1980). *The sounds of speech communication: A Primer of Acoustic Phonetics and Speech Perception*. Austin: University Park Press.

Podesva, R. (2007). Phonation type as a stylistic variable: the use of falsetto in constructing a person. In *Journal of Sociolinguistics*, 11, 478-504.

Podesva, R. (2013). Gender and the social meaning of non-modal phonation types. In Cathcart, C., Chen, I., Finley, G., Kang, S., Sandy, C.S. & Stickles, E. (Eds.), *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society*, 427-448.

ROBB, M., GILBERT, H. & LERMAN, J. (2005). Influence of gender and environmental setting on voice onset time. In *Folia Phoniatrica et Logopaedica*, 57(3), 125-133.

ROMITO, L., CIARDULLO, M.A. & TARASI, A. (2015). Analisi acustica delle occlusive sorde aspirate del dialetto di San Giovanni in Fiore (CS). In VAYRA, M., AVESANI, C. & TAMBURINI, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio. Studi AISV 1*. Milano: Officinaventuno, 169-186.

RYALLS, J., ZIPPRER, A. & BALDAUFF, P. (1997). A preliminary investigation of the effects of gender and race on Voice Onset Time. In *Journal of Speech and Hearing Research*, 40(3), 642-645.

SCHARF, G., MASUR, H. (2002). Voice onset time in normal speakers of a German dialect. Effects of age, gender and verbal material. In WINDSOR, F., KELLY, M.L. & HEWLETT, N. (Eds.), *Investigations in Clinical Phonetics and Linguistics*. Mahwah: Lawrence Erlbaum, 327-339.

SILVERSTEIN, M., (2003). Indexical order and the dialectics of sociolinguistic life. In *Language & Communication*, 23, 193-229.

SINGH, R., KESHET, J., GENCAGA, D. & RAJ, B. (2016). The relationship of voice onset time and voice offset time to physical age. In *Proceedings of the 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016)*, Shanghai, China, 5390-5394.

SKOOG WALLER, S., ERIKSSON, M. & SÖRQVIST, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. In *Frontiers in Psychology*, 6, 978.

SMITH, B.L. (1978). Effects of Place of Articulation and Vowel Environment on "Voiced" Stop Consonant Production. In *Glossa*, 12(2), 163-175.

SORIANELLO, P. (1996). Indici fonetici delle occlusive sorde in cosentino. In *Rivista Italiana di Dialettologia*, XX, 123-159.

SOUKUP, B. (2013). On matching speaker (dis)guises – revisiting a methodological tradition. In KRISTIANSEN, T., GRONDELAERS, S. (Eds.), *Language (de)standardisation in late modern Europe: Experimental studies.* Oslo: Novus, 267-285.

STEVENS, K.N. (1999). *Acoustic Phonetics*. Cambridge: MIT Press.

SUMMERFIELD, A.Q. (1975). Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. In *Speech Perception*, Series 2, 4, 61-72.

SWARTZ, B.L. (1992). Gender difference in voice onset time. In *Perceptual and Motor Skills*, 75, 983-992.

SWEETING, P.M., BAKEN, R.J. (1982). Voice onset time in a normal-aged population. In *Journal of Speech and Hearing Research*, 25, 129-134.

TITZE, I.R. (1994). *Principles of Voice Production*. Englewood Cliffs: Prentice Hall.

TRUDGILL, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. In *Language in Society*, 1(2), 179-195.

WHITESIDE, S.P., HENRY, L. & DOBBIN, R. (2004). Sex differences in voice onset time: a developmental study of phonetic context effects in British English. In *Journal of the Acoustical Society of America*, 116(2), 1179-1183.

Whiteside, S.P., Irving, C.J. (1997). Speakers' sex differences in voice onset time: Some preliminary findings. In *Perceptual and Motor Skills*, 85(2), 459-463.

Whiteside, S.P., Irving, C.J. (1998). Speaker's sex differences in voice onset time: A study of isolated word production. In *Perceptual and Motor Skills*, 86, 651-654.

Whiteside, S.P., Marshall, J. (2001). Developmental Trends in Voice Onset Time: Some Evidence for Sex Differences. In *Phonetica*, 58, 196-210.

Willer, R., Rogalin, C.L., Conlon, B. & Wojnowicz, M.T. (2013). Overdoing Gender: A Test of the Masculine Overcompensation Thesis. In *American Journal of Sociology*, 118(4), 980-1022.

Yu, V.Y., De Nil, L.F. & Pang, E.W. (2015). Effects of age, sex and syllable structure on voice onset time: Evidence from children's voiceless aspirated stops. In *Language and speech*, 58(2), 152-167.

Yuan, J., Liberman, M. & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Proceedings of the 9th International Conference on Spoken Language Processing* (Interspeech-ICSLP), 541-544.

DOI: 10.17469/O2103AISV000006

CLAUDIA ROBERTA COMBEI, OTTAVIA TORDINI

A pilot sociophonetic study on open-mid vowels uttered by young male and female speakers of the Pisan variety

Based on a small dataset of spontaneous speech produced by 6 speakers aged 18-20, this study aims to acoustically describe vowel quality in the Pisan variety of Italian with respect to male-female differences. The Pisan variety is characterized by the lowering of [ɛ] and [ɔ], which is said to be a sociophonetic cue of vernacular speech. The initial aim was to explore whether gender-related sociophonetic differences influence the production of [ɛ] and [ɔ] in young speakers, and which of the two groups (males or females) were more likely to adopt this vernacular feature. However, we did not find consistent evidence of [ɛ] and [ɔ] lowering in the analyzed corpus. Nevertheless, there was variation in the production of [ɛ], which differentiated the male Pisan speakers from their Florentine counterparts. For this reason, this study also tested an automatic classification system of [ɛ] vowels, which could predict with decent accuracy (77%) the variety to which a specific vowel token belonged[1].

Key words: sociophonetics, acoustics, Pisan, open-mid vowels.

1. *Introduction*

> Exploring how physiological and social factors interact in the speech of both men and women is important in understanding the linguistic principles governing language variability.

As Byrd (1994: 41) highlights, gender-related features are considered fundamental to a proper understanding of language variability. However, a large number of studies in the field of experimental phonetics have placed more emphasis on male speakers, since they represent the unmarked choice[2]. On the other hand, female speakers have received less attention in phonetic research, both in acoustics and speech perception, for various reasons.

For example, Fant's early work (1960), which showed the source-filter interaction and paved the way for subsequent studies, was mainly based on the examination

[1] Authors' note: although the contribution is the result of the joint research activity of both authors, the responsibility for drafting the article and carrying out specific analyses has been distributed as follows: § 1, § 2.1 OT; § 2.2 and 3 CRC and OT; orthographic and phonetic transcriptions, segmentations and feature extractions were performed by both authors (vowel [a]: OT; vowels [ɛ] and [ɔ]: CRC and OT); in §4, the plot representations and their interpretations were made by CRC, while OT explained their sociophonetic implication; § 4.1, 5, 5.1 are to be attributed to CRC, § 6 to OT and CRC, and bibliography to OT.

[2] For a more specific and detailed discussion about this topic, see Ferrero, Magno Caldognetto & Cosi (1996).

of male physiological characteristics. Another reason behind the choice to study males rather than females may lie in the difficulty involved in identifying women's higher fundamental frequencies. In fact, this often hinders an accurate analysis of the location of formant frequencies, making it more difficult to detect phonetic contrasts (Klatt, Klatt, 1990).

As far as Italian regional varieties are concerned, until now, only a few studies have focused on gender differentiation related to speech production (Filipponio, Cazzorla, 2016; Nodari, 2016). More specifically, phonetic features of regional varieties of the Tuscan Western areas have been thoroughly investigated, but most of these analyses have not regarded gender as a relevant factor of variation (Calamai, 2001; Nocchi, Calamai, 2009).

2. *The study*

2.1 A brief overview of lowering of open-mid vowels in Western Tuscany

From a phonetic-phonological point of view, the Western area of Tuscany is characterized by specific phenomena which contribute to separating West (especially Pisa and Leghorn) from the rest of the region. In fact, the lowering of stressed open-mid vowels [ɛ] to [æ] and [ɔ] to [Λ] has long been noted in the literature as a diatopically marked feature in this area (Giannelli, 2000; Calamai, 2001, 2003, 2004, 2009; Marotta, Calamai & Sardelli, 2004; Calamai, Ricci, 2005; Nocchi, Calamai, 2009). It can thus be considered relevant for the identification of the above-mentioned regional varieties of Italian, as well as being a cue of vernacularity (Giannelli, 2000; Calamai, 2001, 2004; Marotta, Calamai & Sardelli, 2004; Nocchi, Calamai, 2009). Many studies contributing to the analysis of the open-mid vowels lowering do not take gender into account (Calamai, 2001; Nocchi, Calamai, 2009). Moreover, most of them investigate isolated words or target words within carrier sentences[3] (Calamai 2001, 2003), excluding spontaneous productions.

2.2 Aims

Based on these assumptions, we carried out an acoustic analysis using both un-normalized and normalized vowel formant values on a small *ad hoc* collection of spontaneous and semi-spontaneous speech produced by young Pisan speakers. In order to provide an original contribution to the interpretation of this phenomenon, we considered gender differences within spontaneous productions, excluding written material, scripted dialogues and carrier sentences (Calamai, 2001). On that account, this study aims to make explicit at the sociophonetic level the link between gender and vernacularity, regarding variation in both diastratic and diatopic contexts. This research also tests a classification system that

[3] One notable exception is Nocchi, Calamai (2009). In this study, the authors made a comparison of spontaneous productions in Pisa, Florence, Arezzo and Leghorn. However, they analyzed the first two formants and the duration of stressed and non-stressed vowels, as well as the position of the token within the word. Our study concentrates only on the young generation of Pisan speakers, including both male and female subjects.

exploits data mining tools, using vowel formant values as features, in order to assess the extent to which vowel discrimination can be automated.

3. *Methods and tools*

In terms of settings and procedures, this kind of research requires a thorough review of various methodological issues. Several questions have been raised: the suitability of speech elicitation techniques; the choice of methods employed for acoustic analysis methods; and the statistical tools for data visualization and interpretation. In addition, it is often problematic to describe the place of articulation of vowels, because they produce a smaller and less consistent narrowing of the vocal tract than consonants (Calamai, 2004).

The first step consisted of building a small and homogeneous *ad hoc* corpus[4]. We recruited 6 young people, aged 18-20 years, who were born and had lived in Pisa all their lives. The sample was evenly divided between male and female speakers. Before each recording, subjects were asked to fill in a sociocultural questionnaire to collect information on their linguistic background, level of education, their parents' level of education and occupation, and their daily use of the dialectal variety. At an early stage of this study, education level and social status were not considered impactful. The Pisan speakers showed homogeneous characteristics, since all of them belonged to the same social network and social class (upper-middle). We chose young subjects for two reasons: firstly, because this approach allowed us to target the lowering of open-mid vowels in a uniform group; secondly, and most importantly, we wanted to examine whether there was experimental evidence supporting that these vowels were indeed lowering in young Pisan speakers.

Concerning the data, the speech was either spontaneous or semi-spontaneous, collected by means of the map-task elicitation technique and unstructured interviews. The use of the map-task was intended to increase the occurrence of the linguistic phenomena we were interested in, and allowed us to perform analyses both at an acoustic and sociolinguistic level. As for the sampling, the speakers were recorded with Praat[5] (Boersma, Weenink, 2016), using a Samson METEOR MIC cardioid pickup microphone (condenser diaphragms: 25mm). The sampling parameters were the following: mono channel, 16-bit, 16,000 Hz, linearly encoded WAV.

Due to numerous creaky voice productions in 2 out of 3 Pisan male speakers, at times formant measurements appeared to be significantly distorted. The open-mid vowels produced by the Pisan speakers were subsequently compared with those of a control group of four Florentine speakers. We carried out a contrastive analysis because the lowering of open-mid vowels has not yet been attested in the Florentine Italian, since it is mainly a feature of the Western area of Tuscany (Calamai, 2001; Sorianello, 2002). The Florentine data was retrieved from the CLIPS[6] corpus (Corpora e Lessici dell'Italiano Parlato e

[4] Differently from Nocchi, Calamai (2009), and because it did not suit the needs of this study, we did not use the material stored in Archivio del Parlato Italiano (API), retrieved from http://www.parlaritaliano.it/index.php/it/dati/40-api-archivio-del-parlato-italiano.

[5] Retrieved from http://www.praat.org/.

[6] Retrieved from http://www.clips.unina.it/it/corpus.jsp/.

Scritto), and consisted of audio recordings of 4 young speakers (2 males and 2 females). It was segmented and further analyzed, following the same procedures as for the Pisan informants. This type of analysis, also employed by Calamai (2004), allowed us to obtain a good comparison between the two varieties.

4. *Acoustic analysis*

The present study solely concerns stressed vowels[7], so as to reduce the number of dependent variables and to focus on the sociophonetic features of interest. The vowel extraction and the acoustic data processing were performed with Praat and ELAN[8]. The audio material was transcribed and segmented into three different Praat tiers: sentences, words and target vowels. We obtained 820 tokens of open-mid vowels: 555 for [ɛ] and 265 for [ɔ], roughly 50 tokens per speaker for [ɛ] and 20 tokens per speaker for [ɔ]. In addition, we also segmented stressed tokens of [a] (320 in total, ca. 30 per speaker), both in the Pisan and Florentine productions. Following the literature on the phonetic phenomena in Western Tuscany, [a] could be characterized by a more posterior place of articulation (Giannelli, 2000; Calamai, 2004). Thus, analyzing the acoustic features of this low vowel would help us estimate the vowel space for each speaker. Furthermore, it could contribute to targeting the lowering phenomenon in a more specific and defined space[9]. Using two *ad hoc* Praat scripts, we extracted the vowel duration as well as the first three formant frequencies and the fundamental frequency (f0) from the vowel midpoint. However, duration and f0 were eventually excluded from the final analysis due to the persistent variation among the spontaneous productions, mostly due to large variability in speech rate (Calamai, 2015), even within the same sentence. After the automatic extraction, we noticed that for a consistent number of tokens (ca. 25%) it was not even possible to detect f0 values. Therefore, the following values were calculated: single-point measurements, means, and standard deviations for the first three formants (F1, F2, F3) of the open-mid vowels. These parameters allowed us to plot the vowels using the R[10] package "vowels" (Kendall, Thomas, 2009).

For the first series of graphs, we used the raw formant values expressed in Hz. The graphs representing all tokens of [ɛ] uttered by all 10 speakers clearly show that the values reported for female speakers, both Florentine (efifem1, efifem2) and Pisan (epifem1, epifem2, epifem3), tend to be dispersed (see Figure 1). On the other hand, the Pisan male speakers appear more homogeneous, since the values reported for this group are noticeably similar. As far as [ɔ] is concerned, the data seem to suggest that the position of [ɔ] in the vowel space is comparable for most speakers, both Pisan and Florentine (see Figure 2). With regards to [a], all Pisan

[7] Differently from Calamai (2001).

[8] Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/.

[9] In the present study, we did not compute the Euclidean distance between [a] and [ɛ] and [a] and [ɔ]. However, this aspect could be further investigated in a next study.

[10] Retrieved from https://www.r-project.org.

male speakers (apima1, apima2, apima3) have similar values, unlike the females. In fact, only two Pisan females (apifem2 and apifem3) have similar productions, while the values for the other female speaker (apifem1) resemble those reported for male speakers (see Figure 3).

Figure 1 - *All tokens of [ɛ] uttered by all 10 speakers*

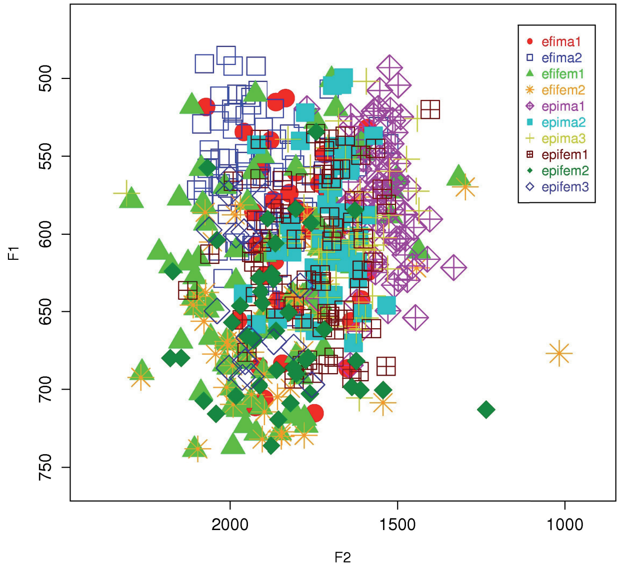


Figure 2 - *All tokens of [ɔ] uttered by all 10 speakers*

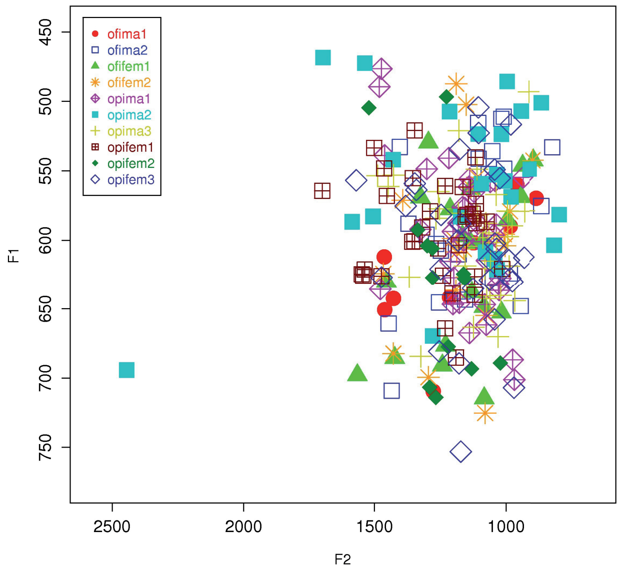
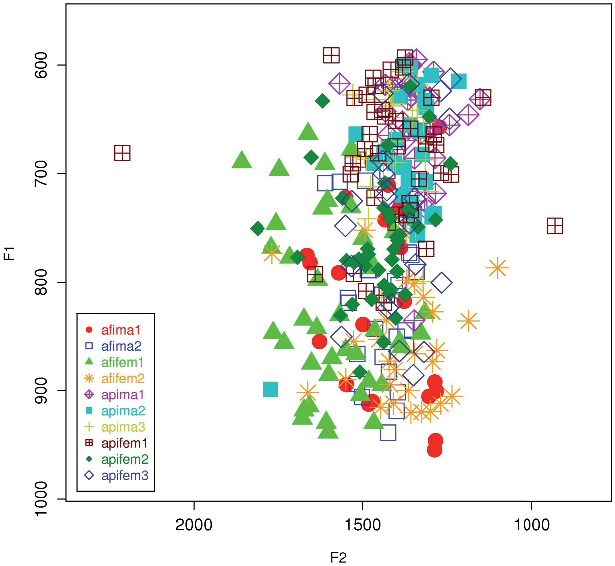
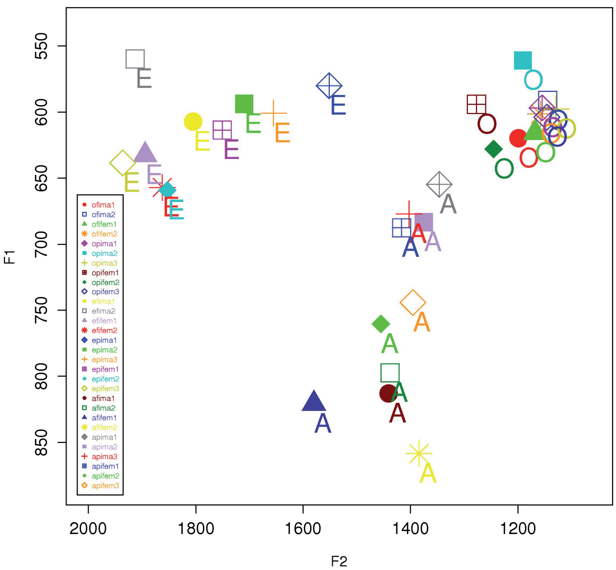


Figure 3 - *All tokens of [a] uttered by all 10 speakers*



The mean values of F1 obtained for the open-mid front unrounded vowel [ɛ] range from 560 Hz to 680 Hz, while the mean F2 values reach values between 1550 Hz and 1900 Hz. The distribution of [ɛ] produced by Pisan male speakers is rather uniform, while there is more variability for their Florentine counterparts. For the female speakers, Pisan and Florentine show similar values.

Figure 4 - *Speakers' means for [ɛ], [ɔ], and [a]*

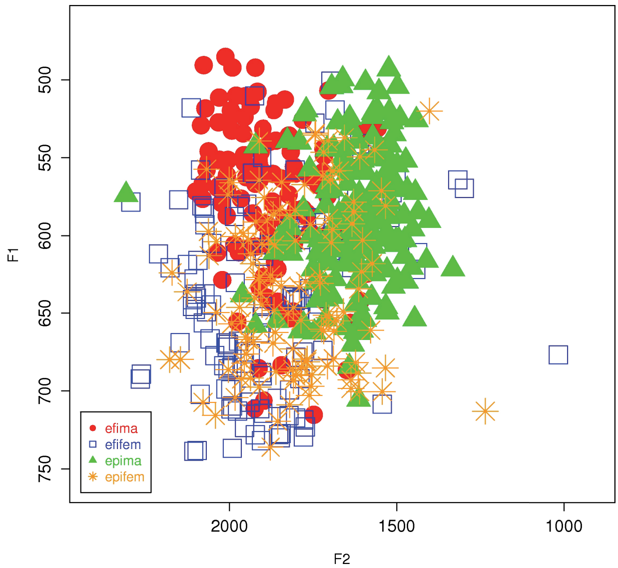


Regarding [ɔ], we noticed that all sets were rather homogeneous, and that the mean values recorded for each speaker and each group were similar, ranging from around 570 Hz to 620 Hz for F1, and from around 1050 Hz to 1250 Hz for F2. The [a] produced by Florentine speakers tended be similar for both males and females (with mean values between 770 Hz to 850 Hz for F1 and 1400 Hz to 1600 Hz for F2). On the other hand, Pisan speakers reveal more variability, especially the female speakers. For example, the average values obtained for one female speaker are comparable to those of the three male speakers. The values for the other two female speakers from Pisa, around 760 Hz for F1 and 1400 Hz for F2, are similar to those produced by their Florentine counterparts. The following graph illustrates the speakers' means for [ɛ], [ɔ], and [a], in support of what has been said in the previous paragraphs (see Figure 4).

The mean values obtained for Pisan [a] (both male and female) tend to be close to the values of [ɔ]. Based on this information, we noticed that perhaps the vicinity between [a] and [ɔ] might not be fortuitous. This could be due to a general shift of the place of articulation of the Pisan vowels. That is, the [ɛ], shifting to lower frequencies, might cause a repositioning of [a], in order to maintain contrast. However, the inconsistent productions for the Florentine variety do not allow us to further explain this behaviour.

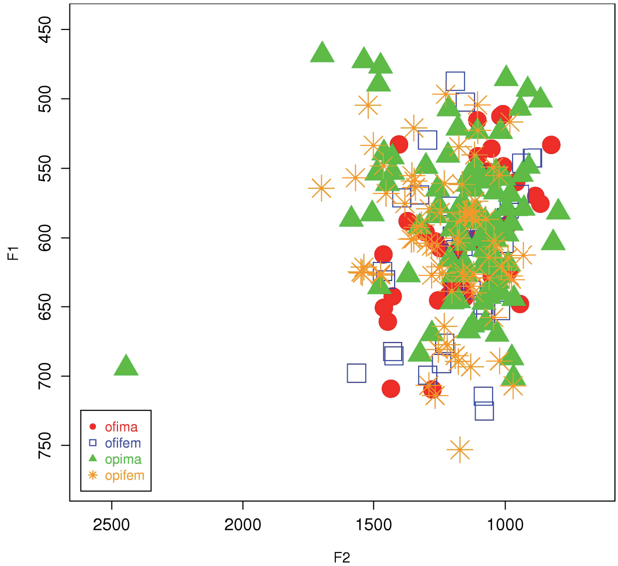
The analysis of the mean values for each speaker suggests that there is a considerable dispersion of the vowels produced by female speakers, both Pisan (epifem) and Florentine (efifem). Plotting as a function of group (Pisan males, Pisan Females, Florentine males, Florentine females), instead of single speakers, provides clearer representations of the vowels' behaviour. It is visible that both male groups (epima and efima) tend to be homogeneous (see Figure 5).

Figure 5 - *Pisan and Florentine groups for [ɛ]*



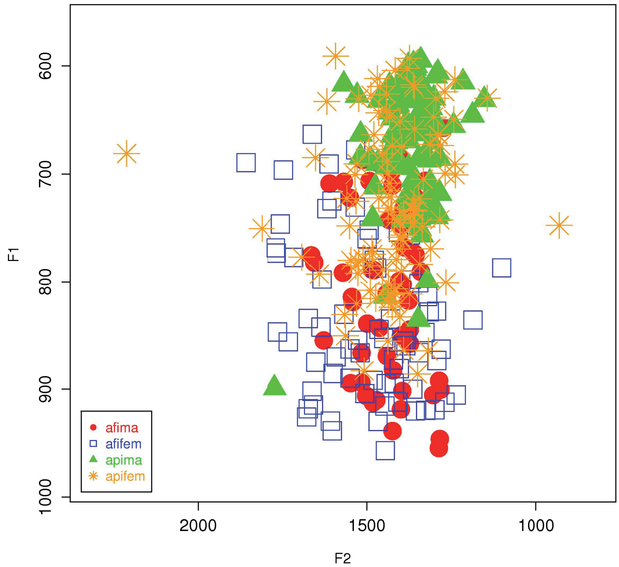
If we consider the graph representing the distribution of [ɔ], it is noticeable that the [ɔ] productions do not allow us to differentiate between the four groups (see Figure 6).

Figure 6 - *Pisan and Florentine groups for [ɔ]*



The tokens of Florentine [a] feature high F1 values. But given the low sample size (4 speakers), it is not possible to establish whether this is a variety-dependent trait.

Figure 7 - *Pisan and Florentine groups for [a]*



The intriguing aspect of this analysis still lies in the vicinity between the values of [ɔ] and [a]. In fact, according to the predictions about the lowering phenomenon, the latter should be closer to [ɛ]. Moreover, the values obtained for [a] display consistent difference only between Pisan males and Florentine males.

So, the results obtained for the four groups were compared with the values reported by Nocchi, Calamai (2009), for [ɛ], [ɔ] and [a] (Table 1). For their research, they analyzed male speakers[11], recorded by means of the map task elicitation technique.

Table 1 - *Nocchi, Calamai (2009) – vowel formants and duration for [ɛ] (Pisan speakers)*

Vowel	*Gender*	*City*	*F1*	*F2*	*F2-F1*
[ɛ]	male	Pisa	538	1693	1155
[ɔ]	male	Pisa	562	1131	569
[a]	male	Pisa	650	1353	703

In the above-mentioned research, the authors described a marked lowering of most of the vowel ellipses and a significant vicinity between [a] and [ɛ]. In fact, it appears that [a] is frequently realized as [æ].

Tables 2, 3 and 4 show the results obtained in our study.

Table 2 - *Mean formant values of [ɛ] as a function of speakers' groups (standard deviations are in brackets)*

[ɛ]		*F1*	*F2*	*F3*	*F2-F1*
Pisa	male	590 (42)	1623 (120)	2625 (138)	1033 (122)
	female	632 (52)	1811 (163)	2779 (300)	1178 (167)
Florence	male	576 (50)	1877 (130)	2673 (150)	1301 (150)
	female	641 (57)	1883 (205)	2746 (298)	1241 (202)

Table 3 - *Mean formant values of [ɔ] as a function of speakers' groups (standard deviations are in brackets)*

[ɔ]		*F1*	*F2*	*F3*	*F2-F1*
Pisa	male	588 (51)	1155 (220)	2654 (201)	566 (229)
	female	604 (53)	1221 (163)	2790 (302)	617 (185)
Florence	male	599 (54)	1150 (131)	2616 (229)	560 (166)
	female	609 (59)	1163 (165)	2624 (232)	554 (155)

[11] Calamai (2004) already presented a thorough review of a cross-gender comparison between Pisan and Leghorn male and female speakers. However, this study analyzed female speakers, employing carrier sentences and controlled speech. Furthermore, it investigated the lowering of open mid-vowels in subjects of different ages.

Table 4 - *Mean formant values of [a] as a function of speakers' groups (standard deviations are in brackets)*

[a]		*F1*	*F2*	*F3*	*F2-F1*
Pisa	male	670 (55)	1372 (81)	2667 (202)	703 (87)
	female	728 (74)	1427 (138)	2669 (296)	699 (150)
Florence	male	804 (82)	1439 (95)	2545 (192)	635 (193)
	female	838 (72)	1498 (160)	2672 (255)	660 (133)

The mean values (Table 2) obtained for the [ɛ] produced by Pisan male speakers (590 Hz for F1, 1623 Hz for F2, 2625 Hz for F3 and 1033 Hz for F2-F1) do not appear to be far from those reported by Nocchi, Calamai (2009) (538 Hz for F1, 1693 Hz for F2, and 1155 Hz for F2-F1), while, on the other hand, the values for Pisan female speakers seem to be vastly different (632 Hz for F1, 1811 Hz for F2, 2778 Hz for F3, and 1178 Hz for F2-F1). However, since this is also the group that displays large internal heterogeneity, it is arguable to what extent we can trust this evidence. As far as [ɔ] is concerned, the means of both Pisan and Florentine groups tend to resemble each other (especially the second and the third formant and the F2-F1 difference), but the standard deviation is rather large. These values are comparable to those found by Nocchi and Calamai (2009). Regarding [a], the mean values for our Pisan male subjects are 670 Hz for F1, 1372 Hz for F2, 2667 Hz for F3 and 703 Hz for F2-F1. These values are perfectly comparable to those found by Nocchi, Calamai (2009) (650 Hz for F1, 1353 Hz for F2, and 703 Hz for F2-F1). Unlike their male counterparts, the Pisan female speakers produced higher mean values (728 Hz for F1, 1427 Hz for F2, 2669 Hz for F3, and 699 Hz for F2-F1), meaning that they are distant from the values presented in Nocchi, Calamai (2009).

4.1 Statistical analysis

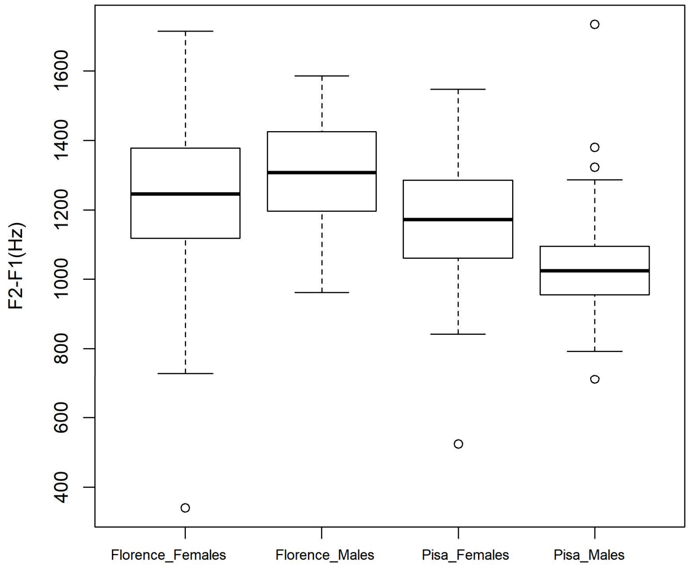
Our *ad hoc* corpus is rather small and not representative of the whole Pisan variety. Indeed, to some degree, it might only be considered representative of the Pisan variety spoken by young educated speakers. Yet, the formant values presented in the previous paragraphs show that there is variability within each group. So, at this point it is questionable whether it is worth carrying out inferential tests. The descriptive statistics (graphs, mean values and standard deviations) introduced earlier might be satisfactory, but a two-way ANOVA with R was also performed to determine whether there was any noteworthy difference between the four groups (Pisan males, Pisan females, Florentine males, and Florentine females).

Based on the results of the pairwise combinations of the continuous variables, and after having analysed the boxplots[12] (see Figure 8, for the mean values of F2-F1 in each of the four groups), we performed an ANOVA using formant values (F1, F2,

[12] Boxplots for [ɔ] and [a] were also analyzed, but for the sake of conciseness this study presents only the [ɛ] boxplot.

and F2-F1) as response variables, and gender and city of origin as factors (categorical variables with two levels each). According to the data presented earlier, the lowering of open [ɛ] and [ɔ] is not significant in the small Pisan corpus investigated here.

Figure 8 - *Boxplots – Pisan and Florentine speakers for [ɛ]*



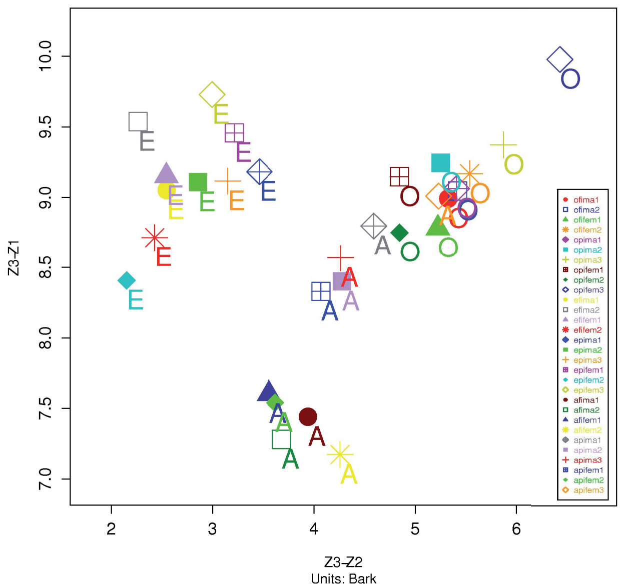
Since the single and the mean values for [ɔ] were similar across the four groups, and the ANOVA did not show any significant outcomes for this vowel, we are only presenting the ANOVA results for [ɛ][13]. We employed R's "aov" built-in function for the analysis of variance in the corpus, having F2-F1[14] as a response variable and gender and city of origin as factors. The results show no significant main effects for the speakers' gender (Df: 1, F value: 2.822, Pr(>F): 0.1440), but there might be a minor effect, even if not it is not solid, for the speakers' origin (Df: 1, F value: 6.847, Pr(>F): 0.0398) and for the interaction between gender and city of origin (Df: 1, F value: 4.168, Pr(>F): 0.0873). Tukey's honest significant difference (HSD) test allowed us to identify the means that were significantly different from one another. In this test, as far as the speakers' origin was concerned, the means reported for the Pisan and the Florentine groups were significantly different (p adj: 0.0398), while in the case of gender-city interaction the most noteworthy difference was between Pisan male speakers and Florentine male speakers (p adj: 0.0607).

[13] The ANOVA for [a] did not reveal any particular phenomenon.
[14] The outcomes are similar for response variables F1 and F2 alone.

5. *Analysis of normalized vowels*

Since only a few studies aim to combine acoustic and auditory analyses[15], we decided to also include the normalized values of the vowels, following the method used by Ferrero *et al.* (1996). For this reason, we transformed the formant values from Hz to Bark using Traunmüller's Bark conversion formula (Traunmüller, 1990, 1997): $Z_i = 26.81/(1+1960/F_i) - 0.53$, where Z_i is the formant value expressed in Bark, while F_i is the formant frequency in Hz. Then we computed the markers: Z3-Z2, Z3-Z1, and Z2-Z1. The diversity in terms of anatomical features among male and female speakers can then be studied and resized, according to whether the aim is to maximize or minimize the differences between the two systems (Maisano, 1996). So, we performed a Bark Difference Metric Normalization, and then we plotted the vowels using Z3-Z2, Z3-Z1 as coordinates. The vowel space was represented by Z3-Z2 (i.e. Bark-converted F3 minus Bark-converted F2), to model vowel advancement, and Z3-Z1 (i.e. Bark-converted F3 minus Bark-converted F1) to model vowel height. For each speaker, and then for each group, we calculated the mean values and the standard deviation.

Figure 9 - *Speakers' means for [ɛ], [ɔ], and [a]*



Following the procedure shown in the first part of this section, we plotted the mean normalized values for all speakers (see Figure 10, 11, 12). We also computed the mean values and the standard deviation for all groups (see Table 5, 6, 7).

[15] Uguzzoni (1988) and Sorianello (2002) provided auditory classifications, for the local varieties spoken in Pavullo (MO) and Siena, respectively.

Figure 10 - *Pisan and Florentine groups for normalized [ɛ]*

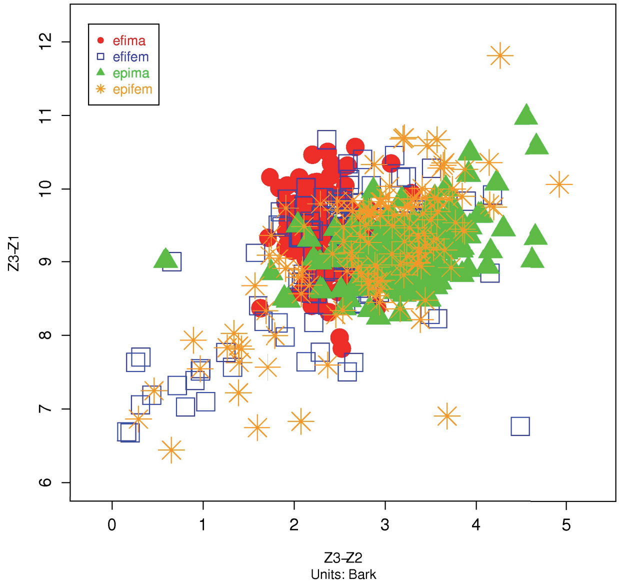


Figure 11 - *Pisan and Florentine groups for normalized [ɔ]*

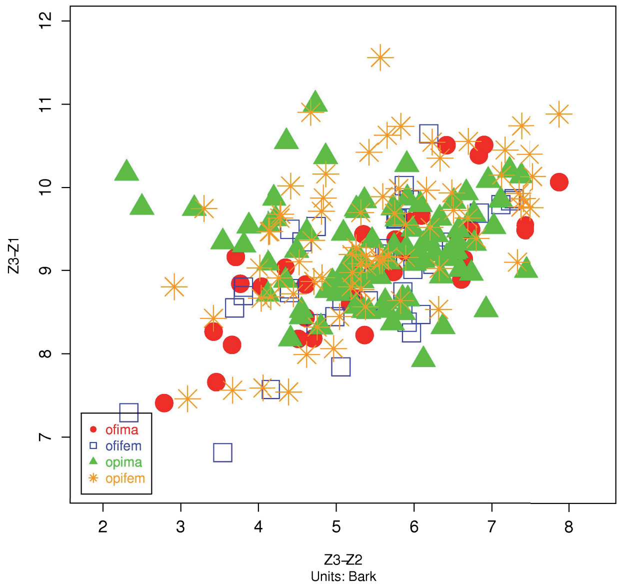
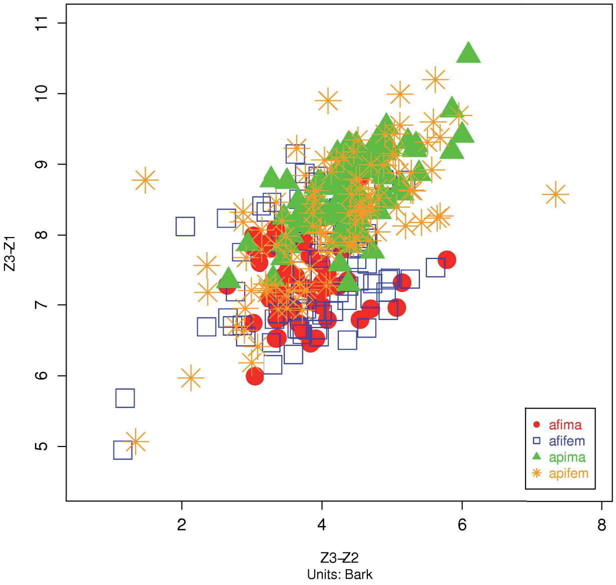


Figure 12 - *Pisan and Florentine groups for normalized [a]*



Tables 5, 6 and 7 show the normalized values of the three vowels obtained for the four groups.

Table 5 - *Mean formant values of [ɛ] as a function of speakers' groups (standard deviations are in brackets)*

[ɛ]		*Z3-Z2*	*Z3-Z1*	*Z2-Z1*
Pisa	male	3.21 (0.56)	9.51 (0.41)	5.93 (0.55)
	female	2.83 (0.84)	9.15 (0.92)	6.32 (0.70)
Florence	male	2.36 (0.34)	9.38 (0.57)	7.02 (0.69)
	female	2.50 (0.85)	9.00 (0.87)	6.49 (0.82)

Table 6 - *Mean formant values of [ɔ] as a function of speakers' groups (standard deviations are in brackets)*

[ɔ]		*Z3-Z2*	*Z3-Z1*	*Z2-Z1*
Pisa	male	5.54 (1.01)	9.22 (0.56)	3.68 (1.16)
	female	5.46 (1.17)	9.40 (0.84)	3.94 (1.01)
Florence	male	5.39 (1.30)	9.03 (0.74)	3.64 (0.86)
	female	5.38 (1.09)	8.97 (0.76)	3.59 (0.83)

Table 7 - *Mean formant values of [a] as a function of speakers' groups (standard deviations are in brackets)*

[a]		*Z3-Z2*	*Z3-Z1*	*Z2-Z1*
Pisa	male	4.40 (0.62)	8.61 (0.56)	4.21 (0.49)
	female	4.14 (0.97)	8.16 (0.90)	4.02 (0.76)
Florence	male	3.79 (0.59)	7.34 (0.57)	3.56 (0.75)
	female	3.85 (0.87)	7.41 (0.82)	3.57 (0.97)

The results confirm that the productions of [ɔ] are similar in all groups, but the standard deviation remains large, since a large number of tokens are distant from the mean. Values for Z3-Z2 range roughly from 5.40 to 5.50 Bark, Z3-Z1 varies between 9.00 and 9.40 Bark, while Z2-Z1 ranges from 3.60 to 3.90 Bark. Regarding [ɛ], the means remain different between the two Pisan groups. Calling to mind the large variability within the female group, the values for Z3-Z2, Z3-Z1, and Z2-Z1 are distributed as follows: 2.83 Bark, 9.15 Bark, and 6.32 Bark for the females, compared to 3.21 Bark, 9.51 Bark, and 5.93 Bark for the males. At the same time, the [ɛ] production by Pisan male speakers is noticeably distinct from the Florentine groups. Finally, the mean values for [a] appear to vary between the four groups, revealing rather substantial differences between Pisan and Florentine male speakers.

5.1 Classification experiment

The intriguing results displayed in the previous sections urged us to test an automatic classification system based on [ɛ]'s specific formant features. The aim was to verify whether this simplified method would allow us to discriminate between the varieties taken into consideration. We performed a series of classification experiments (for gender and city of origin, respectively) with the data mining tool Orange (2013)[16], using the Neural Network technique (sampling: Cross-validation – 10 folds) and the following parameters: Z3-Z2, Z3-Z1, Z2-Z1 expressed in Bark.

Table 8 - *Sampling method and evaluation results for the classification of Pisan [ɛ]*

Groups	*Sampling method*	*Classification method*	*Classification accuracy*
Female_Pisan [ɛ] Male_Pisan [ɛ]	Cross-validation Number of folds: 10	Neural Network	85.94%

[16] Retrieved from https://orange.biolab.si.

Table 9 - *Confusion matrices – Proportions of true and proportions of predicted for the classification of Pisan [ɛ]*

Proportions of true	*Female_Pisan [ɛ]*	*Male_Pisan [ɛ]*
Female_Pisan [ɛ]	75.60%	24.40%
Male_Pisan [ɛ]	6.80%	93.20%

Proportions of predicted	*Female_Pisan [ɛ]*	*Male_Pisan [ɛ]*
Female_Pisan [ɛ]	88.60%	15.50%
Male_Pisan [ɛ]	11.40%	84.50%

Table 10 - *Sampling method and evaluation results for the classification of Florentine and Pisan [ɛ]*

Groups	*Sampling method*	*Classification method*	*Classification accuracy*
Female_Florentine [ɛ] Female_Pisan [ɛ] Male_Florentine [ɛ] Male_Pisan [ɛ]	Cross-validation Number of folds: 10	Neural Network	77.27%

Table 11 - *Confusion matrices – Proportions of true and proportions of predicted for the classification of Florentine and Pisan [ɛ]*

Proportions of true	Female_ Florentine [ɛ]	Female_Pisan [ɛ]	Male_ Florentine [ɛ]	Male_Pisan [ɛ]
Female_Florentine [ɛ]	60.90%	10.20%	16.40%	12.50%
Female_Pisan [ɛ]	3.30%	68.30%	0.00%	28.50%
Male_Florentine [ɛ]	11.40%	4.40%	81.60%	2.60%
Male_Pisan [ɛ]	0.00%	7.40%	0.00%	92.60%

Proportions of predicted	Female_ Florentine [ɛ]	Female_Pisan [ɛ]	Male_ Florentine [ɛ]	Male_Pisan [ɛ]
Female_Florentine [ɛ]	82.20%	11.30%	18.40%	7.40%
Female_Pisan [ɛ]	4.20%	73.00%	0.00%	16.10%
Male_Florentine [ɛ]	13.70%	4.30%	81.60%	1.40%
Male_Pisan [ɛ]	0.00%	11.30%	0.00%	75.10%

The evaluation results for [ɛ] showed that this minimalist Neural Networks classification system based only on Z3-Z2, Z3-Z1, and Z2-Z1 values can predict with reasonable accuracy (over 85%) whether the vowels were uttered by Pisan male or Pisan female speakers.

If we extend this investigation by adding the two Florentine groups the classification accuracy is slightly lower (77%), but still promising. This evidence supports the previous impressions, namely that, for [ɛ] productions, the most significant difference is between the two male groups. Nevertheless, due to the limited number of productions analyzed here, further investigations should be carried out.

6. *Conclusions and further research*

This pilot study aimed to detect and describe gender-specific variation in vernacular speech, by examining the lowering of open-mid vowels in a group of young Pisan speakers. In general, the data did not reveal commensurable differences in terms of vowel quality ascribable to the lowering phenomenon. In regards to gender differences, the Pisan and Florentine female groups showed more internal variation, compared to their male counterparts. On the other hand, the lowering of [ɛ] detected in Pisan male speakers appears clear when investigating the F2-F1 difference. However, the numerous creaky voice productions of the two Pisan male subjects (Calamai, 2015; Melvin, Clopper, 2015) in some cases hindered accurate measurements of the vowel quality of this group[17].

In order to discern these preliminary results, we carried out an automatic classification experiment based on [ɛ]'s specific formant features. The results indicated that it is possible to differentiate between [ɛ] as uttered by the four groups, but the experimental conditions (i.e. limited number of speakers and the type of system employed) are not sufficient to speculate on the reasons behind these differences. Based on this classification experiment, we can only assert that the Bark-converted values Z3-Z2, Z3-Z1 and Z2-Z1 could represent fairly robust features for automatic vowel discrimination.

On a different note, we believe that the level of education and social class (Nodari, 2016) might have a greater effect than gender on discriminating between speakers. In particular, one could assess the role of education in triggering or hindering the lowering of [ɛ] and [ɔ] in the Western area of Tuscany. Arguably, the lowering of open-mid vowels was not attested in our corpus because all speakers belong to the same social network, have comparable family compositions (upper-middle class) and the same level of education (five of them are enrolled at the University). For these reasons, the study should be extended to a larger and more wide-ranging group of subjects in order to obtain more reliable results. Finally, vowel duration and stress could also be considered in future sociophonetic investigations.

Acknowledgments

We would like to thank Prof. Giovanna Marotta for the constructive suggestions and the supervision of our work. Her recommendations have been helpful for the interpretation of the overall vocalic shift that might occur in the Pisan variety. Her proposal to use education as a discriminant sociolinguistic factor has also opened new dimensions for this research in the future.

[17] The use of creaky voice is also highlighted in the literature (Calamai, 2015; Sorianello, 2006) as a typical male feature in low registers of both Tuscan and Roman varieties.

Bibliography

Adank, P., Smits, R. & Van Hout, R. (2004). A Comparison of Vowel Normalization Procedures for Language Variation Research. In *Journal of the Acoustic Society of America*, 116(5), 3099-3107.

Albano Leoni, F. (coordinator) (1999). *Archivio del Parlato Italiano (API)*. http://www.parlaritaliano.it/index.php/it/dati/40-api-archivio-del-parlato-italiano.

Albano Leoni, F. (2007). CLIPS – Corpora e Lessici di Italiano Parlato e Scritto. In *Linguistica computazionale: ricerche monolingui e multilingui*. http://www.clips.unina.it/it/corpus.jsp/.

Berruto, G. (2005). *Fondamenti di sociolinguistica*. Bari: Laterza.

Boersma, P., Weenink, D. (2016). *Praat: doing phonetics by computer.* http://www.praat.org/.

Byrd, D. (1994). Relations of sex and dialect to reduction. In *Speech communication*, 15, 39-54.

Calamai, S. (2001). Il vocalismo atono della varietà pisana. Prime evidenze sperimentali. In Regnicoli, A. (Ed.), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia.* Roma: Il Calamo.

Calamai, S. (2003). Vocali fiorentine e vocali pisane a confronto. In Albano Leoni, F., Cutugno, F., Pettorino, M., Savy, R. (a cura di), *Atti del Convegno Nazionale Il parlato Italiano*. Napoli; D'Auria, 1-25.

Calamai, S. (2004). *Il vocalismo tonico dell'area pisana e livornese. Aspetti storici, percettivi, acustici*. Alessandria: Edizioni dell'Orso.

Calamai, S. (2015). *Introduzione alla sociofonetica*. Roma: Carocci.

Calamai, S., Ricci, I. (2005). Sulla percezione dei confini vocalici in Toscana: primi risultati. In Cosi, P. (Ed.), *Proceedings of the 1st AISV Conference*, Torriana: EDK Editore, 63-87.

Chambers, J.K. (2002). Patterns of variation including change. In Chambers, J.K., Trudgill, P., Schilling-Estes, N. (Eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell Publishing.

Cosi, P., Ferrero, F. & Vagges, K. (1995). Rappresentazioni acustiche e uditive delle vocali italiane. In *Proceedings of the 23rd National Conference of Associazione Italiana di Acustica*, Bologna, 12-14 Settembre 1995, 151-156.

Eckert, P. (1989). The whole woman: sex and gender differences in variation. In *Language variation and change*, 1(3), 245-267.

Elan. https://tla.mpi.nl/tools/tla-tools/elan/. Nijmegen: Max Planck Institute for Psycholinguistics.

Fabricius, A. (2008). *Vowel Normalization in Sociophonetics: When, Why, How?* Seminar for Sociolinguistics Circle. Copenhagen University.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Ferrero, F.E., Magno Caldognetto, E. & Cosi, P. (1996). Sui piani formantici acustici e uditivi delle vocali di uomo, donna, bambino. In Peretti A., Simonetti P. (Eds.), *Proceedings of the 24th National Conference of Associazione Italiana di Acustica.* Padova: Arti Grafiche Padovane, 169-178.

Filipponio, L., Cazzorla, S. (2016). The vowels of Bari. A comparison between local dialect and regional Italian. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio.* Milano: Officinaventuno, 59-71.

Giannelli, L. (2000). *Toscana*. Pisa: Pacini.

Kendall, T., Thomas, E. (2009). *Vowels: Vowel Manipulation, Normalization, and Plotting in R. R package*, https://cran.r-project.org/package=vowels.

Klatt, D.H., Klatt, L.C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. In *Journal of the Acoustic Society of America*, 87(2), 820-857.

Maisano, L. (1996). Sistemi vocalici maschili e femminili a confronto. In Peretti, A., Simonetti, P. (Eds.), *Proceedings of the 24th National Conference of Associazione Italiana di Acustica.* Padova: Arti Grafiche Padovane.

Marotta, G., Calamai, S. & Sardelli, E. (2004). Non di sola lunghezza. La modulazione di f0 come indice socio-fonetico. In De Dominicis, A., Mori, L. & Stefani, M. (Eds.), *Costituzione, gestione e restauro di corpora vocali*. Roma: Esagrafica, 210-215.

Melvin, S., Clopper, C. (2015). Gender Variation in Creaky Voice and Fundamental Frequency. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow. Paper number 320.

Nocchi, N., Calamai, S. (2009). Durata e strutture formantiche nel parlato toscano: Indagini preliminari su un campione di dialoghi semispontanei. In Schmid, S., Schwarzenbach, M. & Studer, D. (Eds.), *Proceedings of the 5th Conference of Associazione Italiana di Scienze della Voce. La dimensione temporale del parlato.* Torriana: EDK, 195.

Nodari, R. (2016). Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio.* Milano: Officinaventuno, 139-153.

Orange (2013). *Data Mining Toolbox in Python*. https://orange.biolab.si.

R Development Core Team (2008). *R: A language and environment for statistical computing*. https://www.r-project.org. Wien: R Foundation for Statistical Computing.

Romito, L., Ciardullo, M.A. & Tarasi, A. (2016). Analisi acustica delle occlusive sorde aspirate del dialetto di San Giovanni in Fiore (CS). In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*. Milano: Officinaventuno, 169-186.

struttura sonora del linguaggio. Milano: Officinaventuno, 169-186.

Sorianello, P. (2002). Il vocalismo dell'italiano senese: un'indagine sperimentale. In Regnicoli, A. (Ed.), *La fonetica acustica come strumento di analisi della variazione linguistica in Italia.* Roma: Il Calamo, 47-52.

Sorianello, P. (2006). Per una rappresentazione uditiva dei segmenti vocalici: il caso del senese. In Agonigi, M., Cioni, L. & Paradisi, E. (a cura di), *Quaderni del Laboratorio di Linguistica della Scuola Normale Superiore di Pisa*, Nuova Serie, 159-173.

Syrdal, A.K., Gopal, H.S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. In *Journal of the Acoustical Society of America*, 79(4), 1086-1100.

Tamburini, F. (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio.* Milano: Officinaventuno, 169-186.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. In *Journal of the Acoustic Society of Am*erica, 88(1), 97-100.

Traunmüller, H. (1997). *Auditory scales of frequency representation*. https://www.ling.su.se/staff/hartmut/bark.htm.

Uguzzoni, A. (1988). Verso un'analisi parametrica del vocalismo di un dialetto frignanese. In *Italia Dialettale*, 25, 86-110.

DOI: 10.17469/O2103AISV000007

DAVIDE GARASSINO, MICHELE LOPORCARO, STEPHAN SCHMID

La quantità vocalica in due dialetti della Liguria

In this paper, we offer an experimental phonetic analysis of vowel length in two dialects of Liguria, those of Genoa and Porto Maurizio. Although Ligurian dialects show a high degree of uniformity, they nonetheless display differences, notably in the distribution of vowel length. This study, based on read speech data collected through dedicated fieldwork, shows some asymmetries in the phonetic implementation of vowel length and its correlates in the two dialects: in Genoese, the phonetic manifestation of the vowel length contrast is more stable than in Portorino, which is in keeping with the fact that the latter is part of Western Ligurian, a subdivision for which contrastive vowel length has been previously reported to be losing ground[1].

Key words: vowel length, Ligurian dialects, sound change.

1. *Introduzione*

Indaghiamo nelle pagine seguenti le manifestazioni fonetiche della lunghezza vocalica distintiva in due dialetti liguri, il genovese e il portorino (parlato a Porto Maurizio, provincia di Imperia). Al § 2 si introduce la problematica collocando i fenomeni qui indagati in prospettiva (italo-)romanza ed esponendo gli scopi della ricerca. Al § 3 si descrivono la procedura di raccolta dei dati ed il metodo dell'analisi, i cui risultati si presentano quindi al § 4. Segue il § 5, conclusivo.

2. *La quantità vocalica distintiva in Liguria*

Buona parte del territorio ligure risulta inclusa nell'area italiana nord-occidentale che conserva la quantità vocalica distintiva (d'ora in avanti QVD) insorta – stando alla spiegazione cui gli autori aderiscono – per la degeminazione delle consonanti intervocaliche (cfr. Loporcaro, 2015, con la bibliografia precedente ivi citata). È noto da numerosi studi (cfr. Ageno, 1957: 12 ss.; Forner, 1975: 50; 1988: 458; Ricciardi, 1975: 60-69; Toso, 1997: 16, 26) come il genovese possegga QVD tanto nell'ossitono romanzo (contesto con minore incidenza nel lessico ligure che non altrove, non essendosi qui prodotta l'apocope delle vocali finali non alte dopo ostruente) quanto nel parossitono (ad es. [ˈpɔːsu] 'mi riposo' vs. [ˈpɔsu] 'posso'), costituendo con ciò la parte occidentale di un'area che include l'Emilia, mentre a nord del Po la QVD risulta prevalentemente – pur con alcune eccezioni – ristretta all'ossitono romanzo;

[1] Lo studio è stato concepito e redatto congiuntamente: a fini accademici sono però da attribuire a ML i §§ 1-2, a SS il § 3 e a DG i §§ 4-5.

ed è anche noto come il genovese mostri tale situazione sin dalle prime attestazioni medievali (Parodi, 1898: 100-102; Formentin, 2002: 102). Nota è pure dalle descrizioni su citate (Toso, 1997: 16; Forner, 1975: 51-52, 250-251) la redistribuzione delle vocali lunghe e brevi in genovese dovuta all'influsso di consonanti seguenti. Vi si cita di norma l'effetto allungante di /ɡ ɡʷ v z ʒ r/ ed inoltre, per dialetti diversi dal genovese, dell'esito della vibrante scempia (mentre /r/ < RR-), scomparsa in genovese ma mantenuta – come /ɹ/ o /ɾ/ – in varietà periferiche (cfr. Ghini, 2001: 183 sul dialetto di Miogliola, in provincia di Alessandria)[2].

Fuori dal genovese, è pure noto come le varietà agli estremi orientale (lunigianino) e occidentale (intemelio[3], con Ventimiglia e dintorni: cfr. Azaretti, 1982: 24-25) dell'area ligure non presentino QVD. Ma un inventario dettagliato della distribuzione sul territorio ligure della QVD resta ancora da produrre. Così, consta che il tipo genovese si estende dal Levante ligure (cfr. Plomteux, 1975: 32) al savonese (cfr. Viglione, 2006: 75, 90), mentre all'estremo occidentale della Liguria, come s'è detto, in ventimigliese (dove [ˈsonu] è sia 'sonno' che 'suono') la QVD manca. Nei dialetti parlati nel mezzo, costituiti in un raggruppamento ligure occidentale nella classificazione di Forner (1988: 458), quest'ultimo dichiara la lunghezza vocalica come pienamente determinata dal contesto, e dunque allofonica. Poiché si ha ragione di ricostruire uno stadio più antico in cui la QVD doveva essere uniformemente presente nell'intera regione, i dialetti intemeli e liguri occidentali che oggi ne sono privi debbono averla perduta.

Principali centri dell'area ligure occidentale sono Albenga ed Imperia; nel dialetto di Porto Maurizio[4] (che, insieme con la vicina Oneglia, è oggi parte di Imperia pur essendo storicamente un borgo a sé) è invece possibile reperire coppie minime, come in genovese, sia pure con alcune differenze[5].

A prima vista, il sistema vocalico di Genova e Porto Maurizio appare omogeneo, almeno per quanto riguarda l'inventario dei fonemi. In entrambe le varietà si

[2] Almeno in alcuni casi quest'effetto non pare essersi esplicato categoricamente, così da lasciar persistere alcune coppie (semi)minime, quali genov. [ˈmazu] 'maggio' ≠ [ˈvaːzu] 'vaso' (cfr. Toso, 1997: 16 e la discussione in Loporcaro, 2015: 93).

[3] Il termine, impiegato nella classificazione dei dialetti liguri di Forner (1988: 455) a indicare le varietà parlate fra Monaco e Taggia, rimanda al nome dei Liguri Intemelii, stanziati in epoca preromana (secc. VIII-VII) sulla sponda destra del Nervia. *Albintimilium* è il nome latinizzato del loro centro, precursore dell'odierna Ventimiglia.

[4] Gli studi dedicati al dialetto di Porto Maurizio non sono numerosi. Ricordiamo Mozio (2008), una grammatica del portorino con obiettivi primariamente didattici, Berardi (2010a), che offre un primo raffronto fra i dialetti di Porto Maurizio e Oneglia (circoscritto alla fonetica e alla morfologia), e Garibbo (1993), che presenta termini e modi di dire del portorino raccolti per aree tematiche. Per il vicino dialetto onegliese si vedano, fra gli altri, Ramella (1997) e Berardi (2010b).

[5] Del resto, il quadro dell'intera metà occidentale della Liguria appare alquanto mosso. Spostandosi infatti nell'interno, si incontrano varietà alpine sia dell'intemelio – al confine con l'occitano – sia del ligure occidentale che conservano la QVD: fra le prime, possiede QVD il roiasco di Breil e Briga (cfr. ad es. [ˈnaːz] 'naso' ≠ [ˈbras] 'braccio'; Dalbera, 1994: 126-129), fra le seconde ad es. il dialetto di Ormea, nel Cuneese: [ˈfryːtu] 'frutto' ≠ [ˈbrytu] 'brutto' (Schädel, 1903: 25).

osservano infatti quattro gradi di apertura e otto vocali[6], ognuna delle quali può presentarsi sia breve sia lunga (Tabella 1).

Tabella 1 - *Il sistema vocalico tonico genovese e portorino*

i	y		u
e			
ɛ	œ		ɔ
		a	

L'asimmetria fra i due dialetti sembra concernere piuttosto la consistenza e la vitalità dell'opposizione fra vocali brevi e lunghe. Il numero delle coppie minime che si distinguono per la lunghezza fonologica della vocale è infatti più ristretto a Porto Maurizio rispetto a Genova (cfr. anche nota 9) e, nei casi attestati anche nella varietà ponentina, il contrasto fra brevi e lunghe sembra espresso più debolmente (da un punto di vista fonetico) a Porto Maurizio. All'illustrazione di tale differenza è dedicato il nostro studio, nel quale ci proponiamo di dar risposta ai quesiti seguenti:

1. Data l'osservazione qualitativa dell'esistenza di QVD nel portorino (dialetto ligure occidentale), è possibile osservare differenze, rispetto al genovese, nella sua realizzazione fonetica? È inoltre possibile osservare differenze in merito ai correlati fonetici della QVD e, in particolare, alla durata della consonante postonica?
2. Se così è, è possibile trarne indicazioni proiettabili in diacronia, ad illustrazione di mutamenti avvenuti o eventualmente in corso in area ligure in quest'ambito della fonologia?

Nel seguito faremo riferimento sinteticamente a queste due questioni come Q1 e Q2. Quanto al secondo quesito (Q2), esso è ispirato a studi precedenti, condotti negli ultimi decenni, nei quali applicando un metodo rodato a partire dai lavori fondamentali di Arianna Uguzzoni (cfr. Uguzzoni, 1971 e vari altri studi sino a Uguzzoni, Busà, 1995), si è mostrato come la microvariazione nella realizzazione fonetica delle opposizioni di QVD in Italia settentrionale fornisca utili indicazioni ricostruttive, permettendo di individuare nei dialetti oggi parlati sul crinale appenninico emiliano ed in poche località alpine (cfr. rispettivamente Loporcaro, Delucchi, Nocchi, Paciaroni & Schmid, 2006 e Loporcaro, Paciaroni & Schmid, 2005, ed il quadro d'insieme in Loporcaro, 2015: 190-194, 219-226) traccia delle fasi di transizione del processo che ha portato infine il tipo italo-romanzo settentrionale a perdere la correlazione di geminazione consonantica ed a fonologizzare la QVD.

Da ultimo notiamo ancora che l'area ligure si segnala, fra le varietà italo-romanze settentrionali, per esser l'unica a mostrare la QVD estesa alla protonia. Tale ricorrenza in protonia contraddistingue il genovese – inclusa la sua varietà d'oltremare

[6] Lo schema presentato nella Tabella 1 è basato su Forner (1988: 458) per il genovese. Nella letteratura non mancano però proposte parzialmente discordanti; per Ricciardi (1975: 60), in genovese, in luogo di /œ ɔ/ si avrebbero invece /ø o/. Secondo Berardi (2010a: 9), in portorino, oltre a /ɔ/ sarebbe disponibile nel sistema anche /o/.

parlata a Carloforte e Calasetta, il tabarchino (cfr. Toso, 2005: 38, 48) – e diversi altri dialetti all'intorno, ad est (cfr. ad es. [kaːˈseta] 'calza' ≠ [kaˈsetu] 'stalla delle pecore' nei dialetti della Val Graveglia, Plomteux, 1975: 32), ad ovest (cfr. Viglione, 2006: 90 sul savonese) ed a nord del capoluogo (cfr. Petrolini, 1983: 232 s. e Vitali, Rulli, 2010: 10 sui dialetti liguri parlati in provincia di Parma). Di questo aspetto particolare della fonologia della QVD in ligure centrale daremo un'illustrazione sperimentale al § 4.3, verificando se essa ricorra anche in portorino.

3. *Raccolta dei dati e metodi di analisi*

3.1 Raccolta dei dati

I dati oggetto di questo studio sono stati raccolti nell'estate del 2016 nell'ambito di un lavoro di ricerca sul campo condotto a Genova e a Porto Maurizio. I parlanti sono stati registrati in ambienti poco rumorosi, benché non insonorizzati, con un registratore digitale Fostex FR_2LE e un microfono a cravatta Sennheiser MKE 2 connessi ad un computer portatile.

Agli informanti è stato chiesto di leggere per tre volte le frasi mostrate sullo schermo del PC all'interno di una presentazione *PowerPoint*. Le frasi erano già state trascritte nei due dialetti (per la grafia abbiamo fatto riferimento primariamente al sistema proposto da Toso, 1997 nella sua grammatica del genovese)[7]. Nell'ordine di presentazione delle parole si è evitato di mostrare le coppie minime in sequenza. Al fine di mantenere l'omogeneità dei contesti prosodici di ricorrenza, si sono inserite tutte le parole in frasi cornice del tipo esemplificato in (1):

(1)
a. Mi o dito *X* pe-a … vota (genovese)
b. A l'o ditu *X* pe-a … vota (portorino)
'Ho detto *X* per la … volta'

Il totale delle parole *target* ammonta a 89 item per il dialetto di Genova e a 79 per la varietà di Porto Maurizio[8]. Per quanto riguarda l'analisi oggetto del presente contributo, abbiamo selezionato un sottocorpus comprendente soltanto le coppie

---

[7] I parlanti intervistati a Genova, allievi di un corso serale di genovese organizzato a Genova Pra dal sig. Nino Durante, avevano già una certa familiarità con le scelte grafiche adottate e possedevano, in genere, una coscienza metalinguistica più sviluppata rispetto agli informanti di Porto Maurizio. Prima dell'inizio delle registrazioni, comunque, a tutti i parlanti è stata presentata la grafia (a questa presentazione è seguita una breve fase di *training*). Nel caso del portorino, abbiamo optato per alcune scelte discordanti rispetto al modello di Toso (1997) al fine sia di cogliere le specificità di questo dialetto rispetto al genovese sia di rappresentare la grafia in accordo con l'uso quotidiano e la sensibilità dei parlanti. Per questo motivo si sono tenute in considerazione anche le riflessioni metalinguistiche e ortografiche presenti nella produzione saggistica locale, in particolare in Garibbo (1993) e Berardi (2010a).

[8] Nel numero totale di item sono confluite anche parole che testimoniano in entrambi i dialetti allungamenti vocalici dovuti alla consonante postonica (come [ʒ] in [ˈlyːʒe] 'luce'), anche in sillaba originariamente chiusa (es. [r] da -ʀʀ- latina, come in [ˈtɛːra] 'terra'), e, al contrario, casi in cui alcune consonanti sono incompatibili con la lunghezza della vocale precedente, anche in sillaba aperta (es. [n]

(semi)minime: 19 per il genovese (incluse le coppie (semi)minime con QVD in protonia, cfr. il § 4.3) e 11 per il portorino[9]. Ne riportiamo alcuni esempi rappresentativi di diversi timbri vocalici:

(2)

a.	[ˈfasu]	'faccio'	≠	[ˈfaːsu]	'falso' (genovese)
b.	[ˈleze]	'(la) legge'	≠	[ˈleːze]	'leggere' (genovese e portorino)
c.	[ˈpɔsu]	'posso'	≠	[ˈpɔːsu]	'(mi) riposo' (genovese; in portorino:
	[ˈpɔʃu]		≠	[reˈpɔːsu])	
d.	[ˈfitu]	'affitto'	≠	[ˈfiːtu]	'presto, velocemente' (genovese)
e.	[ˈduze]	'dodici'	≠	[ˈduːse]	'dolce' (genovese e portorino)
f.	[ˈbrytu]	'brutto'	≠	[ˈfryːtu]	'frutto' (genovese e portorino)

Alla ricerca hanno partecipato dodici informanti, di cui sei per il genovese e sei per il dialetto di Porto Maurizio. Tre soltanto fra essi sono donne (due a Porto Maurizio e una a Genova). Nella Tabella 2, a fianco della sigla identificativa di ogni informante, sono riportati fra parentesi il sesso e l'età al momento dell'intervista[10].

Tabella 2 - *Gli informanti scelti per i due dialetti*

Genova	*Porto Maurizio*
CaGe (m, 77)	AcTo (m, 68)
DaGe (m, 66)	BoLu (f, 76)
DuNi (m, 68)	GeMi (m, 93)
FeGi (f, 62)	LaMa (f, 66)
ToFi (m, 54)	LuTo (m, 69)
VaFe (m, 38)	TeAn (m, 37)

3.2 Metodi di analisi

I dati del corpus ottenuto sono stati segmentati tramite il software *Praat* (versione 6.0.20, Boersma, Weenink, 2016). Per l'analisi statistica sono state usate per tutti i parlanti le segmentazioni relative alla seconda ripetizione di ogni lessema. Per ogni item si sono etichettate le vocali toniche e la successiva consonante postonica

o, in genovese, [ŋ] in [ˈlyna]/[ˈlyŋa] 'luna'), cfr. Loporcaro (2013: 95). Per una lista esauriente di tali contesti fonetici, si rimanda a Forner (1988: 458). Si veda anche il § 2 sopra.

[9] Il numero di coppie minime disponibili nel dialetto di Porto Maurizio è inferiore rispetto al genovese a causa, soprattutto, di mutamenti nel consonantismo che hanno impedito lo sviluppo della QVD in alcuni contesti. Per esempio, nel caso di [ˈfasu] ≠ [ˈfaːsu] (genovese) a Porto Maurizio troviamo [ˈfasu] ≠ [ˈfaɫsu]/[ˈfau̯su], con velarizzazione di -L- (rimasta inalterata nell'italiano 'falso').

[10] Fra gli informanti troviamo alcune piccole differenze di carattere diatopico: quattro dei parlanti genovesi sono originari di Genova Pra (e ivi residenti), mentre due sono originari del (e residenti nel) vicino comune di Arenzano. Nel caso dei parlanti di portorino, si tratta di informanti per i quali Porto Maurizio rappresenta, o ha rappresentato, il principale centro degli interessi familiari e lavorativi. Due di essi sono però originari della Val Prino (più precisamente, di Dolcedo e Villatalla, quest'ultima frazione del comune di Prelà), nell'entroterra di Porto Maurizio.

(se disponibile)[11]. In fase di segmentazione dei foni, abbiamo considerato in primo luogo i valori delle formanti (F1 e F2), e in caso di dubbio, la forma d'onda e il dato uditivo.

Le misurazioni in millisecondi sono state estratte automaticamente ed esportate su un file *Excel* grazie ad uno *script* di *Praat* sviluppato da Dieter Studer-Joho al laboratorio di fonetica dell'Università di Zurigo (cfr. anche Bernardasci, 2015).

La successiva analisi statistica è stata condotta per mezzo del software R (R Core Team, 2016) e dei pacchetti *ggplot2* (Wickham, 2009), per la realizzazione dei grafici, e *llme4 / lmerTest* (Bates, Maechler, Bolker & Walker, 2014), per l'analisi con modelli lineari misti (per una introduzione al loro impiego nella ricerca linguistica, cfr., fra gli altri, Levshina, 2015: 192-196).

Questi ultimi hanno il vantaggio di permettere l'esame del contributo sia degli effetti fissi (come il fattore 'dialetto' o 'quantità vocalica') sia di quelli contingenti (come la variazione fra i parlanti). La significatività (o meno) degli effetti principali nei modelli utilizzati è stata ottenuta automaticamente grazie al pacchetto *lmerTest*.

4. *Analisi dei dati*

4.1 La realizzazione fonetica della quantità vocalica nei due dialetti

La robustezza della realizzazione fonetica della QVD può essere espressa dal rapporto fra la lunghezza in millisecondi delle vocali fonologicamente brevi e di quelle lunghe[12] (cfr. Uguzzoni, Busà, 1995) (Tabella 3).

Tabella 3 - *I rapporti fra vocali brevi e lunghe nei due dialetti*[13]

	genovese	*portorino*
V/Vː	0,61	0,72

[11] I dati relativi alla coppie (semi)minime contengono parole parossitone e ossitone (per esempio [ˈda] '(lui) dà' ≠ [ˈdaː] 'dare'). Queste ultime sono state però escluse dall'analisi a causa dell'alto numero di errori prodotti dai parlanti durante le interviste. Tale confusione potrebbe essere tanto un portato del disegno sperimentale impiegato in questo studio quanto un segnale di instabilità, da interpretarsi come un indebolimento della QVD che avrebbe intaccato gli ossitoni. Infatti, le coppie minime con ossitoni terminanti in vocale, nei quali cioè la lunghezza della vocale non è influenzata da una consonante postonica, sono uno degli indizi usati per stabilire la forza della QVD in un sistema linguistico (v. Martinet, 1975: 205; si veda anche la breve discussione in Bernardasci, 2015: 122 a proposito del dialetto emiliano di Piandelagotti). Una risposta alla questione può giungere soltanto da uno studio specifico avente come obiettivo lo status della QVD negli ossitoni nei due dialetti liguri in esame.

[12] Il numero totale delle vocali e delle consonanti su cui sono stati calcolati i rapporti è 139 per il genovese e 82 per il portorino. Dall'analisi si sono esclusi gli ossitoni.

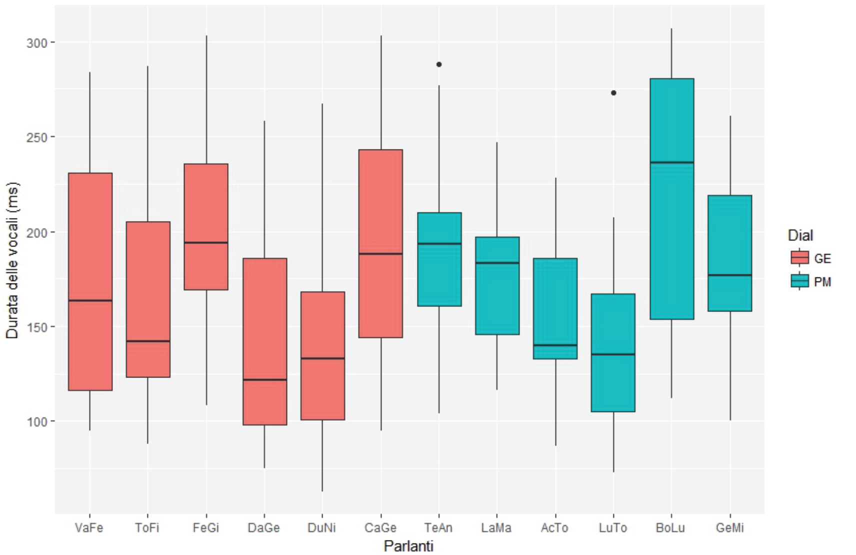
[13] I rapporti calcolati nella Tabella 3 (così come quelli nella Tabella 5) si possono confrontare con quelli disponibili per numerose altre varietà dialettali italo-romanze settentrionali (si vedano Filipponio, 2012: 246-247; Bernardasci, 2015: 124; Loporcaro, 2015: 190-194) con l'avvertenza che i risultati offerti da differenti lavori sono stati spesso ottenuti tramite metodologie di elicitazione non sempre comparabili.

I dati così ottenuti suggeriscono la stabilità della QVD a Genova (varietà in cui le vocali lunghe hanno una durata quasi doppia rispetto alle brevi), che invece appare segnalata in maniera meno pronunciata a Porto Maurizio.

Questo risultato ha un valore soltanto preliminare, poiché non ci fornisce alcuna informazione in merito alle molteplici variabili cui è soggetta la realizzazione fonetica delle vocali. Negli ultimi decenni la letteratura sperimentale ha infatti mostrato, sulla base di analisi dedicate a numerose lingue, l'influsso sulla durata delle vocali sia di fattori intrinseci, come il grado di apertura (cfr. Peterson, Lehiste, 1960; Lehiste, 1970: 18), sia di altri fattori dipendenti invece dai singoli parlanti, come la velocità di eloquio (Cooper, Paccia-Cooper, 1980).

La variabilità nella durata delle vocali prodotte dai singoli parlanti, siano esse brevi o lunghe, sembra in effetti notevole nel nostro corpus, come mostrano le Figure 1-3 (gli informanti, separati per dialetto, sono disposti in ordine progressivo di età all'interno di ciascun gruppo)[14].

Figura 1 - *Variabilità fra parlanti in relazione alla durata delle vocali brevi e lunghe*



[14] Ritorneremo sul possibile effetto dell'età sulla durata delle vocali nel § 4.4.

Figura 2 - *Variabilità fra parlanti in relazione alla durata delle vocali lunghe*

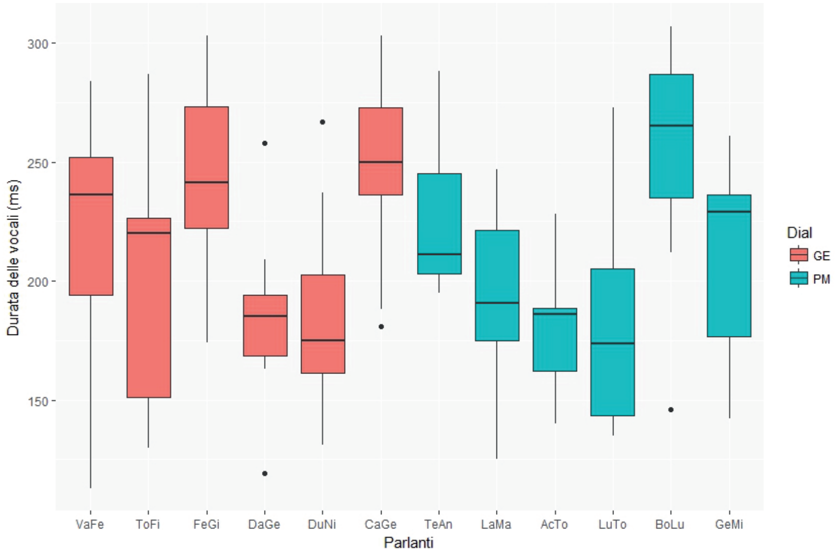
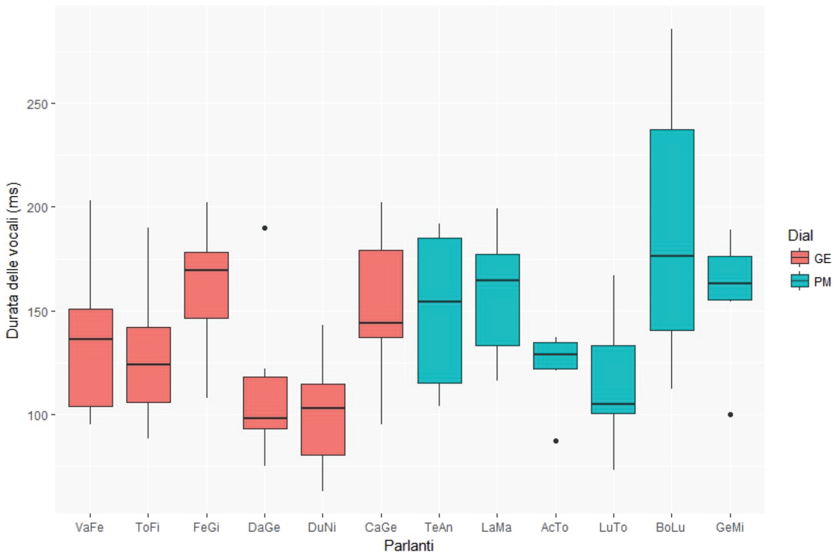


Figura 3 - *Variabilità fra parlanti in relazione alla durata delle vocali brevi*



Dai tre *box plot* emerge un quadro, oltre che eterogeneo, anche poco informativo: è infatti molto difficile tentare di ricostruire sulla base dei grafici uno schema di variazione nel nostro corpus. Se però accorpiamo i parlanti a seconda della varietà dialettale di appartenenza (cfr. Figura 4), diventa più agevole mettere a fuoco alcuni aspetti, come l'apparente preferenza (su cui cfr. già la Tabella 3) mostrata dai parlanti genovesi per il marcare in modo più netto rispetto ai portorini la diversa durata delle vocali brevi e lunghe.

Figura 4 - *La durata delle vocali brevi e lunghe nei dialetti di Genova e Porto Maurizio*

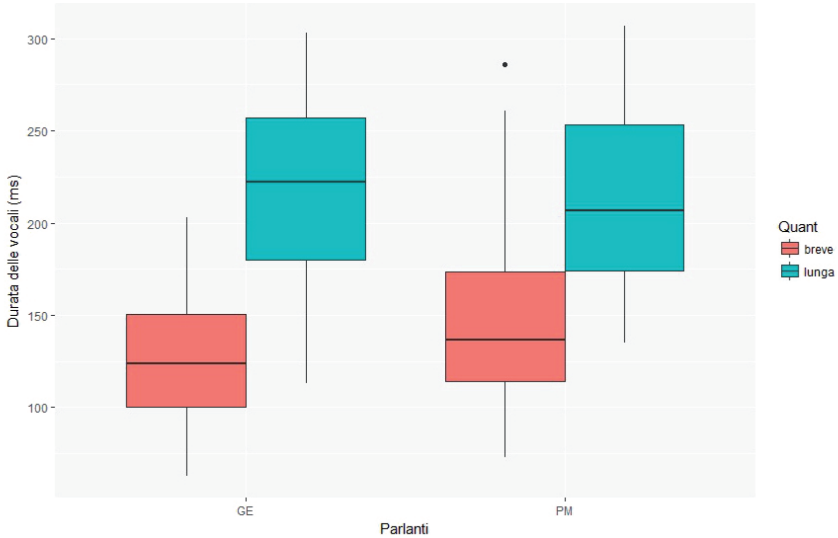


Tabella 4 - *Risultati dell'analisi con modelli lineari misti (fattori fissi)*

	β	*ES*	*T*	*p*
Intercetta	110,13	12,92	8,48	***[18]
Dialetto (portorino *vs.* genovese)	6,19	14,99	0,41	-
Quantità vocalica (breve *vs.* lunga)	85,12	8,85	9,61	***
Timbro (vocali basse *vs.* alte)	40,93	11,21	3,65	**
Timbro (vocali medioalte *vs.* alte)	29,44	10,71	2,74	**
Timbro (vocali mediobasse *vs.* alte)	25,97	13,74	1,89	-
Quantità vocalica (lunga)* Dialetto (portorino)	-21,76	8,91	-2,44	*

Per verificare la significatività statistica di questa osservazione intuitiva, si è scelto di ricorrere ad un'analisi basata su modelli lineari misti (cfr. § 3.2). La nostra analisi multifattoriale presenta come fattori fissi le variabili 'dialetto (con le modalità: genovese / portorino)', 'quantità vocalica (breve / lunga)', 'timbro (alto / medioalto / mediobasso / basso)' e, come fattori contingenti, la variazione individuale fra parlanti (variabile 'parlante', osservabile nelle Figure 1-3) e quella fra gli stimoli lessicali

[15] Seguendo la convenzione proposta in Gries (2013: 29), utilizziamo '***' per un valore p < 0,001, '**' per un valore 0,001 ≤ p < 0,01, e '*' per un valore 0,01 ≤ p < 0,05. Il simbolo '–' indica infine un risultato statisticamente non significativo.

elicitati (variabile ‘parole’)[16]. La variabile dipendente è la realizzazione fonetica delle vocali misurata in millisecondi. I risultati concernenti le stime dei fattori fissi nel modello massimale[17] sono riportati nella Tabella 4[18].

L’interpretazione dei risultati conferma dati noti da tempo nella letteratura fonetica (cfr. Lehiste 1970: 18 e ss.), quali per esempio l’influsso significativo del fattore ‘timbro’ sulla durata delle vocali. Le vocali basse e, in misura minore, quelle medioalte sono più lunghe rispetto a quelle alte.

Anche il fattore ‘quantità vocalica’ è rilevante: la differenza fonologica fra brevi e lunghe si riflette sulla durata in entrambe le varietà, confermando la presenza della QVD sia in genovese sia in portorino. Tuttavia occorre interpretare con attenzione il dato relativo all’interazione fra dialetto e quantità vocalica[19]. L’interazione è stata inserita nel modello per verificare uno degli aspetti centrali dell’ipotesi Q1 (v. il § 2), ossia l’effetto dell’interdipendenza fra variazione diatopica e asimmetria fra vocali fonologicamente brevi e lunghe nella realizzazione fonetica delle vocali nel nostro corpus. L’interazione è significativa (ß = -21,76; ES = 8,91; p = 0.015 [*]): le vocali lunghe del portorino sono mediamente più brevi di circa 22 millisecondi rispetto alle analoghe vocali del sottocorpus genovese. L’opposizione di durata fra brevi e lunghe è dunque più pronunciata nel dialetto di Genova; si tratta di un sostegno piuttosto forte a favore dell’ipotesi di un maggior radicamento della QVD in quest’ultima varietà.

4.2 La durata della consonante postonica nei due dialetti

Passando all’analisi dei correlati della QVD (cfr. la seconda domanda di Q1 nel § 2), il rapporto fra le consonanti postoniche che seguono una vocale breve e quelle che seguono una vocale lunga sembra rivelare un’asimmetria fra i due dialetti (il valore più alto a Genova indicherebbe una differenza minore fra esse rispetto a quella riscontrata a Porto Maurizio) (Tabella 5).

---

[16] Dall’analisi acustica dei dati è emerso anche il possibile ruolo dei correlati spettrali della QVD. Infatti, nei dialetti esaminati da Uguzzoni, Busà (1995: 30) le vocali brevi e lunghe non sono distinte soltanto da una diversa durata, ma anche da differenze timbriche/spettrali.

[17] L’analisi è stata condotta con il metodo di massima verosimiglianza (con approssimazione di Satterthwaite) su un numero totale di 221 vocali (139 per il genovese e 82 per il portorino) prodotte da dodici parlanti. Per quanto riguarda i fattori contingenti, la variabilità fra parlanti e i diversi stimoli lessicali sono responsabili della maggior parte della varianza nel modello.

[18] Nella Tabella 4 e in quelle seguenti, usiamo ‘ß’ per indicare la stima dei coefficienti dei fattori fissi e l’abbreviazione ‘ES’ per designare l’errore standard.

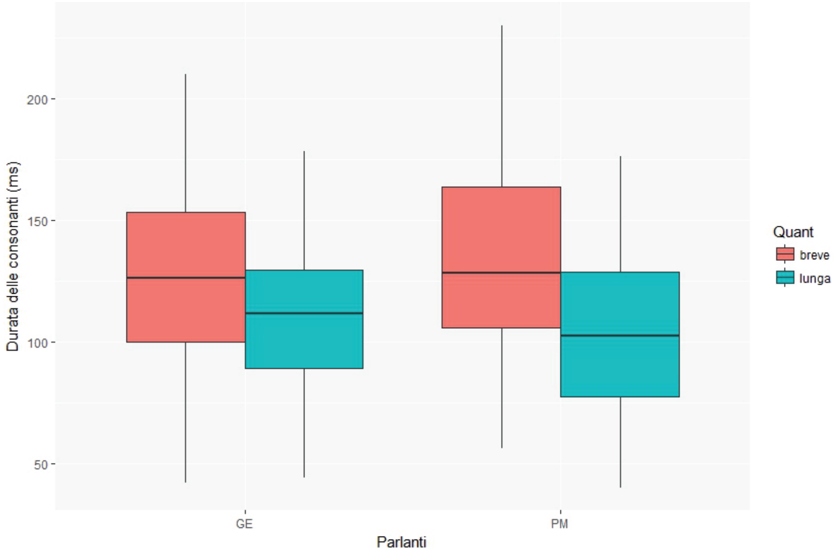
[19] In presenza di un’interazione, i singoli fattori interagenti devono essere trattati con cautela (“they are no longer main effects. They represent the estimates for the combinations of the specified level with the reference level of the interacting variable”, Levshina 2015: 195). Il fattore ‘dialetto’, al di fuori dell’interazione con ‘quantità’, non risulta comunque significativo neppure in un modello privo di interazioni.

Tabella 5 - *I rapporti fra consonanti postoniche nei due dialetti*

	genovese	*portorino*
(V:)C/(V)C	0,87	0,76

Tale asimmetria nelle due varietà pare osservabile anche nella Figura 5.

Figura 5 - *La durata delle consonanti postoniche nei dialetti di Genova e Porto Maurizio*



Al fine di verificare statisticamente questa differenza, si è fatto ricorso alla stessa tecnica (modelli lineari misti) utilizzata nel paragrafo precedente. In questo caso, però, la variabile dipendente è costituita dalla durata della consonante postonica, mentre le variabili indipendenti sono 'dialetto' e 'quantità vocalica' (nel modello sono computati anche i fattori contingenti 'parlanti' e 'parole') (Tabella 6).

Tabella 6 - *Risultati dell'analisi con modelli lineari misti (fattori fissi)*

	ß	*ES*	*t*	*p*
Intercetta	124,20	9,88	12,56	***
Dialetto (portorino *vs.* genovese)	13,61	11,09	1,22	-
Quantità vocalica (vocale lunga *vs.* breve)	-27,34	8,84	-3,09	**

La quantità vocalica fonologica ha un effetto significativo sulla durata della consonante postonica. Più precisamente, le consonanti nel nostro corpus sono significativamente più brevi dopo una vocale lunga.

A differenza di quanto atteso alla luce della Tabella 5 e della Figura 5, non troviamo però un'interazione significativa fra dialetto e quantità vocalica (in un modello alternativo, in cui si è inserita l'interazione fra le due varabili, il risultato si colloca infatti sotto la soglia della significatività statistica: β = -13,82; ES = 8,61; p = 0,11 [*]): la differenza di durata fra le consonanti postoniche dopo vocale breve e quelle che seguono una vocale lunga nel dialetto di Porto Maurizio non differisce significativamente rispetto al dialetto di Genova.

In entrambi i dialetti la differenza nella durata delle consonanti dopo vocale breve o lunga può essere interpretata come un correlato della QVD a livello segmentale e dunque come manifestazione della sua vitalità. In prospettiva interdialettale, la disparità nei rapporti mostrati nella Tabella 5, pur non risultando significativa nel nostro corpus, lascia comunque intravedere una tendenza da verificare in base ad un numero più ampio di dati e di parlanti.

4.3 Quantità vocalica distintiva in protonia

Come si è ricordato nel § 2, una delle proprietà tipiche del genovese – ed un vero *rarum* nel panorama delle lingue romanze – è la presenza di QVD in protonia, testimoniata dalle seguenti coppie (semi)minime:

(3)

a.	[kaˈseta]	'mestolino'	≠	[kaːˈseta]	'calzetta'[20]
b.	[fiˈsaː]	'fissare'	≠	[iːˈsaː]	'alzare'
c.	[œˈjaː]	'oliare'	≠	[vœːˈgaː]/[œːˈgaː]	'vogare'
d.	[byˈtaː]	'buttare'	≠	[fryːˈtaː]	'fruttare'

Questo fenomeno, come detto al § 2, è invece del tutto assente nella maggior parte dei dialetti liguri. Per verificarne un'eventuale presenza nel dialetto centro-occidentale di Porto Maurizio, abbiamo confrontato quattro coppie (semi)minime[21] protoniche del genovese con analoghe, potenziali, coppie (semi)minime nel dialetto di Porto Maurizio (Figura 6).

Il seguente modello[22], nel quale la realizzazione della vocale protonica è la variabile dipendente e i fattori fissi 'dialetto' e 'quantità vocalica' sono le variabili indipendenti (con 'parlanti' e 'parole' inseriti come fattori contingenti), conferma

[20] In questo caso la vocale lunga è risultato di un allungamento di compenso per caduta della laterale preconsonantica (cfr. Loporcaro, 2013: 95, nota 36). Come già osservato nella nota 9, nella maggior parte dei dialetti liguri, fra cui quello centro-occidentale di Porto Maurizio, l'indebolimento della -L- preconsonantica ha avuto esito differente ([l] > [ɫ]/[u̯]), non risultando dunque in un allungamento della vocale precedente.

[21] Nel portorino sono state usate soltanto due delle quattro coppie minime disponibili per il genovese.

[22] L'analisi statistica è stata condotta su un totale di 58 vocali (di cui 37 per il dialetto di Genova e soltanto 21 per quello di Porto Maurizio) prodotte da dodici parlanti. L'esiguità di tale dato deve naturalmente invitare alla cautela nell'interpretazione dei risultati.

la significatività dell'interazione, osservabile nella Figura 6, fra varietà diatopica e quantità vocalica fonologica in protonia (β = -50,36; ES = 21,10; *p* = 0,02 [*]).

Figura 6 - *La durata delle vocali protoniche nei dialetti di Genova e Porto Maurizio*

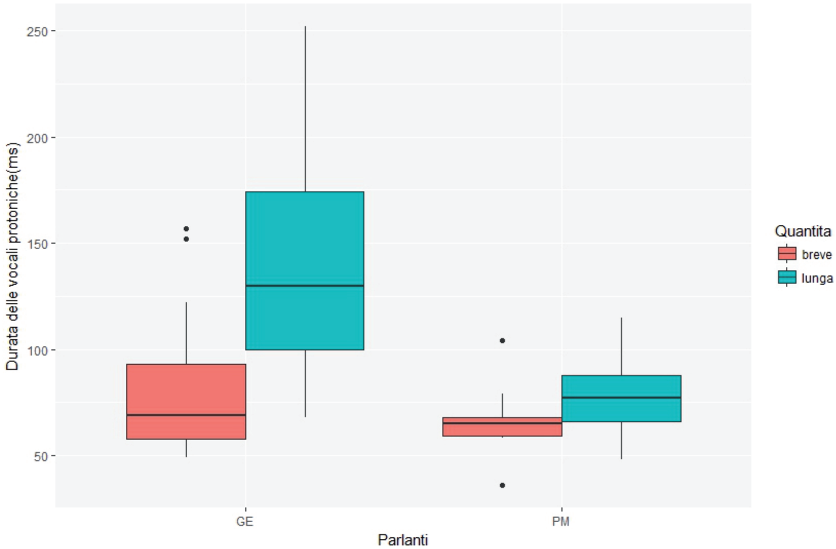
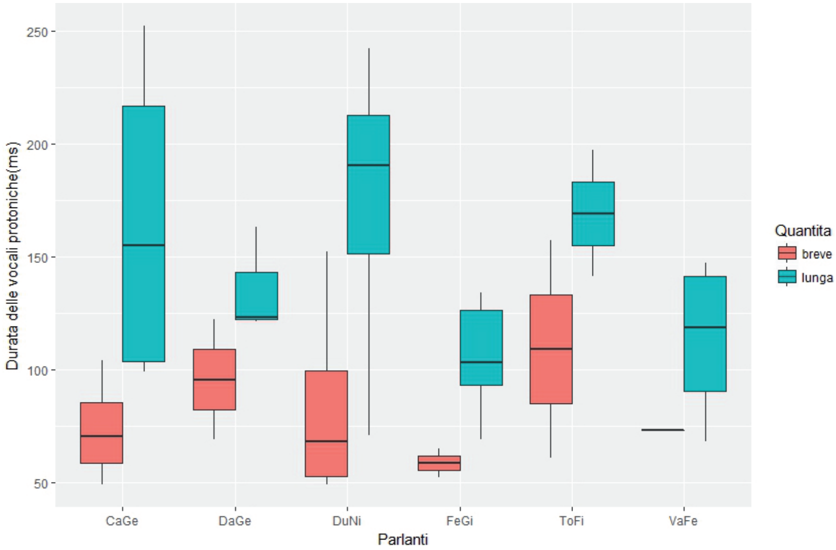


Tabella 7 - *Risultati dell'analisi con modelli lineari misti (fattori fissi) con interazione fra i fattori 'dialetto' e 'quantità vocalica'*

	β	*ES*	*t*	*p*
Intercetta	83,24	12,67	6,59	***
Dialetto (portorino)	-7,49	17,40	-0,43	-
Quantità vocalica (lunga)	51,94	14,94	3,47	**
Dialetto (portorino)* Quantità (lunga)	-50,36	21,10	-2,38	*

Possiamo affermare che nella varietà di Porto Maurizio le vocali corrispondenti alle lunghe protoniche del genovese hanno durata significativamente inferiore rispetto a queste ultime e dunque non si distinguono per durata dal resto delle vocali (brevi) atone. Ciò conferma che in portorino l'opposizione di QVD non si estende alla protonia. A Genova, invece, la QVD in protonia è ben salda e, come si vede nella Figura 7, è realizzata da tutti i parlanti del nostro corpus.

Figura 7 - *La durata delle vocali protoniche in genovese*



4.4 Possibili indizi di un mutamento in atto?

I dati esaminati nei paragrafi precedenti rivelano alcune differenze fra i due dialetti in esame. La QVD è presente in entrambe le varietà, anche se essa è segnalata in modo più marcato a Genova che non a Porto Maurizio. Per quanto riguarda la durata delle consonanti postoniche dopo vocali brevi o lunghe, invece, non si riscontrano differenze significative fra le due varietà.

Questi dati possono essere interpretati come la spia di un indebolimento della QVD a Porto Maurizio (e più in generale nei dialetti liguri centro-occidentali), che spingerebbe sempre più questa varietà verso i vicini dialetti di area intemelia (cfr. Azaretti, 1982), nei quali non esistono opposizioni fra vocali brevi e lunghe (§ 2)[23].

Per rinvenire indizi di un mutamento linguistico in atto in un corpus di dati sincronici è possibile analizzare la variabile sociolinguistica 'età', secondo la modalità laboviana dello studio in tempo apparente del mutamento linguistico (cfr. Labov, 1963; 1966; 1994: 43-72), per cui le differenze intergenerazionali possono riflettere cambiamenti in atto (cfr. anche Calamai, 2015: 67). Il nostro corpus, come risulta evidente dalla Tabella 2, non è purtroppo ideale per una simile analisi, data l'esiguità del numero dei parlanti e la scarsa differenziazione fra essi in termini generazionali. È tuttavia possibile operare un confronto preliminare fra gli informanti: in entrambi i

[23] Non si può escludere che il parziale allontanamento del portorino dal tipo italo-romanzo settentrionale (§ 2) sia anche una conseguenza dell'influsso dell'italiano standard su questa varietà. Se questa ipotesi fosse corretta, si dovrebbe ammettere che il genovese offre una resistenza maggiore all'italiano, conseguenza, forse, del più alto prestigio sociolinguistico del dialetto di Genova rispetto a quello di Porto Maurizio. Soltanto uno studio incentrato sulle differenze linguistiche fra i più giovani (termine con cui intendiamo le persone di circa 35-40 anni, data la nota difficoltà di reperire dialettofoni al di sotto dei 25-30 anni) e gli anziani potrebbe illuminare la questione.

dialetti abbiamo infatti un parlante sotto i 40 anni, un gruppo centrale di parlanti fra i 60 e 70 anni e, infine, almeno un informante di età vicina o superiore agli 80 anni.

Con la dovuta cautela, si è quindi condotta una ulteriore analisi statistica, in cui, a partire dal modello presentato nella Tabella 4 nel § 4.1, è stato inserito anche il fattore fisso 'età'. Tale variabile non risulta però avere un effetto principale significativo ($\beta = 0{,}10$; ES = 0,48; $p = 0{,}82$ [-]). Le differenze anagrafiche nel gruppo dei parlanti non hanno dunque un ruolo nella realizzazione fonetica delle vocali nel nostro corpus; una risposta soddisfacente a Q2 (v. § 2) potrà però essere fornita soltanto sulla base di un insieme di parlanti più ampio, equilibrato e rappresentativo.

5. *Conclusione*

Il nostro studio sperimentale ha confermato la stabilità dell'opposizione di QVD nel dialetto di Genova e ne ha altresì mostrato la presenza nel portorino, anche se con manifestazione meno marcata.

In quest'ultima varietà, non solo la QVD è osservabile in un numero inferiore di coppie minime, ma anche la durata delle vocali brevi e lunghe è significativamente differente rispetto al genovese. Tuttavia, a causa dei limiti nella composizione del corpus, non abbiamo potuto ricavare una diagnosi più precisa in merito ad un eventuale mutamento linguistico in atto nella varietà dialettale di Porto Maurizio[24].

Uno dei prossimi passi della ricerca, oltre all'impiego di un maggior numero di dati e al coinvolgimento di più parlanti differenziati per età, dovrebbe contemplare anche lo studio di più varietà liguri (dal tipo genovese a quello intemelio, passando per l'area centro-occidentale), per offrire una descrizione sperimentale più dettagliata della QVD e del suo graduale arretramento lungo la fascia costiera da Genova a Ventimiglia.

6. *Ringraziamenti*

Procedendo a ritroso, teniamo a ringraziare i tre revisori anonimi che con i loro commenti e suggerimenti hanno permesso di migliorare il testo; ovviamente, restiamo responsabili degli eventuali aspetti problematici rimasti allo stato attuale di questa ricerca. Ringraziamo inoltre Dieter Studer-Joho per la messa a disposizione dello *script* di *Praat* che ha permesso di velocizzare la misurazione delle durate vocaliche e consonantiche. Siamo grati anche a Sandra Schwab, per la preziosa consulenza statistica, e a Lorenzo Filipponio, per la discussione di questioni teoriche e di metodo. Grazie poi a Fiorenzo Toso e a Nino Durante per aver stabilito i contatti con i parlanti genovesi e a Tommaso 'U Luvu' Lupi per aver facilitato gli incontri

---

[24] È auspicabile anche una disamina più attenta di tutte le coppie minime e di tutte le vocali prodotte dai parlanti portorini per verificare in quali casi la QVD si manifesta con maggiore o minore chiarezza. Tale informazione sarebbe di grande utilità per ricostruire le varie fasi dell'indebolimento della QVD in questo dialetto.

con gli informanti di Porto Maurizio. Il nostro ringraziamento più sentito va infine a tutti gli informatori di Genova e di Porto Maurizio, i quali con la loro disponibilità e pazienza hanno reso possibile il presente studio.

Riferimenti bibliografici

AGENO, G.C. (1957). *Studi sul dialetto genovese*. Genova: Istituto internazionale di studi liguri, Sezione di Genova.

AZARETTI, E. (1982). *Evoluzione dei dialetti liguri esaminata attraverso la grammatica storica del ventimigliese*. Sanremo: Edizioni Casablanca.

BATES, D., MAECHLER, M., BOLKER, B. & WALKER, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. In *Journal of Statistical Software*, 67(1), 1-48.

BERARDI, E. (2010a). *Dialetti onegliese e portorino: convergenze fonetiche e morfologiche*. Imperia: Centro Editoriale Imperiese.

BERARDI, E. (2010b). *Fonetica storica del dialetto di Oneglia*. Imperia: Centro Editoriale Imperiese.

BERNARDASCI, C. (2015). Aspetti quantitativi del vocalismo tonico del dialetto di Piandelagotti. In VAYRA, M., AVESANI, C. & TAMBURINI, F. (Eds.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*. Milano: Officinaventuno, 113-127.

BOERSMA, P., WEENINK, D. (2016). *Praat: Doing phonetics by computer*, http://www.praat.org/Accessed 27.03.2017.

CALAMAI, S. (2015). *Introduzione alla sociofonetica*. Roma: Carocci.

COOPER, W.E., PACCIA-COOPER, J. (1980). *Syntax and Speech*. Cambridge MA: Harvard University Press.

DALBERA, J.-P. (1994). *Les parlers des Alpes-Maritimes, Etudes comparative, essai de reconstruction*. London: Association internationale d'études occitanes.

FILIPPONIO, L. (2012). *La struttura di parola dei dialetti della Valle del Reno. Profilo storico e analisi sperimentale*. Bologna: Arnaldo Forni.

FORMENTIN, V. (2002). L'area italiana medievale. In BOITANI, P., MANCINI, M. & VÀRVARO, A. (Eds.), *Lo spazio letterario del Medioevo. 2. Il medioevo volgare.* Vol. 2. *La circolazione del testo*. Roma: Salerno Editrice, 97-147.

FORNER, W. (1975). *Generative Phonologie des Dialekts von Genua*. Hamburg: Buske.

FORNER, W. (1988). Areallinguistik I: Ligurien. In HOLTUS, G., METZELTIN, M. & SCHMITT, C. (Eds.), *Romanistischen Linguistik* (LRL), vol. IV. Tübingen: Niemeyer, 453-469.

GARIBBO, D. (1993). *Parlâ du Portu*. *Raccolta di termini e modi di dire del dialetto portorino.* Imperia: Dominici.

GHINI, M. (2001). *Asymmetries in the Phonology of Miogliola*. Berlin-New York: De Gruyter Mouton.

GRIES, S.T. (2013). *Statistics for Linguistics with R. A practical introduction*. Berlin-New York: De Gruyter Mouton.

LABOV, W. (1963). The social motivation of a sound change. In *Word* 19, 273-309.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington DC: Center of Applied Linguistics.

Labov, W. (1994). *Principles of linguistic change. Volume 1: Internal factors*. Oxford, UK-Cambridge, USA: Wiley-Blackwell.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge MA: MIT press.

Levshina, N. (2015). *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam & Philadelphia: John Benjamins.

Loporcaro, M. (2013). *Profilo linguistico dei dialetti italiani*. Roma-Bari: Laterza.

Loporcaro, M. (2015). *Vowel length from Latin to Romance*. Oxford: Oxford University Press.

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, sistemi di annotazione*. Torriana: EDK, 491-517.

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.* Brescia: EDK, 579-618.

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

Mozio, F. (2008). *Gramàtica pè musciâ a scöra u dialéttu du Portu. Nozioni di morfologia e grafia*. Imperia: Centro Editoriale Imperiese.

Parodi, E.G. (1898). Studi liguri. § 1. Le carte latine. § 2. Il dialetto nei primi secoli. In *Archivio Glottologico Italiano*, 14, 1-110.

Peterson, G.E., Lehiste, I. (1960). Duration of syllable nuclei in English. In *Journal of the Acoustical Society of America*, 32(6), 693-703.

Petrolini, G. (1983). Sul carattere ligure delle parlate altovaltaresi. In Coveri, L., Moreno, D. (Eds.), *Studi di etnografia e dialettologia ligure in memoria di Hugo Plomteux*. Genova: Sagep Editori, 229-247.

Plomteux, H. (1975). *I dialetti della Liguria orientale odierna. La Val Graveglia*. Bologna: Pàtron.

R Core Team (2016). *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing. http://www. R-project. org/Accessed 27.03.2017.

Ramella, L. (1997). *U següu u nu loccia: il dialetto di Oneglia e della Vallata*. Imperia: Dominici.

Ricciardi, J.S. (1975). A brief phonology of three varieties of Ligurian Romance. Ph.D. Dissertation, Cornell University.

Schädel, B. (1903). *Die Mundart von Ormea*. Halle a.S.: Niemeyer.

Toso, F. (1997). *Grammatica del genovese: varietà urbana e di koinè*. Recco: Le Mani.

Toso, F. (2005). *Grammatica del tabarchino.* Recco: Le Mani.

Uguzzoni, A. (1971). Quantità fonetica e quantità fonematica nell'area dialettale frignanese. In *L'Italia Dialettale*, 34, 313-333.

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

Viglione, E. (2006). *Il puro vernacolo sabazio. Storia ed evoluzione bimillenaria della tipica parlata della nostra gente*. Savona: Sabatelli.

Vitali, D., Rulli, E. (2010). *Grammatica del dialetto di Compiano. Parlata ligure della Valle del Taro*. Compiano: Compiano Arte e Storia/Rupe Mutevole.

Wickham, H. (2009). *ggplot2*: *Elegant Graphics for Data Analysis*. New York: Springer.

DOI: 10.17469/O2103AISV000008

CHENG CHEN, CHIARA CELATA, IRENE RICCI

An EPG + UTI study of syllable onset and coda coordination and coarticulation in Italian

This study has two purposes. The first is that of implementing a system for the processing of multi-level speech data, which allows the synchronized acquisition and processing of electropalatographic, ultrasound tongue imaging and acoustic data. The second is that of testing the multi-level speech data system on a subset of Italian CV and VC sequences to ascertain how intersegmental coordination and coarticulation are related. We analyse onset-nucleus and nucleus-coda temporal relationships for three different consonants (/s/, /l/ and /k/) adjacent to either /a/ or /i/. Based on multi-level phonetic data on constriction location and overall tongue configuration, we show that the three consonants have different coarticulation properties and that these properties may in part explain the differences observed in the domain of temporal coordination between the consonant and the vocalic nucleus.

Key words: multi-level phonetic analysis, speech gesture coordination, coarticulation, syllable, Italian.

1. *Introduction*

This work explores the temporal coordination of articulatory gestures within the syllable in Italian, by comparing CV and VC sequences, where the consonant is, respectively, in onset and coda position. Articulatory models of syllable structure assume that the coordination between the vocalic gesture and the consonantal gesture may differ in the two cases. Based on previous literature on different languages, we expect to find differences in the temporal coordination of onset and coda consonants in Italian as well. In addition, recent literature suggests that the articulatory and coarticulatory properties of the segments play an important role in determining the details of the coordination patterns, and that not all segments or segmental sequences behave in the same way as far as gestural coordination within the syllable is concerned. Thus, an additional aim of this work is to compare consonants with different coarticulatory properties (in the sense of modifications of C articulation in varying vocalic contexts) and seek for possible relations between coarticulation and coordination patterns.

The methodology used is new. We establish an original system for the acquisition, real-time synchronization and analysis of acoustic, electropalatographic (EPG) and ultrasound tongue imaging (UTI) data, called *SynchroLing*. EPG and UTI instrumental techniques provide complementary information on, respectively, linguo-palatal contact patterns in the anterior vocal tract and midsagittal profiles of the whole tongue, including postdorsum and root. *SynchroLing* allows real-time inspection of contacts in the artificial palate and tongue midsagittal movements, coupled with acoustics. A preceding version of the multi-level system has been used for the analysis of rhotic variation in Tuscan (Spreafico, Celata, Vietti, Bertini & Ricci, 2015; Celata, Vietti & Spreafico, in press). One of the goals of this paper is that of testing the validity of this new instrumental environment for the analysis of coordination among segments and coarticulation. The experimental setting described here is challenging and beneficial at the same time. Among the challenges is the question of how we can identify landmarks (or anchor points) for temporal measurements starting from information about whole tongue or tongue-palate contact configurations. Among the benefits, a multi-level experimental environment provides fine-grained information about lingual movements, including areas of the oral cavity that are involved in the primary constriction and areas that are not, thus allowing the analysis of coarticulatory activity for the selected landmarks as well as for the temporal lags between them.

This paper is part of the first author's doctoral research project, which has a larger scope and includes the analysis of onset and coda consonant clusters and pitch accent variation. The current paper focuses on the following aspects only: the theoretical background (briefly sketched in § 2), the description of the methodology, with particular emphasis on the procedure for the identification of temporal landmarks for C and V gestures based on EPG and UTI evidence (§ 3), the coarticulatory patterns of /s/, /l/ and /k/ adjacent to /a/ and /i/ according to EPG indices and UTI profiles (§ 4.1) and the coordination patterns of /s/, /l/ and /k/ in onset vs. coda position (§ 4.2). Due to space requirements, acoustic data are not reported here, although they are an essential part of the current project.

2. *Background and hypotheses*

In articulatory models of syllable structure, the syllable is considered as a domain of articulatory timing. Different syllable structures may correspond to different characteristic patterns of phasing (Browman, Goldstein, 1988). In particular, syllable-onset positions are associated with tighter articulatory constrictions and greater stability than syllable-coda positions (since Krakow, 1989, 1999). Additionally, onset clusters are said to be globally aligned to the tautosyllabic vowel along their temporal midpoint in such away that an increase in onset complexity results in an increase of CV overlap ("c-center effect"); by contrast, coda clusters are sequentially organized (since Browman, Goldstein, 2000; De Jong, 2003; Hermes, Mücke & Grice, 2013 for Italian).

However, it has been recently shown that the articulatory and coarticulatory properties of the segments play an important role in determining the details of the coordination patterns. For instance, onset-vowel timing interacts with coarticulatory resistance, since the articulatory composition of the cluster predicts the degree of overlap between the final consonant of an onset cluster and the following vowel (e.g. Pastätter, Pouplier, 2015 on Polish). Similarly, not all coda clusters are sequentially organized with respect to the vocalic nucleus, since V shortening and an increased overlap between the vocalic nucleus and the following consonant have been observed for some complex coda clusters (e.g. Marin, Pouplier, 2010 for American English laterals; Marin, Pouplier, 2014 for Romanian rhotics).

This evidence suggests that the articulatory properties of segments (and primarily their coarticulatory resistance, e.g. Recasens, Espinosa, 2009) influence the timing of syllables and motivates a closer investigation of language-specific and consonant-specific coordination patterns.

A production experiment was run to investigate the difference between syllable onsets and codas in their temporal coordination with the syllable nucleus. Two hypotheses in particular were tested.

According to the first hypothesis, onset C gestures were expected to be more tightly coordinated to the vocalic nucleus than C gestures in syllable codas. Testing this hypothesis not only served the purpose of replicating a well-known articulatory timing effect on a different speech dataset, but – most importantly – served to establish if the procedure for the identification of temporal landmarks in *SynchroLing* (see below, § 3.4) was reliable or not. We anticipate that it was; by looking at *SynchroLing*-based gestural stability areas, it was possible to correctly identify the onset and offset of consonantal and vocalic gestures for an analysis of intergestural coordination that mirrors EMA-based coil tracking procedures.

The second hypothesis concerned the possibility that this coordination pattern varied as a function of the specific consonants involved. This hypothesis was suggested by findings reviewed above, related to the c-center effect in onset and coda clusters and the effects of coarticulatory resistance on shaping cluster-specific coordination patterns. Since the c-center effect in consonant clusters is one of the consequences of tight vs. loose coordination of syllable onsets vs. codas, respectively (since Krakow, 1999), we hypothesized that the degree of coarticulatory resistance could in principle affect also the way in which singletons are coordinated to the following or preceding vocalic nucleus. We tested three consonants, that were assumed to vary for their degree of resistance to vowel-induced coarticulatory modifications: /s/, /l/ and /k/. The former two are produced with the tongue tip as primary articulator, whereas /k/ with the tongue dorsum as primary articulator. Their degree of coarticulatory resistance was established by measuring the articulatory modifications of each consonant adjacent to /i/ as compared to adjacent to /a/. If coarticulatory resistance only affects the magnitude of the c-center effect (possibly because clusters are more susceptible than singletons to the influence of varying articulatory conditions), then the three consonants should show a similar difference between tight onset-nucleus coordination (i.e., a short latency between the consonantal and the vocalic gesture in CV sequences) and loose onset-nucleus coordination (i.e., a comparatively longer latency between the vocalic gesture and the consonantal gesture in VC sequences). If, on the contrary, consonantal properties such as coarticulatory resistance influence the onset-coda coordination pattern in singletons, the three consonants are expected to vary in the way they are timed to the vocalic nucleus as onsets or codas. In particular, spatial (coarticulatory) and temporal effects of vowel variation over adjacent consonants were expected to be positively correlated. Less resistant consonants (i.e., those that vary a lot as a function of vocalic context) were expected to be influenced by the vocalic gesture both in the sense of modifying their constriction location (spatial coarticulation) and in the sense of an increased temporal overlap between the vocalic gesture and the consonantal gesture. The effect was expected to be particularly strong for consonants in onset position. Thus, less resistant consonants in onset position were expected to show not only more variability of the constriction location in the /a/ vs. /i/ contexts, but also a stronger anticipation of the vocalic gesture. Consequently, they were also expected to show larger onset-nucleus coordination differences, compared to more resistant consonants. In the case of the consonants of the present study, /s/ was expected to be the most resistant consonant and to show the smallest onset-coda coordination difference; /k/ was expected to be the least resistant consonant and to show the largest onset-coda coordination difference; /l/ was expected to occupy an intermediate position between the two, as far as resistance to coarticulation is concerned, and therefore to show an onset-coda coordination difference smaller than /k/ and larger than /s/.

The reasons why we expect /s/ to be the most, /k/ the least and /l/ the intermediate resistant consonant are the following. Coarticulatory resistance (e.g. Recasens, Pallarès & Fontdevila, 1997; Recasens, Espinosa 2009) is known to depend on several factors, such as whether a given lingual region is involved in the constriction, jaw height, and the severity of the manner of articulation requirements. With respect to jaw height, /s/ having a higher jaw is known to be more resistant to coarticulation than both /l/ and /k/ (Recasens, 2012). Note that /l/ in Italian is always clear, i.e. with no involvement of the tongue dorsum in articulation. Frication with turbulence is also the most demanding manner of articulation, which reinforces the view that /s/ should show the comparatively highest degree of coarticulatory resistance. However, /l/ as a lateral continuant also has more demanding requirements than stops. In addition, the places of articulation of dorsal /k/ and /g/ are known to vary to a very large extent with vowel fronting (since Öhman, 1966; see also Fowler, Brancazio, 2000).

3. *Methodology*

3.1 Materials

For the current study, the corpus was composed of 12 disyllabic pseudo-words. They are shown in Table 1. The target consonants were /s/, /l/ and /k/. These consonants were included in non-word or very low-frequency word stimuli either as onsets (e.g. /saba/) or as codas (e.g. /bas/). The vowel was /a/ in one series, /i/ in another series.

Each stimulus was included in the carrier sentence *Pronuncia … molte volte* ("He pronounces … a lot of times"). In this sentence, the stimuli are produced with broad focus. The stimuli with /a/ were additionally produced in a prosodically prominent position in which the target word bore a corrective focus. In this case, the carrying sentence was structured as to favouring the emergence of a contrastive pitch accent on the experimental stimulus (e.g. *Pronuncia seba? No, pronuncia SABA molte volte!* "Does he pronounce seba? No, he pronounces SABA a lot of times!"). However, in this paper we deal with broad focus stimuli only. The total number of stimuli elicited was therefore 228 (3 consonants x 2 vowels x 2 syllable structures x 19 repetitions; see below, § 3.2, for details about repetitions). Due to mispronunciations (e.g., VC stimuli produced with a final schwa) or recording errors, the total number of stimuli analysed in the current paper was 216.

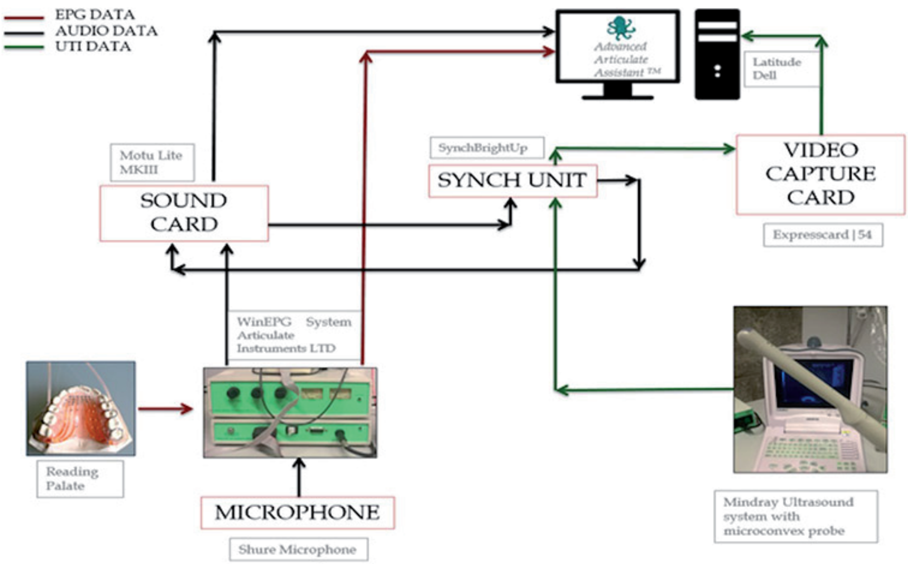
Table 1 - *Experimental stimuli used for the present experiment. For each stimulus, the orthographic and the phonetic transcriptions are given*

	/s/		/l/		/k/	
	CV	VC	CV	VC	CV	VC
/a/	saba[ˈsaba]	bass [bas]	laba [ˈlaba]	bal [bal]	capa [ˈkapa]	pac [pak]
/i/	siba [ˈsiba]	bis [bis]	liba [ˈliba]	bill [bil]	kipa [ˈkipa]	pic [pik]

3.2 *SynchroLing*: the audio-EPG-UTI synch system and the experimental procedure

This sub-section provides a detailed description of how *SynchroLing* works. The innovative aspect of the system lies in the real-time, automatic synchronisation of the three channels, which allows obtaining multi-level phonetic information simultaneously.

Figure 1 - *The instrumental setting of SynchroLing*



The system combines three parallel channels – one for the audio signal, one for the electropalatographic data (EPG) and one for the ultrasonographic data (UTI) (Figure 1). The EPG, UTI and audio signals were synchronized using a synchronization unit

(*SynchBright*) controlled by the Articulate Assistant Advanced software (AAA; version 2.16.14).

The acquisition platform was based on hardware and software components provided by Articulate Instruments Ltd[1]. The data were collected in the sound-proof studio in the linguistics laboratory of Scuola Normale Superiore. UTI data were collected at 60 Hz via a Mindray ultrasound machine with a microconvex probe (Mindray 65EC10EA 6.5 MHz). EPG data were collected at 100 Hz via the WinEPG™ (SPI 1.0) system.

Four subjects were recorded for this study. Each subject was recorded separately, according to the following procedure.

In the first place, the participant was asked to wear her/his personal artificial palate, and to practice with the word stimuli contained in the prompt sentences, by reading them aloud a sufficient number of times. The familiarization phase also served as an adaptation phase to reduce salivation and improve the overall comfort of the subjects in wearing the artificial palate. After 10-15 minutes, the ultrasonic transducer was fixed beneath the mandible of the participant with the help of a stabilizing helmet. The position of the probe was supposed to be orthogonal to the tongue surface so that the ultrasonic wave emitted from the probe was able to catch the entire lingual configuration during the movement of lingual gestures. Approximately half an hour after the wearing of the palate, the recording session began.

During the recording session, the speaker read the prompt sentences one by one, as they appeared in the AAA window on the computer screen in front of her/him. Five repetitions of the prompt list were recorded, except in one case, when one of the participants repeated the prompt list only four times. The order of the prompts was different across participants, but identical across repetitions for the same participant. The total duration of each recording session was never longer than 30 minutes.

The speech material was segmented and annotated according to one tier, i.e. a segmental tier, using the Praat software (version 6.0.29).

3.3 From gestural stability areas to segmental landmarks

3.3.1 Definition of segmental landmarks

For the analysis of onset-coda coordination difference, four articulatory landmarks were defined: Ctarget, Vonset and Vtarget for CV syllables; Vtarget, Conset and Ctarget for VC sequences in CVC syllables. We defined:

- Ctarget as the timepoint at which the consonantal gesture reaches its maximum constriction in the relevant lingual and palatal areas and bears minimal influences of coarticulation from the adjacent vowels;
- Vtarget as the timepoint at which the vocalic gesture reaches its target configuration and bears minimal influences of coarticulation from the adjacent consonant;
- For CV syllables, Vonset as a timepoint within the acoustic interval of the preceding consonant at which the nuclear vowel gesture begins;

[1] We acknowledge the contribution and technical support of Alan Wrench (Edinburgh) as well as the collaboration of Chiara Bertini (SNS Pisa) during the implementation procedures.

– For VC syllables, Conset as a timepoint within the acoustic interval of the preceding vowel at which the constriction gesture for the coda consonant begins.

To define the relative position of the segmental landmarks defined above, gestural stability areas were first detected on the two articulatory channels (EPG and UTI) separately. Then the two sources of information were combined to identify one reference gestural stability area and the landmarks for coordination measurements. The procedure is described below.

3.3.2 Gestural stability area: UTI evidence

The analysis of ultrasound tongue images was based on the idea that the target lingual configuration of a consonantal gesture differs maximally from the target lingual configuration of the adjacent vocalic gesture. The analysis of ultrasound tongue images was divided into three steps:

(i) Selection of a reference spline for the characterization of the vocalic gesture. The spline that happened to be the closest to the acoustic midpoint of the vowel was selected as the reference.

(ii) Identification of relevant fan radii for tongue vertical displacement measurement. Depending on the different articulatory properties shown by the lingual sagittal profile, two radii – the 7th (called "first front") and the 14th (called "first middle") from the right – were identified to maximally capture the articulatory differences between vocalic and consonantal gestures in terms of tongue vertical displacement in a relevant area of the oral cavity. The first front radius was seen to be responsible for the difference between the /a/ gesture and the consonantal gestures of /s/ and /l/, as well as for the difference between the /i/ gesture and the consonantal gesture of /k/. The first middle radius was seen to be responsible for the difference between the /i/ gesture and the consonantal gestures of /s/ and /l/, as well as for the distinction between /a/ and /k/;

(iii) Identification of a gestural stability area for the consonant involved and of the relevant Ctarget landmark. Once the reference spline for the vocalic gesture and the fan radius for capturing the C-V difference were established, the distance between the tongue position in the reference spline and the tongue position in all the splines included in the acoustic interval of the consonant along the selected fan radius was calculated (in millimetres). The temporal interval during which this distance reached its maximum was defined as the UTI gestural stability area of the consonant. So the UTI gestural stability area corresponded to the interval during which the tongue was maximally different (in terms of vertical displacement along the selected fan radius) from the vowel.

3.3.3 Gestural stability area: EPG evidence

A traditional procedure based on the contact index method (Fontdevila, Pallarès & Recasens, 1994) was elaborated for the identification of vocalic and consonantal gestures from EPG data. Since EPG allows the detection of the linguo-palatal contact during speech production, it is possible to get different kinds of information for various regions on the artificial palate by referring to different EPG indices, implemented in the AAA software.

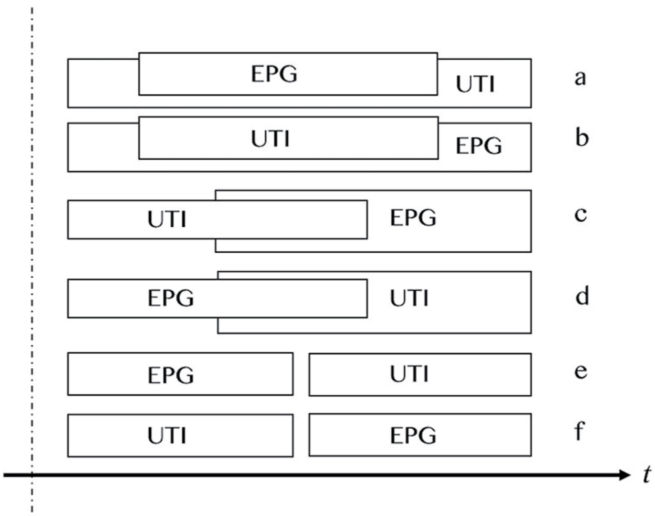
Two of these indices were used to detect the EPG gestural stability area. Specifically, the CAa index (contact anteriority in the anterior palate) was calculated on the three most anterior rows of each EPG frame included in the acoustic intervals of each /s/ and /l/ token. CAa served as an indicator of the anteriority of the linguopalatal contact of /s/ and /l/: the higher the CAa index, the more apical contact on the most anterior region of the palate. Consequently, the temporal intervals characterized by the maximum CAa values within the acoustic intervals of /s/ and /l/ were identified as the gestural stability areas of /s/ and /l/. For /k/, the Qp index (percentage of activated electrodes in the four back rows of the palate) was used: the higher the Qp value, the more contacted the posterior palate for the realization of the consonant. Thus the temporal intervals characterized by the maximum Qp values within the acoustic intervals of /k/ were identified as the gestural stability areas of /k/.

3.3.4 Integration of the two information sources: UTI + EPG gestural stability areas and calculation of gestural coordination latencies

We finally put the two gestural stability areas together, the intersection of them being identified as the stability area for the consonantal gesture, because both lingual profiles and palatal contact maximally distinguished such gesture from the adjacent vocalic gestures.

Six different conditions of temporal overlap of UTI and EPG stability areas were found in the data. They are schematically represented in Figure 2. The shadowed regions represent the final gestural stability areas (a-d). In only two items, the UTI and EPG stability areas were found not to overlap (e-f). In those cases, the EPG gestural stability area was taken as the gestural stability area.

Figure 2 - *Six idealised patterns of EPG-UTI gestural overlap, and resulting gestural stability areas*

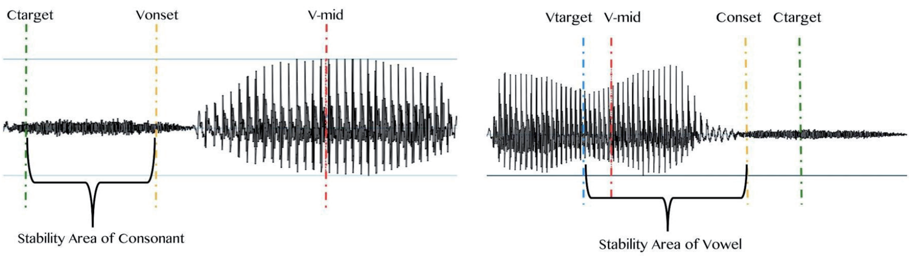


In CV sequences, the time point corresponding to the beginning of the gestural stability area was identified as Ctarget, whereas its offset was identified as Vonset (because we assume that when the position of the tongue starts to move from the consonantal target, the gesture for the following vowel begins). As anticipated in § 3.3.2, Vtarget corresponded to the acoustic midpoint of the vowel (Figure 3, left).

As to the VC sequence, the time point corresponding to the beginning of the gestural stability area was identified as Ctarget. A further step was therefore needed to identify Vtarget and Conset in VC sequences. In these cases, the UTI spline of Ctarget previously obtained was treated as the reference for the identification of the stability area for the vocalic gesture. The C-V distance along the same fan radius was calculated within the acoustic interval of the vowel. Similar to the third step described in § 3.4.1, the distance values plotted as a function of the temporal development gave a plateau representing the gestural stability area of the vowel involved. Then the same criteria were applied: the time point corresponding to the first spline of the gestural stability area was identified as Vtarget, the last one as Conset (Figure 3, right).

Once all the relevant gestural landmarks were identified for both CV and VC sequences, the gestural coordination latencies were calculated. The latency between Ctarget and Vonset was taken as a cue of CV coordination in CV syllables, whereas the latency between Vtarget and Conset was taken as a cue of VC coordination in CVC syllables.

Figure 3 - *Selection of temporal landmarks and identification of the gestural stability area in CV (left) and VC (right) sequences*



3.4 Analysis

The analysis of temporal coordination was based on the quantitative evaluation of gestural coordination latency variations, as illustrated above; the dependent variable was the milliseconds.

To evaluate if the three consonants differed for their coarticulatory properties, qualitative inspection of UTI profiles and quantitative evaluation of EPG indices variation over the acoustically defined consonantal interval were also run. As for EPG indices, CAa (contact anteriority in the anterior palate) for /s/ and /l/ and Qp (percentage of activated electrodes in the posterior palate) for /k/ were the dependent variables.

For both spatial analyses (temporal coordination and coarticulatory properties), the factors under evaluation were Syllable (CV vs. VC), Consonant (/k/ vs. /l/ vs. /s/) and Vowel (/a/ vs. /i/). The statistical analyses were univariate analyses of variance and the non-parametric Mann-Whitney test for distribution differences in independent samples. Non-parametric statistics was run whenever the requirements for ANOVA were not met (i.e., asymmetry and kurtosis higher than |1| and/or unequal variances, as attested by the Levene's test). The statistics were computed using SPSS (22.0.0).

4. *Results*

The results are presented in the following order: first, we present the results of the coarticulation analysis (lingual profiles and EPG indices); second, we present the results of the temporal coordination analysis.

4.1 Spatial coordination: coarticulatory patterns

4.1.1 Lingual profiles

Figures 4 to 6 display the average lingual profile for the three different consonants of the study (average of all lingual profiles included in the acoustic interval of the consonant of each stimulus) when adjacent to /a/ and to /i/, on a subject-by-subject comparison. In each figure, tongue root is to the left and tongue tip to the right of the images. The average tongue profile is represented by the darker line, whereas the upper and lower lighter lines identify the variance (standard deviation of the curves).

Figure 4 - *Comparison between average tongue profiles (with standard deviations) of /s/ adjacent to /a/ and /i/ in CV (upper half) and VC syllables (lower half) for the four speakers of the study. Upper graph: subject's palate in violet; /si/ in dark green. Lower graph: subject's palate in light blue; /is/ in black*

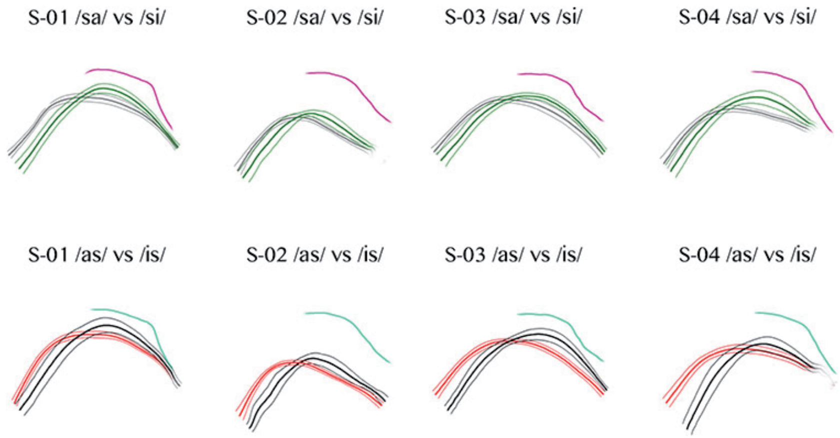


Figure 4 displays average tongue profiles during the production of /sa/ and /si/ (upper part) and /as/ and /is/ (lower part) across speakers. For all of them, the midsagittal profile is higher and more advanced in /si/ as opposed to /sa/; the difference is visible in the predorsal and dorsal areas, whereas in three out of four subjects the lingual profiles do not differ in the region of the anterior tongue (blade and tip). Only S04 appears to neatly differentiate the production of the sibilant in /sa/ vs. /si/ as far as the anterior tongue is concerned, with a higher position in /si/ than in /sa/. As far as coda /s/ is concerned, overall tongue advancement in the /i/ context is visible in all subjects, whereas anterior tongue advancement is present in one or maybe two out of four subjects (S03 and possibly also S02). Note also that for some of the speakers, the tongue tip is raised to a very limited extent, in both /a/ and /i/ contexts. These results therefore show that, although /s/ is produced with the lateral sides of the tongue raised towards the alveolar ridge, and the medial part of the tongue and the tongue tip slightly lowered to allow the air to escape from the front, there are nevertheless significant differences in the midsagittal profile as far as the dorsal and pre-dorsal areas are concerned. Additionally, there may be differences also in the case of the anterior tongue, both when /s/ is in syllable onset (one subject) and when it is in coda (two subjects). In one subject (S03), vowel-induced variation is larger in coda than in onset /s/.

Figure 5 - *Comparison between average tongue profiles (with standard deviations) of /l/ adjacent to /a/ and /i/ in CV (upper half) and VC syllables (lower half) for the four speakers of the study. Upper graph: subject's palate in violet; /li/ in dark green. Lower graph: subject's palate in light blue; /il/ in black*

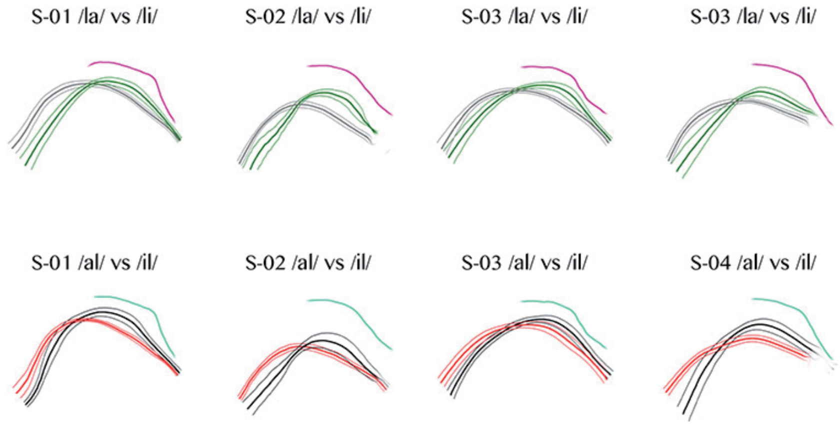
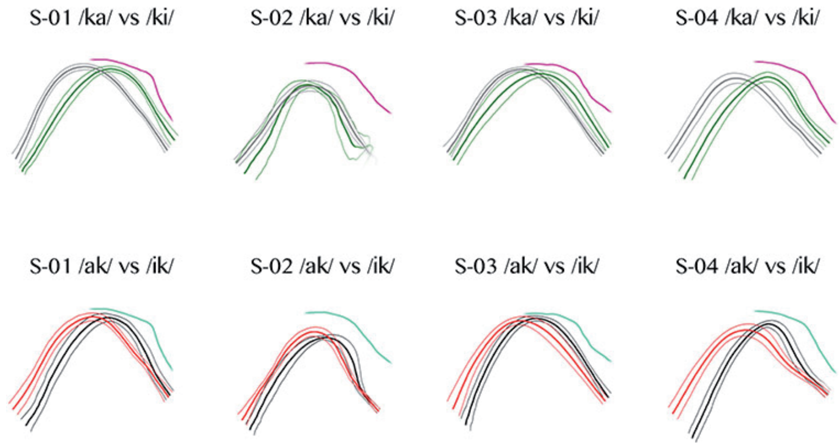


Figure 5 displays average tongue profiles during the production of /la/ and /li/ (upper part) and /al/ and /il/ (lower part) across speakers. As in the case of /s/, the tongue tip gesture is sometimes hardly visible. However, for all of the speakers, the onset consonant is produced with the tongue in a higher and more advanced position before /i/ than before /a/. However, the tongue tip (which is revealing of the constriction location for this consonant) is significantly higher in only two of the subjects (S02 and S04). In coda, tongue tip differences are visible in only one of the subjects (i.e., S04), thus suggesting that there is slightly less coarticulatory variation in coda than in onset position.

Figure 6 - *Comparison between average tongue profiles (with standard deviations) of /k/ adjacent to /a/ and /i/ in CV (upper half) and VC syllables (lower half) for the four speakers of the study. Upper graph: subject's palate in violet; /ki/ in dark green. Lower graph: subject's palate in light blue; /ik/ in black*



Finally, Figure 6 displays average tongue profiles during the production of /ka/ and /ki/ (upper part) and /ak/ and /ik/ (lower part) across speakers. As far as the onset consonants are concerned, three subjects (i.e., S01, S03 and S04) displayed neatly different tongue configurations, with overall tongue advancement before /i/, whereas S02 displayed undifferentiated tongue profiles in the two phonetic contexts. The pattern for coda /k/ is similar to that for onset /k/. Thus we can conclude that in three out of four subjects, the dorsal consonant displayed a more advanced lingual configuration when adjacent to /i/ than /a/, independently of its syllabic status.

4.1.2 EPG indices

The CAa index was expected to be higher when the consonant is adjacent to an /i/ than when it is adjacent to an /a/, because there is more fronting in the production of a high vowel than a low vowel. Table 2 shows the average CAa values of /s/ as a function of vocalic context and position in the syllable (onset vs. coda /s/). The data did not confirm the expectation regarding higher CAa values in /i/ than in /a/ sequences, since the Vowel factor did not differentiate among groups (CV items, Mann-Whitney test: $U = 201$, $z = 1.854$, $p > .05$; VC items: $U = 155.500$, $z = -.206$, $p > .05$). Thus the gesture for /s/ as measured by the CAa index turned out to be insensitive to variations in the vocalic context. However, there was a significant effect of Syllable, with coda /s/ exhibiting higher CAa values in coda than in onset position ($U = 817.500$, $z = 2.156$, $p < .05$). Thus the linguopalatal contact was overall significantly more fronted for coda than for onset /s/.

Table 2 - *Average CAa values for /s/ in CV and VC sequences as a function of vowel quality (/a/ vs. /i/)*

Syllable Type	*Vowel*	*N*	*average CAa*	*st. dev.*
CV	/a/	17	.918	.063
	/i/	18	.949	.032
	all	35	.934	.051
VC	/a/	18	.958	.024
	/i/	18	.953	.039
	all	36	.955	.032

Table 3 shows the average CAa values of /l/ as a function of vocalic context and position in the syllable (onset vs. coda /l/). The data confirmed the expectation regarding higher CAa values in /i/ than in /a/ sequences for onset /l/ ($U = 221$, $z = 2.657$, $p < .05$). By contrast, CAa did not change significantly across vocalic contexts in the case of coda /l/ ($U = 144$, $z = -.297$, $p > .05$). Thus, differently from /s/, the anteriority of the linguopalatal contact for /l/ was sensitive to variations in the vocalic context with more spatial overlap with the /i/ gesture when the consonant was in onset position. However, similar to /s/, there was also a significant effect of Syllable and the linguopalatal contact for /l/ was significantly more fronted in coda than in onset position ($U = 770.500$, $z = 2.107$, $p < .05$).

Table 3 - *Average CAa values for /l/ in CV and VC sequences as a function of vowel quality (/a/ vs. /i/)*

Syllable Type	*Vowel*	*N*	*average CAa*	*st. dev.*
CV	/a/	18	.951	.036
	/i/	16	.984	.017
	all	34	.967	.033
VC	/a/	17	.985	.025
	/i/	18	.982	.023
	all	35	.983	.024

The Qp index was expected to be higher when /k/ was adjacent to an /i/ than when it was adjacent to an /a/, because there is more dorsal activity in the production of a high vowel than a low vowel. Table 4 shows the average Qp values as a function of vocalic context and position in the syllable (onset vs. coda /k/). The data showed that, according to the expectations, Qp values were significantly higher close to /i/ than close to /a/ ($F(1,66) = 208.985$, $p < .001$). The interaction Syllable*Vowel was not significant ($F(3, 66) = 1.997$, $p > .050$), thus suggesting that the difference was equally present in CV and VC sequences.

Table 4 - *Average Qp values for /k/ in CV and VC sequences as a function of vowel quality (/a/ vs. /i/)*

Syllable Type	*Vowel*	*N*	*average Qp*	*st. dev.*
CV	/a/	18	.105	.036
	/i/	16	.283	.054
	all	34	.189	.101
VC	/a/	16	.105	.035
	/i/	17	.252	.050
	all	33	.181	.086
Total	/a/	34	.105	.036
	/i/	33	.267	.054

4.1.3 Summary of coarticulation results

The cumulative evidence of UTI lingual profiles and EPG contact indices suggested that there were varying patterns of coarticulatory activity and varying degrees of coarticulatory resistance across vowel contexts. However, we found variable coarticulation across syllable contexts and (limited to UTI data) across subjects as well. This provided additional information about the dynamics of spatial overlap of /s/, /l/ and /k/in different /a/ and /i/ contexts, proving that the way in which the articulatory gestures of an onset consonants vary to anticipate the lingual properties of the following vowel may be different from the way in which the same consonant, when in coda, accommodates to the lingual properties of the vowel, which precedes.

The sibilant was expected to show a comparatively higher degree of coarticulatory resistance, i.e., to vary the least as a function of vocalic variations. The analysis of linguopalatal contact in the anterior palate during /s/ production confirmed the hypothesis, inasmuch as no variation was found in the /a/ vs. /i/ contexts. However, the midsagittal lingual configuration was slightly different in the two vocalic contexts, according to the UTI analysis. Thus the tongue accommodated to the position required for the vowel (by reaching an overall more anterior position in the /i/ context and, for some of the speakers, a higher blade and tip), at the same time warranting the achievement of the target constriction with central groove in the anterior palate. There was, however, syllable-induced variation, with coda /s/ more fronted (both on average and for some specific speakers in particular) than onset /s/.

The lateral consonant was expected to vary more than /s/ as a function of vocalic variation, thus exhibiting less coarticulatory resistance than the sibilant. The expectation was met, inasmuch as the linguopalatal contact was found to vary in a statistically significant way, at least when the consonant was in onset position. However, the lateral was similar to the sibilant in two aspects. First, both consonants showed vowel-induced variation in the UTI midsagittal view, with an overall more anterior tongue configuration and – for a subset of the subjects – a higher blade and tip in the /i/ context. This variation affected the anteriority of the linguopalatal contact in /l/ only, consistently with the observation that /l/ is produced with a central apical contact. Second, both consonants exhibited syllable-induced variation, being more fronted in coda than in onset position.

Finally, the dorsal consonant showed a consistently variable lingual configuration and a statistically significant change in anteriority values as a function of changing vocalic contexts. The tongue during the production of /k/ was overall more advanced when adjacent to an /i/ than to an /a/. The fact that the vocalic nucleus precedes or follows the consonant did not impact on its coarticulatory modification pattern, which was strong in both syllabic contexts.

We could therefore conclude that, on the basis of additive evidence from mid-sagittal tongue imaging and linguopalatal contact analysis, a hierarchy can be confirmed for the three consonants of the study, with /s/ exhibiting the highest, /k/ the lowest and /l/ an intermediate degree of coarticulatory resistance.

4.2 Temporal coordination of C and V gestures

In this section, we verified if there is variation in the consonant-vowel temporal coordination patterns according to the different syllabic status of the consonant (onset vs. coda). In addition, we verified whether the coordination patterns are affected by the articulatory properties of the consonants involved, with specific reference to the degrees of coarticulatory resistance outlined above.

Table 5 shows the gestural coordination latencies across stimuli and subjects.

For the onset-nucleus coordination (CV syllables), the mean value of the Ctarget-Vonset interval across all valid tokens was 38 msec. For the nucleus-coda coordination (VC sequences in CVC syllables), the mean value of the Vtarget-Conset interval across all valid tokens was 105 msec. The temporal interval occurring between the consonantal gesture and the vocalic gesture in onset-nucleus sequences was therefore much shorter than the temporal interval between the vocalic gesture and the consonantal gesture in nucleus-coda sequences (negative difference: 67 msec), which suggested an opposition between tight coordination (quasi-simultaneous activation) and loose coordination (sequential activation) in CV vs. VC sequences. In an ANOVA with C-V latency as dependent variable and Syllable as between subject factor, the difference between CV and VC was statistically significant ($F(1, 205) = 523.869$, $p < .001$).

Table 5 - *Gestural coordination latencies in CV and VC sequences across subjects, vocalic contexts and prosodic conditions*

Syllable Type	*N*	*mean (ms)*	*st. dev.*
CV	103	38	20
VC	102	105	20.4

Table 6 additionally shows the gestural coordination latencies as a function of consonantal (/k/, /l/, /s/) and vocalic variations (/a/, /i/).

The mean values confirmed that the sharp distinction between CV and VC coordination patterns was present in all three consonants of the study, with a negative difference between the two conditions of 55 msec for /k/, 67 msec for /s/and 77 msec for /l/.

The interaction between Syllable and Consonant turned out to be statistically significant ($F(5, 201) = 5.906$, $p < .05$), thus confirming that the coordination pattern was different across consonants. In particular, post-hoc Tukey HSD test revealed that /l/ was significantly different from both /s/ ($p < .010$) and /k/ ($p < .050$), whereas /s/ and /k/ were not significantly different. Average latency values for VC sequences were similar across consonants, whereas for CV sequences there was a much lower latency value for /l/ (27 msec on average) than for the two other consonants (45 msec for /k/ and /42 msec for /s/).

The difference between CV and VC coordination was equally present in /a/ (70 msec) and /i/ stimuli (65 msec), as confirmed by the non-significant Syllable by Vowel interaction ($F(3, 201) = 0.713$, $p > .05$).

We could therefore conclude that the properties of the consonantal gesture influenced the gestural coordination patterns, whereas vowel quality did not.

Table 6 - *Gestural coordination latencies in CV and VC sequences as a function of consonantal and vocalic variations*

		CV			*VC*		
C	*V*	*mean (ms)*	*st. dev.*	*N*	*mean (ms)*	*st. dev.*	*N*
k	a	48	14	18	109	16	16
	i	41	10	16	91	13	17
	Total	45	13	34	100	17	33
l	a	29	13	17	104	22	16
	i	24	13	16	104	27	17
	Total	27	13	33	104	24	33
s	a	34	19	17	108	14	18
	i	51	29	18	109	16	16
	Total	42	26	35	109	15	34
Total	a	37	17	52	107	17	50
	i	39	22	50	101	21	50
	Total	38	20	102	104	19	100

5. *Discussion*

This study had two purposes. The first was that of implementing a methodologically challenging experimental setting, designed to advance our knowledge of intersegmental coordination and coarticulation and the relationship between the two; the second was that of verifying whether intersegmental coordination is influenced by the articulatory properties of the segments in a selection of stimuli produced by Italian speakers.

Concerning the first purpose, the experimental recordings were successfully run in *SynchroLing* and the speech materials were easily and accurately analysed within the AAA software environment, which allowed us to manage complex data sources in a unified environment. Therefore, *SynchroLing* proved itself a fruitful instrument for the processing of multi-level speech data. It invites further explorations in the domain of speech instruments synchronization and development.

The data provided by *SynchroLing* were used to characterize the dynamics of coarticulation and coordination between adjacent segments. This enriched window over speech movements was exploited here in two different ways. On one hand, it was used to determine the temporal dynamics of the consecutive constriction and opening gestures in sequences of consonants and vowels. This allowed us to identify onset and offset gestures from multiple sources of information, which cover together ample areas of the articulatory organs (as opposed to e.g. EMA-based tracing of individual coil movement). On the other hand, the multi-level analysis provided us with information on spatial adjustments of consonant and vowel lingual gestures from a variety of perspectives, e.g. by showing the effects of variable constriction points for those lingual regions that are not involved in the constriction themselves.

This pilot study was very limited in scope and additional work will be needed to improve methodological aspects and develop more sophisticated techniques and measures to map UTI-based tongue profile information onto EPG-based palate surface information (and viceversa). This will provide advantages both for articulatory phonetics (e.g. to quantify cross-instrument reliability of instrument-specific measures) and for phonetic theory (e.g. to determine the accommodation patterns of the tongue back to movements in the tongue front etc.) (see also Recasens, Rodriguez 2016; Spreafico *et al.*, 2016 for similar considerations).

A production experiment was run to investigate the difference between syllable onsets and codas in their temporal coordination with the syllable nucleus. The hypothesis that we wanted to test with Italian CVCV and CVC word stimuli was based on well-known effects of syllable structure on gestural coordination (see above, § 2). In particular, we tested whether onset C gestures are more tightly coordinated to the vocalic nucleus than C gestures in syllable codas. The hypothesis was confirmed in the present study, by showing that the temporal interval between Ctarget and Vonset in CV sequences is much smaller than the temporal interval between Vtarget and Conset in VC sequences. We use this evidence to suggest that the procedure for the identification of temporal landmarks in *SynchroLing* (see above, § 3.4) was reliable and that, by looking at EPG and UTI stability areas, it was possible to correctly identify the onset and offset of consonantal and vocalic gestures in the mono-and disyllabic experimental stimuli used here.

An additional hypothesis concerned the possibility that this coordination pattern varied as a function of the specific consonants involved. This hypothesis was suggested by the finding that, in some languages, different consonants may have different temporal behaviours within the syllable and the onset-coda opposition can show up differently according to the articulatory properties of the segments involved. As reviewed above (see § 2), the most frequently investigated coordination pattern in this respect is the so-called c-center effect for consonant clusters. The c-center effect is a consequence of tight vs. loose coordination of syllable onsets vs. codas and its magnitude and regularity have been shown to be influenced by segmental properties such as the degree of resistance to coarticulation with neighbouring segments.

The analysis presented here was limited to singleton onsets and codas; as already mentioned, clusters will be dealt with in a prosecution of the study. We hypothesized that different consonants may be timed to the vocalic nucleus as onsets or codas in different ways. In particular, less resistant consonants were expected to show a larger onset-coda coordination difference than more resistant consonants. To test this hypothesis, /s/, /l/ and /k/ were compared; /s/ was expected to be the most resistant, /k/ the least resistant and /l/ to occupy an intermediate position between them. Consequently, /s/ was expected to show a smaller onset-coda coordination difference than /l/, and /l/ was expected to show a smaller onset-coda coordination difference than /k/.

We discuss the results of the coarticulation analysis first. The multi-level approach adopted in this study provided us with information on variation in both constriction location and uncontacted portions of the tongue (midsagittal profile). Constriction location was predominantly evaluated on the basis of EPG index values. In this respect, /s/ was found to maintain the same linguopalatal constriction location before /a/ and /i/, in both onset and coda positions; /l/ showed fronting before /i/ in onset position only; /k/ changed in both onset and coda positions. EPG evidence appeared therefore to support our hypothesis concerning coarticulation degrees across consonants. The target constriction for /s/ notoriously requires fine control over the articulators in order to narrow the air channel and sustain turbulence during frication, which can explain the absence of vowel-induced variation. However, the same EPG indices also revealed that /s/ and /l/ were overall more fronted in coda than in onset. This effect of syllable position on constriction location did not generalize to /k/. According to Recasens (2004: 451-452), in Catalan /VC_1#C_2V/ stimuli /s/ and /l/ are more fronted when they are in coda position (C_1) than when they are in onset (C_2), provided that the adjacent consonant is articulatorily less constrained at tongue front. By contrast, if the adjacent consonant is more constrained at tongue front (e.g. /ʃ/), then the pattern is reversed (i.e., /s/ and /l/ have higher fronting values in onset than in coda). Since our stimuli include intervocalic /s/ and /l/, and vowels have a less constricted lingual configuration than the two apical consonants, we can conclude that our results are consistent with what has been observed for Catalan and lend further support to the generalizations stemming from an account of coarticulation patterns in terms of segments coarticulatory resistance.

Besides constriction location, tongue dorsum and postdorsum as revealed by UTI images were found to vary more for /l/ than for /s/. For both consonants, tongue tip position changed as a function of vowel quality only in a minority of the cases, thus indirectly confirming the results of the EPG analysis. However, there were nevertheless significant differences in the midsagittal profile as far as the dorsal and pre-dorsal areas were concerned. This clearly revealed that, even for strongly coarticulation-resistant consonants, such as /s/ (and slightly less resistant consonants, such as /l/), areas of the tongue that are not directly involved in constriction realization may accommodate to the articulatory requirements of the adjacent vowel, without substantially modifying the position and the amount of linguo-palatal contacts. In the UTI analysis, /s/ and /l/ were therefore similar with respect to coarticulation (whereas EPG data differentiated more clearly between the two). However, while /s/ showed (in one of the subjects) more vowel-induced variation in coda than in onset, /l/ showed (in two of the subjects) the opposite pattern, i.e. less vowel-induced variation in coda than in onset. Whether and how this finding is important for the overall picture of /s/ and /l/ coarticulatory behaviour sketched so far should be ascertained in future analyses, using quantitative measures of tongue profile variations and possibly comparing the behaviour of a larger number of informants. The tip and blade of the tongue were also found to consistently and significantly vary before /a/ as compared to /i/ for the dorsal consonant /k/. This finding suggested that variation in constriction location additionally implies a change in the overall tongue configuration. Thus UTI-based analysis apparently highlights a basic distinction between dorsal /k/, whose place of articulation changes a lot as a function of a modification of the vocalic context, and the two anterior consonants, that still change, but not necessarily for constriction location, and to a smaller extent.

In sum, the multi-level approach adopted here was found to increase the amount of information on vowel-induced modifications in lingual consonant production, thus showing up as a potentially relevant way of improving our understanding of coarticulation patterns. There are major and minor differences among consonants. Major differences (e.g. the difference between dorsal and apical consonants) are highlighted by both EPG and UTI. Minor differences (such as the extent to which the overall configuration of the tongue accommodates to varying constriction locations) may emerge differently, according to the specific source of articulatory data from which the information is drawn. More sophisticated ways of establishing correspondences and overlapping functions between the two instrumental outputs (as well as between the articulatory output and the acoustic data) have to be developed, to fully exploit the advantages of the multi-level analysis of speech movements.

We additionally tested the hypothesis that the onset-coda coordination difference varied as a function of the specific consonants involved. We found that the three consonants did vary in the way they differentiate onset-nucleus from nucleus-coda coordination, but not in the expected direction, which was a larger latency difference for less resistant consonants and a smaller latency difference for more resistant consonants. As a matter of fact, our findings met the expectations to the extent that the latency difference was bigger for /l/ than for /s/, but the finding that the latency difference for /k/ was even smaller than for /s/ did not fit our initial hypothesis. In a comparison between /s/ and /l/, the lateral showed a very short CV latency in onset position compared to /s/. This finding can be interpreted that, consistently with our initial hypothesis, not only the articulation of the following vowel influences the constriction location of the lateral more than it influences the constriction location of the sibilant, but also its temporal anticipation is stronger in the case of an onset /l/ than of an onset /s/. However, if this was the case, onset /k/ should have shown an even stronger anticipation of the vowel and a consequently shorter CV latency than /l/, which was not the case.

We think that there are at least two possible explanations for these findings. The first is that the onset-coda coordination effect is actually *not* affected by the coarticulatory resistance of singleton consonants. Resistance to coarticulation has been found to influence the c-center effect in clusters and maybe singletons are not as prone as clusters to suffer the influence of varying articulatory conditions. The second possibility is that the onset-coda coordination effect *is* affected by the coarticulatory resistance of the singletons, but only for a subset of them. In particular, one could hypothesize that it is only within the range of apical consonants that the degree of articulatory constraint determine the temporal effects of segments coordination; dorsal consonants could be hypothesized to be influenced by adjacent vowels only in terms of variable tongue configuration and changing constriction location. This difference might be due to the fact that dorsal consonants are produced by the movement of a less flexible and therefore slower part of the tongue, which might obscure the effects of tight C-V coordination in syllable onsets. The finding that /k/ showed the smallest difference between onset C-V and coda V-C coordination latencies (55 msec, as opposed to 67 msec for /s/ and /77/ msec for /l/) might bring support to this view, since it shows that the vocalic gesture is not strongly anticipated during the production of onset /k/. However, to be fully ascertained this hypothesis should be tested by comparing two dorsal (or non-apical) consonants with different degrees of coarticulatory resistance, to verify if the same hierarchy found for /l/ and /s/ applies to consonants produced more in the back in the oral cavity. Additional experiments would therefore be needed.

Bibliography

Browman, C.P., Goldstein, L. (1988). Some notes on syllable structures in Articulatory Phonology. In *Phonetica*, 45, 140-155.

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

Celata, C., Vietti, A. & Spreafico, L. (in press). An articulatory account of rhotic variation in Tuscan Italian: synchronized UTI and EPG data. In Gibson, M., Gil, J. (Eds.), *Romance Phonetics and Phonology*. Oxford: Oxford University Press.

De Jong, K.J. (2003). Temporal constraints and characterizing syllable structure. In Local, J., Ogden, R. & Temple, R. (Eds.), *Papers in laboratory phonology*, 6. Cambridge: Cambridge University Press, 253-268.

Fontdevila, J., Pallarès, M.D. & Recasens, D. (1994). The contact index method of electropalatographic data reduction. In *Journal of Phonetics*, 22, 141-154.

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

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.

Krakow, R.A. (1989). The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures. PhD Dissertation, Yale University.

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

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

Marin, S., Pouplier, M. (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. In *Journal of Phonetics*, 42, 24-36.

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

Pastätter, M., Pouplier, M. (2015). Onset-vowel timing as a function of coarticulation resistance: Evidence from Articulatory Data. In The Scottish Consortium for ICPhS (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK, the University of Glasgow. Paper number 783.

Recasens, D. (2004). The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters). In *Journal of Phonetics*, 32, 435-453.

Recasens, D. (2012). A study of jaw coarticulatory resistance and aggressiveness for Catalan consonants and vowels. In *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, 2288-2298.

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, 544-561.

Recasens, D., Rodríguez, C. (2016). A study of coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. In *Journal of Phonetics*, 59, 58-75.

Spreafico, L., Celata, C., Vietti, A., Bertini, C. & Ricci, I. (2015). An EPG+UTI study of Italian /r/. In The Scottish Consortium for ICPhS (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK, the University of Glasgow. Paper number 775.

DOI: 10.17469/O2103AISV000009

BARBARA GILI FIVELA, MASSIMILIANO M. IRACI

Variation in intonation across Italy: The case of Palermo Italian

The present work addresses the main characteristics of Palermo Italian intonation with the aim to extend previous investigations on the same variety of Italian and, even more importantly, with the final goal of improving the knowledge on variation in intonation throughout Italy. The ideas behind the present study are that, first, a better knowledge on intonation may be reached by focusing on a wide set of communicative contexts and by considering different speech styles, and, second, cross-variety variation may be better pointed out by adopting the very same methods used to investigate other varieties. Materials were then collected and analyzed by adopting the same criteria followed in a recent, wide study on Italian varieties. Results show that the adopted methodology allowed to obtain a better knowledge on Palermo Italian, in relation to both the analysis of sentence modalities which had been investigated before and new modalities and pragmatic contexts collected for the present study; moreover, cross-variety comparison was enriched by adding Palermo Italian, and confirming the lack of homogeneity in intonation patterns used within isoglosses traditionally proposed in the literature on vernaculars in Italy.

Key words: Intonation, Palermo Italian, variation, Autosegmental-Metrical framework.

1. *Introduction*

Linguistic variation throughout Italy is well-known even when prosody, and in particular intonation, is taken into account. Intonation patterns and their implementation may indeed vary considerably depending on the variety of Italian considered, as shown by means of works adopting both a phonetic and a phonological perspective (among the former, see Magno-Caldognetto *et al.*, 1978; Endo, Bertinetto, 1997; Romano, 2003; among the latter, e.g., Savino, 2012; Crocco, 2013; for an overview, see Gili Fivela *et al.*, 2015; Gili Fivela, 2008).

Phonological analyses are particularly interesting, given the effort in identifying functional categories out of the phonetic variability and the fact that analyses are often accompanied by descriptions of the main phonetic characteristics (e.g., alignment of tonal events with segmental landmarks), which may also be of interest. A phonological approach, such as that followed within the Autosegmental-Metrical framework (Bruce, 1977; Pierrehumbert, 1980; for an overview see Ladd, 1996), is characterized by the effort in finding categories out of the phonetic continuous variability, keeping in mind linguistically-induced variation such as that related to sociolinguistic factors. Very schematically, in Autosegmental-Metrical analyses high and low tones represent the target levels in a phonological system and are selected within the specific units called pitch accents and edge tones. Among the former, monotonal or bitonal accent, which associate to stressed syllables and add sentence level prominence (e.g., H*, L+H*, where the '*' indicate association to the tone bearing unit; tritonal accent are disfavored); among the latter, phrase accents and boundary tones (e.g., H-, H%, LH%), which are associated to the edge of prosodic constituents and represent relevant cues to prosodic boundaries of different levels. The phonological analyses are usually accompanied by a description of main expected phonetic implementation characteristics and are carried out by considering the main sources of linguistic variation.

In this respect, so far few works have been proposing wide and deep analyses of the Italian language, well-known to be characterized by a high degree of variability. Some of them focused on specific modalities or structures. For instance, working on Map-Task dialogues (corpus CLIPS[1]), Savino (2012) focused on yes/no questions and Crocco (2013) also described right dislocated constituents as produced in fifteen varieties of Italian (those spoken in Turin, Bergamo/Brescia, Milan, Venice, Genoa, Parma, Florence, Perugia, Rome, Cagliari, Naples, Bari, Lecce, Catanzaro, and Palermo). Only recently a larger number of sentence types and communicative contexts have been taken into account, that is statements, exclamations, yes/no questions, wh questions, imperatives and vocatives, by Gili Fivela *et al.* (2015) who analyzed speech material collected by means of both a Discourse Completion Task (DCT, Blum-Kulka *et al.*, 1989) and a read speech task. In particular, the authors collected and analyzed speech material representative of thirteen varieties of Italian, i.e. Turin, Milan, Florence, Siena, Pisa, Lucca, Rome, Pescara, Naples, Salerno, Bari, Cosenza and Lecce. The added value of such works, in comparison to the previous literature on the phonology and phonetics of Italian intonation (such as Grice at al., 2005) is the parallel analysis of speech data from a large number of varieties. Similar investigations were carried out, in fact, on the basis of the same conventions of transcription, as opposed to an *a posteriori* comparison of results that, moreover, was often obtained by different investigators who, even when referring to the same framework, may have not necessarily agreed on the analysis and transcription of specific phenomena. Moreover, such parallel analyses are characterized by the reference to the same speech style(s), which is not warrantied in works on different varieties/languages.

As far as the variation in the phonology of intonation is concerned, the widest comparative study performed so far was presented in June 2011 in Tarragona (Spain), during the Romance Tones and Break Indices (ToBI) workshop, and discussed in Gili Fivela *et al.* (2015). The authors pointed out that the variation observed in intonation patterns does not allow to identify isoglosses that resemble those traditionally proposed on the basis of vernaculars ("*dialetti*") synchronic structural differences or geographical and historical issues. This observation is in line with what independently pointed out by Savino (2012) for intonation in yes/

[1] Corpora e Lessici di Italiano Parlato e Scritto – Corpora and Lexicons of Spoken and Written Italian – www.clips.unina.it.

no questions, that is "the contour type is not geographically related". In particular, Gili Fivela *et al.* (2015) showed that variation may be either marked or limited, depending both on the specific function and on the pattern considered. In fact, in some cases it is possible to find one intonation pattern that can be used by speakers of different varieties of Italian (such as in broad-focus statements, lists, wh-questions, counterexpectational wh-questions, disjunctive questions, and vocatives). However, in other cases, strong variation is found in relation to both the intonation inventory selected by speakers of different varieties and the specific functions associated with nuclear configurations. As for the former case, a good example is represented by the use of L+H* vs. H*+L to express narrow-correction focus in Florence and Pisa Italian, respectively; an example of the latter case, is the pattern L+H* L% to signal yes/no questions in Cosenza and contrastive correction focus in Florence Italian. Anyhow, it is clear that, as far as intonation is concerned, it is not possible to identify across Italy the same isoglosses found in analysing vernaculars ("*dialetti*") as for other aspects of the grammar. That is, a representation such as that by Pellegrini (1977) is not useful to refer to variation in intonation.

As far as Palermo Italian is concerned, previous works focused on yes/no questions (early and late focus and conveying incredulity), statements (broad and narrow focus in contradictory sentences) and lists in read speech corpora (Grice, 1995, who compared British style analyses and Autosegmental ones). Specific investigations on right dislocation in both polar questions and declaratives (Crocco, 2013) and polar questions in general (Savino, 2012) were carried out on semi-spontaneous speech. However, no work has considered so far a wider set of sentence types and pragmatic contexts in different speech styles.

Goal of the investigation described here is achieving a better and wider knowledge on Palermo Italian, being also able to compare its features with those found in other varieties of the language. Our hypotheses are that 1) a better knowledge on intonation may be reached by focusing on a wide set of communicative contexts and by considering various speech styles; 2) cross-variety variation may be better pointed out by adopting the very same methods used to investigate other varieties. Thus, we decided to adopt the same methods and transcription conventions that have been used to perform the widest investigation available so far on intonation in Italian varieties (Gili Fivela *et al.*, 2015), always comparing analyses with relevant works on Palermo Italian.

2. *Previous works on Palermo Italian*

As already mentioned, there are basically three previous works on Palermo Italian proposing phonological inventories for such variety: Grice's (1995) work, whose data were also reported in Grice *et al.* (2005), Savino's (2012) and Crocco's (2013) analyses.

Grice (1995) and Grice *at al.* (2005) proposed the existence of pitch accents such as H+L* and H*+L in statements, to express broad and narrow-correcting focus respectively; L*+H in yes/no questions for both early and late focus, and L+H* to convey continuation in lists. As for the edge tones, L-L% are reported in the contexts mentioned above, that is after all the pitch accents listed so far, while L-H% is reported to be possibly found in yes/no questions conveying incredulity.

Later, Savino (2012) confirmed that the L*+H L-L% pattern is the most frequent one which is used for expressing yes/no questions in Palermo Italian, though "an extra terminal rise after the accentual rise has been occasionally observed [...] when yes-no question utterances are read aloud" (2012: 26). In the data she analysed, the low boundary tones are found in 83.6% of cases, while a final rise is observed in 8.5% of utterances (2012: 33). The analysis was basically left unvaried by Crocco (2013), who analysed the prosody of right dislocated constituents.

Table 1 - *Patterns described in previous works on Palermo Italian intonation (see discussion in text)*

	utterance type	*pattern*	
Statements	broad focus	H+L*	L-L%
	narrow-corrective focus	H*+L	L-L%
	lists	L+H*	L-L%
Yes/no questions	early and late focus	L*+H	L-L%
	incredulous (or read aloud)	L*+H	L-H%

3. *Palermo Italian: The extended model*

3.1 Methods

Along the lines of Gili Fivela *et al.* (2015), data were collected by audio recording 5 Palermo Italian speakers (3F, 2M, aged 20-30 years). All speakers had been continuously exposed to Palermo Italian, used it for everyday conversation, and had a similar educational level, that is high-school to university degree. Speakers were asked to perform a Discourse Completion Task (Blum-Kulka *et al.*, 1989), including 57 situation/contexts presented in pseudo-randomized order. The situation/context was used to have the speaker immersed in the intended pragmatic circumstance and induced to refer to specific lexical words. The speaker was indeed asked to spontaneously react to the given context/situation and, later, to read aloud a sample sentence, in which both lexical entries and sentence structure were controlled.

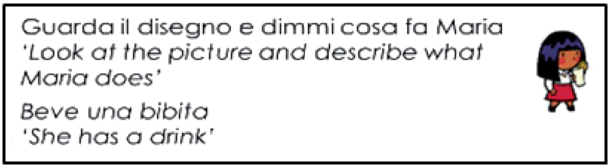
Usually, situations/contexts were included to elicit at least two target sentences for each sentence type, mainly to facilitate the collection of patterns realized on target words showing different stress positions. For instance, for broad focus declaratives, the contexts eliciting both *Mangia I mandarini* 's/he eats tangerins' and *Beve una bibita* 's/he drinks a soft drink' are included in the corpus (having pictures/drawings showing someone eating a tangerine/drinking a soft drink and asking the subjects to state what s/he is doing); lists, to offer another example, are elicited by means of two contexts, one requiring the subject to list the days of the week and the other one asking him/her to list the favourite fruits. However, in few cases only one situation is part of the corpus. This is the case, for instance, of exclamations (see Figure 1) and counter-expectation polar questions (where the context offered to elicit the sentence *Loredana un ingegnere?!* 'Loredana an engineer?!' asks the subject to think to be informed that a friend, who has never been good at math, is now an engineer; the subject is supposed not to believe in the information and to ask for confirmation while explicitly communicating s/he disbelieves it).

In particular, each time a situation/context and an example of response were proposed (see Figure 1), speakers were asked to:

1) read carefully and understand a written text describing a context/situation, presented over the PC screen;
2) produce a spontaneous utterance which would fit with the situational context presented;
3) read as spontaneously as possible the target sentence proposed by the experimenters as suitable for the same context.

The whole set of target utterances was presented twice, that is we collected 4 target utterances for each situation/context. Interviews were carried out by the second author, which is speaker of Palermo Italian.

Figure 1 - *Example of slide used to elicit the speech material (adapted from Gili Fivela et al. 2015)*



The analysis was framed within the Autosegmental-Metrical theory (Bruce, 1977; Pierrehumbert, 1980; for an overview see Ladd, 1996). Auditory analysis and inspection of fundamental frequency tracks were performed by the authors in order to highlight the main phonological and phonetic features of Palermo Italian intonation. Specific attention was devoted to spontaneous renditions, though alignment characteristics were nevertheless confirmed by means of read speech productions (i.e., subtask 3 above). Importantly, if not specified, the analyses are considered not to depend on speech style.

In line with annotation convention established by Gili Fivela *et al.* (2015: 149), "combinations of equal tones are collapsed and represented by one symbol only (e.g., L-L% becomes L%) and sequences of different edge tones are reported with no intermediate hyphen".

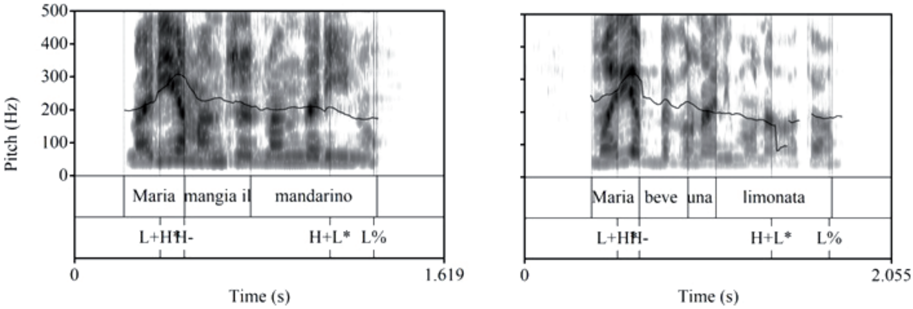
3.2 Results

3.2.1 Statements

Broad focus statements are realized by means of a H+L* L% pattern in line with what reported in the literature on Palermo Italian (see section 1) and in line with what observed in other varieties of Italian. As already observed by Gili Fivela *et al.* (2015) a high variability in H+ scaling is found, though in the case of Palermo Italian data it seems to be possible to relate it to inter-subject or style variability. Indeed, most of the time (72.5%, tot. n. 40) no clearly high leading tone target is identified, and 73% of the remaining 27.5% instances (i.e., those showing a clearly high target for the leading tone) are produced by one single speaker (see Figure 2, speaker 1 vs. speaker 2).

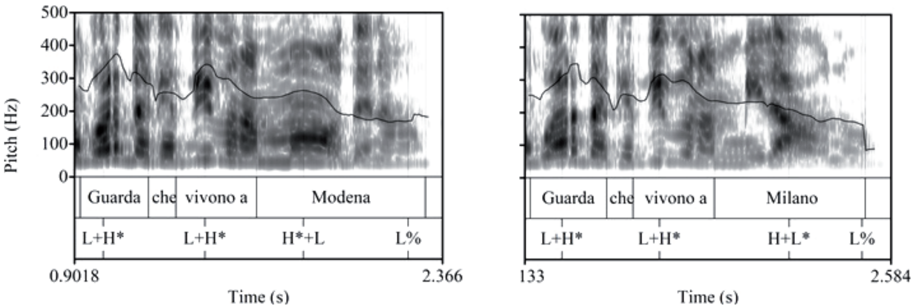
As for lists, they are usually characterized by either H+L* or L+H* on all the items but the penultimate, which usually bears a L+H*, and the last one, which rather carries a H+L*. In some cases, a delayed peak for the L+H* accent seems to be realized. The final edge tone is usually low, though a high one may also be found.

Figure 2 - *Broad focus statements* Maria mangia il mandarino *'Maria has a mandarin' and* Maria beve una limonata *'Maria is drinking a lemonade', speaker 1 (left) and 2 (right)*



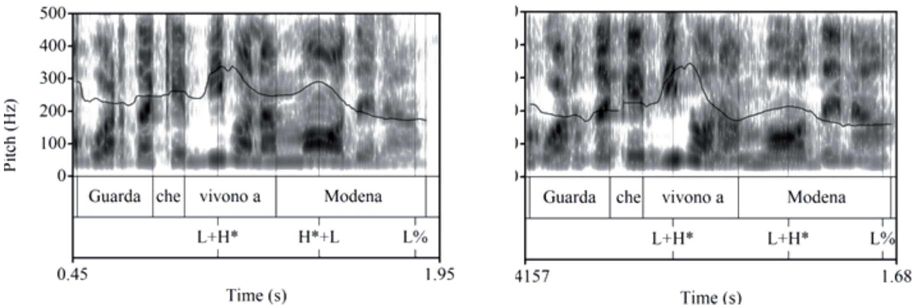
Narrow contrastive-corrective focus in Palermo Italian statements is realized with a H*+L pitch accent (67.5%,) in line with what reported in the literature on Palermo Italian (Grice, 1995, Grice *et al.*, 2005) – see Figure 3, left panel. In about 1/3 of cases, the realization implies a level rather than a rise to the H* target within the H*+L. Thus, given the two main pitch accents used to express corrective focus in Italian, Palermo Italian shows the same phonological categories found in Rome, Pisa, Pescara, Cosenza, Bari and Lecce (and different from the L+H* found in Milan, Turin, Florence, Siena, Lucca, Naples and Salerno).

Figure 3 - *Narrow contrastive-correction focus* Guarda che vivono a Milano *'They live in Milan': example of H*+L and H+L* pitch accent*



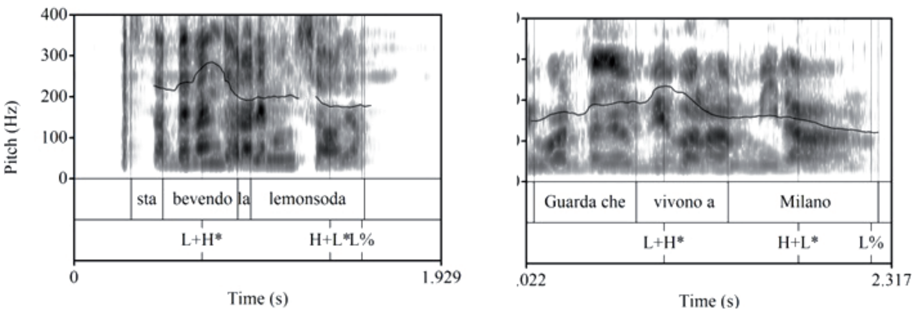
However, in line with what observed for other varieties of Italian, more than one pitch accent is found in the expression of focus: in some cases H+L* is used (20% of cases – Figure 3, right panel) and in some other cases a rising L+H* accent is found (around 12% of cases, but in just two speakers – speaker 1 and 4 – and always during the second repetition, as if a different nuance could relate to the reduced novelty of the context or to variation in either illocutionary force or politeness – see Figure 4).

Figure 4 - *Narrow contrastive-correction focus* Guarda che vivono a Modena *'They live in Modena': example of L+H* pitch accent*



However, a typical feature of Palermo Italian intonation and 'local accent' seems to be related to the presence of a prenuclear L+H* accent that shows a particularly high (and apparently slightly delayed) peak: the scaling relation between the preceding prenuclear and the nuclear accent seems to characterize Palermo Italian statement intonation (see Figure 5). The height and prominence of such accent seems also to be enhanced by the lack of pitch accent on other potential preceding tone bearing units in the phrase (in fact, a deaccentuation or prominence downgrading of items that would usually be pitch accented in other varieties seems to take place in such contexts) – see Figure 5, right panel.

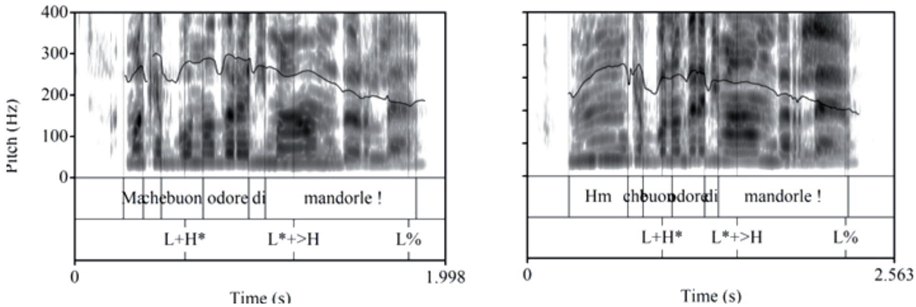
Figure 5 - *Broad focus statements* Sta bevendo una limonata *'She is drinking a lemonade'; Narrow contrastive-correction focus* Guarda che vivono a Milano *'They live in Milan'*



3.2.2 Exclamatives

In Palermo Italian, in line with other varieties (Gili Fivela *et al.*, 2015) exclamatives are expressed by means of a L*+>H pitch accent (80% of cases – see Figure 6), usually followed by a L% boundary tone[2].

Figure 6 - *Exclamatives* Ma che buon odore di mandorle! *'What a good smell of almonds!': examples of L*+>H*



As already observed in Turin, Florence and Siena, this option alternates with a L+H* L% pattern (around 20% of cases – see Figure 7)[3]. Palermo Italian exclamatives show a globally wider pitch range or a higher register than usual, as already observed for exclamatives in other varieties of Italian. As suggested in Gili Fivela *et al.* (2015) it is still a matter of debate if an initial %H boundary tone could be appropriate to describe these contours.

3.2.3 Yes/no questions

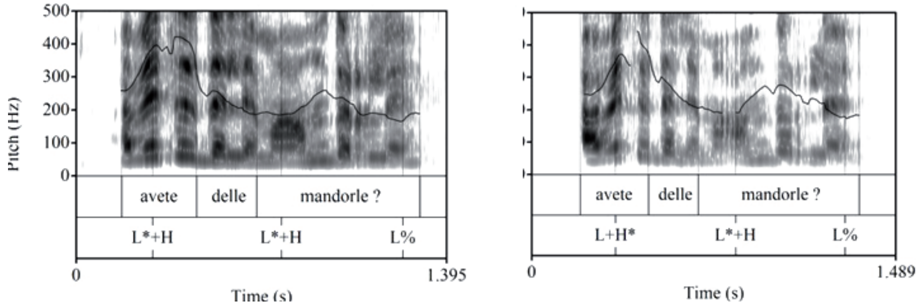
3.2.3.1 Information seeking yes/no questions

In line with previous investigations, the pattern used to express information seeking yes/no questions in Palermo Italian is found to be L*+H L% most of times (82.5% of cases, but as an average of 70% of patterns found on paroxiton and 95% observed on proparoxiton target words). Variability is detected in the alignment of the pitch accent peak (which, however, is usually aligned within the syllable, though in the second half of the coda consonant) and in the alignment of the L% target (which may range from the post-tonic syllable to the end of the phrase) – see alignment difference in the two plots in Figure 7.

[2] One H% was found in this corpus.

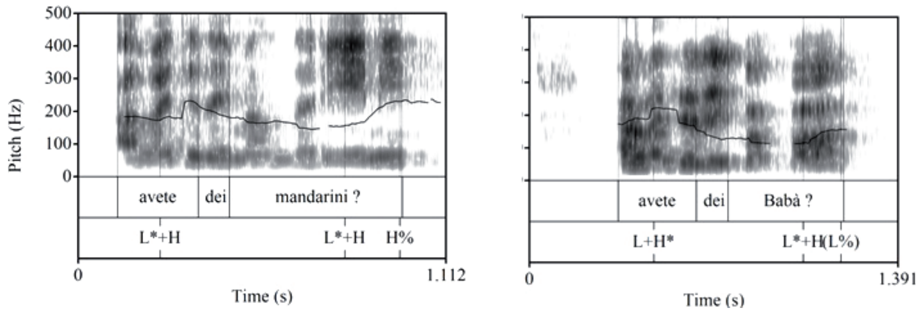
[3] Actually, productions by speaker 5 would be compatible with a H*+L analysis too, given the realization within an extremely lengthened syllable.

Figure 7 - *Yes/no information seeking questions* Avete delle mandorle? *'Do you have almonds?'*



Another option is apparently represented by a L*+H H% pattern, i.e., involving a final high boundary tone target – see Figure 8, left panel. In the corpus analysed here, this pattern is observed in a smaller percentage of cases in comparison to the main pattern (15% of the total amount of target words), and mainly in the case the nuclear pattern is realized on paroxiton (25%) rather than proparoxiton words (5%). This fact, together with the observation of the 100% of total truncation cases in yes/no questions involving an oxiton target word (*Avete dei Babà?* 'Do you have Babà?', in line with finding in Grice, 1995) – see Figure 8 right panel, suggests that the presence of a final high boundary tone may be not only related to a more controlled speech style (as suggested in the literature, e.g., for Bari Italian by Savino, Refice, 1997; see discussion in Savino, 2012), but also to a general tendency towards truncation of the final low tone specification.

Figure 8 - *Yes/no information seeking questions* Avete dei mandarini?
'Do you have mandrins?' (left panel) and Avete dei Babà? *'Do you have Babà?' (right panel)*



3.2.3.2 Echo, counterexpectational and confirmation seeking yes/no questions

The most common pattern used for confirmation seeking yes/no question is, again, L*+H L% (75% of cases – see Figure 9, left panel, plus one instance that is realized with a final high boundary tone), though quite a number of instances of H+L* L% are found

(15%) in confirmation seeking question that are actually realized by the following tag question (e.g., *vero?* "isn't it?")[4].

Echoes are realized by means of a L*+H L% pattern too (80% – see Figure 9, right panel), with the variant involving a high boundary tone realized by one speaker only (20%).

Figure 9 - *Yes/no confirmation seeking questions* Vuoi le mandorle? *'Do you want almonds?' and echo question* Le nove? *'nine o'clock?'*

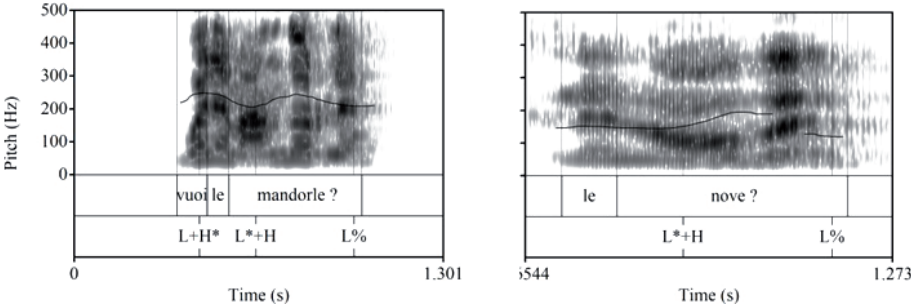
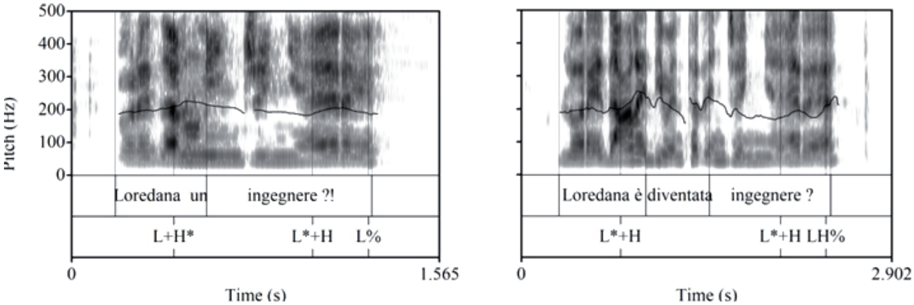


Figure 10 - *Counterexpectational yes/no questions* Loredana un ingegnere!? *'Loredana an engineer!?'*



As in other varieties, Palermo Italian counterexpectational yes/no questions are expressed by means of the same phonological pattern found in echo yes/no questions, although the phonetic implementation may imply differences in syllable lengthening, tonal alignment, and scaling. This means that in Palermo Italian, such questions are realized by means of a L*+H L% pattern (71%) in which, though the pitch accent peak is anticipated within the nuclear syllable to the extent to resemble the L*+>H used in exclamatives, though usually it is more compressed in range (Figure 10, left panel; in 5% of cases no peak anticipation is realized and the production resemble a regular check rather than a counterexpectational question). In the case of one speaker, the final boundary tone includes a high tone (24% of total instances) – see Figure 10, right panel. The latter pattern was already observed by Grice (1995).

[4] In one case a H*+L is also realized.

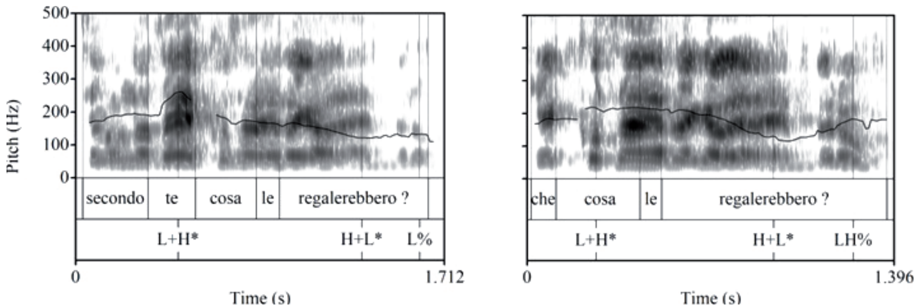
Thus, in Palermo Italian as in many other varieties, one pattern is common to many (sub)functions (e.g. information seeking, confirmation seeking and echo) and more than one pattern may play a specific function, possibly for stylistic reasons. However, in the case of Palermo Italian, the pattern involving a final high boundary tone is clearly speaker-related and, at least in some cases, may rather signal a tendency toward the final low truncation.

3.2.4 Wh-questions

3.2.4.1 Information seeking wh-questions

The most frequent pattern found in the corpus to express information seeking wh-question is H+L* L% (65% – Figure 11, left panel), while that involving a final high boundary tone is less attested (35% – Figure 11, right panel[5]). Palermo Italian, in this respect, shows features that are shared with the other varieties of Italian.

Figure 11 - *Wh information seeking questions* Cosa le regalerebbero? *'What would they gift her?'*



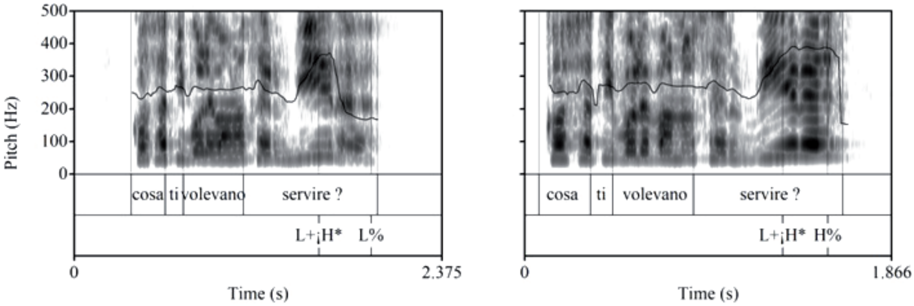
3.2.4.2 Echo, disjunctive and counterexpectational wh questions

In line with other varieties of Italian, echo questions are realized by means of the same pattern found in information seeking yes/no questions, that is L*+H L% (100% of cases), and disjunctive questions are realized by means of a L+H* pitch accent on the first item (eventually followed by a low phrase accent) and a H+L* on the final item followed by either a low or a high boundary tone (in this corpus, L% is attested in 55% and H% in 45% of cases).

As already observed for many other varieties of Italian, counterexpectational wh questions are expressed by means of a rising pitch accent which is characterised by a wide pitch excursion and is therefore labelled as L+¡H.

[5] The postnuclear syllable does not need to be obligatorily implemented at low tone value, but so far the pattern has been analysed as LH% for homogeneity with the transcription chosen for other varieties and for the observed variability in L- alignment.

Figure 12 - *Counterexpectational wh questions* Cosa ti volevano servire? *'What did they want to serve you?'*



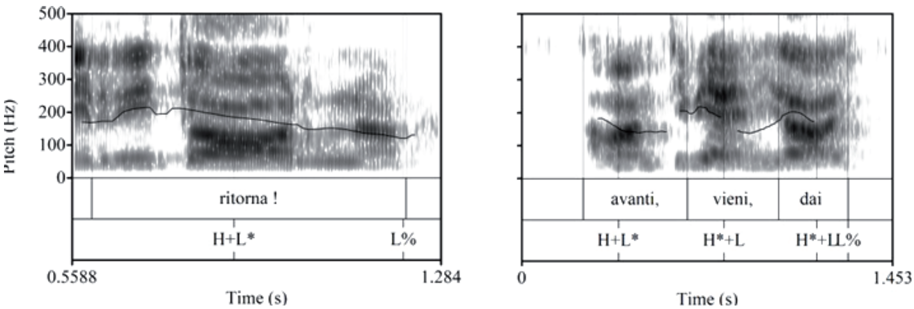
In Palermo Italian, such pitch accent is followed more often by a low boundary tone (L+¡H L% in 61% of cases) than by a high boundary tone (L+¡H H% in 39% of cases). Moreover, the choice of a high boundary tone appears to be quite subject specific. In fact, only two speakers use H% in our corpus and just one of them alternates it with the L% boundary tone – see Figure 12. Thus, the main option found in the other varieties of Italian, that involving the high boundary tone, is the least frequent in Palermo Italian.

3.2.5 Imperatives: commands and requests

Commands are usually realized by a H+L* L% nuclear pattern preceded by a rising L+H* accent, in line with what observed in other varieties (100%, see Figure 13, left panel)[6].

In line with what observed in other varieties on the same materials, imperative requests are usually realized by means of a contrastive corrective pattern, that is a H*+L L% in Palermo Italian (100%, see Figure 13, right panel).

Figure 13 - *Imperative comand* Ritorna! *'Come back here', left panel, and imperative request* Avanti, vieni, dai *'Come on, join us, come on', right panel*



[6] Actually the two spontaneous productions by one of the female speaker are both realized with a L+H* pitch accent on the verb, but they sound as requests rather than commands and they were then considered not representative of the intended function.

3.2.6 Vocatives[7]

3.2.6.1 Initial call

The pattern attested in vocative initial calls is mainly L+H* H!H% (75%), that is the same found in the other varieties investigated so far (i.e., Milan, Turin, Pisa, Lucca, Rome, Pescara, Naples, Salerno, Cosenza, Lecce, Florence, Siena). Moreover, one of the speakers also realize a rising accent followed by a high (level) edge tone (L+H* H%; 15% of total cases) or by a low edge tone (5% cases; finally, one extra pattern fond in the corpus sounds interrogative).

3.2.6.2 Insistent call

The insistent call may be realized by repeating the same phonological pattern used for the initial call, that is L+H* H!H%, though usually produced at a higher fundamental frequency (75% of cases) – see Figure 14, left and central panel; the right panel shows that the postnuclear syllable may also be already downstepped. Another option, found in other varieties too, corresponds to the use of a L+H* L% pattern (20% of instances, showing a slightly earlier pitch accent peak, around the middle of the vowel) – see Figure 15. One speaker also realizes a rising accent followed by a high (level) edge tone (L+H* H%; 5% of total cases).

Figure 14 - *Vocative initial and insistent call* Domenico! *'Domenico': main pattern*

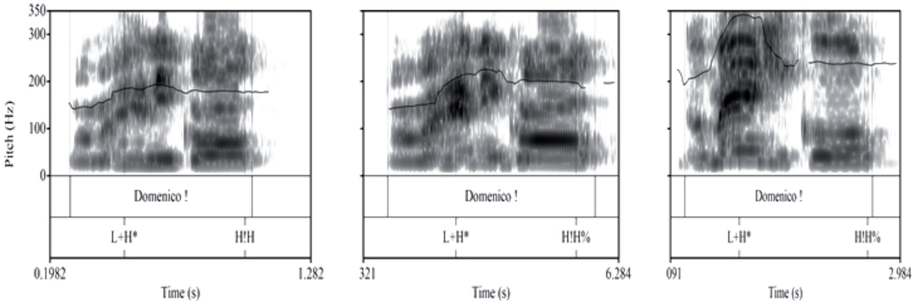
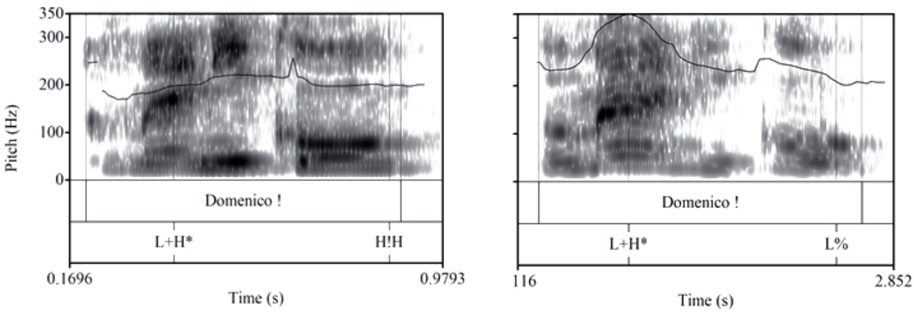


Figure 15 - *Vocative initial and insistent call* Domenico! *'Domenico': alternative pattern*



[7] Due to technical reasons, the productions by one of the male speakers were discarded. Another male speaker (matching all the selection criteria) was recorded to replace the missing speech material.

4. *Variation in Italian intonation and Palermo Italian in its "isogloss"*

Results of the analysis reported in section 2.2 showed that the phonological inventory of Palermo Italian is composed by pitch accents and edge tones that are also found in other varieties, though the patterns in which such basic units combine and their association to specific language functions may change. In this section, the features found in Palermo Italian will be compared with those reported for other varieties of Italian (see Appendix 1), with specific attention to those found in the isogloss corresponding to the extreme-South (Pellegrini, 1977), which in Gili Fivela *et al.* (2015) was represented by Lecce and Cosenza.

As for statements, in broad focus contexts Palermo Italian shows the pattern corresponding to H+L* L%, which is found in many other varieties of Italian. However, the well-known high variability in H+ scaling in our corpus on Palermo Italian seems to be due to inter-subject more than speech style variability. The falling H+L* accent is also found in lists, where it is mandatory in final position. In other positions, it alternates with L+H*, which is though mandatory in penultimate position. As for narrow contrastive-corrective focus, a H*+L accent is found in Palermo, though, in line with what observed for other varieties too, in some cases H+L* may also be used to express the same function. However, a typical feature of Palermo Italian intonation seems to be related to the scaling relation between the prenuclear and the nuclear accent, with the former sounding very prominent and reaching a quite high fundamental frequency value in comparison to the latter. Exclamatives are realized by means of patterns found in other varieties of the language too, that is L*+>H L% or, in fewer cases, a L+H* L% pattern. Thus, as far as statements and exclamatives are concerned, phonological patterns found in Palermo Italian are similar to those observed in other varieties, not only those spoken in the same isogloss.

As for yes/no questions, differences are more marked, though, again, they do not allow us to identify features that are shared within one specific isogloss – see Table 2. First of all, as in many other varieties, one pattern is common to many (sub)functions (e.g. information seeking, confirmation seeking and echo). In Palermo Italian, such pattern is L*+H L%. However, this pattern is typical of the variety and, interestingly, it is not even shared with Lecce and Cosenza Italian which are spoken in the same isogloss (e.g., compare Figure 7, left panel, to the one reported in Figure 16, showing the pattern used in Lecce – see Gili Fivela *et al.*, 2015 for discussion). As observed in the other varieties of Italian, however, more than one pattern may play a specific function, possibly for stylistic reasons. In Palermo Italian, the other option in polar questions is quite often represented by a change in the boundary tone, that is L*+H H%. However, rather than being related to speech style, as suggested in the literature (see Savino, 2012 for a review), the choice of the boundary tone seems mainly to be speaker-related in the corpus analysed here, and, at least in some cases, to be possibly linked to a tendency toward the final low truncation. In fact, the percentage of H% boundary increases in the case the pattern is realized on a paroxiton word in comparison to cases in which it is realized on a proparoxiton. Moreover, no final low is realized when L*+H L% has to be implemented on an oxiton word (truncation in 100% of cases). Truncation in varieties of Italian is well known (Palermo, Grice, 1995, Bari, Savino 2000, Pisa, Gili Fivela, 2008) and is not an all-or-nothing process (Gili Fivela *et al.*, 2015), but, in the case of Palermo and the corpus analysed here, the process could also be related to the proportion of high vs. low edge tones, as the high tones would represent a sort of facilitating factor for, or even a first step toward, truncation.

Table 2 - *Information-seeking yes/no-questions: transcription of nuclear patterns found in the varieties of Italian (left table) and their stylization (right schemes); motives indicate possible groupings on the basis of nuclear tones; varieties are represented by abbreviations: Milan (MI), Turin (TO), Florence (FI), Siena (SI), Pisa (PI), Lucca (LU), Rome (RO), Pescara (PE), Naples (NA), Salerno (SA), Cosenza (CS), Bari (BA), Lecce (LE) and Palermo (PA) – adapted and updated from Gili Fivela et al. (2015)*

	LH%	H%	HL%	L%
H+L*	MI TO LU SA CS LE		PI LU	
H*+L	MI PI RO PE SA LE			
L+H*	TO SA CS BA			SA CS BA
H*	SI FI			
L*+H			TO NA	PA

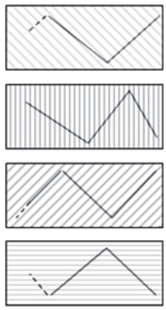
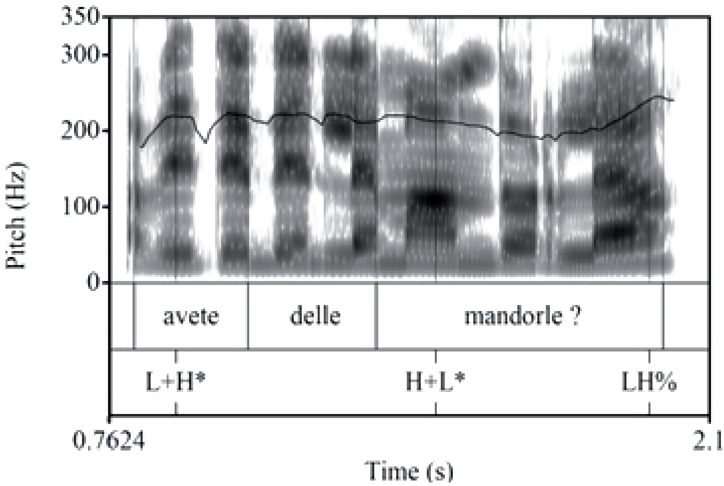


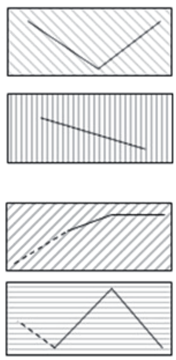
Figure 16 - *Yes/no information seeking questions* Avete delle mandorle? *'Do you have almonds?' by a speaker from Lecce*



In wh- questions, the most frequent pattern in the Palermo Italian corpus is the same found in other varieties, i.e. H+L* L% with some instances involving a final high boundary tone – Table 3. Echoes are realized by means of the same pattern found in information seeking yes/no questions, confirming the general tendency observed in other varieties, though the L*+H L% pattern characterizes Palermo Italian. In counterexpectational wh questions the pitch accent found is the same observed in other varieties, confirming the homogeneity in the intonational features related to this function across Italy. However, in the corpus analysed here, the pitch accent is more often followed by a low than by a high boundary tone, and this preference seems to be peculiar in comparison to what observed in other varieties. Finally, despite the identification of the pattern which is the most widespread through Italy for expressing inform, it is not possible to state that an isogloss may be clearly delimited, as specific patterns are also found and the alternative pattern found in Palermo (e.g. Figure 11) is different from that used, for instance, in Cosenza (see the following Figure, n. 17, and see Gili Fivela *et al.*, 2015 for discussion).

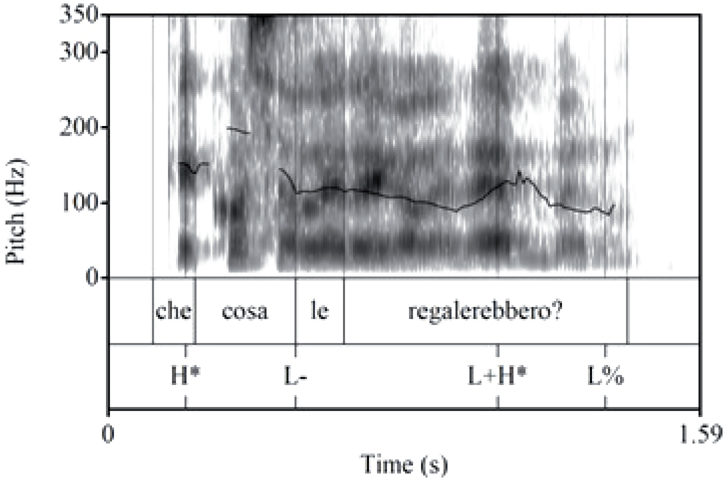
Table 3 - *Information-seeking wh questions: transcription of nuclear patterns found in the varieties of Italian (left table) and their stylization (right schemes); motives represent possible groupings on the basis of nuclear tones; varieties are represented by abbreviations: Milan (MI), Turin (TO), Florence (FI), Siena (SI), Pisa (PI), Lucca (LU), Rome (RO), Pescara (PE), Naples (NA), Cosenza (CS), Salerno (SA), Bari (BA), Lecce (LE) and Palermo (PA) – adapted and updated from Gili Fivela et al. (2015)*

	LH%	H%	HL%	L%
H+L*	MI TO LU FI SI RO SA BA CS PA			MI TO PI LU SI RO NA PE CS SA BA LE PA
H*+L				
L+H*		RO CS		CS
H*				
L*+H		PE SA		PE



Commands and imperative requests are usually realized by patterns which are similar to those observed in other varieties (i.e., L+H* H+L* L% and a contrastive corrective pattern, that in Palermo Italian is represented by H*+L L%). Similarly, vocative initial calls show a quite standard pattern, that is L+H* H!H%, and, similarly to what observed in other varieties, the insistent call may be realized by repeating the same phonological pattern, though usually produced at a higher F0 frequency, or by realizing a L+H* L% pattern.

Figure 17 - *Wh information seeking questions* Cosa le regalerebbero? *'What would they gift her?' by a speaker from Cosenza*



5. *Discussion and conclusion*

The analysis of Palermo Italian intonation performed in this work was carried out within the Autosegmental Metrical framework and allowed to highlight both the phonology and the main phonetic characteristics of Palermo Italian. The speech corpus analyzed here was collected by means of a variant of the Discourse Completion Task, which was indeed used to collect both spontaneous rendition, in line with the original DCT methodology, and read speech productions of a wide set of communicative contexts. Results confirmed our two hypotheses, that is that 1) a better knowledge on intonation may be reached by focusing on a wide set of communicative contexts and by considering various speech styles, and 2) cross variety variation may be better pointed out by adopting the very same methods used to investigate other varieties.

The first hypothesis was confirmed as improved knowledge related both to the analysis of sentence modalities which had investigated before and new modalities and pragmatic contexts collected for the present investigation. Indeed, in comparison to previous works taking into account the phonology of Palermo Italian (Grice, 1995; Grice *et al.*, 2005; Savino, 2012; Crocco, 2013), data showed new patterns or new pattern-to-function pairs, as well as shedding a different light on previously reported patterns. For instance, speakers were shown here to have other options to express correction focus (e.g. by means of H+L* pitch accent), and the choice in yes/no questions boundary tone (L*+H L% vs. H%) turned out to be possibly speaker dependent rather than only speech style dependent. Moreover, the present investigation showed that, in line with what observed by Gili Fivela *et al.* (2015), in Palermo Italian too it was possible to find a pattern shared by different functions (e.g., in confirmation and information seeking questions). Finally, data showed which patterns are used in communicative contexts that were analyzed here for the first time, as in confirmation seeking and echo yes/no questions, various types of wh questions, disjunctive questions, exclamations, commands and both the initial and the insistent call in vocatives (cfr. Table 1 and Appendix 1). In fact, the composition of the tonal inventory of Palermo Italian was enriched by inserting new tonal events (L*+>H, L+¡H*, H!H%) and patterns (e.g., L*+>H L% for exclamatives, H+L* L% and H+L* H% for information seeking wh questions, L+¡H* L% and L+¡H* H% for counter-expectational wh questions, L+H* H!H% and L+H* L% for vocatives).

Moreover, as for the second hypothesis, cross-variety comparison was facilitated and enriched by adding Palermo Italian, as discussed in section 3 and in line with our main sub-goal. In particular, the addition of Palermo Italian confirmed previous observations, to start with the lack of homogeneity in intonation patterns used within isoglosses traditionally proposed in the literature on vernaculars in Italy, in this case those identified in the extreme-South of Italy. It was confirmed that in most cases a high variability is found across Italy and within a specific isogloss (both in relation to the intonation inventory selected by speakers of different varieties and to the specific functions associated with nuclear configurations). Nevertheless, Palermo Italian data also confirmed that in some cases it is possible to find one intonation pattern that can be used by speakers of different varieties, as in broad-focus statements, information seeking wh questions, counterexpectational wh questions and vocatives, as well as in lists and disjunctive questions.

However, Palermo Italian data also clearly suggest that the specificity of a single variety does not always regard the phonological pattern *per se*, but it can be related to other characteristics. In Palermo Italian statements, for instance, quite characterizing features seem to lay in the scaling relation between prenuclear and nuclear accents (see Figure 5 and related discussion), in the distribution, at least in the corpus considered here, of intonation patterns chosen in relation to the target word stress structure (see Figure 8 and related discussion) and on the speaker related (more than speech style related) choice of pattern (see Figure 2 and related discussion). In these respects too, referring to the same methods and material is taken to facilitate cross-variety and cross-language comparison and the identification of peculiarities that are not strictly related to phonological inventories.

Bibliography

Blum-Kulka, S., House, J. & Kasper, G. (1989). Investigating crosscultural pragmatics: an introductory overview. In Blum-Kulka, S., House, J. & Kasper, G. (Eds.), *Cross-Cultural Pragmatics: Requests and Apologies*. Norwood, NJ: Ablex, 1-34.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Gotab Malmö: C.W.K. Gleerup.

Crocco, C. (2013). Is Italian clitic right dislocation grammaticalised? A prosodic analysis of yes/no questions and statements. In *Lingua*, 133, 30-52.

Endo, R., Bertinetto, P.M. (1997). Aspetti dell'intonazione in alcune varietà dell'italiano. In Cutugno, F. (Ed.), *Atti delle VII Giornate di studio del Gruppo di fonetica Sperimentale*. Roma: Collana degli Atti dell'Associazione Italiana di Acustica vol XXIV, 27-49.

Gili Fivela, B. (2008). *Intonation in Production and Perception: The Case of Pisa Italian*. Alessandria: Edizioni dell'Orso.

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M. & Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian. In Frota, S., Prieto, P. (Eds.), *Intonation in Romance*. Oxford: Oxford University Press, 140-197.

Grice, M. (1995). *The Intonation of Interrogation in Palermo Italian: Implications for Intonation Theory*. Tübingen: Niemeyer.

Grice, M., D'Imperio, M., Savino, M. & Avesani, C. (2005). Strategies for intonation labelling across varieties of Italian. In Jun, S.A. (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing.* Oxford: Oxford University Press, 362-89.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Magno-Caldognetto, E., Ferrero, F., Lavagnoli, C. & Vagges, K. (1978). F0 contours of statements, yes/no questions, and wh-questions of two regional varieties of Italian. In *Journal of Italian Linguistics*, 3, 57-68.

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

Pierrehumbert, J. (1980). The phonetics and phonology of English intonation. Ph.D Dissertation, Massachusetts Institute of Technology.

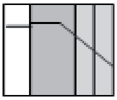
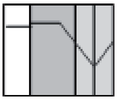
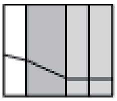
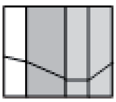
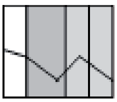
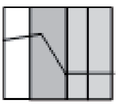
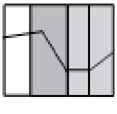
Romano, A. (2003). Accento e intonazione in un'area di transizione del Salento centromeridionale. In Radici Colace, P., Falcone, G. & Zumbo, A. (Eds.), *Storia politica e storia linguistica dell'Italia meridionale: Atti del Convegno internazionale di studi parlangeliani*. Messina: Scientifiche italiane, 169-81.

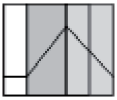
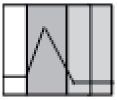
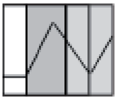
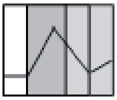
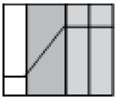
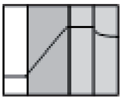
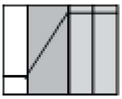
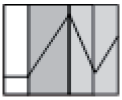
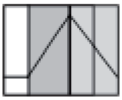
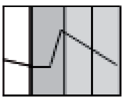
Savino, M. (2000). Descrizione autosegmentale-metrica di alcune tipologie intonative dell'italiano di Bari. In Burr, E. (Ed.), *Tradizione & Innovazione. Linguistica e Filologia Italiana alle soglie del nuovo millennio.* Firenze: Cesati Editore, 163-178.

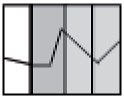
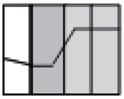
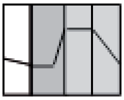
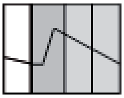
Savino, M. (2012). The intonation of polar questions in Italian: where is the rise? In *Journal of the International Phonetic Association*, 42, 23-48.

Savino, M., Refice, M. (1997). L'intonazione dell'italiano di Bari nel parlato letto e in quello spontaneo. In Cotugno, F. (Ed.), *Atti delle VII Giornate di Studio del Gruppo di Fonetica Sperimentale*. Roma: Collana degli Atti dell'Associazione Italiana di Acustica vol. XXIV, 79-88.

Appendix 1 - *Inventory of nuclear configurations found in fourtheen varieties of Italian, their schematic representations and their use in main sentence types (adapted from* Gili Fivela *et al.* 2015): *updating due to Palermo Italian data are underlined*

Nuclear Configuration	*Sentence types where it is used*
 H* L%	Exclamatives (Cosenza).
 H* LH%	Yes/no questions (Florence and Siena).
 H+L* L%	Broad focus statements, intermediate and final item in lists, narrow informational focus (e.g., Firenze and Siena); contrastive-corrective narrow focus statements (in Pescara Italian, when realized as a high pretonic pitch accent that in long constituents corresponds to a high plateau, described as L+H* H+L*; as a second option in some varieties); exclamatives (Lucca, Milan, Salerno, Lecce); wh questions (Milan, Turin, Pisa, Lucca, Rome, Pescara, Siena, Naples, Cosenza, Salerno, Bari, Lecce, Palermo); final item in disjuctive questions, commands (Milan, Turin, Florence, Siena, Lucca, Pisa, Rome, Salerno, Pescara, Lecce, Palermo); imperative requests (Lucca, Rome, Naples, Pescara; in the latter two, the high pretonic pitch accent is found); vocative initial call (Naples and Pescara, where the high pretonic pitch accent is found).
 H+L* LH%	Yes/no questions (Milan, Turin, Lucca, Salerno, Cosenza, Lecce); wh questions (Milan, Turin, Rome, Florence, Siena, Lucca, Salerno, Bari, Cosenza, Palermo); possibile in lists.
 H+L* HL%	Yes/no questions (Pisa and Lucca).
 H*+L L%	Contrastive-corrective narrow focus statements (Pisa, Rome, Pescara, Bari, Cosenza, Lecce, Palermo); yes/no questions (Milan, Pisa, Rome, Pescara, Salerno, Lecce); counterexpectational yes/no questions, exclamatives (Pisa, Lucca, Rome, Pescara, Salerno, Lecce); commands (Cosenza, Lecce, Pescara); imperative requests (Pisa, Cosenza, Pescara where the high pretonic variance is found, and Palermo), vocative insistent call (Pisa, Pescara, Lecce).
 H*+L LH%	Yes/no questions (Milan, Pisa, Rome, Pescara, Salerno, Lecce).

	L+H* L%	Not final item in lists, early narrow focus (Pisa, Lecce); wh questions (Cosenza).
	L+H* L%	Contrastive-corrective narrow focus statements (Milan, Turin, Florence, Siena, Lucca, Naples, Salerno); exclamatives (Turin, Florence, Siena, Palermo); yes/no questions (Salerno, Cosenza, Bari); counterexpectational yes/no questions; commands (Turin); imperative requests (Milan, Turin, Florence, Siena, Salerno); vocative initial call (Pisa, Lucca, Salerno, Cosenza); vocative insistent call (Milan, Turin, Florence, Pisa, Siena, Cosenza, Palermo). Alternative pattern for narrow-correction focus in Palermo Italian.
	L+H* LH%	Yes/no questions (Turin, Salerno, Cosenza, Bari).
	L+H* L!H%	Counterexpectational yes/no questions (Lecce).
	L+H* H%	Wh questions (Rome, Cosenza), possible on intermediate item in lists.
	L+H* H!H%	Vocative initial call (Milan, Turin, Florence, Siena, Pisa, Lucca, Rome, Pescara, Naples, Salerno, Cosenza, Lecce, Palermo) and insistent call (e.g., Pisa, Pescara, Salerno, Cosenza, Palermo).
	L+¡H* H%	Counterexpectational wh questions (Milan, Turin, Florence, Siena, Pisa, Lucca, Rome, Salerno, Pescara, Cosenza, Palermo).
	L+¡H* LH%	Echo yes/no questions (Lucca).
	L+¡H* L%	Counterexpectational yes/no questions (Bari); counterexpectational wh questions (Lecce, Salerno, Pescara, Palermo).
	L*+H L%	wh questions in Pescara; yes/no questions, confirmation-seeking and echo yes/no questions in Palermo (where the peak seems to be slightly delayed in comparison to the schema offered to the left); couterexpectational yes/no questions in Palermo (though the peak is as anticipated to resemble a L*+>H L% pitch accent).

	L*+H LH%	Alternative patterm in couterexpectational yes/no questions (Palermo)
	L*+H H%	wh questions in Pescara and Salerno.
	L*+H HL% L*+H HL-L%	Yes/no questions (Turin and Naples, although in the latter the low target in the pitch accent is aligned earlier and a bitonal phrase accent is found; see discussion in Gili Fivela *et al.* 2015).
	L*+>H L%	Exclamatives (Turin. Milan, Lucca, Rome, Lecce, Palermo).

DOI: 10.17469/O2103AISV000010

GLENDA GURRADO, PATRIZIA SORIANELLO

Sulla percezione del tratto di dominanza/sottomissione in un campione di voci femminili

According to the Frequency Code proposed by Ohala, there is a close relationship among the physical size of many animals (large vs. small), the vocal pitch height of their vocalizations (low vs. high sounds) and their degree of dominance/submissiveness. Also in humans, low-pitch voices are generally associated with high level of dominance, while high-pitched voices normally convey a meaning of submission and politeness. This research aims to explore the possible relationship between female voice pitch and its paralinguistic interpretation. To this purpose, we conducted an auditory test by asking to a group of Italian listeners to evaluate the voices in terms of five semantic scales reproduced on a *Visual Analogue Scale*. The findings show a confusing trend: low-pitch voices increased ratings of dominance and self-confidence than high-pitched ones; nevertheless, the transmission of these paralinguistic meanings is affected by other phonetic aspects, such as formant structure, speech rate and intensity[1].

Key words: Frequency Code, female voice, pitch, dominance/submissiveness dimension, Visual Analogue Scale.

1. *Il dimorfismo della voce*

Il presente lavoro ha inteso indagare la relazione esistente fra il *pitch* della voce e alcuni *markers* della personalità. Punto di partenza della ricerca è il Codice della Frequenza definito in chiave etologica da John Ohala (1983, 1984, 1994). Si tratta di un codice che evidenzia il ruolo svolto dalla voce nella definizione delle relazioni di potere, in uno scambio comunicativo fra emittente e ricevente. Le caratteristiche acustiche della voce, in particolar modo la frequenza fondamentale (f0), variano sensibilmente a seconda del sesso del parlante. Le differenze anatomiche riscontrabili fra uomini e donne coinvolgono, difatti, anche gli organi preposti alla fonazione, in primo luogo il tratto vocale, lungo in media circa 17-18 cm nell'uomo e 14-14,5 nella donna. La laringe dell'uomo, inoltre, è circa il 50% più larga rispetto a quella della donna, una differenza più marcata nella dimensione antero-posteriore e meno visibile in quella laterale. Ciò comporta la presenza di corde vocali di diversa lunghezza, mediamente 17-18 mm per gli uomini e 13-14 mm per le donne (cfr. Titze, 1989). Tale difformità fa la sua comparsa in età puberale; prima di allora, maschi e femmine possiedono laringi di dimensioni simili. A partire dalla fase adolescenziale, la laringe dell'uomo

[1] Questo studio è il risultato di una collaborazione continua fra le autrici. Tuttavia, GG è responsabile dei §§ 1, 2, 3, 4; PS dei §§ 6, 6.1, 6.2. 6.3, 7.1, 8. I paragrafi 5, 5.1, 5.2, 5.3 e 7 sono comuni, come pure lo svolgimento dell'analisi acustica e del test percettivo.

comincia a evolversi con maggiore possibilità di allungamento, essendo posizionata più in basso nella gola; la voce dell'uomo, sarà per questo tipicamente più grave rispetto a quella della donna. La frequenza fondamentale, da cui dipende il *pitch* della voce, è determinata dalla velocità con cui le pliche vocali, durante il processo di fonazione, compiono il ciclo di adduzione e abduzione al momento del passaggio dell'aria. Le pliche vocali vibreranno più lentamente se grandi e più velocemente se piccole: i suoni gravi sono quindi determinati da una f0 bassa, mentre i suoni acuti sono il risultato di una f0 alta. Generalmente, la f0 media degli uomini adulti si attesta attorno ai 110 Hz, laddove quella delle donne si aggira attorno ai 205 Hz (Titze, 1989; Simpson, 2009). Questi valori sono peraltro influenzati anche dall'età, oltre che dal sesso, essendo fortemente correlati alla produzione ormonale. Infatti, nell'uomo la f0 si abbassa progressivamente fino a 35 anni, per poi ricominciare a innalzarsi verso i 55 anni di età; nelle donne, al contrario, la f0 è stazionaria fino al periodo della menopausa, a quel punto comincia a decrescere raggiungendo, intorno ai 70 anni, il suo valore minimo. Dalla grandezza delle pliche vocali dipende anche l'estensione melodica, ossia l'intervallo in frequenza calcolato fra il valore massimo e il valore minimo in Hz, rilevabili nel contorno intonativo dell'enunciato: nel parlato, l'escursione melodica di un maschio adulto è compresa fra gli 80 e i 160 Hz, laddove quella di una donna adulta si attesta fra i 160 e i 320 Hz.

2. *La comunicazione emotiva animale*

I filoni di ricerca a cui è possibile fare riferimento sono ampi e diversificati, nonché saldamente ancorati allo studio della comunicazione nel mondo animale. A tal proposito, il riferimento a Charles Darwin (1872) è imprescindibile. L'autore si interrogò sugli effetti che le emozioni producono sulle caratteristiche acustiche dei suoni emessi dagli animali. In una situazione di particolare coinvolgimento emotivo, gli animali subiscono una contrazione involontaria dei muscoli del petto e della glottide che determina l'emissione di suoni vocali non abituali e visibili cambiamenti fisici. Ad es. nel cane, la rabbia induce versi tipicamente gravi, come il ringhio, e l'innalzamento dei peli della schiena. Le spinte emotive possono essere positive o negative, a seconda che il contesto tranquillizzi o metta in allerta l'animale: nel primo caso l'animale tende a instaurare un rapporto amichevole fondato sulla remissività, mentre nel secondo sarà portato a trasmettere segnali ostili. In questa circostanza, l'intento, perseguito mediante i versi prodotti e i cambiamenti fisici assunti nella fase di pre-attacco, sembrerebbe identificarsi con il desiderio di apparire il più grande e temibile possibile alla presenza del nemico, al fine di incutere timore e imporre la propria supremazia.

Appare evidente lo stretto rapporto esistente fra il *pitch* e lo stato emotivo. Difficoltosa è però l'attribuzione di un verso a una particolare sensazione; molto spesso i cambiamenti sonori possono risultare, all'orecchio umano, quasi impercettibili: un rapporto biunivoco è difficilmente intuibile nei contesti emotivi realizzabili nel mondo animale. Un ruolo fondamentale è svolto dall'abitudine, elemento fuorviante in alcuni casi, quanto utile in altri: produzione e percezione del suono potrebbero essere fortemente influenzate proprio dal fattore abitudine, che avrebbe favorito la creazione, poi consolidatesi nel tempo, di relazioni fra versi ed emozioni.

Molti versi appaiono distinti sia dal punto di vista acustico che motivazionale, ma spesso la percezione dell'intento soggiacente a essi non è così semplice e immediata. L'analisi delle regole sulla comunicazione animale ha affascinato numerosi studiosi del mondo della scienza e della linguistica. Fra essi, spicca Eugene Morton (1977), il quale, partendo dall'osservazione del comportamento comunicativo di numerosi mammiferi e volatili, ha elaborato le *Motivation-Structural Rules*, regole che influenzano la produzione di versi animali caratterizzati da una specifica struttura acustica. Nelle situazioni comunicative intraspecifiche o interspecifiche, il dato che permette ai soggetti di percepire le sfumature motivazionali che sottostanno all'uso dei suoni emessi è senza dubbio la prossimità: è facile intuire come uno scambio comunicativo realizzato a distanza conduca inevitabilmente alla perdita di sfumature acustiche cruciali ai fini della trasmissione dei cambiamenti motivazionali. Tali intenti possono essere espressi, quindi, solo mediante una gradualità sonora che si inserisce all'interno di uno spettro acustico avente come punti estremi da un lato i suoni a bassa frequenza, utilizzati dagli animali prevalentemente in contesti ostili, e dall'altro i suoni ad alta frequenza, usati al fine di approcciare il nemico in modo amichevole, predisponendolo positivamente nei propri confronti. Morton sottolinea come i poli di tale spettro siano in realtà usati raramente dagli animali, sono invece i gradi frequenziali mediani a esprimere meglio le mutazioni d'intento.

Gli *endpoints* dello spettro acustico corrispondono a suoni ad alta e a bassa frequenza, valori che rappresentano motivazioni diametralmente opposte. Eugene Morton attribuisce tale correlazione a una giustificazione di tipo fisico-anatomico: il primo strumento utilizzato dagli animali per comunicare l'informazione sulla propria mole è la voce. Fisiologicamente, difatti, un animale caratterizzato da una corporatura ampia produce suoni a bassa frequenza, mentre un animale di dimensioni ridotte produce suoni ad alta frequenza. Di conseguenza, in uno scambio comunicativo, il soggetto ricevente, partendo dai soli versi, potrà ricavare l'informazione relativa alla taglia dell'animale emittente, anche in assenza del contatto visivo, e decidere, in base a questa informazione, se procedere all'attacco o ritirarsi prima che inizi lo scontro. È facile comprendere quanto tale correlazione abbia giocato un ruolo decisivo a livello evoluzionistico, visto che, come è noto, un animale grande tenderà sempre a vincere su uno più piccolo e a dettarne l'esclusione dal branco. Tuttavia, in numerose situazioni è proprio l'utilizzo dei suoni ad alta frequenza a determinare la sopravvivenza dell'animale: ad es., al fine di evitare lo scontro, il soggetto, alla stregua di un infante, emette suoni acuti trasmettendo un messaggio di sottomissione e una richiesta di protezione. Appare evidente quanto lo strumento voce permetta agli animali di gestire gli incontri in base alle relazioni di potere da essi impostati, modulando la frequenza dei versi emessi e dando vita a uno scontro verbale in cui le armi in gioco sono la dominanza e la sottomissione.

3. *La voce della dominanza*

Fatta questa premessa, non c'è da stupirsi che la voce sia uno strumento di trasmissione degli intenti, anche con riferimento all'uomo.

Diversi linguisti si sono interrogati sulla correlazione esistente fra le caratteristiche acustiche dei contorni intonativi degli uomini e i messaggi sociali da essi trasmessi, riconducibili a un certo grado di sottomissione o dominanza. In particolar modo, il linguista americano John Ohala (1984) ha indagato il ruolo svolto dalla frequenza fondamentale negli scambi comunicativi fra uomini adulti, elaborando il Codice della Frequenza. Tale modello assume che un contorno intonativo avente una f0 particolarmente alta, o in progressivo innalzamento, trasmetterà sottomissione e insicurezza; diversamente, un contorno intonativo caratterizzato da una f0 bassa, o in progressiva discesa, invierà un messaggio di dominanza e sicurezza di sé. Secondo Ohala, tale correlazione ha un carattere innato e non va assunta come risultato dell'evoluzione della comunicazione. È sufficiente menzionare l'anatomia del tratto vocale maschile e femminile che, come già descritto, differenzia anche la struttura acustica dei suoni prodotti che, pertanto, saranno acuti per le donne e gravi per gli uomini. Tale difformità comporta una relazione fra la dimensione corporale e i significati veicolati, dovuta all'associazione del tratto vocale grande a una maggiore robustezza e prestanza fisica (dominanza) e del tratto vocale piccolo a una ridotta fisicità (sottomissione). Questa differenza, pur essendo innata, si è consolidata nel tempo innescando stereotipi sociali di discriminazione sessuale. Si tratta di universali cross-culturali e cross-linguistici, fra i quali rientra, sul piano propriamente linguistico, anche la contrapposizione fra il contorno intonativo ascendente delle frasi interrogative, che trasmette la dipendenza informativa di colui che la produce, e il contorno discendente delle frasi affermative, autosufficienti dal punto di vista informativo poiché complete: una differenza rilevabile nella maggior parte delle lingue.

John Ohala evidenzia che il Codice della Frequenza ha carattere innato o inconscio se connesso al dimorfismo sessuale dell'apparato fonatorio di uomini e donne, carattere affettivo se connesso all'uso consapevole dei significati trasmessi dai diversi livelli di f0. Al fine di dimostrare le proprie argomentazioni, Ohala raccolse degli stimoli audio di parlato spontaneo, realizzati da 4 adulti angloamericani (2 femmine e 2 maschi). Gli stimoli sono stati successivamente sintetizzati, per eliminare le informazioni sul sesso e sul contenuto semantico della frase e allo stesso tempo innalzati e abbassati in frequenza. I segnali, abbinati in coppia, sono stati proposti a un gruppo di ascoltatori, i quali hanno dovuto giudicare quale fra i due stimoli della coppia fosse più dominante rispetto all'altro. Il test ha dimostrato che il 92% degli ascoltatori ha percepito come più dominante lo stimolo audio caratterizzato da una frequenza fondamentale media ($f0\bar{x}$) più bassa; inoltre, si è rilevato che la discesa finale dei livelli di f0 è il tratto che più contribuisce alla percezione della dominanza.

Gussenhoven (2002) riprende la teoria del Codice della Frequenza, o *Size Code*, definendo altri due codici biologici: il Codice delle Sforzo (*Effort Code*) e il Codice della Produzione (*Production Code*). Il primo riguarda la quantità di energia impiegata dal parlante durante la produzione dei suoni; l'impiego di uno sforzo maggiore non solo rende più preciso il movimento articolatorio e più ampia l'escursione melodica, ma è anche la causa principale dei processi di prominenza. Il Codice della Produzione evidenzia invece il ruolo della dinamica respiratoria nella resa del contorno intonativo di un enunciato assertivo, notoriamente caratterizzato da declinazione intonativa. I picchi intonativi sono per questo più alti in frequenza all'inizio dell'enunciato, quando l'energia è massima, mentre il livello frequenziale più basso si riscontra alla fine, in corrispondenza del momento di minima pressione laringea. In relazione al *Size Code*, Gussenhoven sottolinea come la correlazione esistente fra la larghezza della laringe e la velocità di vibrazione delle pliche vocali permetta l'instaurazione delle relazioni di potere fra i parlanti, in base alla percezione di una dominanza fisica e sociale. I tre codici possono essere letti attraverso interpretazioni affettive, che attribuiscono caratteristiche al parlante, e interpretazioni informazionali, che associano caratteristiche al messaggio linguistico. Con riferimento al *Size Code*, le interpretazioni affettive riguardano l'opposizione tra i valori afferenti alla sfera della sottomissione (es. gentilezza, vulnerabilità), veicolati da alte frequenze, e i valori della dominanza (es. sicurezza di sé, aggressività) trasmessi da basse frequenze. Le interpretazioni informazionali evidenziano la contrapposizione tra il senso di incertezza convogliato da un messaggio linguistico prodotto ad alte frequenze e il senso di certezza trasmesso da un messaggio a basse frequenze, nonché il confronto tra frase affermativa e frase interrogativa, caratterizzata l'una da un contorno intonativo discendente, l'altra da un contorno finale ascendente. Poiché tali corrispondenze si ravvisano nella maggior parte delle lingue del mondo, è lecito supporre che gli usi del Codice della Frequenza abbiano subito una sorta di grammaticalizzazione.

4. *Gli studi*

In questa direzione, una serie ampia e diversificata di ricerche svolte in ambito linguistico e psicolinguistico ha indagato il ruolo della voce come vettore di un messaggio pregno di significati 'altri' rispetto a quello linguistico, sconfinando nel campo della trasmissione dei marcatori della personalità del parlante. Oltre agli studi incentrati sull'origine dello stereotipo sociale (Scherer, 1972; Henton, 1989; Jiang, 2011), diverse ricerche hanno preso in esame anche la relazione fra la voce e la percezione della dominanza. Dagli studi non emerge un quadro interpretativo sempre nitido. Apple, Streeter & Krauss (1979) e Scherer (1979) rilevano una correlazione positiva fra il *pitch* e la percezione di alcuni tratti della personalità, giungendo però a risultati fra loro discordanti. Tale relazione sarebbe assente per Aronovitch (1976), mentre Tusing, Dillard (2000) e Collins (2000), basandosi su un test di formulazione di giudizio da parte di ascoltatrici donne su voci maschili, affermano che le voci caratterizzate da un *pitch* più basso trasmettono un maggior grado di dominanza. In modo speculare, Collins, Missing (2003) hanno dimostrato come voci più alte in frequenza denotino sottomissione e subordinazione del parlante. I principi del Codice della Frequenza ohaliano sono confermati anche da uno studio svolto da

Chen, Gussenhoven e Rietveld (2004) sulla lingua inglese e olandese. Gli autori, attraverso la sintesi di specifici contorni intonativi e la somministrazione di un test percettivo, elaborato mediante una *Visual Analogue Scale,* attestano come al decrescere dei livelli di f0 e di estensione tonale corrisponda un aumento della percezione della dominanza. Fra le due lingue indagate, emergono comunque delle sostanziali differenze percettive, un dato che mette in luce la presenza di una certa connotazione linguo-specifica dei significati paralinguistici.

Rimanendo nell'ambito degli studi sui marcatori della personalità, spiccano le ricerche di Feinberg, Jones, Little, Burt & Perret (2005) e di Puts, Gaulin & Verdolini (2006), inerenti alle influenze che la manipolazione, tramite sintesi, di f0 e delle formanti determina sul grado di attrazione esercitato sulle donne da maschi adulti. A questo proposito, Feinberg et al. (2005) osservano che negli uomini le frequenze basse trasmettono un senso di mascolinità e di capacità riproduttiva. Gli autori elaborano un test di sintesi vocale; a un gruppo di dodici donne si chiede di giudicare il grado di mascolinità di voci maschili, precedentemente impiegate in 12 *task* linguistici. I risultati dimostrano che le voci caratterizzate dall'abbassamento di f0 sono più spesso associate a significati di grandezza, mascolinità e maturità rispetto alle voci caratterizzate da una f0 alta. Parimenti, anche il grado di attrazione risulta positivamente connesso ai valori della frequenza fondamentale. Sempre con riferimento alla percezione di mascolinità, Puts et al. (2006) esaminano la relazione fra *pitch* e dominanza (sociale e fisica), durante un *dating-game,* indagando anche la possibile correlazione fra f0 e il successo degli uomini con l'altro sesso. Il test ha previsto la registrazione di voci maschili e la formulazione del giudizio di dominanza percepito da parte di maschi adulti. I risultati hanno rivelato che voci gravi sono associate a un più alto grado di dominanza rispetto a voci acute; inoltre, in un contesto competitivo, gli uomini che ritengono di essere fisicamente più dominanti rispetto al proprio concorrente tendono ad abbassare il *pitch* della loro voce. Tuttavia, sebbene gli uomini che hanno dichiarato di avere avuto un maggior numero di partner sessuali siano accomunati da una voce particolarmente bassa in frequenza, la relazione fra f0 e livello di successo con le donne non è statisticamente significativa.

Rilevante è infine il filone di studi dedicati ai tratti acustici e prosodici della voce carismatica, per lo più incentrati sull'analisi del parlato politico. Fra tutte, merita di essere menzionata la ricerca di Rosenberg, Hirschberg (2009). Lo studio ha preso in esame i discorsi realizzati da politici democratici americani. I segnali audio sono stati sottoposti al giudizio percettivo di otto ascoltatori nativi, ai quali è stato chiesto di valutare il grado di carisma presente nei discorsi. Per il giudizio, gli autori si avvalgono di una scala Likert a 5 punti e dell'ausilio di una gamma di aggettivi, a scelta dell'ascoltatore, da associare a ciascuna voce, in modo da pervenire a una valutazione più affinata della propria percezione di carisma. I risultati provano che livelli alti di $f0\bar{x}$ conferiscono maggior carisma al parlante, lo stesso vale per la variazione dell'estensione tonale e dell'intensità. Sulla stessa scia, altre ricerche hanno approfondito tale tematica (Signorello, D'Errico, Poggi & Demolin, 2012; Signorello, Demolin, 2013); per meglio definire le varie 'dimensioni del carisma', gli autori hanno analizzato i discorsi di politici francesi e brasiliani, anche con riferimento contrastivo alla lingua italiana (D'Errico, Signorello & Poggi, 2013).

Da questa breve sintesi è possibile dedurre un aspetto centrale. I campioni impiegati nelle ricerche citate sono per lo più costituiti da voci maschili, mentre i test sono per gran parte composti da un uditorio femminile chiamato a giudicare voci maschili in termini di dominanza sociale. Le lingue maggiormente esplorate in questa direzione sono inglese e angloamericano. Diversamente, la lingua italiana, oggetto di studio sulla variazione delle frequenze formantiche di uomini, donne e bambini e sulle modalità di normalizzazione (Ferrero, Magno Caldognetto & Cosi, 1995; Ferrero, Magno Caldognetto & Cosi, 1996), non è stata indagata in questa specifica direzione. Recentemente, la correlazione tra voce e tratti della personalità è stata presa in esame da Sorianello & Gurrado (in stampa). Partendo dalla letteratura disponibile, con esplicito riferimento ai principi del Codice della Frequenza, le autrici hanno cercato di verificare se il rapporto fra i parametri prosodici e la percezione della personalità trovasse riscontro anche all'interno di un campione di sole voci femminili. La finalità della ricerca, improntata su un corpus di parlato letto prodotto da cinque locutrici italiane, era quella di accertare se voci femminili con caratteristiche timbrico-frequenziali molto diverse tra loro potessero essere associate a quei tratti della personalità afferenti alle sfere della dominanza o della sottomissione. L'analisi acustica e la successiva raccolta di giudizi percettivi, formulati da cinquanta ascoltatori per mezzo di una *Visual Analogue Scale*, producono risultati disomogenei. Non sempre la correlazione fra voce (acuta vs. grave) e tratti della personalità percepiti (sottomissione vs. dominanza) appare infatti verificata, sebbene il test, nel complesso, riveli una certo collegamento tra la presenza di un *pitch* grave e la trasmissione di un senso di dominanza e sicurezza di sé. A discapito dello stereotipo sociale, è stato inoltre rilevato che anche la voce femminile può trasmettere un messaggio di dominanza e sicurezza, se contrassegnata da specifiche caratteristiche acustiche, e se confrontata con una voce più alta in frequenza di pari genere. Tuttavia, come atteso, nella valutazione percettiva delle voci, diversi parametri fonetici come ad es. l'intensità e la velocità elocutiva, e non solo il *pitch*, interagiscono ai fini della trasmissione dei marcatori della personalità. Il quadro che emerge da questo recente studio non è comunque nitido, dato che alcuni fattori potrebbero aver confuso i giudizi. Da un lato il numero elevato di *speakers* potrebbe aver destabilizzato gli uditori, dall'altro lo stile imposto dalla lettura potrebbe aver annullato, o almeno ridotto, alcune caratteristiche peculiari delle voci del campione.

5. *La ricerca*

La presente ricerca trae origine proprio dalle criticità metodologiche rilevate da Sorianello, Gurrado (in stampa). Partendo dai risultati già conseguiti, abbiamo operato una duplice modifica al protocollo di ricerca: da un lato abbiamo ridotto il campione delle locutrici da cinque a tre, lasciando solo le voci più rappresentative, cioè una voce acuta (CI), una voce media (GM) e una grave (RV) (*ultra*), dall'altro abbiamo ampliato la gamma dei materiali, affiancando, in un'ottica contrastiva, il parlato letto a un corpus di parlato spontaneo.

Lo studio intende esplorare più aspetti. In primo luogo, vuole verificare se un numero più ristretto di voci produca una migliore polarizzazione dei giudizi percettivi in rapporto alla dimensione di 'dominanza/sottomissione'. L'ipotesi di lavoro, fondata sui principi del Codice della Frequenza, è che la voce grave (RV) sia, a una valutazione percettiva, più spesso associata ai tratti della dominanza, mentre la voce acuta (CI) si disponga in senso opposto, classificandosi come voce non dominante. In secondo luogo, sarà possibile verificare se i giudizi degli uditori risentano del genere stilistico (parlato letto vs. parlato spontaneo). Il perseguimento di tali obiettivi è subordinato alla verifica acustica dei materiali che pertanto ha costituito il primo stadio della ricerca.

5.1 I partecipanti

Per dare fondatezza alle nostre ipotesi, la fase preliminare dello studio ha riguardato la selezione dei parlanti. A tal fine, sono state scelte informatrici di sesso femminile, aventi, a un riscontro uditivo, voci molto diverse fra loro e particolarmente marcate in frequenza. Il campione analizzato si compone di tre partecipanti, da ora indicate con le iniziali del loro nome (CI, GM, RV), di pari provenienza ed estrazione socioculturale. Le informatrici sono nate e vissute a Bari, hanno un'età compresa tra 22 e 25 anni e sono in possesso di laurea. Nel dettaglio, la voce CI è percepita come acuta, la voce RV mostra un timbro vocale particolarmente grave, mentre la voce GM si colloca infine in una posizione intermedia tra le due. Le partecipanti hanno compilato un breve questionario su provenienza, età, grado di istruzione, contenente pure specifiche domande su peso, altezza e principali interessi.

5.2 I materiali

Ogni partecipante è stata coinvolta in due diversi *task* linguistici, la lettura di un testo e la descrizione di una ricetta gastronomica[2]. La registrazione dei materiali è avvenuta in un ambiente insonorizzato con strumenti digitali (WAV. format, 44100 Hz, 32 bit). I soggetti hanno accettato il consenso informato sul trattamento dei loro dati personali e sull'utilizzo dei materiali audio, ai soli fini scientifici.

5.3 Il protocollo

Il disegno sperimentale dello studio prevede due fasi complementari coincidenti la prima con una analisi acustica, la seconda con una analisi percettiva. La prima fase ha un valore preliminare, poiché consente di caratterizzare la produzione vocale delle tre informatrici sul piano spettro-acustico, un dato imprescindibile per poter poi interpretare i risultati del test percettivo e dunque l'associazione di ciascuna voce con la dimensione di 'dominanza/sottomissione'.

[2] Si tratta di un testo giornalistico, tratto da *La Repubblica* dal titolo *Perché i Bronzi di Riace non andranno all'Expo di Milano?*, pubblicato il 10/04/2014.

Per la verifica acustica delle registrazioni abbiamo considerato i seguenti parametri: 1) valore medio, massimo e minimo della frequenza fondamentale (f0x̄, f0min, f0max), 2) escursione melodica convertita in semitoni (EM, ST), 3) intensità media (INT, dB), 4) velocità di eloquio (VE, sill/sec), 5) computo della *Vowel Space Area* (VSA)[3], previa estrazione dei valori in Hz di F1 e F2 delle 3 vocali cardinali toniche /i a u/[4]. Questo parametro, il cui risultato è espresso in Hz^2, definisce l'area di un poligono irregolare. Si tratta di un indice utilizzato per valutare l'ampiezza dello spazio vocalico, anche in caso di disturbi linguistici, ma è soventemente sfruttato in ambito dialettologico per cogliere, in modo contrastivo, le differenze globali della dimensione tra sistemi vocalici marcati sul piano sociolinguistico, ad es. per stile o diatopia (in questa direzione, Jacewicz, Fax & Salmons, 2007). VSA si calcola applicando la formula riportata in (1).

(1) VSA: [(F1i (F2a-F2u) + F1a (F2u-F2i) + F1u (F2i-F2a)]/2

L'analisi è stata condotta mediante il *software Praat* (versione 5.0.16, Boersma, Weenink, 2013). La significatività statistica è stata testata mediante t-test e ANOVA, per le comparazioni multiple è stato impiegato il test *post-hoc* di *Tukey*, la significatività statistica è stata impostata per $p < 0.05$.

6. *Risultati*

Gli esiti acustici avvalorano l'impressione iniziale, in quanto le informatrici manifestano più tratti fonetici differenzianti che non riguardano solo il *pitch*, ma coinvolgono, in modo globale, anche gli altri parametri esaminati. Di seguito, esporremo i risultati ottenuti per la frequenza fondamentale (6.1), il timbro vocalico (6.2) nonché per l'intensità e la velocità elocutiva (6.3).

6.1 La frequenza fondamentale

Come atteso, il *pitch* si conferma il parametro più robusto ai fini della distinzione delle voci delle tre informatrici, le quali effettivamente presentano valori di f0 differenziati. Molte delle tendenze rilevate mostrano, inoltre, una buona sistematicità a un confronto tra parlato letto e spontaneo. Per quantificare tali differenze, abbiamo assunto quale riferimento la f0x̄ e abbiamo computato lo scarto tra le singole voci, adottando quale unità di misura il semitono. Tale confronto è sempre significativo. Precisamente, nel parlato letto la distanza tra la f0x̄ di CI e RV, le voci più rappresentative, è di 4,2 ST (dev.st. 0,2), mentre nello spontaneo è pari a 4 ST (dev.st. 1,1). Minore è la distanza tra queste voci e GM, rispettivamente, per la lettura, CI/GM = 2,3 (dev.st. 0,4) e RV/GM = 1,9 (dev.st. 0,7); valori simili si attestano anche nello spontaneo (CI/GM = 2,2, dev.st. 0,5; RV/GM = 1,8, dev.st. 1,6).

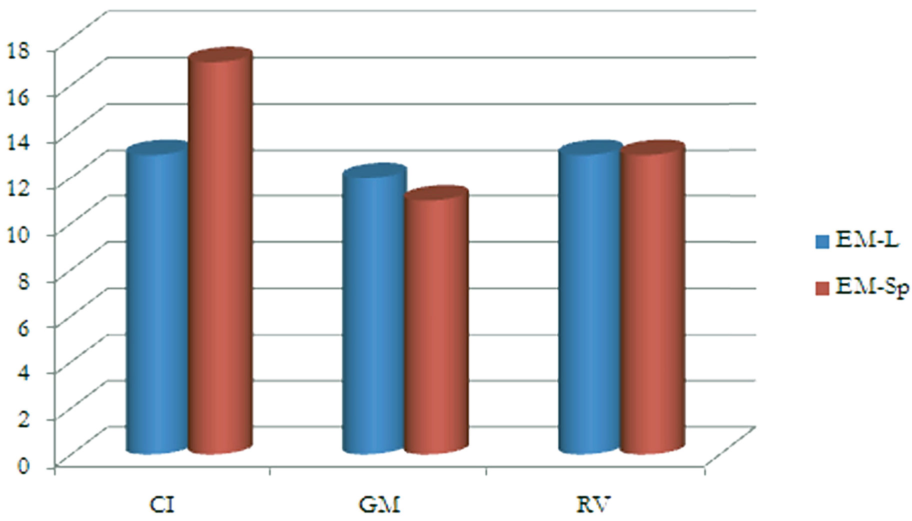
[3] Le ricerche che hanno applicato VSA sono numerose, fra tutte si vedano gli studi di Higgins, Hodge (2002); Vorperian, Kent (2007); Flipsen, Lee (2012).

[4] L'estrazione delle formanti è stata effettuata manualmente mediante tecnica FFT, in corrispondenza dello *steady state* delle vocali.

Nel complesso, il comportamento di f0 è più regolare nel parlato letto che non nello spontaneo, sebbene le differenze siano sempre significative [Lettura, $F(2,33) = 117{,}099$, $p = .000$; Spontaneo, $F(2,33) = 36{,}305$, $p = .000$]. Nello spontaneo, malgrado ciò, alcune differenze si riducono: è quanto avviene per f0min e f0max di GM e RV ($p = .88$), si veda la rappresentazione della Figura 1.

Figura 1 - *Valori medi dei 3 parametri di f0 (in Hz) per le tre locutrici nel parlato letto (L) e spontaneo (Sp)*

Figura 2 - *Valori medi relativi all'Escursione Melodica in semitoni per le tre locutrici nel parlato letto (L) e spontaneo (Sp)*



Per quanto concerne l'escursione melodica, nel parlato spontaneo le parlanti presentano valori più bassi rispetto allo stile di lettura, ma anche più variegati, la dispersione intorno alla media è di circa 2 ST per tutte le locutrici (Figura 2). I dati sull'EM, inoltre, essendo piuttosto simili, non acquisiscono uno *status* differenziante ai fini della caratterizzazione delle tre voci; l'analisi della varianza in merito a questo parametro non è, infatti, significativa, fatta eccezione per l'EM di CI estratta per lo spontaneo.

6.2 Il timbro vocalico

Dai risultati ottenuti per le frequenze formantiche si desume che il timbro concorre a differenziare le locutrici. Nei due stili considerati (letto e spontaneo), l'assetto vocalico mostra diversi elementi distintivi che riguardano l'articolazione delle vocali nella loro duplice dimensione: altezza e grado di anteriorità/posteriorità, come si vede dai sistemi vocalici lineari (in Hz) rappresentati nelle Figure 3 e 4.

È tuttavia emblematico osservare che, a parità di stile, le informatrici si comportano diversamente. Nel parlato letto, con riferimento alle parlanti GM e RV, l'aspetto più cospicuo riguarda lo slittamento verso il centro della vocale /a/, una contrazione che riduce sensibilmente l'area vocalica nel suo complesso, con specifica ripercussione sulla dimensione dell'altezza vocalica. Diversamente, CI presenta uno spazio vocalico ampio e ben definito, dalla classica forma triangolare. Tali effetti sono ragionevolmente da ascriversi a differenze anatomiche, come ad esempio la dimensione della cavità orale, la grandezza della mandibola, ma anche a una certa postura articolatoria che potrebbe condizionare il grado di protrusione labiale o determinare la faringalizzazione delle vocali.

Figura 3 - *Valori medi di F1/F2 (in Hz) relativi alle tre locutrici nel parlato letto*

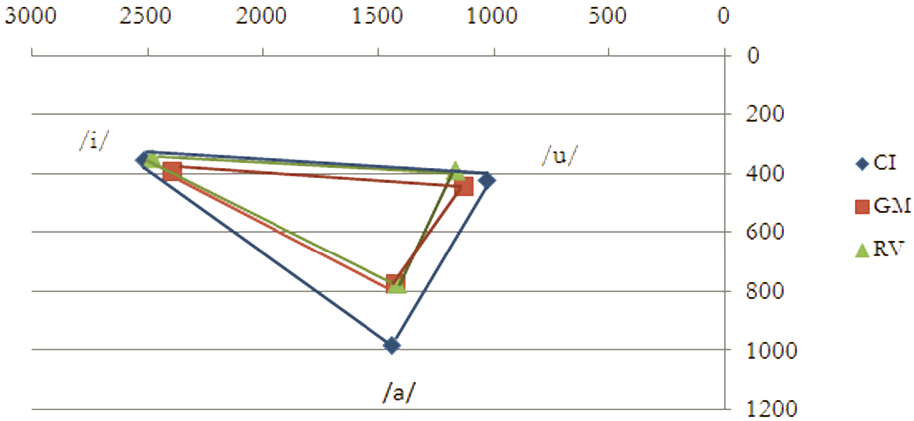
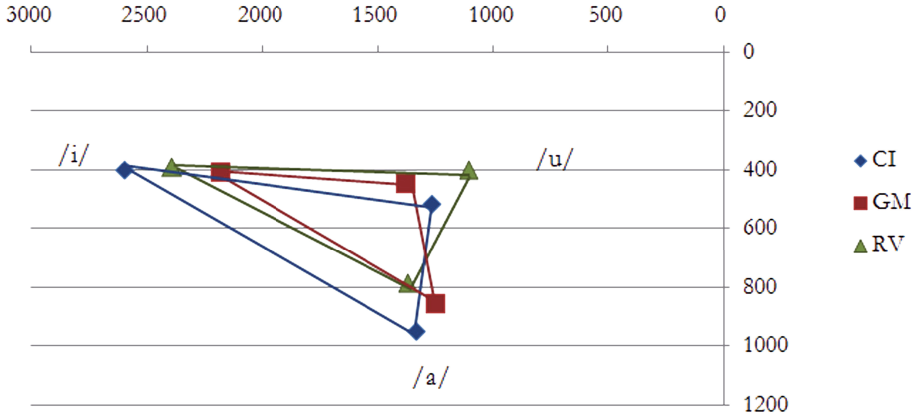


Figura 4 - *Valori medi di F1/F2 (in Hz) relativi alle tre locutrici nel parlato spontaneo*



Nel parlato spontaneo il quadro è meno nitido. Uno stile meno controllato e più espressivo determina i ben noti processi di riduzione e di centralizzazione timbrica. Tuttavia, i parlanti non si comportano allo stesso modo. Al pari del parlato letto, GM presenta un assetto vocalico centralizzato, il posizionamento dei timbri cardinali causa una complessiva riduzione dello spazio. Più contenuta è invece la contrazione del sistema vocalico di RV; in questa parlante sia /i/ che /u/ mantengono una posizione periferica, mentre in CI i timbri più stabili, meno centralizzati, sono /i/ e /a/. Il comportamento vocalico osservato per le tre parlanti potrebbe indurre un effetto timbrico particolare, non indifferente ai fini della successiva valutazione percettiva delle voci.

Per quantificare con maggiore precisione le dinamiche articolatorie delle tre parlanti, e quindi giungere a una stima quantitativa delle differenze vocaliche, abbiamo ritenuto utile calcolare anche l'Area dello Spazio Vocalico (in Hz^2) o *Vowel Space Area*, da ora VSA.

I valori medi relativi a VSA sono riportati nella Figura 5. Nella Tabella 1 abbiamo invece calcolato per le tre locutrici le variazioni percentuali che interessano VSA. I risultati sono congruenti con quanto già descritto, dato che lo scarto più cospicuo concerne CI e GM, tanto nella lettura (± 41%) quanto nello spontaneo (± 33,7%). In quest'ultimo stile, lo scenario è poco sistematico, in CI e RV la dimensione di VSA è approssimabile (± 11%), contraddicendo il *trend* riscontrato per la lettura.

C'è inoltre un condizionamento dipendente dal registro stilistico (Tabella 2), sebbene gli effetti prodotti siano eterogenei. Prevedibilmente, la dimensione di VSA è più ampia nella lettura che nello spontaneo; la distanza tra le due forme di parlato è rilevante, in special modo per CI e GM. Nel dettaglio, in RV, VSA subisce una riduzione del 16,7%, nelle locutrici CI e GM la variazione di VSA dal passaggio tra letto e spontaneo è maggiore.

Figura 5 - *Valori medi di VSA (in Hz²) per le tre locutrici nel parlato letto (L) e spontaneo (Sp)*

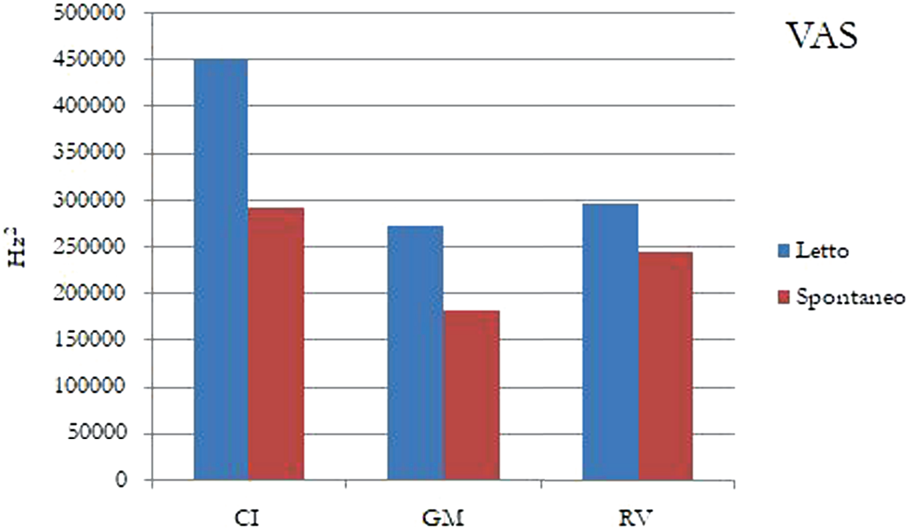


Tabella 1 - *Differenze percentuali tra la dimensione della VSA nei due stili*

Soggetti	*Letto vs Spontaneo*
CI	± 47,8%
GM	+ 36,7%
RV	+ 16,7%

Tabella 2 - *Differenze percentuali tra la dimensione della VSA tra parlanti*

Soggetti	*Letto*	*Spontaneo*
CI/RV	+ 39,8%	+ 11%
CI/GM	+ 42%	+ 33,7%
GM/RV	+ 11,2%	+ 22%

6.3 Intensità e velocità elocutiva

L'intensità ha un comportamento nel complesso significativo, [Lettura, $F(2,33)$= 14,595, p = .000; Spontaneo $F(2,33)$ = 10,586, p = .000]. I confronti multipli dimostrano che nella lettura lo scarto tra CI e GM si pone sulla soglia di accettabilità (p = .05); tutte le altre comparazioni sono però rilevanti. Nel parlato spontaneo, solo il confronto tra GM e RV non è significativo (p = .201). Non si evincono apprezzabili differenze dipendenti dallo stile: nel complesso, CI e GM mostrano valori di poco più elevati nel parlato spontaneo, per cui lo scarto tra i due stili per le due locutrici non è significativo, (rispettivamente, p = .40 e p = .12), mentre in RV la distanza tra letto e spontaneo è trascurabile (p = .42), si veda in merito la Figura 6.

La velocità elocutiva è più sostenuta nella lettura, un risultato atteso (Figura 7), considerato il numero maggiore di pause silenti che caratterizza il parlato spontaneo. Tutte le differenze sono significative sia a un confronto tra parlanti, sia a un confronto tra stili [Lettura, $F(2,21) = 7{,}988$, $p = .003$; Spontaneo $F(2,21) = 31{,}790$, $p = .000$]. Il test post-hoc di Tukey raggiunge sempre la soglia di significatività, ad eccezione del confronto tra CI e RV nel letto ($p = .824$).

Figura 6 - *Valori medi dell'Intensità (in dB) per le tre locutrici nel parlato letto (L) e spontaneo (Sp)*

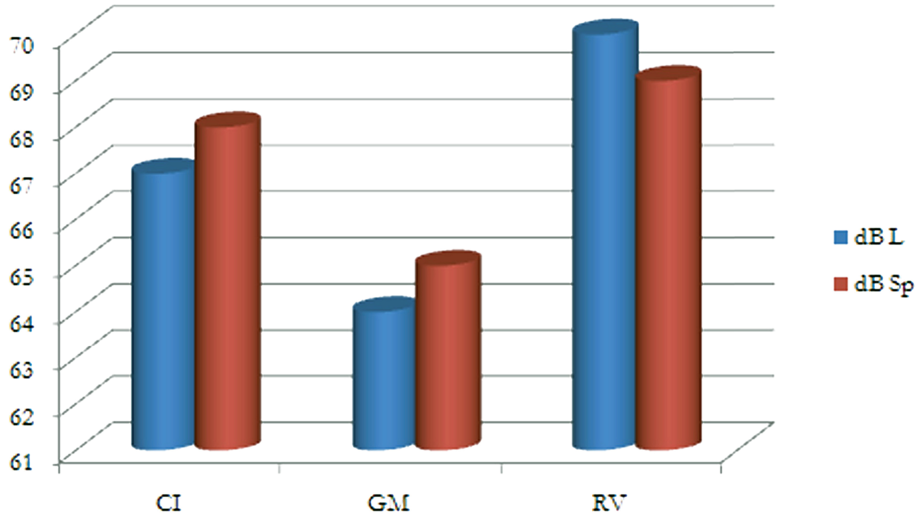
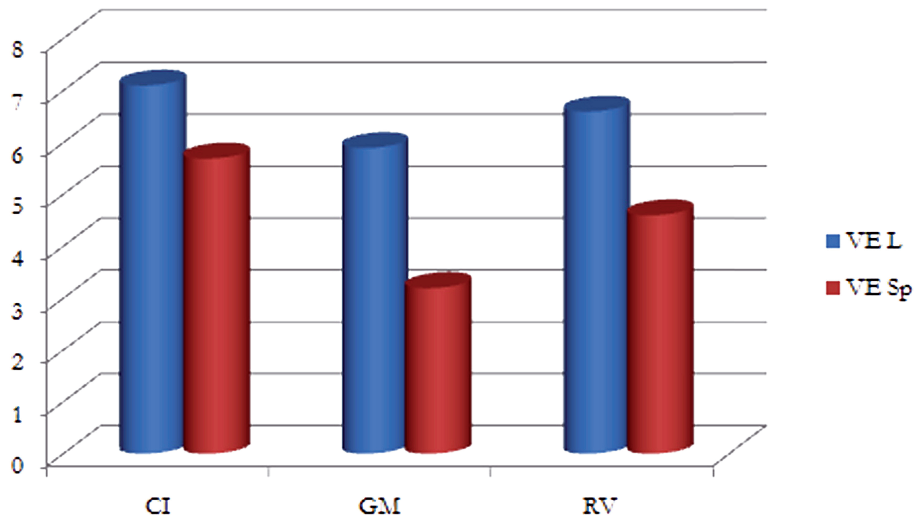


Figura 7 - *Valori medi della Velocita di Eloquio (in sillabe/secondo) per le tre locutrici nel parlato letto (L) e spontaneo (Sp)*

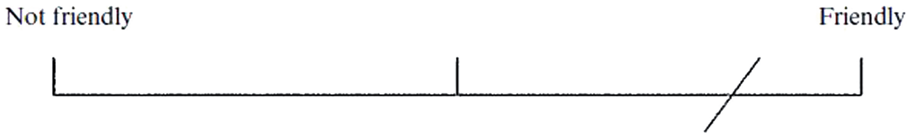


7. *Il test percettivo*

Il test ha avuto lo scopo di verificare l'esistenza di una relazione tra la costituenza acustica delle voci, in primo luogo il *pitch*, e alcuni marcatori della personalità. Il test è stato somministrato a un gruppo di 50 studenti universitari (36 F e 14 M) provenienti dal territorio barese e peri-barese, aventi un'età media di 23.8 anni. Agli uditori è stata inizialmente fornita una scheda cartacea contenente cinque coppie di tratti semantici rappresentati mediante una *Visual Analogue Scale* (VAS)[5], (Chen *et al.*, 2004). Per la percezione dei tratti della personalità, abbiamo scelto una scala concepita in modo continuo, la VAS per l'appunto, che consiste di un asse graduato utilizzato prevalentemente in ambito clinico per misurare quegli stati variabili come l'umore o il dolore. Nel dettaglio, la VAS è una scala di 10 cm ai cui estremi, a sinistra 0 e a destra 10, sono posizionati in modo antitetico gli elementi lessicali di una coppia semantica associata a un tratto della personalità o della fisicità. Un esempio di VAS è riportato nella riproduzione grafica della Figura 8. Diversamente da quanto effettuato da Chen *et al.* (2004), all'estremità sinistra della VAS abbiamo posto i termini denotanti la sfera della sottomissione, mentre a destra quelli relativi alla sfera della dominanza. Gli aggettivi selezionati per il test sono i seguenti: *Piccolo/Grande*, *Insicuro/Sicuro di sé*, *Sottomesso/Dominante*, *Vulnerabile/Aggressivo*, *Desideroso di benevolenza/Deciso*.

Agli uditori è stato chiesto di ascoltare un sottocampione di stimoli audio delle tre voci, ordinati in modo casuale, e di formulare un giudizio sulla parlante, assumendo quale riferimento le coppie di aggettivi rappresentate. Durante l'ascolto, ogni giudice è stato invitato a marcare il punto della VAS in cui a suo avviso si collocava, nella voce proposta, il tratto semantico considerato. I punteggi sono stati ottenuti sommando i millimetri intercorrenti tra l'estremità sinistra, ossia il polo contrassegnato con il valore 0, e il punto marcato dall'ascoltatore.

Figura 8 - *Rappresentazione grafica della Visual Analogue Scale (fonte: Chen et al., 2004: 320)*



L'ipotesi sperimentale è che le voci, essendo differenziate per *pitch*, suscitino negli uditori giudizi diversi. Precisamente, ci aspettiamo che la voce grave (RV) sia associata più frequentemente, nella valutazione percettiva, a quei tratti della personalità che trasmettono dominanza e sicurezza e che pertanto superi sulla VAS il punteggio

[5] Diversamente, la scala Likert, a 5 o a 7 punti, originariamente impiegata in campo sociologico per la misurazione delle opinioni, si serve di una batteria di affermazioni rispetto alle quali i soggetti devono dichiarare il loro grado di accordo (Marradi, Gasperoni, 2005). Avendo carattere discreto, la scala Likert ci è sembrata quindi meno adatta rispetto alla VAS, almeno in questa fase esplorativa della ricerca.

semanticamente neutro, cioè cinque. Diversamente, per la voce acuta (CI), ci attendiamo che i giudizi, inferiori a cinque, si collochino in modo rilevante nella sezione associata ai tratti di benevolenza e sottomissione.

7.1 I giudizi percettivi

I giudizi raccolti non si prestano a una lettura univoca. Com'è evidente dai dati riportati nella Tabella 3, tanto nel parlato letto che nello spontaneo, non c'è una decisa polarizzazione degli esiti. In ambedue gli stili, sebbene in modo più evidente nello spontaneo, si nota una concentrazione consistente dei giudizi intorno all'area centrale della VAS. I risultati medi computati per ogni locutrice e per ogni coppia semantica non sono sempre significativamente differenziati. Nel parlato letto, la situazione è nel complesso più definita, i giudizi sono statisticamente significativi [F(2,736) = 102,629, p = .000], come pure tutte le comparazioni multiple. Limitandoci al confronto tra CI e RV, osserviamo che nello stile lettura le tendenze attese sono per buona parte rispettate, anche se le differenze sono spesso di modesta entità. Una parte dei giudizi espressi per la voce più grave RV si colloca, sulla VAS, nella sezione dei valori superiori a 6 (media 6,4, dev.st. 2); la valutazione propende, anche se non in modo netto, verso il riconoscimento in questa voce dei tratti di sicurezza e dominanza. Al contrario, CI detiene i giudizi meno elevati, in media 5,6 (dev.st. 1,7). Ciò nonostante, nelle coppie *insicuro/sicuro di sé* e *desideroso di benevolenza/deciso* gli *scores* percettivi, a un confronto tra CI e RV, sono statisticamente irrilevanti (t-test), rispettivamente: *insicuro/sicuro di sé* Lettura: p = .89, Spontaneo: p = .30; *desideroso di benevolenza/deciso*, Lettura: p = .17, Spontaneo: p = .72.

Tabella 3 - *Valori medi e deviazione standard (tra parentesi) dei giudizi percettivi ottenuti per ogni coppia di tratti semantici, nel parlato letto (L) e spontaneo (Sp)*

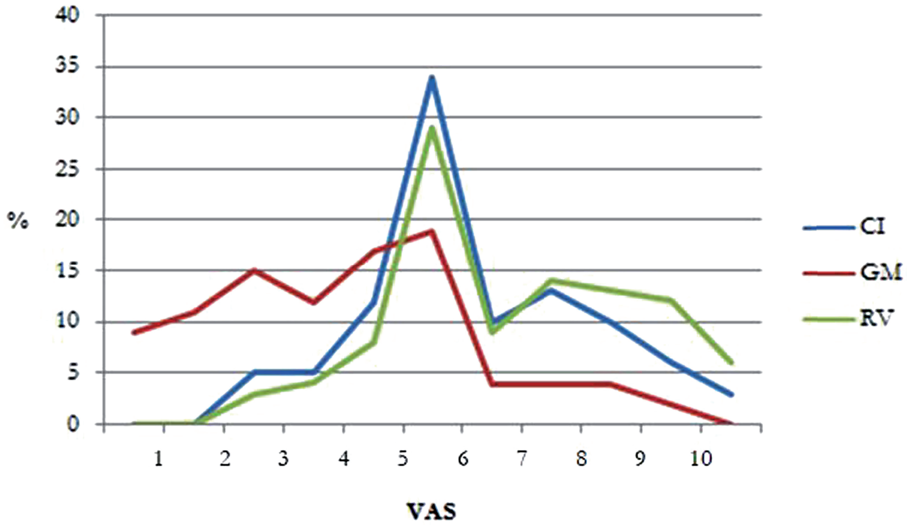
	Piccolo/ Grande		*Insicuro/ Sicuro di sé*		*Sottomesso/ Dominante*		*Vulnerabile/ Aggressivo*		*Desideroso Benevolenza/ Deciso*		*Media*	
	L	Sp	L	Sp	L	Sp	L	Sp	L	Sp	L	Sp
CI	5 (1,7)	4,6 (1,9)	6,6 (1.8)	5,5 (2,7)	5,4 (1,5)	6 (1,9)	4,5 (1,3)	5,1 (2,1)	6,5 (2)	5,6 (1,7)	5,6 (1,7)	5,4 (2,1)
GM	5,3 (2,5)	5,9 (2,6)	3,2 (2)	3,4 (2,5)	3,3 (1,6)	4 (2,3)	2,8 (1,7)	3,7 (2,5)	3,1 (1,7)	5,4 (2,7)	3,5 (1,9)	4,1 (2,4)
RV	6,1 (2)	5,8 (1,9)	6,7 (1,9)	6 (2,3)	6,8 (1,7)	5,5 (2)	5,8 (1,7)	5 (1,9)	6,7 (2,3)	5,8 (2,4)	6,4 (2)	5,6 (2,1)

Nel parlato spontaneo, la situazione è nell'insieme significativa [F(2,737) = 28,64, p = .000]; dai confronti multipli emerge però che le differenze tra CI e RV sono trascurabili (p = .591). Se si esclude la dicotomia *piccolo/grande*, i giudizi mostrano una sostanziale omogeneità, il loro confronto non è infatti statisticamente significativo. Mediamente, CI ottiene sulla VAS giudizi percettivi pari a 5,4 (dev.st. 2,1), un indice di poco inferiore rispetto a quello attribuito durante lo stile lettura (p = 0.41), mentre RV ha un valore pari a 5,6 (dev.st. 2,1); in quest'ultimo caso, la differenza tra letto e spontaneo è significativa (p = 0.01). C'è un aspetto che comunque non passa inosservato: i dati dello spontaneo sono complessivamente più vicini al valore mediano della VAS, ossia 5. Ciò sembra denotare la presenza di una sorta di incertezza da parte degli uditori nel valutare quanto richiesto, che pertanto scelgono, molto spesso, di apporre un valore pressoché neutro, per questo anche poco indicativo.

In decisa controtendenza si pone la voce dell'informatrice GM che, pur non essendo la più acuta, ottiene, nei due stili di parlato, i valori più bassi, rispettivamente 3,5 (dev.st. 1,9) nella lettura e 4,1 (dev.st. 2,4) nello spontaneo (p = 0,01), indice dell'attribuzione in questa voce di quei tratti della personalità che convogliano indecisione, sottomissione e vulnerabilità. Lo scarto è significativo, sia rispetto allo stile (letto vs. spontaneo: p = .011) sia rispetto alle altre voci (p = .002). È plausibile pensare che i giudizi degli ascoltatori siano stati in qualche modo deviati, in questa parlante, dalla concomitante presenza di una più bassa intensità sonora e di una velocità elocutiva rallentata (cfr. Figura 6 e 7), elementi che hanno finito per conferire al suo eloquio un certo alone di insicurezza, un risultato che conferma quanto già emerso in Sorianello, Gurrado (in stampa).

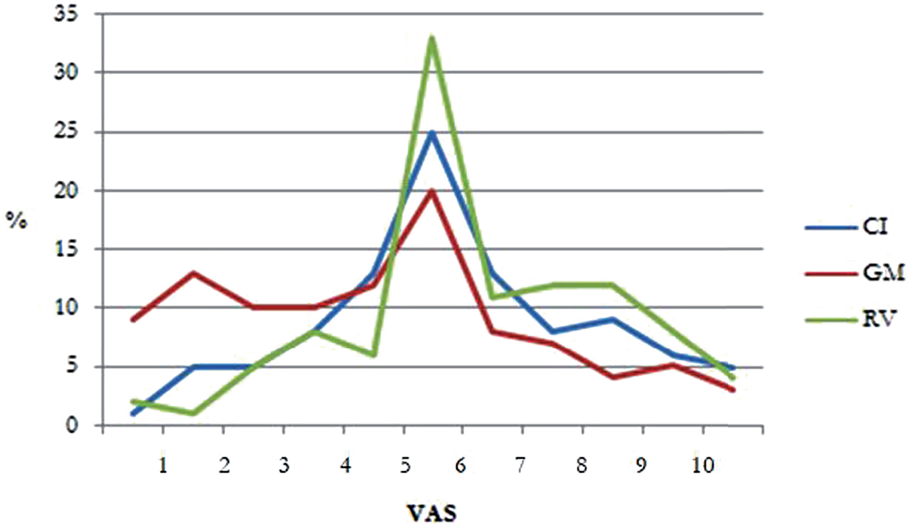
Con l'intento di accertare la distribuzione dei giudizi rispetto ai punteggi della VAS, abbiamo computato gli indici di frequenza e li abbiamo rappresentati graficamente in termini percentuali (v. Figura 9 e 10).

Figura 9 - *Parlato Letto: distribuzione percentuale dei giudizi percettivi assegnati alle tre voci*



Da ciò risulta che nel parlato letto (Figura 9), la distribuzione dei giudizi tra CI e RV mostra una certa diversificazione solo nella sezione sinistra della VAS; precisamente il 34% (CI) e il 29% (RV) di tutti i giudizi ottenuti per queste due locutrici coincide con il valore mediano di 5. Tuttavia, il 13% dei giudizi formulati per RV si colloca tra i punteggi 7 e 9, una percentuale che, per i medesimi indici numerici, scende al 10% in CI. Una simile tendenza si ravvisa anche nello spontaneo (Figura 10).

Figura 10 - *Parlato Spontaneo: distribuzione percentuale dei giudizi percettivi assegnati alle tre voci*



Nuovamente, i giudizi pari a 5 sono maggiori in RV (33%) che non in CI (25%), laddove quelli superiori a 5 sono pari al 12% in RV, ma all'8% in CI, limitatamente al punteggio 8. La rappresentazione grafica mette in luce anche la posizione particolare dell'informatrice GM. Tanto nella lettura che nello spontaneo, i giudizi riguardanti GM che ricadono nella parte inferiore della VAS sono decisamente maggiori, in termini percentuali, se confrontati a quelli assegnati dagli uditori alle altre due parlanti: precisamente, nella lettura i punteggi 2 e 3 sono rappresentati con percentuali pari al 12 e 15%, nelle altre locutrici tali punteggi, invece, raggiungono appena il 5%, nello spontaneo. Va inoltre osservato che i giudizi percettivi formulati per GM si mantengono intorno al 10%, con una punta del 13% per il punteggio 1, diversamente da CI e RV che mostrano percentuali più basse. In aggiunta, si noti che gli indici ottenuti da GM rientranti nella sezione mediana sono percentualmente minori, 19% nella lettura, 20% nello spontaneo.

8. *Riflessioni*

La domanda che ha ispirato l'intera ricerca è se voci femminili diverse siano percettivamente valutabili, in modo altrettanto diverso, con riferimento alla sfera della dominanza fisica e sociale della parlante. La caratterizzazione acustica delle voci ha costituito un tassello preliminare, al fine di poter determinare quanto le voci fossero effettivamente diverse e quindi procedere con un'interpretazione più consapevole dei giudizi percettivi. Un primo aspetto degno di rilievo conferma che l'f0 media dei soggetti considerati è realmente differente. Ciò è sufficiente a conferire alle voci una propria identità percettiva. Tuttavia, i dati sulle frequenze formantiche e sulla dimensione complessiva della VSA tendono solo in parte nella medesima direzione. La parlante avente la frequenza fondamentale più alta, CI, mostra anche la maggiore superficie vocalica, nettamente al di sopra rispetto all'area rinvenuta per RV, la locutrice dalla f0 media più bassa. Due aspetti confondono però il quadro interpretativo: nel parlato spontaneo, le differenze timbriche a volte si attenuano (CI vs. RV), a volte si rinforzano (GM vs. RV), probabilmente per via di quella elasticità espressiva, poco controllata, che distingue il parlato spontaneo. Inoltre, in GM si registra l'area vocalica più piccola, a prescindere dallo stile, sebbene tale locutrice abbia una $f0\bar{x}$ perfettamente in linea con la media prevista per una femmina adulta. Nel complesso, il *pitch* e il timbro vocalico sembrano prefigurarsi dei solidi *discrimen* acustici, ma non sono gli unici. A intricare il quadro, va aggiunto che altri parametri concorrono alla diversità vocale delle informatrici, come il volume e la velocità elocutiva. Il peso svolto da tutti questi indici non sembra risentire in modo incisivo dell'effetto stile, giacché tutte le tendenze emerse nel parlato letto si ripresentano, seppure in misura diversa, anche nello spontaneo. Né d'altra parte possiamo escludere che altri fattori, da noi non esplorati, condizionino ulteriormente la situazione descritta; è il caso dei contorni intonativi oppure della qualità della voce[6].

L'analisi acustica ha dimostrato che le voci del nostro campione sono diverse e che tale 'diversità' sia riconducibile a più parametri. Ciò nonostante, durante il test percettivo, i giudizi degli ascoltatori tesi a saggiare il carattere più o meno dominante di una voce non si distribuiscono in modo uniforme. L'associazione di un *pitch* grave con una personalità dominante, e viceversa, di un *pitch* acuto con una personalità non dominante, assume, infatti, solo un carattere tendenziale. I punteggi si sfrangiano, distribuendosi su tutta la lunghezza della VAS, sebbene alcune sezioni siano emblematicamente più rappresentate. A un confronto con la precedente ricerca (Sorianello, Gurrado, in stampa), limitata solo al parlato letto, si osserva che la riduzione del campione delle informatrici, da cinque a tre, non ottimizza i risultati. Gli ascoltatori sembrano spesso indecisi, non propendendo in modo netto verso l'uno o l'altro tratto della personalità proposto attraverso le cinque scale semantiche. Questa ipotesi sembra avvalorata anche dall'alta densità di giudizi che ricadono nell'area mediana della VAS, rappresentata dal valore 5, un punteggio che, nei casi implicati, sembra neutralizzare il giudizio formulato dagli ascoltatori che pertanto lo sospendono, assegnando alla personalità della parlante un significato né dominante, né sottomesso. L'ipotesi che vi sia una correlazione tra il *pitch* della voce e il suo significato in termini di 'dominanza/sottomissione' trova dunque solo una debole conferma.

La questione affrontata in questa seconda parte della ricerca è per alcuni versi insidiosa. Dire che le voci delle tre locutrici analizzate sono acusticamente diverse non

[6] In merito a quest'ultimo aspetto, abbiamo già avviato una prima verifica, analizzando, per un campione di vocali pronunciate in modo prolungato, i parametri di *shimmer*, *jitter* e *HNR*. Le locutrici non sembrano tuttavia manifestare differenze rilevanti, ad eccezione di GM in cui il valore di *jitter* è più alto di quello attestato negli altri soggetti.

presume automaticamente che ognuna di esse richiami nel giudizio di un ascoltatore gradazioni della personalità altrettanto differenti. Eppure sappiamo che proprio attraverso la voce, strumento semiotico molto potente, 'generatrice di sensi' come direbbe Albano Leoni (2009), viaggiano informazioni di varia natura, affettive e sociali, oltre che squisitamente linguistiche. Si tratta di fondamentali indizi sociofonetici relativi, ad esempio, alla provenienza geografica del parlante, alla sua appartenenza sociale. Accanto a questi, troviamo pure indizi di natura paralinguistica, spesso involontari e incontrollabili, si pensi agli stati d'animo, e finanche informazioni extralinguistiche, come quelle inerenti al sesso, all'età o allo stato di salute. Ora, da questa complessa mescidanza di elementi, stabilire quanta parte l'ascoltatore decida di trattenere, assegnandovi un precipuo significato, è un'operazione tutt'altro che semplice. Pur restringendo il campo di osservazione alle componenti della personalità, e nello specifico alla dimensione della dominanza, l'aspetto che ci riguarda più da vicino, resta da capire come avvenga, nella mente dell'ascoltatore, l'associazione tra una categoria sonora, la voce, e la scelta di una o più qualità della personalità dello stesso locutore. C'è da chiedersi, dunque, cosa induca un ascoltatore a classificare una voce come più o meno dominante, in assenza di contatto visivo o di conoscenza diretta del parlatore. È d'altra parte plausibile ritenere che i significati affettivi e sociali che affiorano dall'ascolto di una semplice voce siano per buona parte guidati da stereotipi culturali e sociali. In questo ambito di indagine, che per molti aspetti sconfina anche verso una dimensione più prettamente psicologica, non sarà fuori luogo menzionare la ben nota tecnica del *matched-guise*, una procedura che coinvolge gruppi di ascoltatori ignari per giudicare più versioni di una stessa voce 'travestita', al fine di esplorare la dinamica degli stereotipi sociali. In questo caso, giudizi diversi, ma in realtà espressi per uno stesso individuo, tradiscono involontariamente la presenza dello stereotipo. Nella nostra ricerca, l'assenza di un confronto diretto tra voci femminili e voci maschili potrebbe aver in parte disorientato l'uditorio che ha cercato, non sempre riuscendoci, di estrapolare dal campione femminile una gamma di significati relativi al tratto di 'dominanza/sottomissione' posizionandoli su un asse continuo. C'è tuttavia un aspetto che potrebbe gettare altra luce su questa complessa tematica, ovvero lo studio degli effetti 'percettivi' prodotti per una stessa voce dalla manipolazione, tramite sintesi, di alcuni parametri acustici. A tale argomento intendiamo dedicarci nel prossimo futuro.

Riferimenti bibliografici

Albano Leoni, F. (2009). *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna: Il Mulino.

Apple, W., Streeter, L.A. & Krauss, R.M. (1979). Effects of pitch and speech rate on personality attributions. In *Journal of Personality Social Psychology*, 37, 715-727.

Aronovitch, C.D. (1976). The voice of personality: stereotyped judgements and their relation to voice quality and sex of speakers. In *Journal of Social Psychology*, 99, 201-220.

Boersma, P., Weenink, D. (2013). *Praat: doing phonetics by computer*, http://www.praat.org.

Chen, A., Gussenhoven, C. & Rietveldt, T. (2004). Language specificity in perception of paralinguistic intonational meaning. In *Language and Speech*, 47(4), 311-349.

Collins, S.A. (2000). Men's voices and women's choices. In *Animal Behaviour*, 60(6), 773-780.

Collins, S.A., Missing, C. (2003). Vocal and visual attractiveness are related in women. In *Animal Behaviour*, 65(5), 997-1004.

Darwin, C.R. (1872). *The expressions of the emotions in man and animals*. London: John Murray (trad. it. *L'espressione delle emozioni nell'uomo e negli animali*, Torino: Bollati Boringhieri, 1992).

D'Errico, F., Signorello, R. & Poggi, I. (2013). The perception of charisma from voice. A cross-cultural study. In *Proceedings of the IEEE Conference on Affective Computing and Intelligent Interaction*. Geneva, Switzerland, 552-557.

Feinberg, D.R., Jones, B.C., Little, A.C., Burt, D.M. & Perret, D.I. (2005). Manipulation of fundamental and formant frequencies influence the attractiveness of human male voices. In *Animal Behaviour*, 69, 561-568.

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

Ferrero, F., Magno Caldognetto, E. & Cosi, P. (1996). Sui piani formantici acustici e uditivi delle vocali di uomo, donna e bambino. In Peretti, A., Simonetti, P. (Eds.), *XXIV Convegno Nazionale dell'Associazione Italiana di Acustica*. Trento: Arti Grafiche Padovane, 169-178.

Flipsen, P., Lee, S. (2012). Reference data for the American English acoustic vowel space. In *Clinical Linguistics and Phonetics*, 26, 926-933.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In Bel, B., Marlien, I. (Eds.), *Proceedings of the International Conference on Speech Prosody 2002*. Aix-en-Provence, Université de Provence, 47-57.

Henton, C. (1989). Fact and fiction in the description of female and male pitch. In *Language and Communication*, 9, 299-311.

Higgins, C., Hodge, M. (2002). Vowel area and intelligibility in children with and without dysarthria. In *Journal of Medical Speech and Language Pathology*, 10, 271-277.

Jacewicz, E., Fax, R.A. & Salmons, J. (2007). Vowel space areas across dialects and gender. In Trouvain, J., Barry, W.J. (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, 1465-1468.

Jiang, H. (2011). Gender differences in English intonation. In Wai-Sum, L., Zee, E. (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong, China, 974-977.

Marradi, A., Gasperoni, G. (2005). *Costruire il dato, 3: Le scale Likert*. Milano: Franco Angeli.

Morton, E.S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. In *The American Naturalist*, 111, 855-869.

Ohala, J.J. (1983). Cross-language use of pitch: an ethological view. In *Phonetica*, 40, 1-18.

Ohala, J.J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. In *Phonetica*, 41, 1-16.

OHALA, J.J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In HINTON, L., NICHOLS, J. & OHALA, J.J. (Eds.), *Sound symbolism*. Cambridge: Cambridge University Press, 325-347.

PUTS, D.A., GAULIN, S.J.C. & VERDOLINI, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. In *Evolution and Human Behaviour*, 27, 283-296.

ROSENBERG, A., HIRSCHBERG, J. (2009). Charisma perception from text and speech. In *Speech Communication*, 51(7), 640-655.

SCHERER, K.R. (1972). Judging personality from voice: a cross cultural approach to an old issue in interpersonal perception. In *Journal of Personality*, 40, 191-210.

SCHERER, K.R. (1979). Personality markers in speech. In SCHERER, K.R, GILES, H. (Eds.), *Social markers in speech*. Cambridge: Cambridge University Press, 147-201.

SIGNORELLO, R., D'ERRICO, F., POGGI, I. & DEMOLIN, D. (2012). How charisma is perceived from speech. A multidimensional approach. In *ASE/IEEE International Conference on Social Computing*. Amsterdam, The Netherlands, 435-440.

SIGNORELLO, R., DEMOLIN, D. (2013). The physiological use of the charismatic voice in political speech. In *Proceedings of Interspeech 2013*. Lyon, France, 987-991.

SIMPSON, A.P. (2009). Phonetic differences between male and female speech. In *Language and Linguistics Compass*, 3, 621-640.

SORIANELLO, P., GURRADO, G. (in stampa). Il Codice della Frequenza: una valutazione acustica e percettiva della voce femminile in rapporto ad alcuni tratti della personalità. In DE MEO, A., DOVETTO, F. (Eds.), *La Comunicazione Parlata*, Napoli, 2016.

TITZE, I.R. (1989). Physiological acoustic differences between male and female voices. In *Journal of the Acoustical Society of America*, 85, 1699-1707.

TUSING, K.J., DILLARD, J.P. (2000). The sounds of dominance: vocal precursors of perceived dominance during interpersonal influence. In *Human Communication Research*, 26, 148-171.

VORPERIAN, H.K., KENT, R.D. (2007). Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. In *Journal of Speech, Language, and Hearing Research*, 50(6), 1510-1545.

DOI: 10.17469/O2103AISV000011

PAOLO BRAVI

Variation in singing voice quality. A case-study from traditional music of southern Marche

In this paper a recording of two traditional *stornellos* performed by two non professional singers from southern Marche (Italy) has been analyzed. Three acoustic measures related to harsh voice quality – jitter, shimmer and harmonics-to-noise ratio (HNR) – have been examined in order to describe by means of quantitative descriptors the difference between the voices of the two singers and to investigate some prospective factors that can be related to the variability of their voice quality in the recorded performance.

Key words: voice quality, singing voice, harsh voices, Central Italy traditional singers, ethnomusicology.

1. *Introduction: voice quality in singing*

In vocal music, one of the main issue regarding score notation is that music symbols can express only the skeleton of a melody, whereas all the subtleties present in the real act of singing are hidden in the partiture (Stockmann, 1989; Ellington, 1992). The movements between and inside the notes may be of greater importance than the basic structure of a melody, but cannot be adequately described by standard musical notation. In particular, the so-called 'timbre' dimension – in the widest sense of the term – of the voice is not included in the score and depends almost completely on the specific characteristics and choices of the singer. One of the essential parts defining the timbral dimension of singing is voice quality, that has a major importance both as part of a musical style and as a crucial trait of a singer's performative art. At the so called *estesic* level (Nattiez, 1987), this aspect has been left for centuries to the impressionistic evaluation of the listeners and singing expert. Nowadays, instrumental analysis allows one to observe and analyze this aspect of vocal performance in detail and in an objective way.

In this paper I will examine one particular characteristic of voice quality that is referred to with different adjectives (the most frequent of which are *harsh*, *rough* and *raspy*) by specialists approaching the voice from different perspectives. In fact, no agreement is nowadays shared on the terminology to be used to describe those voices that, according to the common sense and use of the language, may be said to have a harsh sound quality.

In the field of phonetics, as John Laver observes, "small cycle-to-cycle variations in fundamental frequency are associated with voice judged to be *harsh*" (Laver, 1980: 127, italic mine). In particular, in fact, *jitter*, *shimmer* and *harmonicity* (or HNR: Harmonics-to-Noise Ratio) are acoustic features correlated with the perception of harshness. The first two are measures of the vocal perturbation: *jitter* refers to small variations in the fundamental frequency; *shimmer* refers to perturbation in the amplitude of cycles. *Harmonicity* refers to the amount of spectral noise in the signal and is measured as the ratio between the energy in the harmonics and in the noise[1].

In the field of logopedics and phoniatrics, anomalies of this kind relevant to the laryngeal activities are usually considered under the term *roughness*. Both GRBAS (Isshiki *et al.*, 1969; Hirano, 1981) and CAPE-V (ASHA, 2006), two scales commonly used by voice pathologists, consider "roughness" as one of the crucial parameters in the diagnosis of voice dysfunctions.

In the field of ethnomusicology, one of the parameters used in Alan Lomax's *cantometrics* to describe the vocal features of singing is defined by the word *raspiness*, but the same author remarks that the words "throatiness" or "harshness" could also have been used (Lomax, Grauer, 1968: 73-74).

This paper aims at examining the effects of factors that may be supposed to be involved in the variation of voice quality that occurs while singing. In particular, it is analyzed how acoustic features related to harsh voice quality (jitter, shimmer and HNR) vary in relation to some prospective factors of change – tone level and movement, duration, vowel type, phonetic context, intensity, metrical position and stress – in a sung performance of two traditional and non professional singers from the southern part of the Marche region (central Italy). A crucial difference between linguistic and musicological analysis that has to be taken into consideration is that vocal harshness may be, in the case of singing, more a matter of stylistic choice (either individual or genre-dependent) rather than a 'mechanical' outcome bound to phonetic, prosodic or metrical factors.

2. *Case-study: Defiantly stornellos from Corridonia (Marche Region – Italy)*

In 1998, the folk revival group "La Macina"[2] published a CD comprising, in addiction to its musical elaboration of traditional songs from the region of Marche (Italy), some recordings of old, non-professional singers of this region (La Macina, 1998). Having been made in a studio, these high-quality recordings are valuable documents of the voices of these traditional singers. Of particular interest are the recordings of the "stornelli a dispetto" (defiantly *stornellos*), a genre of singing typical of this area (Arcangeli, 1982). The

[1] It has to be noted that variations in these acoustic correlates may refer not only to harsh voices, but also to creaky ones. In articulatory terms, the two registers are distinguished by a major constriction of the ventricular folds in harsh voice. According to Edmondson and Esling (2006: 169): "[c]reaky vocal register [...] generally occurs at low pitch, with constricted Valve 3 ['sphincteric compression of the arytenoids and aryepiglottic folds forwards and upwards by means of the thyroarytenoid muscle complex'], but with a loose enough glottis that vibrations are slow and undulating. Harsh vocal register at high pitch [...], when viewed laryngoscopically, shows Valves 1 ['glottal vocal fold adduction and abduction'] and 2 ['partial covering and damping of the adducted glottal vocal fold vibration by the ventricular folds'] engaged, with clear Valve 3 engagement. Noticeable are the ventricular folds, which cover a substantial portion of the vocal folds [...], compressing them and damping oscillation".

[2] La Macina is a folk music revival group devoted – as expressed by the complete name of the group – to the "research and singing" of traditional music of Marche. The group was founded by its current leader, Gastone Pietrucci, some forty years ago and is still active today (La Macina, 2016).

stornellos (track 9 of the CD) were recorded in 1995 and were performed by two singers of traditional songs, both from the village of Corridonia, in the province of Macerata - Marche (Italy): Nazzareno Saldari (hereafter NS), male (1912-2004), and Lina Marinozzi Lattanzi (hereafter LML), female (1925-2010)[3]. The two voices documented in the recording, though clearly different from one another, share some basic characteristics of the traditional singing voices of the Centre and South of Italy. Those voices have been described as "strozzate" (throttled) and "forzate" (forced) by Diego Carpitella and Tullia Magrini, two well-known Italian ethnomusicologists of the recent past (see Carpitella, 1955: 23; Magrini, 1990: 23).

Figure 1 - *The two singers of traditional songs from Marche Lina Marinozzi Lattanzi (left) and Nazzareno Saldari (right)*

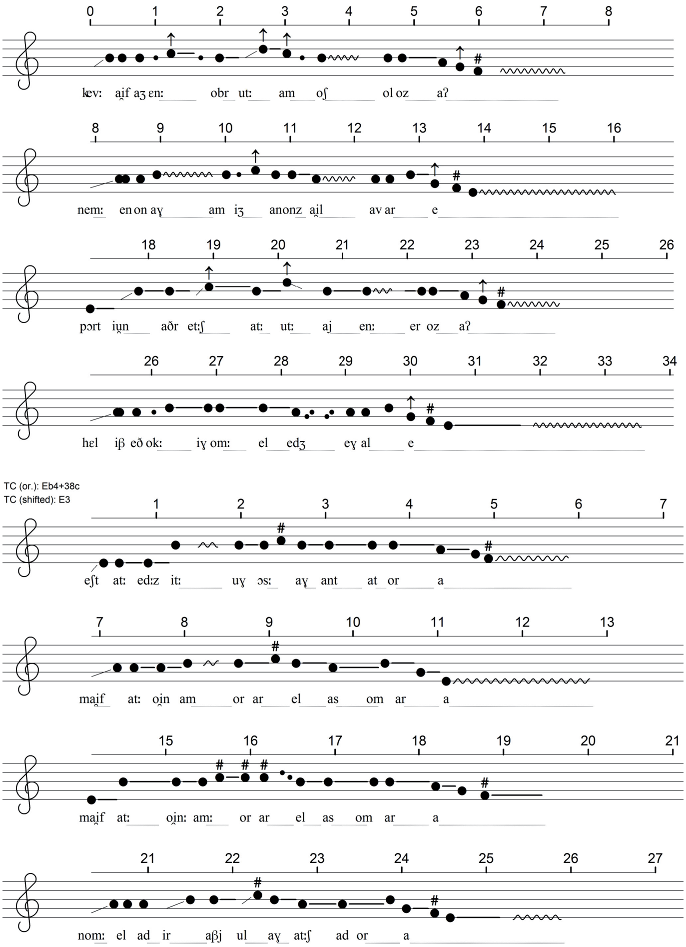
A musical analysis of the recording is out of the scope of the present work. However, a basic description of the free rhythm melodies through staff notation, together with a broad phonetic transcription of the texts, is given in Figure 2[4]. The two musical transcriptions have been reported to the same tonal centre (E4) to facilitate their

[3] Little information is available on the two non professional singers. The booklet that accompanies the CD recording reports that Lina Marinozzi Lattanzi was a farmer and later worked in a shoe factory (La Macina 1998: 13). Nazzareno Saldari was also a farmer (direct information). Further information and a short biography of Lina Marinozzi Lattanzi are in Arcangeli, Pietrucci & Bravi, 2016.

[4] The verbal texts of the *stornellos* and their translation as reported in the booklet are the following: "[NS] Che vai facendo brutta mosciolosa / nemmeno unà camigia non sai lavare / porti una treccia tutta jennerosa / e li pedocchi comme le cicale [What are you doing, filthy woman / You are not able to wash even a shirt / Your hair is so much dirty / That the lice are big as cicades] [LML] E statte zittu co' ssa cantatora / m'hai fatto innamorare la ssomara / m'hai fatto innamorare la ssomara / non me la tira più la cacciatore [Shout up, with this song / You delight my sheass / You delighted my sheass / So that she doesn't pull the hand-cart anymore]" (La Macina, 1998: 13).

comparison, but in fact they have different ranges, being LML's performance one perfect fourth higher than NS's one[5].

Figure 2 - *Musical transcriptions of the stornellos sung by NS (top) and LML (bottom)*

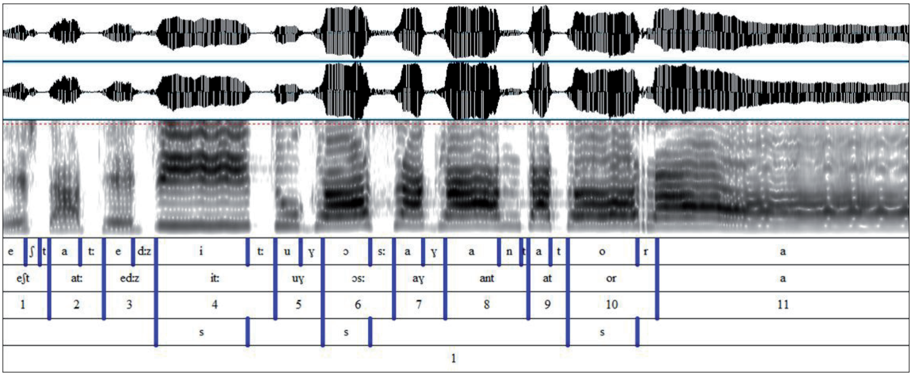


In order to analyze the acoustic features of the two voices, a manual multilevel segmentation of the recording was carried out by means of the software Praat (Boersma, Weenink, 2016). Segmentation tiers were relevant both to phonetic (phono, V-to-V interval) and metrical (line, position, stress[6]) aspects (Figure 2).

[5] *Tonus finalis* – i.e. the tone at the end of a melodic phrase or section, also referred to some times as 'cadential tone' (Agamennone, 1991: 158-163) – has a mean value of 240 Hz for NS and 320 Hz for LML.

[6] The three words referring to metrical notions may require to be clarified and referenced. 'Line' is a basic concept in poetry and in metrical analysis: "[w]hat distinguishes all poetry from prose is that poetry is made up of lines (verses). [...] In metrical poetry [...] lines must satisfy requirements on length and on the location in the lines of marked syllables, and different conditions are met by different kinds of non-metrical poetry" (Fabb, Hall, 2008: 1). The "metrical position" is assumed to be the basic unit of the line in some modern approaches

Figure 3 - *An example of multilevel segmentation performed through Praat TextGrid*



3. *Analysis*

Based on an impressionistic evaluation, the two voices have different voice qualities: while LML's voice has a strong harsh quality, NS's one may be defined – using a term that is often used in the field of singing voice analysis – as "pressed"[7]. The first part of the analysis has aimed to better define this course-grained evaluation through an instrumental analysis of the three acoustic features – jitter, shimmer and HNR – that are considered to be associated with the perception of harshness in voice quality. Measurements of the values of these three features in the 96 vowels sung in the *stornellos* by the two singers (see Table 1) have been carried out via Praat[8].

to metrical analysis, particularly those referring to the generative approach. As observed by Morris Halle and Samuel J. Keyser, "[a]position is normally occupied by a single syllable, but under certain conditions it may be occupied by more than one syllable or by none. [...] Two vowels may constitute a single position provided that they adjoin, or are separated by a liquid or nasal or by a word boundary which may be followed by *h*-, and provided that one of them is a weakly stressed or unstressed vowel. [...] An unstressed or weakly stressed monosyllabic word may constitute a single metrical position with a preceding stressed or unstressed syllable" (Halle, Keyser, 1966: 197; for the concept of metrical position used with reference to the Italian metrical tradition, Di Girolamo, 1983: 22-24, and Beltrami, 1991: 39-40). 'Stress' is here considered not as referring to the prominent syllable in a word (lexical stress), but as a concept related to the metrical/musical structure of the sung line (Beltrami, 1991: 27-29; Fraisse, 1979 [1974]: 65-68).

[7] As observed by Johann Sundberg, "[p]ressed phonation is characterized by a high subglottic pressure combined with a strong adduction force" (Sundberg, 1987: 80; Fussi, Magnani, 2010: 200). Closing quotient of the vocal folds, subglottal pressure, sound level and glottal area are the main objective features related to it: "[w]hen the adduction force is high as in pressed phonation, the glottogram amplitude becomes small, the closed phase long, subglottic pressure is high, sound level is low, and the glottal area is small" (Sundberg, 1987: 83-84).

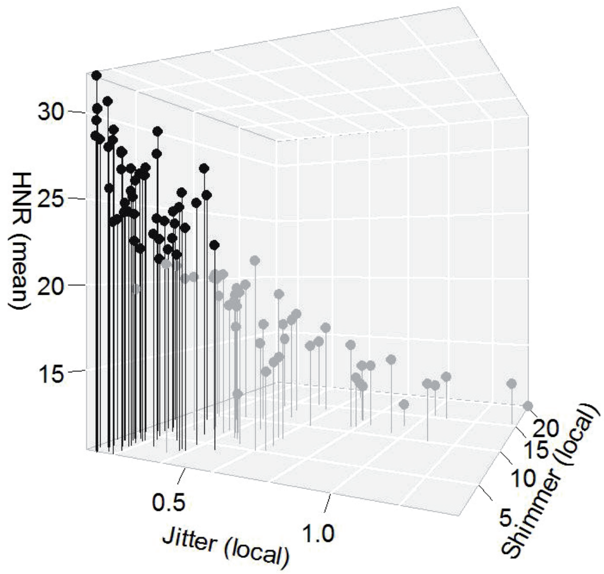
[8] Different types of measurement of jitter and shimmer are used to assess voice quality. Here the most common types of measure – local jitter and shimmer – have been utilized (see "Voice" section in the Praat manual for details).

Table 1 - *Table of counts of vowels in the two stornellos sung by singers LML and NS*

Singer		*Vowel type*							
		i	*e*	*ɛ*	*a*	*ɔ*	*o*	*u*	
	LML	6	5	0	25	1	9	2	48
	NS	6	12	2	15	1	9	3	48
		12	17	2	40	2	18	5	96

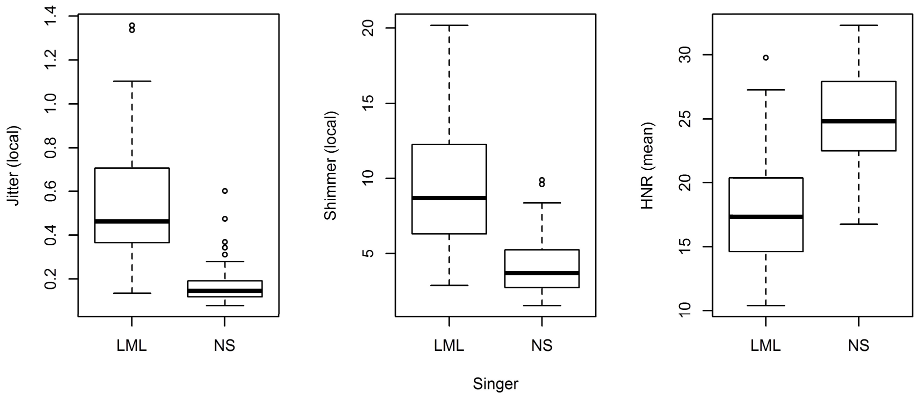
The distribution of the values of the three acoustic features confirms the subjective evaluation of a clear distinction of the quality of the two singing voices. The 3D plot in Figure 4 shows that the area of distribution of the points relevant to the three features in the vowels sung by LML is clearly set apart from that of NS. A closer inspection of the value distributions for each acoustic feature proves that a significant difference between the two singers is present (see Figure 5 for a graphical representation)[9]. While jitter and shimmer levels are higher in LML sung vowels, HNR is lower. In particular, the values of shimmer are extremely high in LML's voice, compared with the threshold value considered for the assessment of pathology[10].

Figure 4 - *3D plot of the values of jitter, shimmer and HNR values in the sung vowels of the stornellos sung by LML (light grey) and NS (dark grey)*



[9] Wilcoxon tests with continuity correction have been carried out with the following results: W = 2140, p-value < 0.001 (jitter); W = 2032, p-value < 0.001 (shimmer); W = 286, p-value < 0.001 (HNR).
[10] This threshold, as reported in the Praat manual, is 3.810% according to the program Multidimensional Voice Program – MDVP (see http://www.fon.hum.uva.nl/praat/manual/Voice_3__Shimmer.html).

Figure 5 - *Boxplots showing the distribution of jitter, shimmer and HNR for the two singers*



As far as the relation among the three features is concerned, in 5 out of 6 cases values are significantly but rather weakly correlated to one another in each singer[11].

The second part of the analysis has aimed to explore what factors may be associated to the variability of voice quality in the two *stornellos.* This part, in turn, is divided in two sections: the first one aims at investigating if the variability of jitter, shimmer and HNR may be explained by continuous variables relevant to duration, fundamental frequency (mean and range) and intensity; the second one aims at verifying if factors vowel type, phonetic context, metrical position and stress have a significant effect in the variation of the three acoustic features.

As far as continuous variable are concerned, a series of correlation tests have been performed. A number of statistically significant correlations (see results in Table 2[12]) have been found, particularly as far as duration and intensity are concerned. However, the effects are usually rather week, being the variance shared by the two variables within the range 11%-45%. Two plots based on data that exhibit statistically significant correlations (shimmer / duration and HNR / intensity) are shown, as example cases, in Figure 6.

[11] Correlation tests (Pearson method) give the results that follow. Singer LML: jitter / shimmer [$r = 0.68$, $df = 46$, $p < 0.001$], jitter / HNR [$r = -0.78$, $df = 46$, $p < 0.001$], shimmer / HNR [$r = -0.74$, $df = 46$, $p < 0.001$]; singer NS: jitter / shimmer [$r = 0.16$, $df = 46$, $p = 0.28$], jitter / HNR [$r = -0.45$, $df = 46$, $p = 0.001$], shimmer / HNR [$r = -0.73$, $df = 46$, $p < 0.001$].

[12] Measures are relative to the vowels sung in the two *stornellos*.

Table 2 - *Correlation tests between numeric variables*

statistic_t	*df*	*p-value*		*r-squared*	*variable1*	*variable2*	*Singer*
-3.285	42	0,0021	**	0,2044	Duration	Jitter (local)	LML
-0,5686	40	0,5728		0,008			NS
-24.014	42	0,0208	*	0,1207		Shimmer (local)	LML
-49.175	40	0	***	0,3768			NS
23.311	42	0,0246	*	0,1146		HNR (mean)	LML
58.256	40	0	***	0,459			NS
-31.319	46	0,003	**	0,1758	f_o (mean)	Jitter (local)	LML
-30.093	46	0,0042	**	0,1645			NS
-19.635	46	0,0557		0,0773		Shimmer (local)	LML
10.133	46	0,3162		0,0218			NS
0,9068	46	0,3692		0,0176		HNR (mean)	LML
-0,1275	46	0,8991		4,00E-04			NS
-0,2344	46	0,8157		0,0012	f_o (range)	Jitter (local)	LML
36.671	46	6,00E-04	***	0,2262			NS
-0,0394	46	0,9687		0		Shimmer (local)	LML
-23.697	46	0,0221	*	0,1088			NS
13.297	46	0,1902		0,037		HNR (mean)	L
0,6587	46	0,5133		0,0093			NS
-61.805	46	0	***	0,4537	Intensity (mean)	Jitter (local)	LML
-14.994	46	0,1406		0,0466			NS
-39.457	46	3,00E-04	***	0,2529		Shimmer (local)	LML
-25.687	46	0,0135	*	0,1254			NS
27.231	46	0,0091	**	0,1388		HNR (mean)	LML
45.291	46	0	***	0,3084			N

Figure 6 - *Correlation between acoustic features and continuous variable: shimmer / duration (left); HNR / intensity (right)*

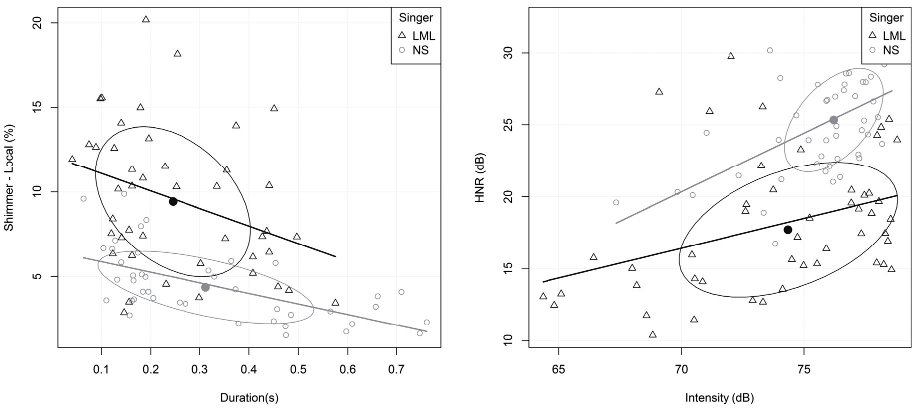


Table 3 - *Mean values of jitter, shimmer and HNR for levels relevant to the factors vowel, phonetic context (pre), phonetic context (post), metrical position, stress*

		factor: *vowel*						
		i	*e*	*ɛ*	*a*	*ɔ*	*o*	*u*
Jitter	LML	0.62	0.50	0.77	0.36	/	0.58	0.23
(local)	NS	0.21	0.14	0.16	0.60	0.31	0.12	0.19
Shimmer	LML	9.39	9.54	12.51	3.76	/	8.58	6.01
(local)	NS	5.68	4.05	3.10	4.99	3.50	4.14	6.93
HNR	LML	17.64	17.84	13.23	25.38	/	18.29	28.50
(mean)	NS	20.61	26.05	26.10	21.36	25.06	26.56	21.07

		factor: *phonetic context (pre)*					
		plosive	nasal	fricative	trill/tap	vowel	others
Jitter	LML	0.52	0.51	0.47	0.51	0.85	1.08
(local)	NS	0.26	0.12	0.15	0.15	0.21	/
Shimmer	LML	8.50	9.95	8.00	9.96	12.18	14.97
(local)	NS	4.64	3.77	3.57	4.03	9.28	/
HNR	LML	19.29	17.21	19.28	17.32	14.36	13.26
(mean)	NS	23.49	25.24	26.20	26.62	18.59	/

		factor: *phonetic context (post)*					
		plosive	nasal	fricative	trill/tap	vowel	others
Jitter	LML	0.53	0.61	0.58	0.41	0.67	0.46
(local)	NS	0.14	0.15	0.18	0.26	0.19	0.16
Shimmer	LML	9.24	8.96	9.75	9.24	10.79	8.15
(local)	NS	3.74	4.72	4.55	3.85	3.55	1.61
HNR	LML	18.76	17.30	17.65	18.13	16.41	20.40
(mean)	NS	25.89	24.93	24.49	24.73	23.92	29.22

				factor: *metrical position*								
		1	2	3	4	5	6	7	8	9	10	11
Jit	LML	0.84	0.67	0.60	0.29	0.44	0.37	0.61	0.36	0.77	0.38	0.46
loc	NS	0.42	0.19	0.13	0.15	0.15	0.15	0.12	0.13	0.17	0.12	0.15
Sh	LML	12.48	10.65	8.37	5.33	8.87	7.40	12.50	8.29	10.79	9.34	8.15
loc	NS	4.03	5.63	4.99	2.92	5.29	3.70	4.57	3.83	5.45	2.84	2.36
HNR	LML	15.82	15.27	17.60	21.60	20.08	21.46	14.47	18.74	15.13	18.92	20.40
mn	NS	22.77	20.79	23.81	27.63	23.91	24.85	26.34	27.06	22.04	29.12	28.29

		factor: *stress*	
		unstressed	stressed
Jitter *(local)*	LML	0.61	0.35
	NS	0.17	0.17
Shimmer *(local)*	LML	10.02	7.48
	NS	4.51	3.48
HNR *(mean)*	LML	17.14	20.65
	NS	24.43	26.48

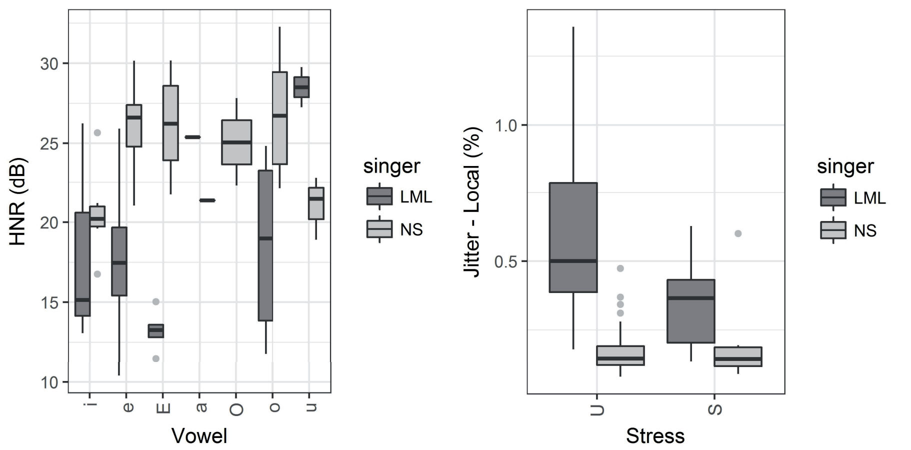
Table 4 - *Kruskal-Wallis tests for factor vowel type, phonetic context (pre and post), metrical position and stress*

chi-sq	*df*	*p-value*		*factor*	*response*	*singer*
82.967	5	0,1406		Vowel	Jitter *(local)*	LML
122.835	6	0,0559				NS
67.463	5	0,2402			Shimmer *(local)*	LML
95.473	6	0,1451				NS
146.951	5	0,0117	*		HNR *(mean)*	LML
174.312	6	0,0078	**			NS
73.818	5	0,1938		Phonetic context *(PRE)*	Jitter *(local)*	LML
103.252	4	0,0353	*			NS
75.239	5	0,1845			Shimmer *(local)*	LML
111.825	4	0,0246	*			NS
65.788	5	0,2539			HNR *(mean)*	LML
124.331	4	0,0144	*			NS
3.132	5	0,6796		Phonetic context *(POST)*	Jitter *(local)*	LML
0,8794	5	0,9717				NS
12.968	5	0,9353			Shimmer *(local)*	LML
58.591	5	0,3202				NS
38.907	5	0,5653			HNR *(mean)*	LML
50.191	5	0,4136				NS
165.123	10	0,0859		Metrical Position	Jitter *(local)*	LML
151.962	10	0,1251				NS
138.848	10	0,1783			Shimmer *(local)*	LML
155.361	10	0,1137				NS
144.953	10	0,1516			HNR *(mean)*	LML
277.778	10	0,002	**			NS

73.673	1	0,0066	**	Stress	Jitter *(local)*	LML
0,0082	1	0,9277				NS
3.449	1	0,0633			Shimmer *(local)*	LML
17.307	1	0,1883				NS
46.944	1	0,0303	*		HNR *(mean)*	LML
28.174	1	0,0932				NS

As far as factor analysis is concerned, a series of Kruskal-Wallis tests have been performed. Results are presented in Table 3 and Table 4. In the case of vowel type, a significant effect occurs on HNR (but not in the other two features); in the case of phonetic context[13], significant effects occur in the case of one singer (NS), but not in the other (LML), and only as far as left context is concerned; in the case of metrical effects, stress has a significant effect in singer LML as regards the features jitter and HNR. Two series of boxplots relevant to factors that exhibit statistically significant effects (HNR by vowel type and jitter by stress) are shown, as example cases, in Figure 7.

Figure 7 - *Vowel type and matrical stress effects. Left: HNR by vowel, SAMPA Symbols; right: jitter by stress, [U]nstressed vs. [S]tressed*



4. *Discussion*

The analysis of voice quality in the sung *stornellos* performed by two traditional singers of the Marche has shown that, despite both singers can be regarded as valuable expressions of the vocal style that characterizes traditional song of central and

[13] The phonetic context – either "pre" or "post" vowel – have been evaluated grouping the phones that precede/follow the vowel into five main categories (plosive, nasal, fricative, trill/tap, vowel), plus a residual group comprising a little number of other different types of phonos. Grouping phonos into a small number of phonetic categories is due to the limitation of the available corpus of sung vowels.

southern Italy, their voices are significantly different from one another. The strong harshness that characterizes Lina Marinozzi Lattanzi's voice is absent in Nazzareno Saldari's one. All three features here examined – and particularly jitter and shimmer – show that sung vowels of the first singer have a different quality with respect to those of the second.

Less conspicuous are the results of the analysis of factors and variables that may be related to the change in voice quality. Short vowel duration – that implies vicinity to consonants and rapid articulatory change – is correlated with increase of jitter and shimmer (and decrease of HNR), but from one side the variance explained is rather low, from the other there is no clear evidence of a specific effect of the phonetic context on the three features. Intensity is correlated with all features (with the exception of jitter in NS), but from one side – again – the variance explained is low, from the other the factor "stress" (which in its turn might have been expected to be associated to intensity) is a significant factor only for LML.

As a whole, the analysis of factors and correlations confirms that the voices of the two singers differ not only in their intrisic characteristics, but also in their response to prospective factors of variability. Tone movement within vowels (here measured in terms of f_0 range within the phono), is significantly correlated with jitter and shimmer in NS, but not in LML. Left phonetic context has a significant effect on all features in NS, but not in LML. Metrical stress, as noted beforehand, has a significant effects in LML's voice, but not in NS's one.

The analysis of the factors that may explain the variability of the voice quality in this recording certainly suffers from the scarsity of available data. This is an obvious but inescapable side-effect of the exceptionality of the recording that we have considered. Further studies, based on a larger corpus of recordings, are likely to be able to shed more light on the issue of the variability of voice quality in this type of traditional singing.

In a wider perspective, two main points have to be highlighted. The first one is that when one deals with singing, criteria and benchmarks that are commonly used for assessing voice quality with regard to speech cannot be mechanically applied. In the case of singing, harshness cannot be regarded as a form of voice disorder or a pathology, but instead has to be seen as a distinctive and remarkable characteristic of particular singing styles, or as an expressive trait of particular vocal passages. Within the traditional song of the Marche region, this is the case of the extra-harsh voice of Lina Marinozzi Lattanzi. Her typical vocal quality is considered a gift of a talented singer, and not a blemish or an unpleasant sign of pathology[14]. Observed under this perspective, the characteristic harshness of her voice might be evaluated, at least in part, as a matter of specific stylistic choice, more (or rather) than an effect – perhaps indesirable and/or unaivodable – of particular factors related to the metrical, prosodic and phonetic settings and conditions.

[14] This is also the case of many renowned rock and flamenco singers, to name just a couple of well known musical genres.

This observation leads to the second aspect that is worth of proper consideration. Differently from what both a romantic view of folk art and an old-fashioned practice of field research in the field of ethnomusicology might lead us to think, usually there is no such things as 'anonymous' voices in traditional music. Or, at least, one cannot find 'impersonal' voices here more than in the field of lyrical song or in other singing styles which are duly based on a formal training. Traditional music is in most cases a practice which involves identifiable performers, each of whom – included the late Nazzareno Saldari and Lina Marinozzi Lattanzi taken into examination here – has his/her own style and imprints a particular 'vocal signature' on his/her singing performances.

Bibliography and Discography

Agamennone, M. (1991). Modalità/Tonalità. In Agamennone, M., Facci, S., Giannattasio, F., Giuriati, G. (Eds.), *Grammatica della musica etnica*. Roma: Bulzoni, 145-200.

Arcangeli, P. (Ed.) (1982). *Marche 1. Musica tradizionale del Maceratese*, LP (SU 5006, recordings by D. Toccaceli). Roma: CETRA.

Arcangeli, P., Pietrucci, G. & Bravi, P. (2016). Ricordando la figura di Lina Marinozzi Lattanzi. In *Quaderni Musicali Marchigiani*, 14, 47-74.

ASHA [American Speech-Language-Hearing Association] (2006). *Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)*. http://www.asha.org/uploadedFiles/members/divs/D3CAPEVprocedures.pdf/.

Beltrami, P.G. (1991). *La metrica italiana*. Bologna: Il Mulino.

Boersma, P., Weenink, D. (2016). *Praat: doing phonetics by computer*, computer program. http://www.fon.hum.uva.nl/praat/

Carpitella, D. (1955). Sulla musica popolare molisana. In *La Lapa*, III, 1/2, 22-23.

Di Girolamo, C. (1983). *Teoria e prassi della versificazione*. Bologna: Il Mulino.

Edmonson, J.A., Esling, J.A. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. In *Phonology*, 23, 157-191.

Ellington, T. (1992). Transcription. In Myers, H. (Ed.), *Ethnomusicology: An Introduction*. New York: Norton, 110-152.

Fabb, N., Hall, M. (2008). *Meter in poetry. A new theory*. New York: Cambridge University Press.

Fraisse, P. (1979). *Psicologia del ritmo*. Roma: Armando.

Fussi, F., Magnani, S. (2010). *Le parole della scena. Glossario della voce del cantante e dell'attore*. Torino: Omega Edizioni.

Halle, M., Keyser, S.J. (1966). Chaucher and the Study of Prosody. In *College English*, XXVIII, 187-219.

Hirano, M. (1981). *Clinical Examination of Voice*. New York: Springer Verlag.

Isshiki, N., Olamura, M., Tanabe, M. & Morimoto, M. (1969). Differential Diagnosis of Hoarseness. In *Folia Phoniatrica*, 21(1), 9-19.

La Macina (1998). *Je se vedea le porte dell'affanno... Canti satirici e licenziosi della cultura orale marchigiana*. Recanati: MCM-Records.

La Macina (2016). *Nel vivo di una lunga storia*. Roma: Squilibri.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Lomax, A., Grauer, V. (1968). The Cantometric Coding Book. In Lomax, A. (Ed.), *Folk Song Style and Culture*. New Brunswick-London: Transaction Publishers, 34-74.

Magrini, T. (1990). Il canto monodico in Italia. In Leydi, R. (Ed.), *Le tradizioni popolari in Italia. Canti e musiche popolari*. Milano: Electa, 19-28.

Nattiez, J.J. (1987). *Musicologie générale et sémiologie*. Paris: Christian Bourgois Éditeur.

Stockmann, D. (1989). Il problema della trascrizione nella ricerca etnomusicologica. In Carpitella, D. (Ed.), *Ethnomusicologica. Seminari internazionali di etnomusicologia*. Siena: Accademia Musicale Chigiana, 209-238.

Sundberg, J. (1987). *The Science of the Singing Voice.* DeKalb, Illinois: Northern Illinois University Press.

PARTE III

SOCIALE E BIOLOGICO NELLA DEVIAZIONE DALLE ‘NORME’: BILINGUISMOE PATOLOGIA

DOI: 10.17469/O2103AISV000012

VIOLETTA CATALDO, RICCARDO ORRICO, RENATA SAVY

Phonetic variation of f_0 range in L1 and L2: A comparison between Italian, English and Spanish native and non-native speakers

This work was carried out with the purpose of investigating the use of language-specific features of pitch span and level in L2. Different languages were investigated: on the one hand, we analysed productions in L2 Spanish and English, uttered by Italian learners with different proficiency levels; on the other hand, we analysed productions in L2 Italian uttered by Spanish and English speakers. The results show a very heterogeneous situation: to some extent, learners seem to be sensitive to f_0 excursion and modulation of the L2 input they receive; however, these intonational features of Target Language speech: i) are out of non-native speaker's control, ii) do not affect all the aspects of L2 productions, and iii) present a high degree of inter-speaker variability.

Key words: Pitch range, L2 prosody, foreign language learning, transfer, topic.

1. *Introduction*

1.1 Theoretical background

Several studies state the relevance of the pitch range parameter and its relative modulations for the prosodic analysis of intonational contours (Mennen, 1998; Mennen, Schaeffler & Doherty, 2007; Busà, Urbani, 2011; Urbani, 2013). The frequency excursion of speech is an element of great importance, which can be influenced by the speaker, the sentence, and the language under investigation. In particular, the speaker's pitch range provides information about his/her biological characteristics (such as age and gender; Traunmüller, Eriksson, 1995; Mennen, *et al.*, 2007); it is also indicative of a series of individual and more specific features of the speaker, such as his/her smoking habit, regional variety, and level of education (Urbani, 2013). Additionally, para-linguistically motivated variation of utterances can be found, which expresses speakers' emotional attitudes and states of mind (i.e. surprise or boredom; Ohala, 1983; Rosenberg, Hirschberg, 2005). Finally, pitch range modulations are closely linked to linguistic functions and language-specific meanings.

A number of recent studies underline the fact that languages differ in the way they use the pitch range. Among others, Mennen *et al.* (2007) refer to the differences in the global pitch range in native speakers of different mother languages (L1s), even if presenting similar physiological features, when producing utterances in their relative L1. Additionally, Zimmerer, Juegler, Andreeva, Moebius & Trouvain (2014), present an overview of the differences among languages, stating that "languages differ with respect to the pitch range they use, their exact pitch contours and the exact placement of pitch changes" (Zimmerer *et al.*, 2014: 1037).

The strong language specificity of pitch range represents a major hurdle for Second Language (L2) learners. In the literature on L2 Acquisition, at least two different phenomena have been attested with regard to pitch range: on the one hand, several studies (Mennen, Schaeffler & Docherty, 2012; Mennen, Schaeffler & Dickie, 2014; Zimmerer *et al.*, 2014) attest the occurrence of prosodic transfer, which means that the language-specific features of pitch range are transferred from the learners' L1 to the Target Language (TL). On the other hand, other studies observe a narrower pitch range excursion compared to both the L1 and TL; in other words, L2 speech is characterized by an overall smaller modulation. This pitch range compression has been observed in a number of cross-linguistic studies investigating and comparing different pairs of languages (Finnish learners of Russian in Ullakonoja, 2007; Italian learners of English in Busà *et al.*, 2011; Busà, Stella, 2012; French learners of German and German learners of French in Zimmerer *et al.*, 2014).

Different explanations to this phenomenon have been proposed: Ullakonoja (2007), for example, finds that pitch compression is stronger in the first stages of L2 learning than in more advanced stages, showing the link between compression and general L2 proficiency. Other studies (Busà *et al.*, 2011; Zimmerer *et al.*, 2014), on the contrary, propose a link between the pitch compression and the state of insecurity of the learners, which means that a less modulated – therefore compressed – pitch contour in L2 speech might be caused by psychological rather than linguistic reasons.

1.2 Object of the study

The object of our study is an analysis of the f_0 range excursion in L2, in order to verify the occurrence of two phenomena largely attested in language learning: *transfer* and *pitch range compression*.

The study makes a comparison between Italian and two other languages (English and Spanish) in order to highlight cross-linguistic differences in the use of pitch range; then, it investigates the use of pitch range in L2 productions of different group of speakers, all having Italian as L1 or L2.

This work is part of a wider research project, which aims to monitor the development of prosodic competences in Italian-speaking learners of different foreign languages. So far, results have been obtained regarding Italian learners of Spanish (Savy, Luque Moya, 2014; Luque Moya, Savy, 2017) and English (Orrico, Cataldo, Savy & Barone, *forthcoming*). These studies have investigated a series of prosodic parameters:

- Global Profile (GP), the overall trend of the curve
- Nuclear Accent (NA), the phonetic characteristics of the nuclear prominence
- Terminal Contour (TC), the phonetic realization of the final portion of the curve
- Topic, the temporal alignment of the f_0 peak in the Topic
- Range, the overall pitch range excursion of the curve.

Generally speaking, results obtained from these studies reveal a strong presence of phonetic transfer from L1 to the TLs, with a fairly low percentage of learners producing intonational contours slightly closer to the TL or adopting other solutions (for a more detailed account, see Savy, Luque Moya, 2014; Orrico *et al.*, *forthcoming*; Luque Moya, Savy, 2017). A different result, however, was registered for the learners' pitch range in English L2, revealing a gradual approach to the TL norm. The explanation proposed is that pitch range has greater perceptual salience for non-native speakers than the other parameters taken into account.

Given the results of the previous studies, we want to test two hypotheses:

- H1: Transfer; language-specific features of pitch range will be transferred from learners' L1 to the TL;
- H2: Pitch range compression; L2 speech will show a narrower pitch range than learners' L1, regardless of the speakers' L1 and TL.

2. *Methodology*

The present section reports a description of the methodology used for our research; it includes specifications of the dataset used for the productions, the recruited speakers, the methods.

2.1 Dataset

The present study uses a limited dataset of data coming from larger corpora specifically built for the previous studies (Savy, Luque Moya, 2014; Orrico *et al.*, *forthcoming*; Luque Moya, Savy, 2017). The dataset consists of yes/no questions in the three languages in analysis; questions present a SVC (Subject-Verb-Complement) syntactical structure and a *Topic+Comment* information structure[1]. Since, as is well known, the prosodic level correlates with the syntactical level by means of pragmatic-informative structures[2], we chose to have comparable informative structures in the languages under investigation, namely the *Topic+Comment* structure.

In Italian and Spanish, the *Topic+Comment* information structure is possible for both declarative and interrogative sentences, due to a similar and free order of constituents. On the contrary, in English, the *Topic+Comment* structure seems to be possible for declarative sentences only, since interrogatives require the inversion between the verb and the subject (VSC[3], e.g. "Is the rectangle coloured?"). However, SVC syntactical structures (*Topic+Comment* information structure) appear to be possible also for interrogative sentences in spontaneous speech. For this reason, we have elicited both canonical questions (SVC) and other forms (SVC), which appear "non canonical", but totally possible and admissible in informal/spontaneous speech.

[1] Phrasal constituents constitute the units of the information structure, covering roles of pragmatic nature. The information structure of the utterances is based on the two categories of Topic and Comment. The Topic is the entity or the entities that the proposition is about, and represents the referent of the proposition (Gundel, 1988; Lambrecht, 1994); the Comment is what is said about the Topic.

[2] For a detailed review see Savy, Alfano (2016).

[3] We use SVC instead of SVO in order to emphasize the fact that the C element does not always designate a direct object.

An example of the selected kind of question in the three languages is reported:
- "La balena è grande?" in Italian
- "¿El albergue es cómodo?" in Spanish
- "The rectangle is colored?" in English.

In these structures the topic is made up of different words for each language, trisyllabic words which differ according to the position of the lexical stress; we chose three paroxytone words (*bambino*, *budino*, *balena* for Italian, *albergue*, *bodega*, *helado* for Spanish, *December*, *September*, *fiancée* for English) and three proparoxytone (*rondine*, *albero*, *dondolo* for Italian, *péndulo*, *cámara*, *águila* for Spanish, *triangle*, *rectangle*, *pullover* for English).

The dataset has been elicited by means of a reading task: speakers had to carefully and silently read the context, trying to imagine themselves in it, and read aloud only the target sentences.

Table 1 - *Example of context from the corpus used for the elicitation*

TRIANGLE
Your neighbour asks you to look after Angelo, her 4-year-old son, while she goes shopping. Angelo is sitting on his bed looking at a book with a lot of pictures and figures. He is shy and not really talkative so you start asking questions about the book to break the ice:
Is there a triangle?
He answers no, but you saw one while he was flipping through the book so you ask:
And that triangle?

2.2 Speakers

The participants of this research are speakers belonging to two groups:
- Italian native speakers (It-L1)
- Spanish and English native speakers (Sp-L1, En-L1).

Selection criteria and groups' characteristics are shown below.

Participants of the first group are 44 Italian (L1) learners of Spanish or English (Sp-L2/En-L2); they are all female students at the University of Salerno. They were selected through sociolinguistic questionnaires in order to control for some common requirements: all the students and their families had to come from and live in the area of Salerno, so as to ensure a certain degree of homogeneity for the L1 variety; the students had to never have spent significant periods abroad, and therefore never have acquired their L2 in Spanish- or English-speaking countries; the students had to specify their current academic year, so that we were able to further classify them on the basis of their linguistic competence in L2. According to the CEFR levels (Council of Europe, 1996), students were divided into three groups: group A (with an A2/B1 level of competence, corresponding to second-year students at University of Salerno), group B (B2 level of competence, corresponding to third-year students), and group C (C1 level of competence, corresponding to fifth-year students). To these, we added a group of 7 advanced learners who have studied for at least six months in a Spanish- or English-speaking country, according to the language they are studying (group E).

Speakers belonging to the second group of participants are Spanish (3) and English (3) native speakers; they have lived in Italy for at least 15 years and they are all foreign language teachers of the Italian learners.

All the participants of the two groups represent both native and non-native speakers of the research; in other words, Italian native speakers represent the Italian L1 model of reference and, since they are foreign language learners, they also constitute the groups Sp-L2 and En-L2[4].

Conversely, Spanish and English native speakers represent the model of reference of their relative L1 and constitute the group of Italian non-native speakers (It-L2). The It-L2 speakers are a subgroup of the native speakers of Spanish (2) and English (1). These speakers should be considered as a control group, since they have been selected by means of different criteria. In particular, these speakers have been selected to collect additional data to support or reject the hypothesis about pitch compression (see § 1.1): a narrower pitch excursion in all the groups would indicate that the phenomenon of compression is a typical characteristic of interlanguages.

All the speakers have produced utterances in both their L1 and L2.

The following summary table reports the exact number of the speakers, divided into the different groups, the number of productions and the total of the utterances that has been actually analysed.

Table 2 - *Number of speakers, number of productions and total of utterances analysed*

	Groups of speakers	*n° of speakers*	*n° of productions*	*total*
L1	IT L1	5	6	28
	SP L1	3	6	18
	EN L1	3	5	11
SP L2	SP L2 – group A	10	6	53
	SP L2 – group B	7	6	36
	SP L2 – group C	8	6	37
	SP L2 – group E	3	6	16
EN L2	EN L2 – group A	8	5	38
	EN L2 – group B	5	5	24
	EN L2 – group C	6	5	25
	EN L2 – group E	4	5	16
IT L2	SP > IT	2	6	12
	EN > IT	1	6	6

[4] The Italian-speaking speakers are learners either of Spanish or of English. Therefore, Sp-L2 and En-L2 are two different subgroups of the Italian native speakers.

2.3 Analysis

The first phase of analysis of the productions has been carried out through the INTSINT international labelling system (Hirst, Di Cristo, 1998); thus, the preliminary prosodic analysis of the curves aims to obtain a phonetic description of the contours. For the intonational analyses of the study we have used the software Praat for the pitch track extraction (Boersma, Weenink, 2007) and the script Prosomarker for the stylization of the f_0 curve (Origlia, Alfano, 2012).

In a second phase, pitch range has been closely examined; in this respect, the analyses have taken into consideration a subdivision of the intonational contours into three prosodic domains[5] and an investigation of pitch range for each of the domains:

- Global Profile (GP), which concerns the entire curve and the overall trend of the contour; it gives a general idea about the overall dynamics and the intonational profile.
- Topic, which requires the description of the alignment of the tonal peak in the Topic[6].

[5] These prosodic domains (see below within the text) have been taken into consideration by other models of intonational analysis, such as the IPO (Cohen, 't Hart, 1967; 't Hart *et al.*, 1990), the MAS (Cantero, 2002; Font-Rotchés, Cantero, 2009), the autosegmental-metrical models (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert & Hirschberg, 1992).
According to the IPO model, pitch movements are classified as "configurations", sequences of single constituents. The GP corresponds to the root, the only mandatory configuration of the contours; the Topic and the TC correspond to the constituents, which can precede or follow the root, respectively. The model requires the definition of a grammar, the intonational inventory of a specific language, built from the combination of pitch movements.
The MAS model carries out an analysis at the linguistic level according to which intonational contours present a structure divided in three elements, different functional constituents of intonational contours. Our individuation of three prosodic domains within intonational contours corresponds to such a tripartition. The three elements are: *anacrusis*, the portion of the curve composed by the unstressed syllables preceding the first peak; *cuerpo* ("body" or "declination"), which consists of the syllables between the first peak and the Nucleus, corresponding to the last stressed syllable of the contour; *inflexión final*, which covers the final portion of the contour, from the Nucleus to the end of the curve. This last domain is regarded as the element which best permits the definition of the contour melody. Despite the tripartition of the curves, the MAS model does not identify constituents according to their linguistic functions, but only on the basis of their phonetic characteristics.
The AM model, on the other hand, does not recognize prosodic domains, but it analyses tonal events which constitute intonational contours and signal different prosodic constituents. Therefore, a first constituent could be represented by an Intermediate Phrase by means of phrase accents, corresponding to the position of the Topic; on the other hand, the GP is described as a sequence of Pitch Accents. Finally, as is well known, the TC, the last portion of the contour, consists of both Phrase Accents and Boundary Tones, representing the edge tones of the curve and conveying the pragmatic information of the utterance.
In most intonational models, a particular status is attributed to the TC, as the portion which carries the overall information needed to distinguish between declaratives and yes/no questions.

[6] In SVC structures the Topic section is found at the beginning of the sentence. Such a constituent can present a specific intonational realisation, representing in such cases an independent prosodic unit (Crocco, Savy, 2007). For this reason, a description of the Topic as a separate prosodic domain provides an interesting parameter to be investigated.

– Terminal Contour (TC), which defines the final portion of the curve, from the nuclear accent to the end.

TC and Topic represent two local events with their precise dynamics.

In the present study, measurements of accentual events (level and excursion of the accent) are not treated, as they are not included within the objectives of this work.

The analyses conducted for each of the three domains employs the Long-Term Distributional (LTD) methodology, which considers the f_0 distribution within the speaker's performance (Patterson, 2000; Mennen *et al.*, 2012). In addition, the pitch range has been investigated according to the two dimensions of level and span; the former, also referred to as "register", concerns the overall relative pitch height, the latter pertains the excursion between f_0 maximum and minimum showing pitch variations within each utterance[7]. A series of measures for each domain has been taken in support of the general interpretation of the pitch range parameter, calculated in Herz (Hz) or semitones (ST).

– GP: f_0 maximum and f_0 minimum (Hz), mean f_0 (Hz), median f_0 (Hz), standard deviation (ST), 100% span (ST), 90% span (ST), 80% span (ST), skew (ST), kurtosis (ST). These measures serve as control data to take account of the imbalances which could characterise the curves; in particular, values of skew and kurtosis specify the degree of distribution of the curve.
– Topic: f_0 maximum and f_0 min (Hz), 100% span (ST).
– CT: f_0 maximum and f_0 minimum (Hz), 100% span (ST), slope[8] (ST).

In addition to Patterson (2000) and Mennen, Schaeffler & Docherty's (2012) measurements for pitch level, we calculated the range of f_0 values mostly exploited by the speaker, in order to distinguish it from f_0 local phenomena (peaks and valleys). In our view, this parameter represents a robust indicator for pitch level.

We also argue that the range of values most frequently exploited by a speaker in an utterance is a language-specific feature, therefore potentially susceptible of transfer.

These measurements were performed by dividing the f_0 excursion of each utterance into five regular strips, each of which contains the 20% of the excursion. The strips were numbered from 1 to 5. according to their position with respect to the median[9]:

– Strip 1: low values;
– Strip 2: mid-low values;
– Strip 3: median values (in which the median value is located);
– Strip 4: mid-high values;
– Strip 5: high values.

For each production we calculated the percentages of f_0 values per strip.

[7] A more in-depth discussion regarding these two dimensions is reported in Ladd (2008); he states that the terms of level and span are often used together "under the catch-all term 'pitch range'" as it does exist a tight covariance between the two aspects.

[8] The slope was calculated as the ratio between the span and the duration of the TC.

[9] The choice to calculate strips of pitch level which are not linked to absolute values represents a form of "rudimentary normalisation", since an actual procedure of normalisation has not yet been possible due to problems regarding the refining of the methodology.

Table 3 summarises by means of an example the total of measurements which have been conducted.

Table 3 - *Example of the list of measurements divided according to the three domains (GP, TC, Topic)*

GP level (Hz)	Max		388,13
	Min		180,04
	Mean		239,83
	Median		244,3504
	Frequency zone	from	221,65
		to	263,27
GP span (ST)	Standard Deviation		2,66
	Range (100% Span)		13,30
	90% Span		7,45
	80% Span		6,46
	Skew		0,31
	Kurtosis		0,44
TC	Span (ST)		10,83
	Slope (ST)		0,04
	Max (Hz)		388,13
	Min (Hz)		207,63
Topic	Max (Hz)		280,16
	Min (Hz)		180,04
	Span (ST)		7,66

As the last phase of analysis, the obtained data have been subjected to an analysis of variance, the ANOVA test, in order to verify their statistical significance among the different groups of speakers.

3. *Results*

In the following sections we present data resulting from the analysis. The first part of results is the outcome of a description of the L1s under investigation, while the second is the outcome of a systematic comparison between the L1 and the L2 of each group of speakers.

3.1 Native productions

Native productions are described in two steps: a detailed description of the characteristics related to pitch range based on the set of measurements explained in § 2.3; a cross-linguistic comparison aimed to identify the parameters that differentiate the three languages.

Table 4 shows the measurements performed for L1 speakers of the three languages, based on the mean of speakers' values.

Table 4 - *Measurement of the L1s based on the mean of the speakers' values*

		It	*Sp*	*En*
GP Level (Hz)	Max	298.92	312.83	415.83
	Min	177.35	159.21	196.39
	Mean	217.83	230.41	262.31
	Median	207.95	230.99	242.31
	Frequency zone	strip 1	strip 1-5	strip 1
GP Span (ST)	Standard Deviation	2.42	3.49	3.59
	Range (100% span)	8.94	11.66	12.94
	90% Span	7.77	10.50	11.32
	80% Span	6.20	8.99	9.37
	Skew	0.73	-0.18	1.02
	Kurtosis	-0.29	-0.84	0.30
TC	Span (ST)	3.73	8.42	7.97
	Slope (ST)	0.01	0.04	0.03
	Max	227.79	284.42	402.32
	Min	182.49	174.44	256.73
Topic	Max (Hz)	298.68	268.39	265.19
	Min (Hz)	187.69	167.85	205.63
	Span (ST)	7.93	8.10	4.44

Before commenting the results in Table 4 a caveat must be given.

The indexes of kurtosis provide information about the extreme values (the highest and lowest ones) of the distribution; the fact that this index is close to zero in the three languages tells us that the distribution is mesokurtic (*i.e.* similar to a normally distributed data set).

In addition, the indexes of skew (indicating the symmetry of the distribution) clearly differentiate Spanish from the other two languages; in particular, Spanish looks much more symmetric than the others, with an index of skew that is closer to zero, and median and mean that are very close to each other. This is confirmed by results about the most frequent values, which are spread out along the five strips. As for English and Italian, they have higher positive values for skew, indicating a higher concentration of the distribution within the lower values. This is confirmed again by the calculation of the frequency zone, located around the lowest values (strip 1). Furthermore, English appears to be somewhat less symmetric than Italian, suggesting an even higher concentration of values in strip 1.

Therefore, in order to avoid redundancy in the presentation of the results, we selected a subset of measurements to be presented in the following sections:

- 100% span: chosen over 90% and 80% because it most significantly signals differences cross-linguistically, since the most frequent values in Italian and English are pooled along the lowest values, which would not be taken into account by 80% and 90% span;
- Frequency zone: examined with respect to the median, since it is reported to be a better indicator for pitch level (De Looze, Hirst, 2008; D'Imperio, Cavone & Petrone, 2014);
- Topic span;

– TC span;
– F_0 peak alignment in the Topic[10].

We present for each language a stylization of the curve and the relative values, which has been built as follows: it is not an actual curve, but rather a symbolic-descriptive interpretation of data presented in Table 4 aimed at providing an overview of values and trends in the L1s.

3.1.1 It-L1

Italian L1 (Figure 1) presents a wide Topic in terms of range excursion (span: 7.9 ST); f_0 peak is aligned with the stressed syllable of the Topic.

With respect to the GP the model is characterized by a gradually declining trend from the top (298.9 Hz), set at the beginning of the curve, to the bottom (177.3 Hz), set in the final portion of the curve, with a global span of 8.9 ST.

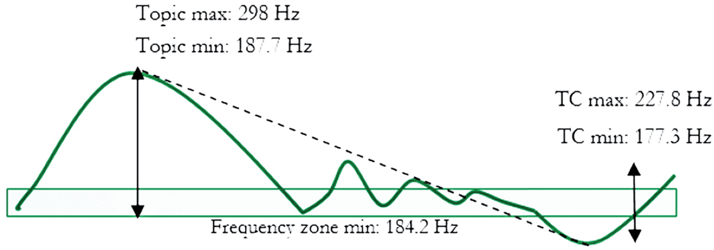
The TC shows a very narrow span if compared with the Topic and the GP domains (3.7 ST).

Moreover, from the analysis of this model, we can notice that the Top (T) of the entire contour corresponds to the f_0 peak aligned with the stressed syllable of the Topic, while the bottom coincides with the TC.

In the light of the above, for this model we can add that most of the curve appears to be modulated around low frequencies; in fact, the frequency zone corresponds to strip 1 (min: 184.2 Hz; median: 207.95 Hz).

In such a situation, the wide span of the Topic appears to be a local phenomenon.

Figure 1 - *Stylization of the curve for It-L1*



Utterance	La ba le na è gran		de?
Domain	Topic		TC
Span	7.9 ST	GP: 8.9 ST	3.7 ST
INTSINT labels	[M T		B H]
Frequency zone	Strip 1		

[10] In the present work we add preliminary results about the tonal alignment of the Topic, which are currently being investigated for another study (see Savy, Alfano & Orrico, *to appear*).

3.1.2 Sp-L1

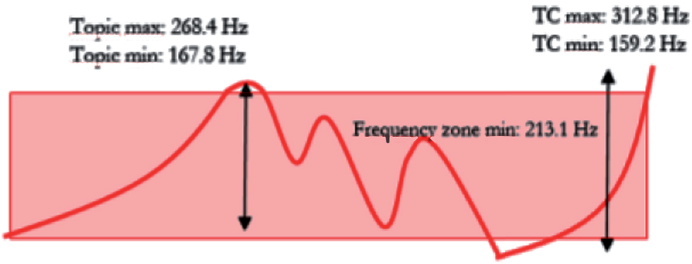
Spanish L1 (Figure 2) is characterised by a wide overall modulation. The span of the Topic measures 8.1 ST; it shows a rising trend which reaches the f_0 peak at the right boundary of the intermediate phrase of the Topic. Such a configuration is described in the literature as "flat-rising" with the f_0 peak aligned with the last syllable (see Savy, Luque Moya, 2014).

As a whole, data concerning the GP confirm the wide range excursion and the high modulation of this model (GP span: 11.6 ST; max: 312.8 Hz; min: 159.2 Hz).

The rising trend toward the right boundary and the wide range excursion observed in the Topic recurs in the TC (TC span: 8.4 ST).

As for pitch level, it appears that the f_0 values are distributed rather homogeneously along the five strips, with a relatively lower percentage of occurrences in the median strips.

Figure 2 - *Stylization of the curve for Sp-L1*



Utterance	El pen du lo	està col ga	do?
Domain	Topic		TC
Span	8.1 ST	GP: 11.6 ST	8.9 ST
INTSINT labels	[M H		B T]
Frequency zone	Strip 1-5		

3.1.3 En-L1

Figure 3 shows the stylisation of En-L1.

En-L1 shows a very narrow excursion and flat profile for the Topic domain (span: 4.4 ST) and the f_0 peak is aligned with the stressed syllable.

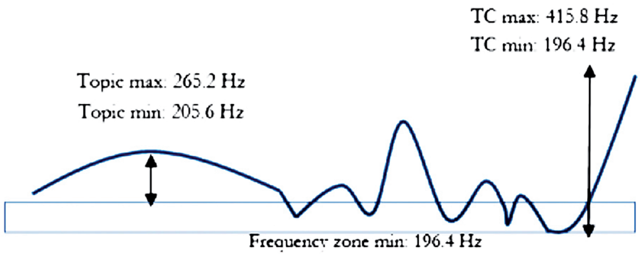
Data regarding the GP show a very wide global excursion (span: 12.9 ST; max: 415.8 Hz; min: 196.4 Hz).

The TC presents a rising trend and a very wide span (8 ST).

Because of the data regarding the narrow span of the Topic, we can state that the overall wide modulation of English concentrates on the central and the final portion of the curve, therefore mostly on the domains of GP and TC.

The frequency zone covers the low frequencies of the curve and it is set at strip 1 (min: 196.4 Hz; median: 242.31 Hz); such a situation shows that the majority of the excursion of En-L1 is made up of episodic peaks within the contour.

Figure 3 - *Stylization of the curve for En-L1*



Utterance	The rec tan gle	is co	loured?
Domain	Topic		TC
Span	4.4 ST	GP: 12.9 ST	8.0 ST
INTSINT labels	[M H		B T]
Frequency zone	Strip 1		

3.1.4 Comparison between It-, Sp- and En- L1

The following is a descriptive summary of the parameters found for the L1s, according to the parameters we chose to use for the modelling (§ 3.1).

It-L1 is described as follows:

- Topic: wide excursion (7.9 ST)
- GP: declining trend from the Topic to the TC (8.9 ST)
- TC: narrow excursion (3.7 ST)
- Alignment of the Topic f_0 peak: with the stressed syllable
- Frequency zone: strip 1.

Sp-L1:

- Topic: wide excursion (8.1 ST)
- GP: wide excursion and high modulation (11.6 ST)
- TC: wide excursion (8.4 ST)
- Alignment of the Topic f_0 peak: with the last syllable (right boundary)
- Frequency zone: strip 1-5

En-L1:

- Topic: narrow excursion and flat profile (4.4 ST)
- GP: wide excursion (12.9 ST)
- TC: rising trend and wide excursion (8 ST)
- Alignment of the Topic f_0 peak: with the stressed syllable

– Frequency zone: strip 1.

For the calculable parameters, an analysis of variance (Anova) has been conducted in order to identify statistical significance; on the contrary, for descriptive data, such as the alignment of the f_0 peak of the Topic and the frequency zone, this has been a qualitative type of analysis.

The comparison has been carried out taking into consideration on the one hand It- and Sp-L1 (see Table 5), and on the other hand It- and En-L1 (see Table 6). A comparison between Sp- and En-L1 falls outside the scope of our work.

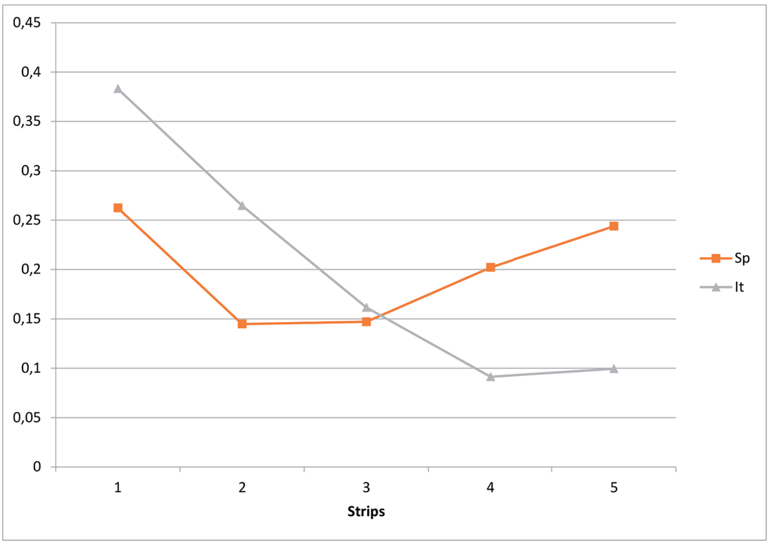
Table 5 - *Contrastive parameters between It- and Sp-L1. In bold the four statistically contrastive parameters are reported*

	It-L1	*Sp-L1*	*Significance*
Topic span	7.9 ST	8.1 ST	$p = 0.82$
GP span	8.9 ST	11.6 ST	$p = 0.001$
TC span	3.7 ST	8.4 ST	$p = 0.009$
Alignment of Topic f_0 peak	stressed syllable	end of Topic	significant
Frequency zone	strip 1	strip 1-5	significant

As can be seen, the parameters which diversify It- from Sp-L1 (Figure 4a, 4b, 4c) are:

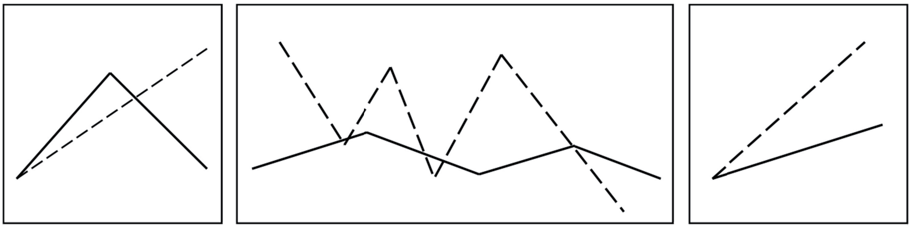
– GP span
– TC span
– Alignment of the Topic f_0 peak with the stressed syllable or to the end of the Topic
– Frequency zone (Graph 1), set at the low values of the curve (strip 1) or distributed along the five strips (strip 1-5).

Graph 1 - *Distribution of the f_0 values in It- and Sp-L1*
Frequency zone



Therefore, the difference between Spanish and Italian lies in the greater modulation of the former, which results from the more homogeneous distribution of the f_0 values along the strips.

Figure 4 (a, b, c) - *Stylized representation of the parameters that diversify It- from Sp-L1. Figure 4a shows the alignment of the Topic f_0 peak with the stressed syllable (It-L1) or with the end of the Topic (Sp-L1, in dotted line). Figure 4b shows differences with regard to both GP span and frequency zone. Figure 4c shows the TC span (narrow span for It-L1, wide span for Sp-L1, in dotted line)*



(a) (b) (c)

Differences between It-L1 and En-L1 are shown in Table 6.

Table 6 - *Parameters diversifying It- from En-L1. In bold the three statistically significant values are reported*

	It-L1	*En-L1*	*Significance*
Topic span	7.9 ST	4.4 ST	p = 0.001
GP span	8.9 ST	12.9 ST	p = 0.001
TC span	3.7 ST	8 ST	p = 0.009
Alignment of Topic f_0 peak	stressed syllable	stressed syllable	non-significant
Frequency zone	strip 1	strip 1	non-significant

The parameters of statistical significance that have been identified (Figure 5a, 5b, 5c) are:

- Topic span, that in En-L1 appears to be reduced by half compared to that of the It-L1
- GP span, since It-L1 is characterized by the narrowest global span
- TC span, since En-L1, as Sp-L1, ends with a wide excursion in that portion.

As can be seen, data regarding the f_0 peak in the Topic and the frequency zone do not show significant differences between this pair of languages. Nevertheless, Graph 2 shows the distribution of f_0 values along the five strips. It appears that both languages make an extensive use of the low values (strip 1); however, En-L1 has a much greater percentage of values in strip 1 than It-L1, which is also confirmed by the indexes of skew in the two languages (see Table 4).

Contrary to what we observed for Italian and Spanish, the differences between Italian and English do not result from the greater modulation of English, but only from a higher excursion.

Graph 2 - *Distribution of the f_0 values in It- and En-L1*
Frequency zone

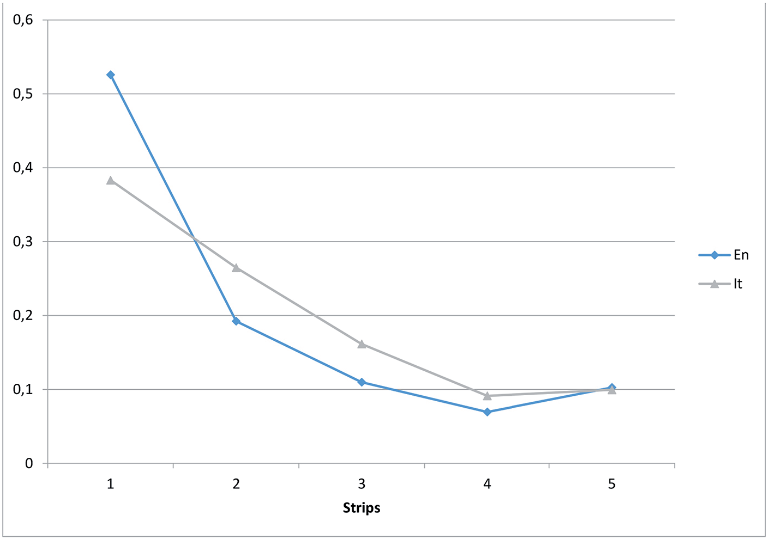
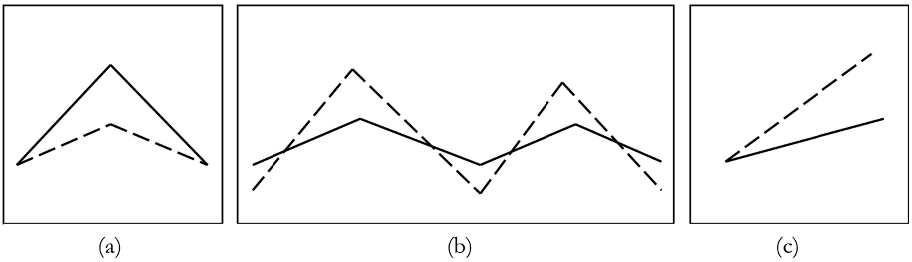


Figure 5 (a, b, c) - *Stylized representation of the parameters that diversify It- from En-L1. Figure 5a shows the Topic span (wide span for It-L1, narrow span for En-L1, in dotted line). Figure 5b shows differences with regard to the GP span (narrow span for It-L1. wide span for En-L1, in dotted line). Figure 5c shows the TC span (narrow span for It-L1, wide span for En-L1 in dotted line)*



3.2 Non-native productions

Non-native productions (L2)[11] have been compared with the modelling profiles of speakers' mother language (L1) and target language (TL) described above, in order to verify if they transfer pitch range features from L1 or if they approach to the TL.

[11] Non-native productions concern the three groups of Italian learners (group A, B, C) and the Experienced students (group E).

3.2.1 Sp-L2 (Italian speakers)

3.2.1.1 Topic

As we have seen in the comparison between It- and Sp-L1, the differences in Topic span did not result as statistically significant, while we have identified significant differences regarding the f_0 peak alignment.

As for L2 productions, we found that Italian learners of Spanish align the f_0 peak with the stressed syllable, following their L1 model (Figure 6a). Moreover, some alternative solutions in the L2 productions are found for learners belonging to level C, as they realize a displacement of f_0 peak toward different directions: in some cases toward the post-stressed syllable (Figure 6b), in others toward the pre-stressed syllable (Figure 6c), in others toward the last syllable, corresponding to the end of the Topic (Figure 6d).

These different solutions adopted by the L2 learners can be interpreted as different "attempts" made by the speakers to reproduce an intonation pattern consistent with the input they hear. At present, we do not have sufficient evidence to treat them as conscious attempts in the TL direction, or as generic efforts to drift apart from the L1 (see also Savy, Luque Moya, 2014; Orrico *et al.*, *forthcoming*).

In general, these results confirm the findings in Savy, Luque Moya (2014).

Figure 6a - *L2 production of an Italian learner of Spanish; the f_0 peak of the Topic is aligned with the stressed syllable*

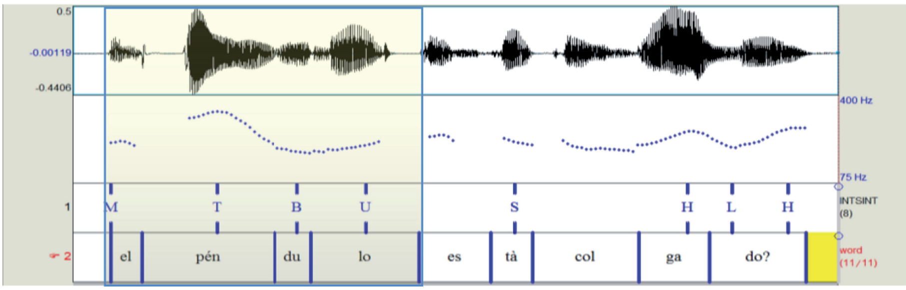


Figure 6b - *L2 production of an Italian learner of Spanish; the f_0 peak of the Topic is aligned with the post-stressed syllable*

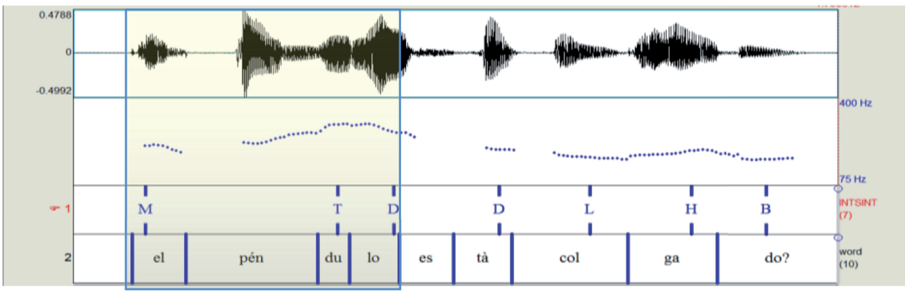


Figure 6c - *L2 production of an Italian learner of Spanish; the f_0 peak of the Topic is aligned with the pre-stressed syllable*

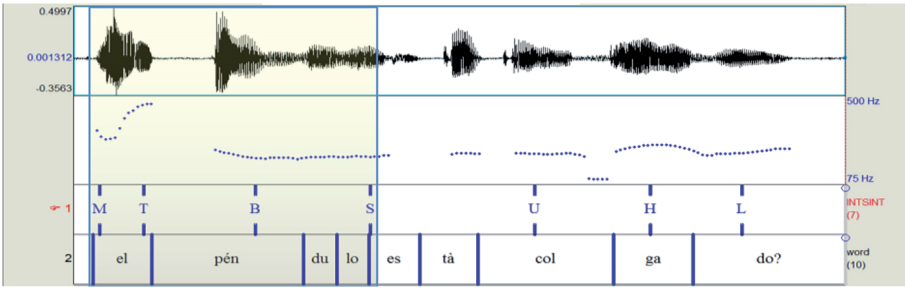
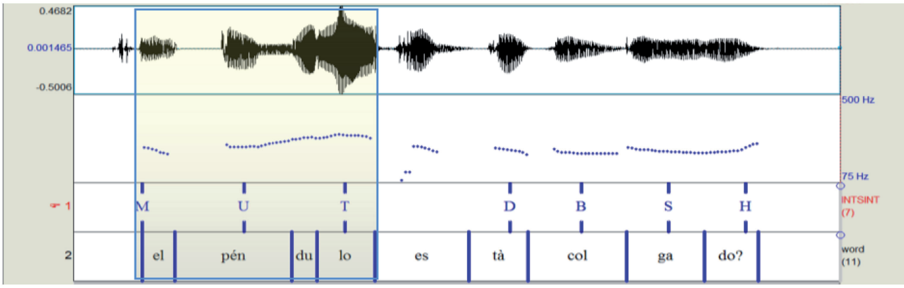


Figure 6d - *L2 production of an Italian learner of Spanish; the f_0 peak of the Topic is aligned with the last syllable of the Topic*



Our results (Table 7) show no significant differences with regard to the distribution of transfer phenomena and attempts among the three groups of learners (groups A, B, C: 64% of transfer, 36% of attempts). By contrast, the difference appears to be more significant as concerns group E, in which percentages of transfer and attempts are reversed (32% and 68%, respectively). However, if we drill down and look at the attempts by groups A, B and C we can see that Italian speakers align the f_0 peak with the pre-stressed syllable only in 2% of cases, to the post-stressed syllable in 7% of cases, to the last syllable of the Topic in a good 27% of cases.

In addition, it is worth noting a better performance of group C with respect to groups A and B in realising the f_0 peak on the last syllable of the Topic (32% vs 26% and 22%; group C reaches the performances of group E).

Table 7 - *Percentages of f_0 peak alignment within the Topic in Sp-L2 productions*

	Group A	*Group B*	*Group C*	*A+B+C*	*Group E*
Transfer	61%	70%	62%	64%	32%
Attempts	39%	30%	38%	36%	68%
Pre-stressed syllable	4%	0%	3%	2%	6%
Post-stressed syllable	9%	8%	3%	7%	31%
Last syllable	26%	22%	32%	27%	31%

Such a result as a whole can lead us to consider these "attempts" as approximations to the TL, rather than mere drifts apart from the L1, also because of the evident "improvement" which involves all the groups of Italian speakers (from group A to E) and with a certain gradual approach.

3.2.1.2 Global Profile and frequency zone

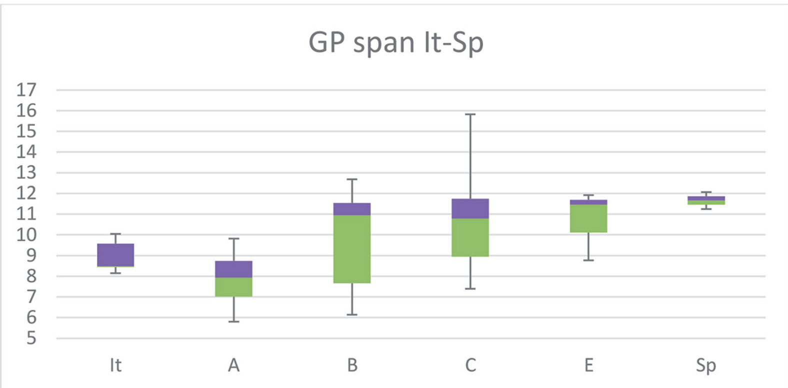
The first statistically relevant parameter according to Anova, concerns the GP span; data of the three L1s and the L2 productions of Italian speakers are reported in Graph 3. The Anova test does not report significant differences among the three groups of learners ($p = 0.05$); despite this situation, we can make some generic observations.

All the L2 productions (groups A, B, C and E) present a high degree of internal variability and significant standard deviations with respect to the L1s.

Group A reports a minor span even compared to It-L1 and little variability (SD = 1.35), showing therefore a certain internal consistency. L2 productions of groups B and C are characterised by a pretty high variability (SD = 2.53 and 2.61. respectively), which could be interpreted as a possible sign of uncertainty; for group C, with respect to groups A and B, we registered a higher mean span. As concerns group E, the low degree of variability (SD = 1.70) and higher span values show a clearer tendency toward Sp-L1.

On the whole, we can note a progressive increase of L2 performance and an interesting transition from uncertainty to consistency. Looking at the groups in sequence, we can outline a gradual improvement as the speakers' L2 competence increases.

Graph 3 - *GP span in Sp-L2 productions, compared with It-L1 (on the left) and Sp-TL (on the right)*

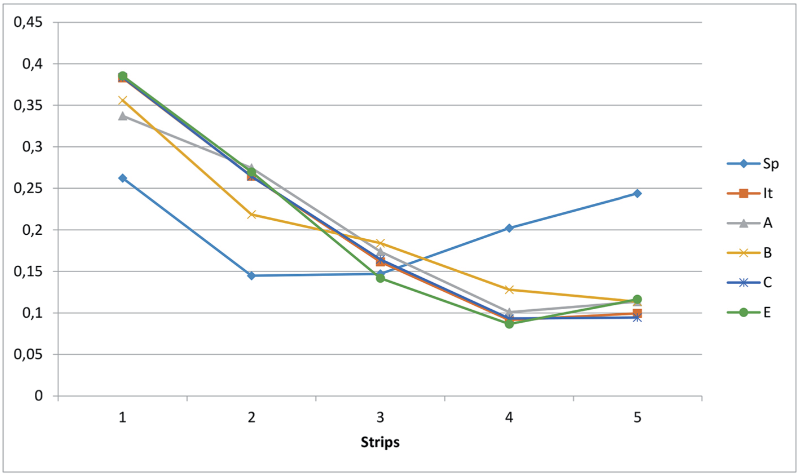


	It-L1	Group A	Group B	Group C	Group E	Sp-L1
GP span	8.9 ST	**7.9 ST**	**9.7 ST**	**10.8 ST**	**11.7 ST**	11.6 ST

As for pitch level, Graph 4 clearly shows that all the Italian speakers, regardless of the proficiency level, follow the Italian model: the highest percentages of f_0 values are located in strip 1.

Unlike the GP span, for which we registered a gradual improvement along the proficiency levels, values for pitch level is clearly transferred from the speakers' L1.

Graph 4 - *Distribution of the f_0 values in Sp-L2 productions*



3.2.1.3 TC span

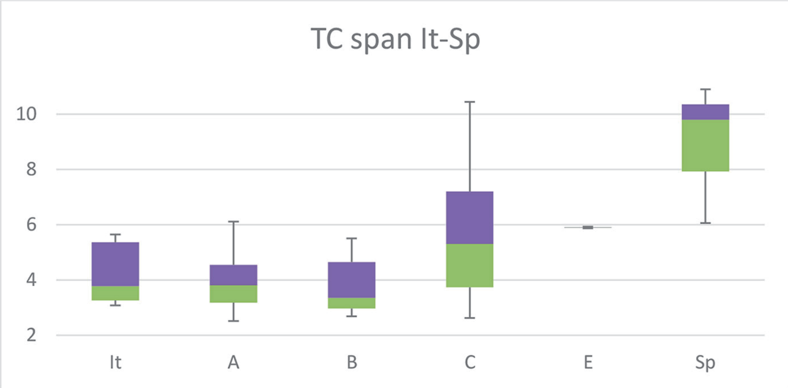
Signs of gradual improvement can be observed from the data regarding the TC span (Graph 5). It-L1 shows a very lower span compared to Sp-L1 (3.7 and 8.4 ST, respectively). Here too, there are no significant differences among the groups of Italian speakers according to Anova ($p = 0.15$); however, a description of the realizations of the TC is given in the following lines.

Groups A and B realize a very low TC span, roughly staying on the level of It-L1 (3.5 ST and 3.3 ST, respectively), with a fairly low variability (SD = 1.15 and 1.12. respectively). Better results regard data of groups C and E: they present a similar mean value of span (5.1 ST and 5.4 ST); while productions of group C show a higher variability (SD = 2.63), group E appears more systematic and regular (SD = 1.42)[12].

Results concerning the TC span as a whole, despite the partial improvements of groups C and E, do not show the evident gradualness observed in the GP domain (Graph 3). They present a situation of some controversy and a clearer evidence of transfer, as Italian speakers do not reach Sp-L1 values.

[12] As group E consists of less speakers than group C, this last result needs to be verified.

Graph 5 - *TC span in Sp-L2 productions, compared with It-L1 (on the left) and Sp-L1 (on the right)*



	It-L1	Group A	Group B	Group C	Group E	Sp-L1
TC span	3.7 ST	3.5 ST	3.3 ST	**5.1 ST**	**5.4 ST**	8.4 ST

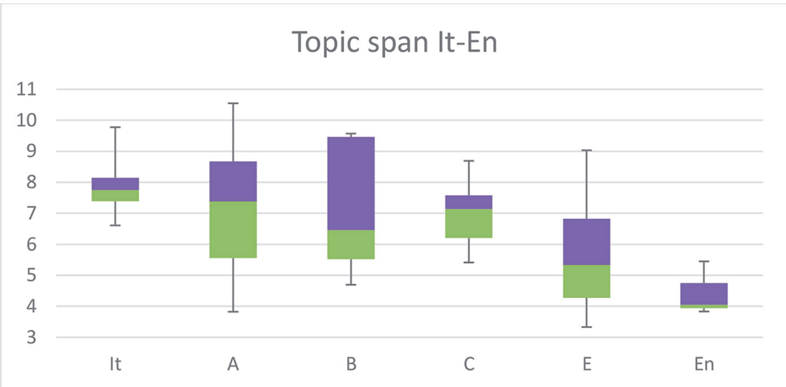
3.2.2 En-L2 (Italian speakers)

The parameters diversifying It- from En-L1 concern the span values of the three domains.

3.2.2.1 Topic span

First of all, in the realization of the Topic in the L2 productions (Graph 6), a first evidence of transfer has been found; compared to the low span of Topic in En-L1 (4.4 ST), groups of learners A, B and C realize a quite higher span (7.3 ST, 7.1 ST and 7 ST, respectively) following the model of their L1 (7.9 ST).

Graph 6 - *Topic span in En-L2 productions, compared with It-L1 (on the left) and En-L1 (on the right)*



	It-L1	Group A	Group B	Group C	Group E	En-L1
Topic span	7.9 ST	7.3 ST	7.1 ST	7 ST	**5.8 ST**	4.4 ST

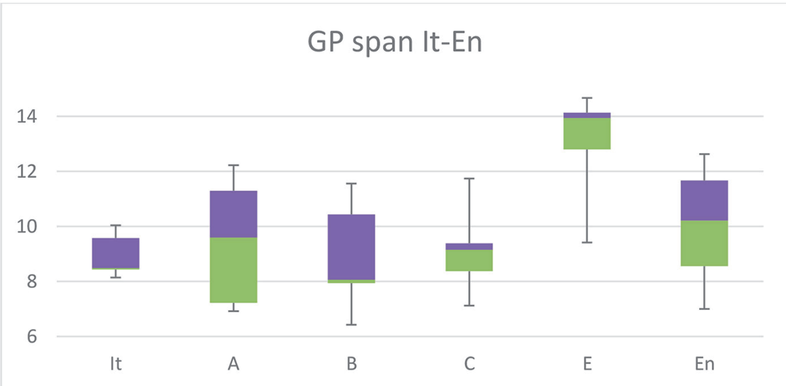
Although differences among the groups are not statistically relevant (p = 0.67), we can nevertheless comment the L2 productions.

Productions of both groups A and B show a certain degree of within-group variability (SD = 2.19 and 2.57. respectively) and a rather high span; even if on they keep the same span level of A and B, group C decreases in variability (SD = 0.99). Finally, in productions of group E the span level appears to be lowered on average (5.8 ST), although they present a high degree of variability (SD = 2.46).

3.2.2.2 Global Profile span

As for the GP span, shown in Graph 7. the analysis of variance shows that differences among the groups of Italian speakers are significant (p = 0.03). Results of groups A, B and C do not appear very different from their L1 model and show almost the same span level with different degrees of variability (SD = 1.98. 2.33 and 1.33. respectively); on the contrary, results of group E, in spite of the high internal variability (SD = 2.41), reach the TL span values (13 ST). Despite the significative differences among the groups, we cannot easily find a clear tendency in the productions by the different groups of learners.

Graph 7 - *Data of the GP span in En-L2 productions, compared with It-L1 (on the left) and En-TL (on the right)*

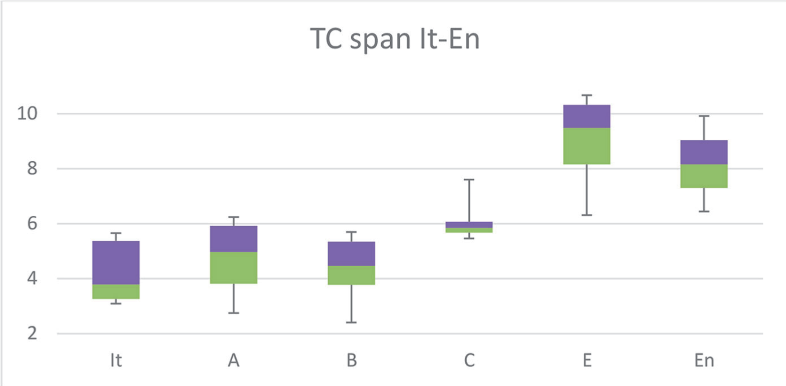


	It-L1	Group A	Group B	Group C	Group E	En-L1
GP span	8.9 ST	9.5 ST	8.9 ST	9.1 ST	**13 ST**	12.9 ST

3.2.2.3 TC span

Results concerning the TC span in En-L2 productions (Graph 8) present significant differences among the groups of Italian native speakers (p = 0.0008). As in the case of TC span in Sp-L2 productions, observations on groups A and B can be merged, as they both stay on the same level of It-L1 (4.3 and 3.8 ST, respectively). Group C occupies a mid position between L1 and TL (5.6 ST), while group E, even if with a high variability (SD = 1.96), approaches and exceeds En-L1 values (8.4 ST).

Graph 8 - *TC span in En-L2 productions, compared with It-L1 (on the left) and En-TL (on the right)*



	It-L1	Group A	Group B	Group C	Group E	En-L1
TC span	3.7 ST	4.3 ST	3.8 ST	5.6 ST	**8.4 ST**	8 ST

3.3 It-L2 (Spanish and English speakers)

In this last section of results, productions in Italian L2 are discussed. As mentioned in § 2.2. It-L2 speakers were not selected according the same criteria as the other groups: they are English and Spanish native speakers who have been living in Italy for at least 15 years, which makes them far more proficient L2 speakers than En- and Sp-L2. Therefore, the reader must bear in mind that the results presented and discussed about this group function as a way to test our H2 (see § 2.2).

Furthermore, we do not present any statistical data for this group of speakers, but only a qualitative analysis, since only few speakers agreed to participate (2 Spanish and 1 English).

3.3.1 Spanish speakers

As concerns Spanish speakers, we focused on the analysis of the Topic structure and the span of GP and TC.

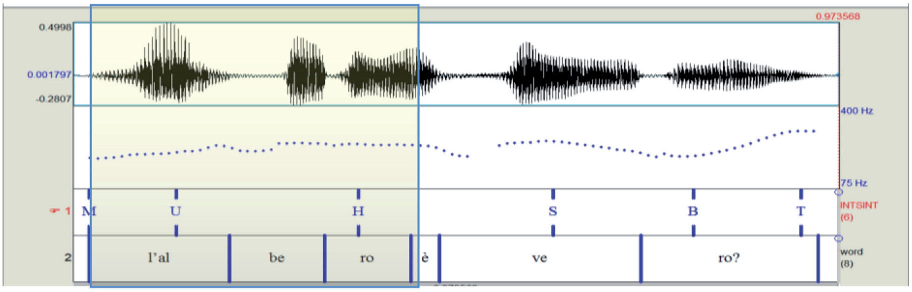
Results regarding the alignment of f_0 peak within the Topic are shown in Table 8. We draw attention to the fact that for It-L2 Spanish speakers we need to consider "attempts" the cases of f_0 peak retraction; we remind the reader that in Sp-L1 f_0 peak is normally aligned with the last syllable.

Table 8 - *Percentages of f_0 peak alignment within the Topic in It-L2 productions by Spanish speakers*

Transfer	58%	
Attempts	42%	
penultimate syllable		25%
stressed syllable		17%

As can be seen, in more than half of the cases (58%), Spanish speakers tend to maintain the alignment of the f_0 peak with the end of the Topic portion (Figure 7): such a realisation corresponds to a clear evidence of transfer.

Figure 7 - *L2 production of a Spanish speaker; the f_0 peak of the Topic is aligned with the last syllable of the Topic*



In the remaining 42% of cases, two different solutions (attempts) can be observed: in 25% of the productions, f_0 peak is aligned with the penultimate syllable, in 17% of the productions it is aligned with the stressed syllable; in this last case, the speakers realize the TL Topic configuration.

As for the GP domain, a movement toward TL is found too, but to a different degree in the two dimensions of level and span.

The span turns out to be so narrow as to overcome values of It-TL.

Table 9 - *Data of the GP span, the frequency zone and the TC span in It-L2 productions by Spanish native speakers*

	Sp-L1	*It-L2*	*It-TL*
GP span	11.6 ST	7.64 ST	8.9 ST

As for pitch level, Graph 9 shows that speakers keep exploiting all the five frequency strips homogeneously, following the pattern of their L1; however, it appears that they make a lesser use of the highest values (strips 4 and 5) and consequently increase the percentage of values in strips 2 and 3. which were not much exploited in Sp-L1.

Unlike the GP span, the TC span of L2 productions is narrower than Sp-L1, but not as narrow as It-TL (Table 10).

Graph 9 - *Distribution of the f_0 values in It-L2 productions by Spanish native speakers*

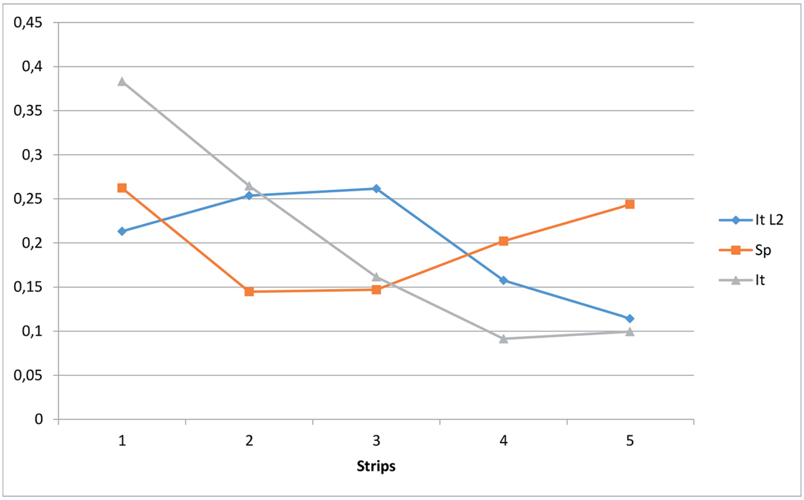


Table 10 - *Data of the TC span in It-L2 productions by Spanish native speakers*

	Sp-L1	*It-L2*	*It-TL*
TC span	8.4 ST	6.73 ST	3.7 ST

3.3.2 English speaker

Table 11 shows a summary of the productions of It-L2 by the English native speaker. Results about the Topic show a situation of partial movement toward TL, with a relative expansion of the Topic span.

As already observed for Spanish speakers, here again the span of the GP experiences a reduction, which overcomes It-TL.

A more evident improvement concerns the span of the TC, which was already particularly contrastive between English and Italian L1; productions show a clear flattening of the En-L1 span matching the values of It-TL.

Table 11 - *Topic span, GP span and TC span in It-L2 productions by a native English speaker*

	En-L1	*It-L2*	*It-TL*
Topic span	4.4 ST	6.4 ST	7.9 ST
GP span	12.9 ST	7.3 ST	8.9 ST
TC span	8 ST	3.7 ST	ST

4. *Discussion*

The data presented in § 3 can be divided in two main parts. The first one deals with the identification of the features of the L1s under consideration; the second part focuses on L2 productions, which we compared to both speakers' L1 and TL.

The analysis and comparison between pairs of L1s has pointed out some features of the curve that diversify Italian from English and Spanish (see § 3.1.4). With regard to It- and Sp-L1, we have identified four of these features. The first one is the alignment of the f_0 peak within the Topic: in It-L1 the peak is aligned with the stressed syllable of the word in position of Topic, while in Sp-L1 it is aligned with its right boundary. In the GP domain, differences concern both the span (narrow span for It-L1, wide span for Sp-L1) and the frequency zone (strip 1 for It-L1, strip 1 to 5 for Sp-L1), reflecting a greater modulation of the overall curve of Spanish compared to Italian (see § 3.1.4). The third parameter regards the TC, which in Italian has a narrower excursion than in Spanish.

The comparison between It- and En-L1 reveals that the two languages show significant differences relating to the span of the three domains under investigation. Specifically, It-L1 presents a wider Topic span and narrower GP and TC spans compared with En-L1.

In the second part of results (comparison of L2 productions with speakers' L1 and TL), we have observed that L2 learners do not behave homogeneously. Instead, different tendencies have been identified; in particular, we found:
- Sensitive parameters, which show a certain degree of learnability;
- Resistant parameters, for which we did not register any progress toward TL.

In other words, we define as "sensitive" all the parameters that are successfully reproduced by learners, or for which we registered some degree of improvement along the levels; conversely, we define as "resistant" those parameters that the learners transfer from their L1.

In Sp-L2 productions, we noted sensitivity with regard to the f_0 peak alignment of the Topic; in particular, groups C and E increase the percentages of f_0 peak alignment with the last syllable of the Topic, as in their TL structure (see § 3.2.1.1).

Table 12 - *Percentages of Topic f_0 peak alignment with the last syllable by Italian learners of Spanish*

It-L1	*Group A*	*Group B*	*Group C*	*Group E*	*Sp-TL*
f_0 peak aligned with the stressed syllable	% of f_0 peak alignment with the last syllable				f_0 peak aligned with the last syllable
	26	22	32	31	

There seems to be a certain degree of sensitivity in the range excursion of the GP domain too (see § 3.2.1.2), with an increase of span values in the different stages of the learning process, despite the high standard deviation, which testifies to a condition of uncertainty (Table 13). Such an increase of span is not accompanied by a greater modulation or pitch level variation: the frequency zone of Italian speakers in Sp-L2 productions stays the same as their L1.

Even if there is no exact continuity of change across the groups, what appears to be particularly relevant is the gradual improvement, which affects the learning process as a whole: although it concerns span values only (therefore local phenomena), such an improvement develops through the three levels of learners and group E.

Table 13 - *Values of the GP span and the frequency zone by Italian learners of Spanish*

	It-L1	*Group A*	*Group B*	*Group C*	*Group E*	*Sp-TL*
GP span	8.9 ST	7.9 ST	9.7 ST	10.8 ST	11.7 ST	11.6 ST
fr. zone	strip 1	strip 1	strip 1	strip 1	strip 1	strip 1-5
SD	0.82	1.35	2.53	2.61	1.70	4.65

As for the resistant parameters, we can call attention to TC (see § 3.2.1.3), in which the span is narrower compared to a Sp-TL (Table 14).

Table 14 - *Values of the TC span by Italian learners of Spanish*

	It-L1	*Group A*	*Group B*	*Group C*	*Group E*	*Sp-TL*
TC span	3.7 ST	3.5 ST	3.3 ST	5.1 ST	5.4 ST	8.4 ST
SD	1.20	1.15	1.12	2.63	1.42	5.31

On the contrary, in En-L2, the only sensitive parameter is the TC (see § 3.2.2.3); in this domain we have highlighted a clear improvement of span values for groups C and E, even if the former group does not reach the TL values (Table 15).

Table 15 - *Values of the TC span by Italian learners of English*

	It-L1	*Group A*	*Group B*	*Group C*	*Group E*	*En-TL*
TC span	3.7 ST	4.3 ST	3.8 ST	5.6 ST	8.4 ST	8 ST
SD	1.21	1.05	1.47	0.87	1.96	1.74

Instead, the span of Topic and GP (Table 16) preserve pitch excursion in accordance with the speakers' L1 e show therefore a persistence of transfer (It > En for the Topic, It < En for the GP; see § 3.2.2.1 and § 3.2.2.2).

Table 16 - *Values of Topic span and GP span by Italian learners of English*

	It-L1	*Group A*	*Group B*	*Group C*	*Group E*	*En-TL*
Topic span	7.9 ST	7.3 ST	7.1 ST	7 ST	5.8 ST	4.4 ST
SD	1.18	2.18888	2.570136	0.996304	2.457264	0.88
GP span	8.9 ST	9.5 ST	8.9 ST	9.1 ST	13 ST	12.9 ST
SD	0.82	1.98	2.33	1.33	2.41	2.25

Lastly, we need to consider that both the f_0 peak alignment of the Topic and the frequency zone of the GP between It- e En-L1 are similar.

With regard to It-L2 productions, it is necessary to discuss the results regarding the span of the GP. Both Spanish and English GP spans are wider than the Italian one: however, Spanish and English native speakers have produced in It-L2 a very narrow span, compared not only to their relative L1 but also to It-TL.

5. *Conclusions*

From the discussion of the results, we can conclude that there is a certain sensitivity and a selective attention to matters of range; such sensitivity:

1. lacks of intra-speaker and intra-group systematicity, therefore appears to be out of L2 speaker's control[13];
2. does not affect all the aspects of pitch range in L2;
3. does not affect the three domains in the same way;
4. presents a high degree of inter-speaker variability.

During the process of L2 learning, it seems that learners do not control the pitch modulations in the whole sentence, but rather they focus on only one of the three domains analyzed. We have only detected some attempts in L2 productions: such attempts involve a change of span (an increase for Italian speakers in Sp- and En-L2 and a reduction for Spanish and English speakers in It-L2). However, this change is mainly due to local phenomena rather than to a different kind of modulation. Indeed, pitch level values show a clear transfer from the speakers' L1s. Data regarding all the L2 speakers (Sp-, En- and It-L2) show that there is neither a generic span compression, nor a change of modulation.

Conversely, learners show a greater sensitivity to the Topic and the TC: in Sp-L2 learning, it prevails the different f_0 peak alignment, while in En-L2 learning an increase of the range excursion in the final portion of the curve has been observed. We think that the improvements found in the realization of TC could result from the attempts of the learner to realize tonal boundaries of the TL, which consequently involve changes of excursion; however, further evidence to support this hypothesis is needed. The deeper sensitivity to these parameters can be due to their distinctiveness between the speakers' L1 and TL: they are perceived by the learner, who succeeds in improving them during his learning process.

The fact that speakers do not realize all the features of the TL pitch range allow us to draw the conclusion that L2 learners never rule all the parameters together, but they concentrate on some more salient points in their productions, that seem to be highly relevant and sensitive to non-native speakers' attention. Moreover, we

[13] In previous works (Savy, Luque Moya, 2014; Orrico *et al.*, *forthcoming*; Luque Moya, Savy, 2017) the lack of control detected in the implementation of L2 productions has been attributed to the lack of prosodic and metaprosodic competence of the learners; in fact, in most cases, their guided learning process does not involve the development of prosodic skills and metaprosodic awareness which would allow them to more easily reproduce and implement L2 intonational features.

must point out that all the observed phenomena result affected by a high degree of inter-speaker variability, reflecting different degrees of awareness and different kinds of attempts of each speaker.

Taking into consideration the discussion of the data and the results of the analyses, we would conclude that pitch range variations in L2 productions seem to be conditioned by the L1/TL combination. In other words, variations produced by non-native speakers are due to actual transfer phenomena from their L1, which affect and play a role in the TL learning process. Therefore, our results do not allow us to confirm the hypothesis of pitch compression in L2 (H1), as found by previous study, but we can accept the hypothesis that pitch range features of the curve are transferred from L1 to L2.

In the light of the conclusions of this work, in the future we intend to conduct a series of perceptual experiments in order to examine in depth and verify the perceptual salience of the range parameter. Furthermore, we believe we should investigate to a much greater extent the inter-speaker variability, relating it to a more detailed analysis of the TL input.

Bibliography

Boersma, P., Weenink, D. (2007). *PRAAT: Doing Phonetics by Computer*. http://www.fon.hum.uva.nl/praat/

Busà, M.G., Stella, A. (2012). Intonational variations in focus marking in the English spoken by North-East Italian speakers. In Busà, M.G., Stella, A. (Eds.), *Methodological Perspective on Second Language Prosody – Papers from ML2P 2012 (Methods in L2 Prosody)*. Padova: CLEUP, 31-35.

Busà, M.G., Urbani, M. (2011). A cross linguistic analysis of pitch range in English L1 and L2. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong, China, 380-383.

Cantero Serena, F.J. (2002). *Teoría y análisis de la entonación*. Barcelona: Ed. Universitat de Barcelona.

Cohen, A., 't Hart, J. (1967). On the anatomy of intonation. In *Lingua*, 19, 177- 192.

Council of Europe (1996). *Modern languages: learning, teaching, assessment. A common European framework of reference*. Draft 2 of a framework proposal. Strasbourg: Council of Europe.

Crocco, C., Savy, R. (2007). Topic in dialogue: prosodic and syntactic features. In *Proceedings of 8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*. Antwerp, Belgium, 114-117.

De Looze, C., Hirst, D.J. (2008). Detecting key and range for the automatic modelling and coding of intonation. In Barbosa, P.A., Madureira, S. & Reis, C. (Eds.), *Proceedings of Speech Prosody 2008.* Campinas, Brazil, 135-138.

D'Imperio, M., Cavone, R. & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. In *Frontiers in Psychology*, 5, 1226.

Font-Rotchés, D., Cantero Serena, F.J. (2009). Melodic Analysis of Speech Method (MAS) applied to Spanish and Catalan. In *Phonica*, 5, 33-47.

Gundel, J.K. (1988). *The role of topic and comment in linguistic theory: Outstanding Dissertations in Linguistics*. New York: Garland.

Hirst, D., Di Cristo, A. (1998). *Intonation systems: a survey of twenty languages.* Cambridge: Cambridge University Press.

Ladd, D.R. (2008). *Intonational Phonology, 2nd ed.* Cambridge: Cambridge University Press.

Lambrecht, K. (1994). *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Luque Moya, J.A., Savy, R. (2017). Un método para la enseñanza asistida de la prosodia en aprendientes italófonos de ELE. In Arroyo, I., Musto, S. & Ripa, V. (Eds.), *Sistema, codificación e interpretación. aproximaciones al análisis de la lengua y a su didáctica en una perspectiva metaoperacional*. Monografías marcoELE: núm. 24. 192-217.

Mennen, I., (1998). Can language learners ever acquire the intonation of a second language? In *Proceedings of STiLL-1998*. Marholmen, Sweden, 17-20.

Mennen, I., Schaeffler, F. & Dickie, C. (2014). Second language acquisition of pitch range in German learners of English. In *Studies in Second Language Acquisition*, 36 (2), 303-329.

Mennen, I., Schaeffler, F. & Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of German and English speakers. In *Proceedings of 16th International Congress of Phonetic Sciences (ICPhS XVI*). Saarbrücken, Germany, 1769-1772.

Mennen, I., Schaeffler, F. & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. In *The Journal of the Acoustical Society of America*, 131(3), 2249-2260.

Ohala, J.J. (1983). Cross-language use of pitch: an ethological view. In *Phonetica*, 40(1), 1-18.

Origlia, A., Alfano, I. (2012). Prosomarker: a prosodic analysis tool based on optimal pitch stylization and automatic syllabification. In Calzolari, N., Choukri, K., Declerck, T., Ugur Dogan, M., Maegaard, B., Mariani, J., Odijk, J. & Piperidis, S. (Eds), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul, Turkey, 997-1002.

Orrico, R., Cataldo, V., Savy, R. & Barone, L. (*forthcoming*). Transfer, Fossilization and Prosodic Drift in Foreign Language Learning. In Savy, R., Alfano, I. (Eds.), *La fonetica nell'apprendimento delle lingue*.

Patterson, D. (2000). A linguistic approach to pitch range modelling. PhD Dissertation, University of Edinburgh.

Rosenberg, A., Hirschberg, J. (2005). Acoustic/prosodic and lexical correlates of charismatic speech. In *Proceedings of 9th European Conference on Speech Communication and Technology (Interspeech'2005 – Eurospeech)*. Lisbon, Portugal, 513-516.

Savy, R., Alfano, I. (2016). La "richiesta di informazione" nei dialoghi task-oriented: aspetti di interfaccia prosodia-pragmatica in prospettiva intra- e inter-linguistica. In Elia, A., Iacobini, C. & Voghera, M., (Eds.), *Livelli di analisi e fenomeni di interfaccia*, *Atti del XLVII congresso internazionale di studi della Società di Linguistica Italiana*. Roma: Bulzoni, 205-230.

Savy, R., Luque Moya, J.A. (2014). Aspectos prosódicos de las interrogativas en aprendientes italianos de ELE. In *Anejos n° 7 de Normas- Revista de Estudios Lingüísticos Hispánicos*, 283-296.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992). ToBI: a standard for labeling English prosody. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 92)*. Banff, Canada, vol. 2, 867-870.

'T Hart, J., Collier, R. & Cohen, A. (1990). *A Perceptual Study of Intonation: an Experimetal Phonetic Approach*. Cambridge: Cambridge University Press.

Traunmüller, H., Eriksson, A. (1995). *The frequency range of the voice fundamental in the speech of male and female adults*. Unpublished manuscript, Department of Linguistics, University of Stockholm, (online at URL: www.ling.su.se/staff/hartmut/f0_m&f.pdf).

Ullakonoja, R. (2007). Comparison of pitch range in Finnish (L1) and Russian (L2). In *Proceedings of 16th International Congress of Phonetic Sciences (ICPhS XVI*). Saarbrücken, Germany, 1701-1704.

Urbani, M. (2013). The pitch range of Italians and Americans. A comparative study. PhD Dissertation, Università degli Studi di Padova.

Zimmerer, F., Juegler, J., Andreeva, B., Moebius, B. & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In Campbell, N., Gibbon, D. & Hirst, D., (Eds.), *Proceedings of Speech Prosody 2014.* Dublin, Ireland, 1037-1041.

DOI: 10.17469/O2103AISV000013

VINCENZO GALATÀ, GIULIA ANGONESE, CLAUDIO ZMARICH

Italian as L2 in Romanian pre-schoolers. Evidence from a perception and production task

Following state-of-the-art research on the linguistic development of children acquiring a second language, the proposed study aims at deepening our understanding of the mechanisms that drive the phonetic and phonological acquisition of Italian as L2 in migrant Romanian pre-school children. By shedding some light on the difficulties faced by a group of 17 Romanian pre-schoolers in Italian L2 production and discrimination tasks, our aim is to provide evidence of their phonological development to support potential educational intervention activities that suit the needs of multilingual children. Moreover, though on a limited sample of children, we provide important evidence that needs to be taken into account when carrying out diagnostic language evaluations and treatments on bilingual children born to immigrant parents[1].

Key words: Italian L2 acquisition, Romanian pre-school children, non-word repetition, non-word discrimination, phonological skills.

1. *Introduction*

The growing rate of immigration in Italy in the last few decades poses serious challenges to the diagnostic evaluations and treatments for bilingual children born to immigrant parents with specific speech and language disorders. While such aspects are out of the scope of our research, understanding how the phonetic and phonological properties of a language affect the acquisition of a second language (L2) in young migrant children remains of crucial importance.

In L2 acquisition, one of the factors influencing the L2 proficiency is the so-called "L1 transfer", namely the cross-linguistic influence, going from the mother tongue to the L2. The L1 influences the acquisition of different linguistic domains of the L2, and this influence is more evident in phonology, phonetics, prosody and lexicon/semantics than in other domains such as syntax and morphology, especially during the initial steps of the acquisition process and mainly in children (Mioni, 2005: 36). In different words, the distance between L1 and L2 can facilitate the acquisition process, or interfere with it. The structural proximity between the two languages can ease the learning process, but it can favour, at the same time, the interference between

[1] Authorship note. Although the paper is the result of a joint collaboration and discussion between the three authors, the main research idea is to be attributed to VG and CZ; the data collection to VG, the data preparation, analysis and results' discussion to VG, CZ and GA. Finally, the paper drafting and preparation is to be attributed to VG.

the two languages; on the other hand, the structural distance requires the acquisition of unfamiliar structures which require more time to be acquired, but reduces the interference (cfr. Anderson, 2004; Mioni, 2005; Valentini, 2005).

We know from literature on adult L2 acquisition that, when learning a new language, speakers face situations in which phonemes of their first language (L1) are absent in the new L2 and *viceversa*; similarly, situations in which positional variants in L1 are phonemes in L2 (or *viceversa*) may also occur. Not to mention other differences originating from the relative frequency of occurrence in each of the two languages, the phonotactics (e.g. syllable position; Hansen, 2004), the prosodic domains larger than words (Young-Sholten, 1994) and stress patterns (Archibald, 1994). This is, for example, the case of adult members of migrant families in which the mother tongue is the main language spoken at home and in everyday communication, and in which the second language is learned only in their adulthood as a foreign language, hence emerging poorly intelligible, syntactically and lexically poor and not fluently spoken (MacKay, Flege, 2004).

This cross-linguistic influence produces acquisition strategies based on 3 factors (Topoliceanu, 2011): "congruency", that is identifying structures and elements which are similar in the two languages; "correspondence", i.e. relating norms and frames of the two languages; "difference", namely detecting in the new language, components which do not exist in the L1.

To explain this "L1 transfer", the Contrastive Analysis Hypothesis (CAH, first presented by Lado, 1957) ascribed to L1 an inhibitive role in L2 speech perception, processing and production. The CAH predicted that those aspects of the L2 sound system that are similar to the L1 will be easy to acquire, while other aspects that are different from the L1 will be more difficult (cfr. Zampini, 2008: 219). However, as pointed out by Zampini (2008), the role of L1 in L2 phonological acquisition is not so straightforward and many other factors affect L2 phonological acquisition thus limiting or reinforcing the role of the L1. These factors can be attributed to linguistic and extra-linguistic characteristics. Among the linguistic characteristics, factors such as the level of markedness (Eckman, 1977) and/or the acoustic closeness of the trait to be acquired can hinder a correct acquisition (because the L2 trait is assimilated to an existing similar, but not equal, L1 trait); whereas relative distance paradoxically makes differences more evident and recognizable, and then more learnable (Flege, 1995; Best, Tyler, 2007). Among the extra-linguistic characteristics, factors such as age and social context are important. These factors interact, as predicted by the Ontogeny Model of Phonological Development (Major, 1986), in the production errors of toddlers acquiring L2 (these errors derive both from normal developmental processes of acquisition, which have a universal nature, and from L1 transfer). In fact, the age of first exposure to L2 is another aspect among the most recognized and significant factors contributing to L2 proficiency (SLM model, cfr. Piske, MacKay & Flege, 2001; Flege, MacKay, 2010; Wei *et al.*, 2015). The concept "earlier is better" in L2 acquisition has been convincingly documented since the last century's '80s (as stated by Flege, MacKay, 2010). Focusing on phonological learning specifically, early second language learners usually display more native-like patterns in L2 perception than late learners (Archila-Suerte, Zevin & Hernandez, 2015), and age of first exposure to L2 strongly influences perception and production abilities in L2, such as perceptive efficacy, degree of foreign accent and speech accuracy (Flege, 2007). The real importance of this factor, however, is its alleged association with "causative" variables thought likely to vary with age of exposure, such as variation in neural maturation (itself thought to influence degree of neural plasticity), the state of development of native language phonetic categories, and the kind and/or amount of L2 input typically received (Flege, MacKay, 2010). Moreover, it has been shown that age of acquisition modulates functional neural activity in several aspect of language processing, with various studies suggesting that age of acquisition of L2 may have important effects on the functional organization of the language system in bilingual brains (see Wei *et al.*, 2015 for a review).

Another determinant of the level of proficiency attained in a second language to be taken into account is the concept of "the more experience, the better": that is, time and amount of exposure to a language (Sebastián-Gallés, Bosch, 2005). It is known that learning language-specific properties requires exposure to relevant speech in learning environments that support interpretation and social interaction; the interesting question is how the environment aids or hinders specific learning outcomes (Carroll, 2017a). Given that experience matters to bilingual language acquisition, it is also necessary to deepen the investigation on how exposure guides specific learning mechanisms towards the solution to a well-defined learning problem (Carroll, 2017b).

The situation described so far applies to the young Romanian children in the present study. These children are systematically exposed to L2 (Italian) only as they enter kindergarten and where they acquire the second language "in the environment where the language being learned is spoken" (Gass, Glew, 2008: 266; see also Bettoni, 2001; Ritchie, Bathia, 2009). On the segmental level, these young children need to cope with cognitive challenges (categorization of phones and their possible combinations), as well as with motor control abilities (learning new articulatory habits). If we then would take into account also the prosodic level, which again is outside the scope of the present contribution, the number of difficulties would increase enormously (see Zampini's review, 2008; Mennen, 2015).

Notwithstanding the enormous difficulties outlined above, when entering kindergarten at age 3-6, these children are in the optimal condition to become simultaneous bilinguals by learning their mother tongue (their parents' language spoken at home) and the second language (the language of the country they live in). The age period from 3 to 6 represents, from a neurobiological point of view, a fruitful and privileged time-window for the acquisition of languages (Bates, 1995; Birdsong, 1999; Ioup, 2008).

As far as the authors are aware of, the present study is one of the firsts to be carried out on a group of Romanian pre-schoolers learning Italian as L2 by addressing specifically their phonetic and phonological development. In addition to this, studies tackling the phonetic and phonological development of Italian pre-schoolers are also very limited (e.g. Zanobini, Viterbori & Saraceno, 2012, who analysed the phonological development of children between 36-42 months of age in relation to other aspects of language acquisition; for a review, see also Zmarich, Pinton & Lena, 2014).

2. *Objectives*

Our main aim is to assess the phonetic-phonological development of a group of pre-schoolers born from Romanian parents and attending the Italian kindergarten in order to contribute to the very sparse literature on the acquisition of Italian in pre-schoolers and children learning Italian as L2.

The research question we try to answer here is whether differences between the phonetic-phonological system of Romanian (ro) and Italian (it) may influence the perception and the production of Italian as L2 in Romanian children. We hypothesize that Romanian children learning Italian as L2 by entering the kindergarten may encounter major difficulties in the production and perception of specific sounds not present in their home language (L1 Romanian). To this end, we specifically focus on the consonantal system of the two languages the children are exposed to.

The Romanian language is very similar to Italian in different linguistic domains (Topoliceanu, 2011), but for the purpose of our study, we limit ourselves to highlight that on the phonological level, with respect to Italian, consisting of 23 consonants, the Romanian language counts 22 consonants (and various allophonic variants; more details in Chitoran, 2001) and lacks the phonemes /ɲ/ (found in words like <gnomo>, Eng. 'goblin'; <ragno>, Eng. 'spider'), /ʎ/ (as in words like <gli>, one of the alternative forms for Eng. 'the'; <aglio>, Eng. 'garlic') and /dz/[2] (quite infrequent, as in words like <zero> Eng. 'zero'; <azzurro>, Eng. 'blue') as well as the consonants' gemination process with distinctive function (e.g. word pairs like <papa>, Eng. 'pope' vs <pappa>, Eng. 'food'). Despite this strong similarity between the two systems, we will show how the acquisition path of Italian L2 in our group of Romanian children may be markedly diverse from their age-matched Italian peers.

In doing this, we will analyse the error patterns and fluency measures in the children's productions as these represent major evidence for the emergence of a phonological system (for Italian: Zanobini *et al.*, 2012; for British-English: Dodd, Holm, Hua & Crosbie, 2003; for a comparison between English and Dutch speech sound development: Priester, Post & Goorhuis-Brouwer, 2011).

We will also attempt to verify possible relationships between direct language measures of language ability and phonological skills with selected measures of language exposure[3] collected by means of parent's reports via questionnaires.

[2] In Italian, these three consonants undergo gemination in intervocalic position.

[3] Although this last point has become the topic of a very recent and vivid debate (see Carroll, 2017a; Carroll, 2017b for a more detailed discussion).

3. *Materials and methods*

3.1 Materials

A Non-Word Discrimination Test (NWDT) and a Non-Word Repetition Test (NWRT) based on the same CV.CV non-words targeting, among others, also Italian phonemes absent in Romanian (that we consider as "critical" consonants), were administered to a group of children speaking Romanian as L1. We choose non-words because of "the potential usefulness of processing-based measures generally, and nonword repetition tasks specifically, in providing culturally nonbiased assessments of linguistic abilities" (Weismer, Tomblin, Zhang, Buckwalter, Chynoweth & Jones, 2000: 874). In fact, the non-words repetition (or discrimination) is a task of speech production (or perception) less dependent from previous lexical knowledge, thus engaging only the phonological memory and/or the articulatory system, but not the lexical/semantic system (Baddeley, Gathercole & Papagno, 1998).

The NWDT is based on the opposition, in ′CV.CV non-word pairs, of Italian consonantal phonemes: the difference among the contrasting phonemes involves one phoneme in each pair whose opposed one differs for a maximum of 3-4 distinctive features (Jakobson, 1966; Muljacic, 1972). Examples of nonword pairs are: /nabi, ɲabi/, /ʎaki, laki/, /nafi, napi/, /daffi, dafi/, /tisa, tissa/, /biti, bitti/ etc.. The NWDT also includes non-word pairs contrasting for sonority and gemination. The purpose of the NWDT is to determine whether the Romanian L1 children could perceive any difference in the two words by asking them to indicate for each proposed pair of non-words whether the two non-words are same or different.

In the NWRT the child is asked to listen and to repeat single non-words in a sort of "parrot game" (for more details on the two tests: Galatà, Zmarich, 2011a, 2011b). The same pre-recorded non-words, uttered by a professional speech therapist, have been used in both the NWDT and the NWRT.

In order to collect more natural speech samples to be added to the phonetic and phonological analysis as well as to evaluate more ecological parameters usually found in spontaneous speech, an additional Narrative Task (NT) has been administered to the children in order to elicit their oral production. This additional task included two painted story strips containing six scenes each that the child had to describe orally. These were "The Nest story" (Paradis, 1987; Mauro, Burelli, 2002; Marini, Marotta, Bulgheroni & Fabbro, 2015) and the "Flower pot" story (Huber, Gleber, 1982; Marini, Andreetta, del Tin & Carlomagno, 2011). For this task, the children were asked to describe the two stories to the experimenter[4].

[4] It is worth mentioning here that we administered the two stories with the aim to collect as much spontaneous speech as possible. We tried to stimulate the children to continue with the story description even in those cases where they provided very short descriptions or even when they did not want to tell them at all. This attempt of ours had, on one side, the desired effect of getting the child engaged to talk thus allowing us to collect more spontaneous speech; on the other side, this operation prolonged pauses and recording time drastically. For this second reason, for example, we could not compare the Narrative Fluency (NF) index as number of words per minute as proposed by Marini *et al.* (2015).

Additionally, the children's parents were further asked to fill-in a questionnaire. The parents' responses provided with the questionnaires (Qs) addressed the child's linguistic biography and the family's socio-demographic factors expected to be relevant for reaching fluency in Italian L2 (Galatà, Zmarich, 2011a, 2011b).

3.2 Participants

The current study focuses on a group of 17 bilingual children (Romanian-Italian) aged between 61 and 83 months (mean age 5;10) attending Italian kindergartens in the town of Padova in the Veneto area (a region in the north-eastern part of Italy). The children are all born from Romanian parents, having Romanian nationality, immigrated to Italy. The children received Romanian language input at home in a family context through their parents and relatives. On the other hand, Italian language input can be considered as a mixed input consisting of a formal one, as the one provided in an educational setting (at kindergarten) taking place only as they enter the Italian school system (between 36 and 48 months of age), and of an informal one provided in the family setting (at home). Technically speaking, the participants may be considered as "early sequential bilinguals" (Marini, 2014; Meisel, 2004; Bettoni, 2001; McLaughlin, 1978).

Table 1 - *Distribution of the ROU participants according to age-group, mean age in months and gender*

Group	*Age group (in years)*	*Mean age in months (SD)*	*Males*	*Females*	*Total*
1	5;0-5;5	63.0 (1.6)	2	3	5
2	5;6-5;11	69.6 (2.2)	3	2	5
3	6;0-6;5	73.0 (1.4)	1	3	4
4	6;6-6;11	82.0 (1.0)	1	2	3
		Total	7	10	17

Some of the children were assessed at their homes in a quiet room, while some others were assessed during the kindergarten hours in a separate room. None of the children had, as reported by the parents' questionnaires, hearing or language problems. Table 1 provides an overview of the distribution of the children involved in the present study. As shown in the table, the children have been grouped according to four age-groups (expressed in years and months) in order to compare their performances to the above mentioned tests with those from Italian peers provided in the BVL 4-12 (Marini *et al.*, 2015).

3.3 Procedures

Both the NWDT and the NWRT have been administered via PRAAT's ExperimentMFC procedure by randomizing the items within each test (Boersma, Weenink, 2016). A pre-test session preceded both tests to make sure the child understood the assignment. The presentation of the stories in the NT has been randomized as well.

In order to assess the children's reliability in answering the NWDT, to verify the children understood and recalled the assignment while doing the test and to exclude they were answering by simple guessing, extra pairs prompting for a very clear "same" as well as a "different" response were included as control items in the test. All of the children answering correctly to at least two thirds of the control items were considered as reliable: all of the children included in the present study were reliable. The children's answers (same/different) to the NWDT have been exported, collected and coded in a unique matrix. The analysis has been carried out both on the overall mean of the correct responses for each child as well as on the mean of the correct responses grouped by age. For ease of comparison between the NWDT's results and those from the NWRT and the NT we will however report the percent errors.

The NWRT, the NT and all of the interactions between child and experimenter were audio recorded for offline transcription, coding and analysis.

For the NT and NWRT, the speech samples of each child have been first acoustically segmented and orthographically transcribed via PRAAT. For the transcription of the speech samples and for the identification of the utterances[5] in the NT, the criteria proposed by Marini *et al.* (2011) and Marini *et al.* (2015) have been followed in order to be able to compare the current data with those present in the BVL 4-12 (Marini *et al.*, 2015).

The pre-processed speech samples for the NT and the NWRT have been further coded via PHON (Rose, MacWhinney, 2014; Rose, Stoel-Gammon, 2015) by matching the children's actual productions to the expected adult target (e.g. IPA actual ~ IPA target). This allowed us to analyse the children's speech samples according to the classical analysis in terms of error patterns and allowed us to qualify and quantify all types of errors and phonological processes (see next paragraphs). The extraction of the raw measures to be used for the calculation of the direct measures of language ability has been carried out both in PRAAT via scripting and in PHON via specific queries.

3.4 Direct measures of language ability and proficiency

After transcribing and coding the data as described in the previous paragraph, different direct measures of speech and language ability were calculated separately for the NT[6] and the NWRT. These are:

[5] There is so far still no consensus on the criteria that define and allow the segmentation of a conversation into utterances. Most of the studies on young talkers, whose speech is highly fragmented and low in fluency, adopt a quantitative approach to temporally identify the utterances of a speech sample: segments separated by pauses below 2 seconds are considered as a single utterance. An "utterance" can be thus defined as a sequence of words, even without having grammatical structure, preceded and followed by a silence (pause) or a conversational turn (Devescovi, D'Amico, 2001).

[6] Where otherwise stated, the indexes calculated for the NT have been calculated without distinction between the two stories.

- Units produced (UP) indicates the total number of words in units (well-formed and not, disfluent words included)[7] produced by each child in the NT.
- Verbal Productivity (VP) indicates the total number of phonologically well-formed words (in units) produced by each child in the NT and it is therefore different from the total number of words (UP). This measure accounts for the child's linguistic development and abilities and it is expected to grow as a function of age and literacy. The VP index has been calculated separately for both stories in the NT and the average values for each age group have then been compared with normative data from age-matched Italian children in the BVL 4-12 (Marini *et al.*, 2015).
- Mean Length of Utterance in words (MLU_w) provides a measure of morpho-syntactic complexity and represents a predictive index for the linguistic abilities of bilingual children. MLU_w is expressed as number of phonologically well-formed words as measured by VP, divided by the number of utterances as defined in note 5 (see also Marini *et al.*, 2011: 1380). As for VP, MLU_w has been calculated separately for both stories in the NT and averaged values per age group have been compared with normative data from the BVL 4-12.
- Articulation Rate (AR) is "*a measure of how quickly a speaker transmits information when speech continuity is uninterrupted*" (cfr. Logan, 2015: 124) and thus represents a measure of articulatory rapidity. AR is expressed as number of syllables per second (syll/sec) of the uttered sequence in relation to the duration of the entire phonic chain (delimited by empty pauses) excluding hesitations and disfluencies (cfr. Zmarich, Magno Caldognetto & Ferrero, 1996)[8]. This index is computed for the NT only. Due to the lack of reference data for age-matched Italian peers, we compare our data with those reported in Logan, Byrd, Mazzocchi & Gillam (2011) for a group of typically developing children from the United States (mean age 6;10, range 5;6-7;7) keeping in mind though that we are dealing with two languages with different temporal organization[9].
- Percentage of Disfluencies (%Disf) represents the number of disfluencies on the total number of words produced. Following the classification by Yairi, Ambrose (2005) we considered as disfluencies the repetition of parts of words, repetition of monosyllabic words (excluding repetition of words with emphatic intent), dysrhythmic phonation (lengthening or blocking of speech sounds), interjections, repetition of polysyllabic words and of utterances, revisions and interrupted utterances.

---

[7] Contracted articles as well as contracted prepositions have been split and counted as two separate words following Marini *et al.*'s (2015) instructions: e.g. Ita. "del" = "di" + "il"; Ita. "col" = "con" + "il"; Ita. "nella" = "in" + "la" etc.

[8] This index is different from Marini *et al.*'s (2015) speech rate where it is calculated as the number of phonologically well-formed words produced per time of narration in minutes (words per minute).

[9] The authors also report on results from previous studies on the same index (see Logan *et al.*, 2011: 131, Table 1) stating that for typically developing children from 3 to 6 of age, values in the range from 2.9 to 4.2 syllables per second are found.

- Percentage of Incorrect Words (%IW), calculated according to Zanobini *et al.* (2012) as the number of incorrect words with reference to the adult target divided by the total number of words produced. This index has been further split in $\%IW_{str}$ accounting for incorrect words referred to the phonotactic structure (simplification of the syllabic structure) and $\%IW_{sys}$ accounting for incorrect words referred to the phonological system (simplification of phonemic oppositions).
- Percentage of Phonological Errors (%PE), expressed as the number of substituted or omitted phonemes on the total number of phonemes present in the phonological target. We preferred this type of measure and not the %PE calculated by Marini *et al.* (2015)[10] because it allowed us to compare the performance of the children in the NT with their performance in the NWRT.

For the NWDT the error rates (in %) in the discrimination of critical and non-critical consonants (+! Cs and -! Cs, respectively) have been calculated. The Romanian children's performances to the NWDT in the age groups 5;0-5;5 and 5;6-5;11, were compared to 2 groups of age-matched Italian children grown up in the same area (respectively: group 1, 20 children, mean age: 62.4 months, SD 1.4; group 2, 8 children, mean age: 67.3 months, SD 1.4) tested with the very same NWDT.

3.5 Extra-linguistic variables

In line with Carroll (2017b), who considers doubtful the possibility to obtain an accurate measure of the children's language use or exposure to a specific language as accounted from parents' reports, we decided at this stage, to pick one of the extra-linguistic variables that we could objectively verify. For this reason and in order to overcome the fact that parents are not always able to quantify the amount of speech they use with their children (Carroll, 2017b) in a reliable way, we used a derived measure accounting for the amount of formal exposure to Italian (expressed in months) as received by the children during their schooling hours. This measure has been obtained by subtracting the age of admission to kindergarten to the children's chronological age (something similar has been done in Goldstein, Bunta, Lange, Rodriguez & Burrows, 2010 to determine the age of acquisition of English in Spanish-English children). Since we also had information on whether the children attended nursery school, we added 12 months to this index[11]. For example, if a child has been recorded at 61 months of age, entered kindergarten at age 36 and attended also nursery school, the months of formal exposure (MFE) to Italian was computed as: 61 - 36 + 12 = 37 MFE. We are aware and acknowledge that this

[10] In Marini *et al.* (2015) the %PE is calculated by dividing the phonological errors (phonological paraphasias, neologisms and false starts in a narrative speech sample) by the number of units (e.g. each word, non-word, or false start uttered by the speaker including phonological errors) and multiplying this value by 100 and is meant as a measure of lexical processing.

[11] For sake of clarity, all of the children entered kindergarten between 36 and 48 months of age, with just one child entering kindergarten at 67 months of age.

measure does not solve the issue of the quality and quantity of exposure to Italian and that its computation may be questionable (as well as its validity). However, with reference to the quantity of exposure, we know from the questionnaires that these children did spend at least 8 hours a day at kindergarten, while with reference to the quality we expect these children received input in "formal" (e.g. standard) Italian as spoken from their kindergarten teachers and nursery school caregivers[12]. Furthermore, there is a high probability that these children most likely would also have had informal contact with Italian through their peers and friends at school, although this kind of contact is difficult to quantify and qualify. Notwithstanding, we think this index can provide us with some general indication for future in-depth analyses.

The question we ask is whether the MFE is able to predict the children's performances and skills with particular reference to the phonetic and phonological development. We expect that children who received more formal exposure to Italian score better on measures of segmental accuracy (e.g. amount of phonological errors in the production tasks) than those who received less exposure. To this aim, we performed regression analyses with MFE as a predictor and the phonetic and phonological measures as dependent variables. The same was done for the age of the children.

4. *Results and discussion*

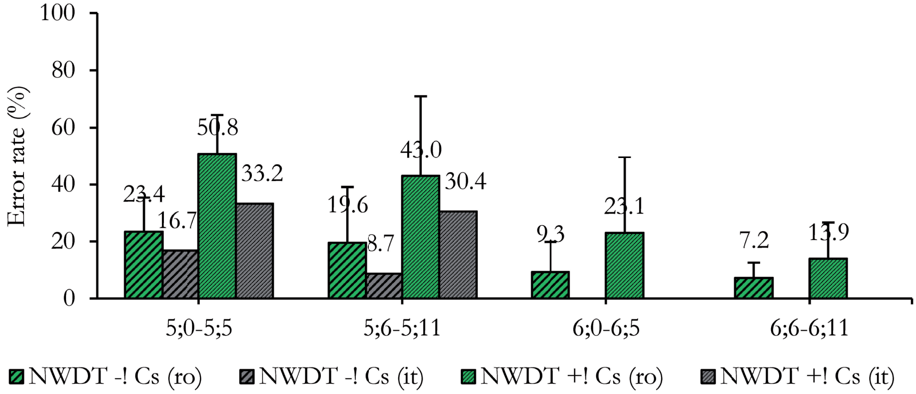
We will start this section by addressing the phonological measures collected by means of the NWDT, the NWRT and the NT. We will first report the results for the NWDT and compare them with the NWRT and the NT where appropriate and possible. We will then compare the two narrative indexes computed over the two stories in the NT with normative data available in the BVL 4-12 (Marini *et al.*, 2015) and we will conclude this section with a short discussion on the SR and %Disf measures.

4.1 Phonological measures of language ability in the NWDT, NWRT and NT

The children's responses to the NWDT across the four age groups are summarized in Figure 1. The figure also shows the performances of two age-matched groups of Italian peers (age-group 5;0-5;5 and 5;6-5;11) to the same NWDT.

[12] We should further state that these children are exposed to a high amount regional speech that, in Veneto as in many other Italian regions, is still strongly pervasive also in formal contexts.

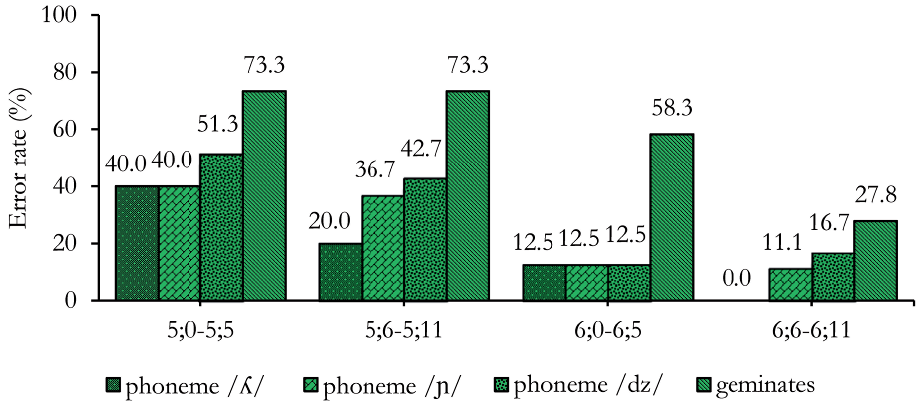
Figure 1 - *Accuracy in the discrimination of non-critical vs critical consonants in the NWDT (ro: Romanian native speakers *in* Italian; it: Italian native speakers *in* Italian)*



Looking more closely at the results, our Romanian children do better at discriminating non-critical consonants as compared to those that we considered critical: for example, the first age group of Romanian children reaches on average an error rate of 50.8±12.0 SD in the discrimination of critical consonants vs 23.4±13.6 SD for those consonants that are not. Looking at the other three age groups, although the error rate decreases with age, this tendency is maintained with the error rate for critical consonants (e.g. consonants not present in Romanian, the children's L1) almost doubling that for non-critical consonants (e.g. consonants shared between Romanian and Italian). However, comparing the Romanian children to the Italian age-matched peers in the first two age groups, we notice that also the Italian children do show some slight difficulty in the discrimination. This is not a surprise, since the consonants that we are considering here as critical for our Romanian children are somehow difficult for the Italian children. We should keep in mind for the discussion that follows, that these consonants are among the latest ones to be acquired by Italian children (Bortolini, 1995; Zanobini *et al.*, 2012). This result confirms the hypothesis that Romanian children would have had a general difficulty with consonants absent in their L1.

If we look at the NWDT's results for the critical consonants into more detail (e.g. by splitting the results for /dz/, /ɲ/ and /ʎ/ as well as for the gemination of consonants), we notice that the major difficulty for the Romanian children is represented by those pairs of non-words involving a contrast between geminates and non-geminates (see Figure 2). The discrimination of geminate consonants reaches the highest error rates in all of the four age groups followed by /dz/.

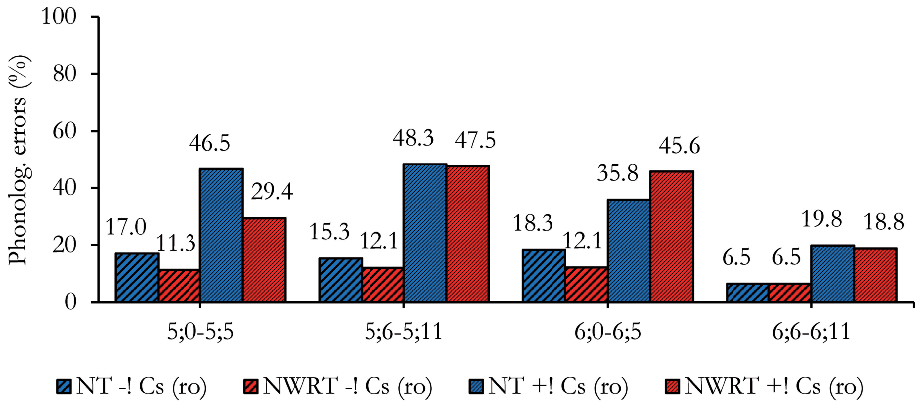
Figure 2 - *NWDT's criticalities for Romanian children*



Overall, by comparing the results in the two figures, there is an improvement (e.g. decreasing error rates) in the discrimination ability for all the consonants considered (both critical and non-critical ones) as the age of the children increases: one could hypothesize that the children's maturation and higher exposure to L2 lead to a better discrimination.

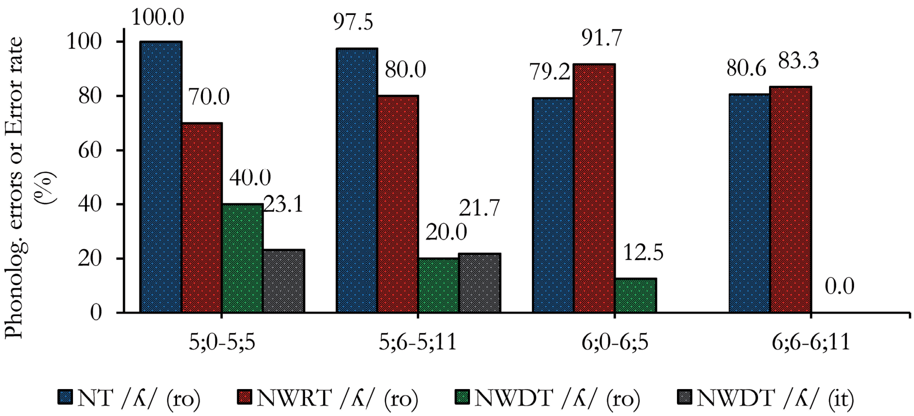
Taking into account only the production level (as measured in the NT and in the NWRT) and separating the results for critical and non-critical consonants, the trend found for the discrimination in the NWDT is confirmed. As shown in Figure 3, the amount of phonological errors (%PE) for the children in our sample is higher for the set of critical consonants as compared to those that are not.

Figure 3 - *Accuracy in the production of non-critical vs critical consonants in the NT and the NWRT*



If we further compare the children's performances (e.g. error rates) in the discrimination of non-word pairs involving contrasts with critical consonants with their ability to produce them spontaneously in the NT and upon repetition in the NWRT, two different trends do emerge[13]. For the phoneme /ʎ/ we see that a good ability in successful discrimination does not match with their ability to correctly produce it. This emerges clearly if we examine the error rates for the children's productions in the NT and the NWRT compared to their discrimination ability in the NWDT in Figure 4. For sake of comparison, we report the error rates available for the Italian children of the first two age groups in the NWDT. The NT shows the lowest accuracy in the production of /ʎ/, especially in the first two age groups with Romanian children producing incorrectly 100% and 97.5% of the expected adult targets. Surprisingly, the two oldest age groups seem to have more problems with this consonant in the NWRT where the error rates are a bit higher than those found for the NT: 91.7% and 83.3% in the NWRT vs 79.2% and 80.6% in the NT, respectively. One possible explanation for this difference in production may be found in the fact that the phoneme /ʎ/, other than being what we called a critical consonant, is also one of the late consonants for 50% of the Italian children at 48 months of age (Bortolini, 1995). These children probably still need to acquire specific articulatory habits that will allow them to utter this specific consonant.

Figure 4 - *Accuracy for phoneme /ʎ/ per age group compared across the different tasks*



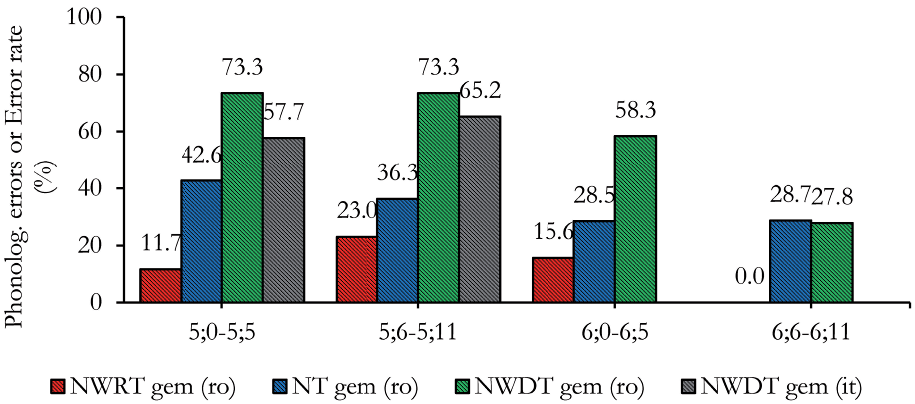
Doing the same comparison across the three tasks for the consonants' length contrast (e.g. geminate consonants) the picture is reverted. Despite the children's ability to produce the expected length of geminate consonants as recorded in the two production tasks (NT and NWRT), the same does not hold for their ability to discriminate them. This is at least the case for the children in the first three age groups (see Figure 5). The Romanian children's performance is however not different from the perfor-

[13] We will report here only on phoneme /ʎ/ and on geminates because for the consonants /dz/ and /ɲ/ the data available specifically for the NT are very sparse. Narrative tasks as the ones proposed here are not, as also highlighted by Zanobini *et al.* (2012), necessarily able to elicit all of the consonants equally. This is even truer for those phonemes having a low frequency of use in spoken Italian, with /dz/ and /ɲ/ being among the less frequent in spoken Italian (Bortolini, Degan, Minnaja, Paccagnella & Zilli, 1978; Tonelli, Panzeri & Fabbro, 1998; Calvani, 2003).

mance obtained by their Italian peers in the age groups 5;0-5;5 and 5;6-5;11. This should not appear surprising as the Italian children are all from the Veneto region where in the dialect (as well as in the Italian) spoken in this area, the consonants' length does not have distinctive function and where geminate consonants are produced with de-gemination (cfr. Zamboni, 1988). As already found in a previous work by Galatà, Meneguzzi, Conter & Zmarich (2012), the Romanian children seem to be insensitive to the length contrast and they produce the length contrast with a duration that is intermediate between geminate and non-geminate consonants[14]. On this point, however, more detailed acoustic analysis is needed in order to confirm what has here been reported based on the phonetic transcription only.

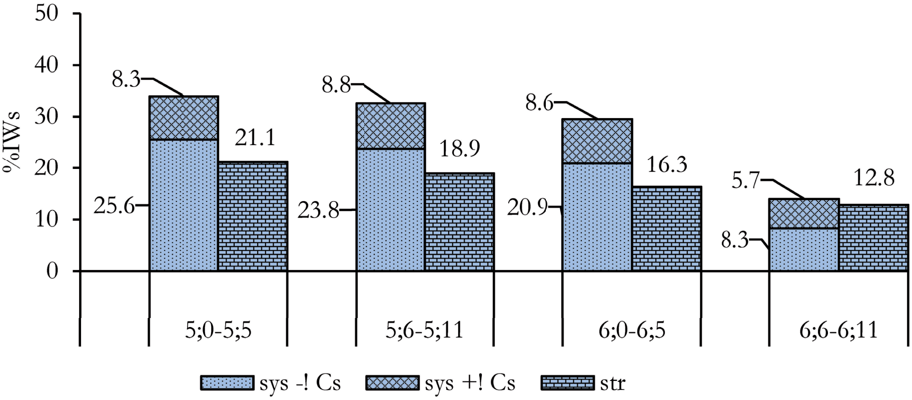
Analysing the amount of incorrect words (%IW) produced by the Romanian children in the NT, we find that the words deviant form monolingual production are: a) in 29.0% (±11.6 SD) of the cases due to the phonological system (simplification of phonemic oppositions, $\%IW_{sys}$); b) in 17.9% (±6.1 SD) of the cases to the phonotactic structure (simplification of the syllabic structure, $\%IW_{str}$). If we compare the two types of errors across the four age groups as done so far (see Figure 6), we notice that the amount of incorrect words affected by substitutions or deletion of phonemes according to the adult target decreases in function of age (and indirectly, in function of an increased exposure to the Italian language). In comparison to the amount of errors affecting the words' syllable structure (e.g. $\%IW_{str}$), the simplification of phonemic oppositions remains always higher.

Figure 5 - *Accuracy for geminate Cs per age group compared across the different tasks*



[14] For the length contrast we considered as errors all cases of: a) de-gemination, where a geminate was expected according to the adult target; b) substitution and/or cancellation of phonemes according to the adult target (also for those cases where the length contrast is correctly realized through phonemes' substitution).

Figure 6 - *Percentage of incorrect words (%IW) per age group and type of error with reference to the phonological system (sys) and the phonotactic structure (str) for non critical (-!) and critical (+!) consonants (Cs) produced by the Romanian children in the NT*



To highlight the differences in our group of Romanian children as compared to what we would or could expect from Italian children, we separated the $\%IW_{sys}$ in incorrect words to be ascribed to the presence of critical consonants for the Romanian children (sys +! Cs), to others that are not (sys -! Cs) and that might be common to Italian children (see the stacked bars in Figure 6). As can be observed from the stacked bars in Figure 6 across the four age groups here considered, respectively 8.3, 8.8, 8.6 and 5.7 percent of the words classified as incorrect contains at least one of the critical consonants for our Romanian children.

Overall, if we compare the Figures from 1 to 6 and focus on the decrease in the phonological errors, we notice that as the children grow older their phonological skills improve.

4.2 Measures of linguistic productivity in the NT

We will now discuss and compare the two narrative indexes VP and MLU_w with normative data available in the BVL 4-12 (Marini *et al.*, 2015) for age-matched Italian peers (cfr. Table 2).

When administering the two stories to the children, we had the impression that some of them were happier with the "Flower pot" story than with "The Nest story"[15]. Therefore, we tested for differences in the total number of words produced in each story as measured by the UP index. A paired-samples t-test disconfirmed our impression by revealing non-significant differences in the children's UP between "The Nest story" (M = 76.88±30.05 SD) and the "Flower pot" story (M = 77.18±31.8 SD); $t(-0.057) = 16$, $p = 0.955$).

[15] They said the first one was a "funny story"; for *The Nest story* they said it was a "sad story".

Table 2 - *Narrative indexes (VP and MLU) and summary statistics (mean and standard deviation) for the two NTs ("The Nest story" and the "Flower pot") per child and age group*

age group	*ID*	*VP (nest)*	*VP (pot)*	*mean VP*	*MLU_w (nest)*	*MLU_w (pot)*	*mean MLU_w*	*mean MLU_{wNorm}*
5;0-5;5	1	89	119	104.0	3.7	4.8	4.2	4.5
	2	32	52	42.0	3.6	3.3	3.4*	3.8
	10	73	40	56.5	2.4	2.9	2.6**	2.8*
	12	17	29	23.0*	1.9	2.4	2.2**	2.8*
	17	48	46	47.0	2.5	3.3	2.9*	3.2*
		51.8 (29.4)	57.2 (35.6)	54.5 (30.2)	2.8 (0.8)	3.3 (0.9)	3.1 (0.8)*	3.4 (0.8)
5;6-5;11	4	28	39	33.5	1.8	1.6	1.7**	1.9**
	8	59	64	61.5	4.5	3.8	4.2	4.5
	9	43	41	42.0	2.7	3.2	2.9*	3.5*
	13	35	56	45.5	4.4	4.3	4.3	4.7
	14	58	49	53.5	3.1	2.7	2.9*	3.3*
		44.6 (13.8)	49.8 (10.4)	47.2 (10.7)	3.3 (1.2)	3.1 (1.0)	3.2 (1.1)*	3.6 (1.1)
6;0-6;5	5	51	60	55.5	2.8	4.0	3.4*	3.8
	6	77	76	76.5	5.1	5.1	5.1	5.3
	15	13	24	18.5**	1.4	2.4	1.9**	2.0**
	16	32	25	28.5*	2.1	2.5	2.3**	3.2**
		43.3 (27.3)	46.3 (26.0)	44.8 (26.3)	2.9 (1.6)	3.5 (1.3)	3.2 (1.4)*	3.6 (1.4)
6;6-6;11	3	101	121	111.0	4.4	4.8	4.6	4.9
	7	59	46	52.5	3.9	4.2	4.1	4.5
	11	48	53	50.5	4.0	5.3	4.7	4.8
		69.3 (28.0)	73.3 (41.4)	71.3 (34.4)	4.1 (0.2)	4.8 (0.6)	4.4 (0.3)	4.7 (0.2)

Data compared to BVL 4-12 (Marini *et al.*, 2015): * ≤ 1 SD; ** ≤ 2 SD.

Moving to VP, a paired-samples t-test revealed no significant differences for the children's VP between "The Nest story" (M = 50.76±24.33 SD) and the "Flower pot" story (M = 55.29±27.87 SD); $t(-1.227) = 16$, $p = 0.238$). Analysing the mean values for each child in Table 2, we see however that there is a very high inter-subject difference with mean values ranging from a very low 18.5 VP for child 15 in the age group 6;0-6;5, to a relatively high 111.0 VP for child 3. According to the normative data used here as reference, almost all the children's VP indexes in the NT are within 1 standard deviation. The only exceptions are: a) the aforementioned child with a very low VP index falling 2 SD below the norm; b) two children with a VP index below 1 SD (23.0 and 28.5, respectively). Considering the overall group means per age, we find VP values that are within the norm if compared to age-matched Italian peers.

Looking at the mean values of MLU_w over the two stories, only 7 children out of 17 scored with values that fall within the norm. Of the remaining children, 5 reported a MLU_w falling below 1 SD, the other 5 below 2 SDs. The children in our sample, with the exception of the three children in the oldest age group, produce short utterances as expressed by the MLU_w. According to Marini *et al.* (2015), MLU_w values below what would be expected may highlight a reduced capacity of the phonological working memory. This is in line with Hipfner-Boucher, Milburn, Weitzman, Greenberg, Pelletier & Girolametto (2015) reporting similar results for narrative abilities in subgroups of English language learners and monolingual peers. However, we should here recall that the MLU_w (as well as the VP) is computed by considering the number of phonologically well-formed words. In fact, the stacked bars in Figure 6 highlight among the Romanian children the presence of words classified as incorrect because they contain at least one of the critical consonants (specifically /ɲ/, /ʎ/ and the gemination of consonants).

In a previous study (see Galatà *et al.*, 2012) in which we compared the production (e.g. repetition) of a selected number of non-words containing the challenging consonants in a group of 10 Romanian children to age-matched Italian peers, we showed the higher difficulty for the Romanian children in the correct realization of these consonants. Note, for example, that as shown in Table 2 in the column referring to MLU_{wNorm} the inclusion of these words, considered as ill-formed because of the presence of challenging consonants for the Romanian children, produces an increase in this measure with some children being classified as falling below 1 SD instead of 2 SD (see for example the children with ID 10 and 12). This is also confirmed by a paired-samples t-test, which shows an overall statistically significant increase in the values associated to the MLU_w index ($MLU_w = 3.37 \pm 1.05$ SD; $MLU_{wNorm} = 3.74 \pm 1.01$ SD; $t(-8.07) = 16, p = < .001$). Unfortunately, this conclusion needs to be further investigated by checking if the assumption that considers these challenging consonants as acquired in Italian age-matched peers holds.

Articulation rate (AR) and disfluencies frequency (%Disf) in the NT are reported for the four groups of Romanian children in Table 3. The overall mean AR of 3.3±0.6 SD for the studied children is comparable with the value of 3.9±0.4 SD reported by Logan *et al.* (2011) and Logan (2015) for a Narration task in children being one year older (average age of 6;10). Analysing the amount of disfluencies as indexed by %Disf, Logan *et al.* (2011) report for the same children a value of 8.9±4.2 SD. In our case we found lower values with an overall mean of 5.6±2.4 SD (range 2.6~9.3) which makes our data more similar to those reported by Yairi, Ambrose (2005) for the conversational speech of the 42 "control" children (5.75 and 6.35 for age-group 5;0-5;11 and 6;0-6;11, respectively)[16].

[16] The amount of disfluencies in Yairi, Ambrose (2005) is obtained as the sum of *Stuttering Like Disfluencies* (SLD) and *Other Disfluencies* (OD).

Table 3 - *Mean and standard deviation of the fluency indexes AR and %Disf in the NT of the Romanian children divided per age group*

age group	*AR*	*%Disf*
5;0-5;5	3.6 (0.4)	6.1 (2.2)
5;6-5;11	3.1 (0.9)	5.1 (2.4)
6;0-6;5	3.3 (0.8)	6.8 (3.1)
6;6-6;11	3.4 (0.3)	4.4 (2.2)

4.3 Months of formal exposure to Italian and chronological age as predictors of segmental accuracy

The results of the regression analyses performed on the phonetic and phonological indexes are presented in Table 4.

The results of the regression analysis yielded a marginally significant effect of MFE on $\%IW_{sys}$ ($p < 0.1$) and a significant effect on %PE in the Narrative Task ($p < 0.05$), and a significant effect of age on $\%IW_{sys}$ ($p < 0.05$) and a marginally significant effect on %Error in the Non-Word Discrimination Test ($p < 0.1$). No significant effect was found in all other cases.

Table 4 - *Effects of months of formal exposure (MFE) and of chronological age on segmental accuracy measures*

	Extra-linguistic variables	
Segmental accuracy measures	MFE	Chron. age (months)
$\%IW_{sys}$ (NT)	$R^2 = .204$, $F(1, 15) = 3.852$ $p = .068$	$R^2 = .277$, $F(1, 15) = 5.752$ $p = .030$
%PE (NT)	$R^2 = .298$, $F(1, 15) = 6.362$ $p = .023$	$R^2 = .152$, $F(1, 15) = 2.693$ $p = .121$
%PE (NWRT)	$R^2 = .049$, $F(1, 15) = 0.770$ $p = .394$	$R^2 = .014$, $F(1, 15) = 0.217$ $p = .648$
%Error (NWDT)	$R^2 = .013$, $F(1, 15) = 0.204$ $p = .658$	$R^2 = .227$, $F(1, 15) = 4.415$ $p = .053$

5. *Conclusions*

The proposed study offers an insight into the main difficulties that a Romanian pre-schooler faces in the acquisition of Italian as L2.

As hypothesized, strong difficulties in the acquisition of L2, as well as the influence of the underlying L1 system, emerged at both the phonetic (through the Articulation rate, the Percentage of Disfluencies) and phonological level (through the Verbal Productivity, the Mean Length of Utterance in words, the Percentage of Incorrect Words and the Percentage of Phonological Errors). The results from the segmental analysis considering the Romanian children's performances in the Non-Word Repetition Test (NWRT) fully confirm our expectations since Romanian children have more difficulty in the production of those non-words containing sounds that are absent in the Romanian language (mean error rate for non-critical Cs: 10.9±7.2 SD; mean error rate for critical Cs: 36.7±16.0 SD). However, a significant improvement in the correctness of the responses was found as the children grew up in age and as the verbal productivity and fluency increased, a clear consequence of the influence exerted by the maturation of higher cognitive functions. The study thus confirms and supports the effectiveness of the adopted methodology (namely the NWRT), that proved to be extremely suitable in capturing such difficulties, especially for those consonants that we considered as "critical" because absent in their Romanian L1 and language spoken at home (e.g. /ɲ/, /ʎ/, /dz/ and the gemination process).

The presence of these criticalities was also found to inflate the amount of incorrect words (follow the discussion relating to Figure 6) with direct effects on the measures of linguistic productivity: Mean Length of Utterance in words (MLU_w) and Verbal Productivity (VP). In fact, for MLU_w we found different children scoring below 2 SD as compared to their monolingual age-matched peers in the BVL 4-12 (Marini *et al.*, 2015). The same issue may have lowered VP, which in turn, seems to be less affected as opposed to MLU_w remaining in both cases on average with respect to what was expected from their age. However, the increase in MLU_w with chronological age of the children proved to be more rapid in the last group, allowing them to reach average performances comparable to those found for Italian monolinguals age-matched peers in the BVL 4-12.

The articulation rate and the number of disfluencies do not deviate significantly from the data we took as reference: the children's speech is characterized by the presence of numerous silent pauses and hesitations between one uttered sentence and another.

The regression analysis taking into account the amount of formal exposure (MFE) and the chronological age of the children did not provide a strong evidence in explaining the investigated measures of segmental accuracy. Some standardization procedure on the measures needs probably to be carried out, but this will be left to future in-depth investigation that will better explore the influence of other extra-linguistic variables, as well as to assess the speech performances of Italian children by means of the same variables.

The study furthermore contributes to highlight how factors depending on individual variables and on the degree of exposure to L2 may influence the linguistic competences shown by the children which result in a high inter-subject variability among the children analysed.

There are obvious limits in the present study that need to be addressed further in the future. Firstly, the limited number of children in our sample does not allow us to generalize over the results and this brings us to the second limitation of the study: the need to complete the analysis also for the other children we recorded within the wider project that includes both Romanian and Italian children also for younger age bands. Once this has been done, a more appropriate comparison of the data from our Romanian children learning Italian L2, with particular reference to the phonetic and phonological development, will be possible by comparing these data to age-matched typically developing Italian monolingual children (as done here, for example, with data from our Italian control group in the NWDT).

Acknowledgments

To the Italian National Research Council for funding the present research through the project "Migrazioni" (2009-2012). To the children and their families for participating in the study. To two anonymous reviewer who helped improving this paper with their useful comments and suggestions.

Bibliography

ANDERSON, R.T. (2004). Phonological acquisition in preschoolers learning a second language via immersion: a longitudinal study. In *Clinical Linguistics & Phonetics*, 18(3), 183-210.

ARCHIBALD, J. (1994). A formal model of learning L2 prosodic phonology. In *Second Language Research*, 10(3), 215-240.

ARCHILA-SUERTE, P., ZEVIN, J. & HERNANDEZ, A.E. (2015). The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds. In *Brain and Language*, 141, 35-49.

BADDELEY, A.D., GATHERCOLE, S.E. & PAPAGNO, C. (1998). The phonological loop as a language learning device. In *Psychological Review*, 105(1), 173-158.

BATES, E. (1995). Conclusioni. In CASELLI, M.C., CASADIO, P. (Eds.), *Il primo vocabolario del bambino. Guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita*. Milano: Franco Angeli, 93-98.

BEST, C.T., TYLER, M.D. (2007). Nonnative and second-language speech perception. In BOHN, O.S., MUNRO, M.J. (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. John Benjamins Publishing Company, 13-34.

BETTONI, C. (2001). *Imparare un'altra lingua*. Bari: Laterza.

BIRDSONG, D. (1999). Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In BIRDSONG, D. (Ed.), *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Lawrence Erlbaum Associates, 1-22.

BOERSMA, P., WEENINK, D. (2016). *Praat: doing phonetics by computer*. http://www.praat.org/retrieved 03.09.2016.

BORTOLINI, U. (1995). *PFLI Prove per la valutazione fonologica del linguaggio infantile*. Venezia: Edit Master.

BORTOLINI, U., DEGAN, C., MINNAJA, C., PACCAGNELLA, L. & ZILLI, G. (1978). Statistics for a stochastic model of spoken Italian. In DRESSLER, W.U., MEID, W. (Eds.), *Proceedings of the Twelfth International Congress of Linguists*. Innsbruck, Austria, 580-586.

Calvani, S. (2003). L'effetto frequenza d'uso dei fonemi della lingua italiana in soggetti afasici con deficit di lessico fonologico di output e/o di buffer. In Aprile, L. (Ed.), *Psicologia dello sviluppo cognitivo-linguistico*. Firenze: Firenze University Press, 125-136.

Carroll, S.E. (2017a). Explaining bilingual learning outcomes in terms of exposure and input. In *Bilingualism: Language and Cognition*, 20(1), 37-41.

Carroll, S.E. (2017b). Exposure and input in bilingual development. In *Bilingualism: Language and Cognition*, 20(1), 3-16.

Chitoran, I. (2001). *The Phonology of Romanian: A Constraint-Based Approach*. Berlin-New York: Mouton de Gruyter.

Devescovi, A., D'Amico, S. (2001). Lo sviluppo della morfosintassi. In Camaioni, L. (Ed.), *Psicologia dello sviluppo del linguaggio*. Bologna: Il Mulino, 117-152.

Dodd, B., Holm, A., Hua, Z. & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. In *Clinical Linguistics & Phonetics*, 17(8), 617-643.

Eckman, F.R. (1977). Markedness and the contrastive analysis hypothesis. In *Language Learning*, 27(2), 315-330.

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*. Baltimore: York Press, 233-237.

Flege, J.E. (2007). Language contact in bilingualism: Phonetic system interactions. In *Laboratory Phonology*, 9, 353-381.

Flege, J.E., MacKay, I.R.A. (2010). "Age" effects on second language acquisition. In Dziubalska-kolaczyk, K., Wrembel, M. & Kul, M. (Eds.), *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech, New Sounds*. Poznan, Poland.

Galatà, V., Meneguzzi, G., Conter, L. & Zmarich, C. (2012). Primi dati sull'acquisizione fonetico-fonologica dell'italiano L2 in prescolari rumeni. In Paoloni, A., Falcone, M. (Eds.), *La voce nelle applicazioni* (Vol. 8). Roma: Bulzoni Editore, 35-50.

Galatà, V., Zmarich, C. (2011a). Le non-parole in uno studio sulla discriminazione e sulla produzione dei suoni consonantici dell'italiano da parte di bambini pre-scolari. In Gili Fivela, B., Stella, A., Garrapa, L. & Grimaldi, M. (Eds.), *Contesto comunicativoe variabilità nella produzione e percezione della lingua* (Vol. 7). Roma: Bulzoni Editore, 118-129.

Galatà, V., Zmarich, C. (2011b). Una proposta per valutare l'influenza fonetico-fonologica della lingua di origine dei bambini figli di immigrati sull'acquisizione dell'italiano. In Bruno, G.C., Caruso, I., Sanna, M., Vellecco, I. (Eds.), *Percorsi migranti*. Milano: McGraw-Hill Companies, Publishing Group Italia, 301-317.

Gass, S., Glew, M. (2008). Second Language Acquisition and Bilingualism. In Altarriba, J., Heredia, R.R. (Eds.), *An Introduction to Bilingualism: Principles and Processes*. New York/London: Lawrence Erlbaum Associates, 265-294.

Goldstein, B.A., Bunta, F., Lange, J., Rodriguez, J. & Burrows, L. (2010). The effects of measures of language experience and language ability on segmental accuracy in bilingual children. In *American Journal of Speech-Language Pathology*, 19(3), 238-47.

Hansen, J.G. (2004). Developmental sequences in the acquisition of English L2 syllable codas. A Preliminary Study. In *Studies in Second Language Acquisition*, 26(1), 85-124.

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J. & Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. In *International Journal of Bilingualism*, 19(6), 677-692.

Huber, W., Gleber, J. (1982). Linguistic and nonlinguistic processing of narratives in aphasia. In *Brain and Language*, 16(1), 1-18.

Ioup, G. (2008). Exploring the role of age in the acquisition of a second language phonology. In Hansen Edwards, J.G., Zampini, M.L. (Eds.), *Phonology and Second Language Acquisition*. Philadelphia: John Benjamins Publishing Company, 41-62.

Jakobson, R. (1966). *Essais de linguistique générale*. Paris: Minuit (trad. it. *Saggi di linguistica generale*, Milano: Feltrinelli, 1966).

Lado, R. (1957). *Linguistics across cultures: applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Logan, K.J. (2015). *Fluency Disorders*. San Diego: Plural Publishing Inc.

Logan, K.J., Byrd, C.T., Mazzocchi, E.M. & Gillam, R.B. (2011). Speaking rate characteristics of elementary-school-aged children who do and do not stutter. In *Journal of Communication Disorders*, 44(1), 130-147.

MacKay, I.R.A., Flege, J.E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. In *Applied Psycholinguistics*, 25(3), 373-396.

Major, R.C. (1986). The ontogeny model: evidence from L2 acquisition of Spanish r. In *Language Learning*, 36(4), 453-504.

Marini, A. (2014). Caratteristiche della elaborazione linguistica in bambini bilingui con disturbi dello sviluppo linguistico. In Caselli, M.C., Marotta, L. (Eds.), *I disturbi del Linguaggio. Caratteristiche, valutazione, trattamento*. Trento: Edizioni Erickson, 65-85.

Marini, A., Andreetta, S., Del Tin, S. & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. In *Aphasiology*, 25(11), 1372-1392.

Marini, A., Marotta, L., Bulgheroni, S. & Fabbro, F. (2015). *BVL 4-12. Batteria per la Valutazione del Linguaggio in bambini dai 4 ai 12 anni*. Firenze: Giunti O.S.

Mauro, F., Burelli, A. (2002). Assessment of the linguistic development of Friulian-Italian bilingual children. In *Friulian Journal of Science*, 2, 95-110.

McLaughlin, B. (1978). *Second language acquisition in childhood*. Hillsdale: Erlbaum.

Meisel, J.M. (2004). The Bilingual Child. In Bhatia, T.K., Ritchie, W.C. (Eds.), *The Handbook of Bilingualism*. Oxford, UK: Blackwell Publishing Ltd, 90-113.

Mennen, I. (2015). Beyond Segments: Towards a L2 Intonation Learning Theory. In Delais-Roussarie, E., Avanzi, M. & Herment, S. (Eds.), *Prosody and Language in Contact. Prosody, Phonology and Phonetics*. Berlin, Heidelberg: Springer, 171-188.

Mioni, A. (2005). *Immigrati e comunicazione interetnica in Italia*. Padova: Università di Padova.

Muljacic, Z. (1972). *Fonologia della Lingua Italiana*. Bologna: Il Mulino.

Paradis, M. (1987). *The assessment of bilingual aphasia*. Hillsdale: Erlbaum.

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

Priester, G., Post, W. & Goorhuis-Brouwer, S. (2011). Phonetic and phonemic acquisition: Normative data in English and Dutch speech sound development. In *International Journal of Pediatric Otorhinolaryngology*, 75(4), 592-596.

Ritchie, W.C., Bhatia, T.K. (2009). *The new handbook of second language acquisition*. Emerald Group Publishing.

Rose, Y., MacWhinney, B. (2014). The PhonBank Project: Data and Software-Assisted Methods for the Study of Phonology and Phonological Development. In Durand, J., Gut, U. & Kristoffersen, G. (Eds.), *The Oxford handbook of corpus phonology*. Oxford: Oxford University Press, 308-401.

Rose, Y., Stoel-Gammon, C. (2015). Using PhonBank and Phon in studies of phonological development and disorders. In *Clinical Linguistics & Phonetics*, 29, 686-700.

Sebastián-Gallés, N., Bosch, L. (2005). Phonology and bilingualism. In Kroll, J.F., Groot, A.M.B. (Eds.), *Handbook of bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press, 68-87.

Tonelli, L., Panzeri, M. & Fabbro, F. (1998). Un'analisi statistica della lingua italiana parlata. In *Studi Italiani Di Linguistica Teorica Ed Applicata*, 28(3), 501-514.

Topoliceanu, H. (2011). Italiano e rumeno a confronto: analisi di alcuni problemi di apprendimento dell'italiano da parte dei madrelingua rumeni. In *Philologica Jassyensia*, VII(1), 243-255.

Valentini, A. (2005). Lingue e interlingue dell'immigrazione in Italia. In *Linguistica e Filologia*, 21, 185-208.

Wei, M., Joshi, A.A., Zhang, M., Mei, L., Manis, F.R., He, Q., Beattie, R.L, Xue, G., Shattuck, D.W., Leahy, R.M., Xue, F., Houston, S.M., Chen, C., Dong, Q. & Lu, Z. (2015). How age of acquisition influences brain architecture in bilinguals. In *Journal of Neurolinguistics*, 36, 35-55.

Weismer, S.E., Tomblin, J.B., Zhang, X., Buckwalter, P., Chynoweth, J.G. & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. In *Journal of Speech, Language, and Hearing Research*, 43(4), 865-78.

Yairi, E., Ambrose, N.G. (2005). *Early childhood stuttering for clinicians by clinicians*. Austin, Tex.: PRO-ED.

Young-Scholten, M. (1994). On positive evidence and ultimate attainment in L2 phonology. In *Second Language Research*, 10(3), 193-214.

Zamboni, A. (1988). 270. ltalienisch: Areallinguistik IV. Veneto. In Holtus, G., Metzeltin, M., Schmitt, C. (Eds.), *Lexicon der Romanistischen Linguistik* (*Vol. IV*). Tübingen: Niemeyer Verlag, 517-538.

Zampini, M.L. (2008). L2 speech production research: Findings, issues, and advances. In Edwards, J.G.H., Zampini, M.L. (Eds.), *Phonology and second language acquisition*. Philadelphia: John Benjamins Publishing Company, 219-249.

Zanobini, M., Viterbori, P. & Saraceno, F. (2012). Phonology and language development in Italian children: an analysis of production and accuracy. In *Journal of Speech Language and Hearing Research*, 55(1), 16-31.

Zmarich, C., Magno Caldognetto, E. & Ferrero, F.E. (1996). Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza. In Cutugno, F. (Ed.), *Atti delle XXIV Giornate di Studio del Gruppo di Fonetica Sperimentale*. Roma: Esagrafica 111-139.

Zmarich, C., Pinton, A. & Lena, L. (2014). Lo sviluppo fonetico-fonologico nell'acquisizione di L1 e L2. In Caselli, M.C., Marotta, L. (Eds.), *I disturbi di Linguaggio. Caratteristiche, valutazione, trattamento*. Trento: Edizioni Erickson, 87-124.

DOI: 10.17469/O2103AISV000014

CINZIA AVESANI, VINCENZO GALATÀ, CATHERINE BEST, MARIO VAYRA, BRUNO DI BIASE, FABIO ARDOLINO

Phonetic details of coronal consonants in the Italian spoken by Italian-Australians from two areas of Veneto

The paper investigates the distribution and the phonetic details of a set of coronal obstruents in the speech of Veneto immigrants who arrived in Australia more than 50 years ago in order to understand how they negotiate the languages that form their linguistic repertoire. We focus on their L2-Veneto Regional Italian, with the aim of ascertaining whether it has crystallized at the stage it was at the time they left Italy; whether it shows signs of attrition due to replacement of features from the L1-dialect or from L3-English; or whether it shows signs of koinization due to the contact with other dialects and regional varieties of Italian. Results on coronal obstruent distribution based on narrow phonetic transcriptions and on their acoustic characteristics (spectral moments analysis) show a change in progress in the fricative set and a general maintenance of the stop consonants' phonetic features. Individual differences in language change are discussed in light of both internal (linguistic) and external (sociolinguistic) factors[1].

Key words: sociophonetics, Italian, language variation and change, heritage languages, Italian-Australian immigrants, koinization.

1. *Introduction*

The present study investigates the maintenance, attrition and/or convergence of language-specific phonetic features in the Regional Italian spoken as a second heritage language[2] by immigrants to Australia from two different dialectal subregions of Veneto centered on the towns of Belluno and Rovigo. The purpose of this research was to investigate how specific phonetic features of Italian may be affected by the native heritage language (i.e. regional dialect) in the context of learning and using a new, third language in the adoptive country (English). This study falls within the

[1] Authorship note: This paper is part of a larger project, designed and conducted in close collaboration by the first five authors. Main responsibility for this paper is divided as follows: § 1: Best, Avesani; § 1.1 Avesani, Di Biase; § 1.2 Avesani; § 1.3: Vayra, Avesani; § 2: Best, Avesani, Vayra; § 3: Galatà; § 4 and 5: Avesani, Galatà, Best, Vayra, Di Biase, Ardolino.

[2] We use the term "heritage language" in its broad connotation to refer to non societal and non majority languages spoken by immigrants and their children (Valdés, 2000; 2005; Montrul, 2013). The speakers we analyse in the present paper are first generation immigrants who arrived in the host country as adults. As they are early sequential bilinguals (local dialect and Regional Italian), we consider both their dialect and their variety of Regional Italian as heritage languages. Consequently, we refer to their L1 dialect as the native heritage language and to their L2 Regional Italian as the second heritage language.

scope of research of contact-induced change in the speech of bilingual (e.g., Major, 1992; De Leeuw, Schmid & Mennen, 2010; Nagy, Kochetov, 2013) and multilingual (Cabrelli Amaro, 2012) speakers.

We will specifically focus on a set of coronal consonants that exhibit interesting differences in the two Veneto's subregions. Coronals 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 space.

Our phonetic-level findings complement and extend the more frequently investigated macro-structures of lexical, syntactic and morphological interactions among immigrants' languages. This study builds upon and complements our previous examination of the phonetic features of these speakers' first language, Veneto Dialect, produced during the same recording session (Avesani, Galatà, Vayra, Best, Di Biase, Tordini & Tisato, 2015). The findings offer new insights into the effects that long-term and long-distance migration to a country with a different dominant language may have on the speech properties of the heritage languages still used by members of the immigrant community.

We begin by positioning heritage language use by Italian-Australians in its broader historical context. From there we consider how that context has influenced use of Dialect and Italian in the English-dominant adoptive country, thus providing key sociolinguistic characteristics of the situation that could impinge on their speech patterns in their heritage languages decades after relocating to Australia.

1.1 Italians in Australia

Italians began to migrate to Australia at the end of nineteenth century, but it was only after World War II that they arrived in large waves to start a new life there[3]. In recognition of an urgent need to "populate or perish", soon after World War II the Australian government began negotiations with Britain and other European countries for assisted migration programs. In 1951 a bilateral agreement for assisted immigration was signed by the Australian government with Italy, that allowed 20,000 immigrants to relocate to the country each year (e.g., Campolo, 2009). By 1961 over 330,000 Italians had settled in Australia, their transport being subsidized by the Australian government (http://john.curtin.edu.au/1940s/populate/; Cavallaro, 2003; Campolo, 2009).

The result was that large Italian communities became established in the principal urban areas of mainland Australia, most rapidly soon after the war but continuing steadily through the late 1970's. This process formed the backbone of today's Italian-Australian community (Cavallaro, 2003)[4].

[3] By 1881 there were 521 Italians in New South Wales, with a sizable Italian community in Sydney of professionals, artists and importers. Among them, a group of 217 migrants from Veneto. (see http://www.migrationheritage.nsw.gov.au/exhibition/journey/pioneers/index.html).

[4] At the last comprehensive census (Australian Bureau of Statistics, 2011), 916,116 members of the population listed Italian ancestry (with at least one of the parents born in Italy), and more than

During that period, Italians left their homeland mainly from southern Italy, but also from two less industrialized regions in the north: Veneto and Friuli. They spread around Australia, usually creating multi-regional communities: people from regions whose local dialects could be mutually unintelligible lived side by side, and started using Italian as their vehicular language to understand each other. However, there was also much "chain" migration[5], in which Italians from a given town or region settled in specific towns in Australia. The latter process created regional nuclei of linguistically homogeneous communities such as, for example, the town of Griffith, in New South Wales, where more than 60% of the population is Italian (or of Italian descent). Among them, more than half originated from Veneto, mostly from Verona (Corazza, Grigoletti & Pellegrini, 2012). In the 1950s, when the Veneti accounted for 75% of Griffith's Italian residents, there was little need for them to learn either English or the dialects from other regions of Italy[6].

Generally, Veneto immigrants, the focus of this paper, originated from small villages with a rural economy with severely limited resources and only rudimentary formal education. The language they learned first and mastered best while in Italy was the local dialect, which served as the exclusive language of their daily communications. Italian, when acquired, was only learned from age 6, when they entered the elementary schools, and was used almost exclusively for formal occasions. From the 1951 census of the Italian National Institute of Statistics (ISTAT), we gather that between the late 1940s and early 1950s 61% of the Italian population spoke mainly dialect and two decades later, in 1974, 51.3% of the population still spoke only dialect (De Mauro, 2014). It is hardly surprising, then, that when those migrants embarked on the ships that brought them to the "other side of the world", only few of them had learned the language of their future land and that the vast majority of them acquired it only by immersion once in Australia[7].

The small group of Veneto immigrants who served as informants for the present study, coming from two different areas of the Veneto region (the North-eastern and the Center areas), left Italy between 1954 and 1961. They reflect the picture we sketched above: they were born in small rural villages, had the local dialect as L1, learned Italian at school as an L2, and did not know any English before arriving in Australia.

180,000 were actually born in Italy. In the Greater Sydney area, the Italian community is the sixth in size, with 4.3% of its residents declaring Italy as their birthplace or land of ancestry.

[5] Observations have shown that the migratory movement, especially to rural areas of Australia, has followed the pattern of a "chain migration", which creates "districts in which emigrants are bound together by shared kinship ties based on a specific village" (Tosi, 1991: 337).

[6] The president of the "Veronesi nel mondo" association acknowledges and reports that, after 25 years of residence in Griffith, she knew only few words of English, while her brother, like most other Italians who settled in Sydney, had to learn English as soon as he arrived (Corazza et al., 2012: 94).

[7] In some cases, classes of English were organized on board, as reported as a personal experience by one of our participants, to provide migrants with at least a basic English vocabulary.

1.2 The Italian language in Australia: Multiple regional origins

Apart from cases of homogeneous immigration such as Griffith, where immigrants could keep using only their local dialect for everyday communication, the picture was quite different. The Veneti in Australia usually found themselves immersed in communities where they lived not only with Italians from the same village/region, but also those from different regions, whose native language was the respective local dialect.

According to Tosi (1991), Italian immigrants who settled in urban areas tended to develop personal and social contacts with people from the same region, either within the original kinship network or within the local regional club. In such circumstances, alongside the development of English used for outside-group purposes, first generation Italians are likely to have used the village dialect in the home as well as outside with members of their heritage community. However, for communicating with other Italians from different regions, Italian was the primary language they used.

The different social conditions dictating the use of their heritage languages on the one hand and the acquisition of the official language of their new land on the other changed the dynamics of immigrants' original use of the local dialect versus the national language, and complicated the management of their multilingual repertoire.

The local dialect, spoken as L1 within the entire local community in Italy in both informal and formal circumstances[8], became restricted to use with family and friends who originated from the same geographical areas.

Italian, the national language learned at school as an L2 used near-exclusively for writing and formal contexts, widened its use in the new land to encompass oral exchanges with Italian-Australians from other regions in Italy. When immigrants left the country in the first half of the twentieth century, speakers of dialect in Italy were engaged in a process of group second language acquisition and the process of unilateral convergence[9] from dialects toward Italian had just started. The imperfect learning of Italian favoured the occurrence of dialect features in the varieties of the language spoken in the different regions and gave rise to Regional Italian (henceforth RI), varieties of the national language that carried different phonetic, phonological and morphosyntactic features as a result of substratum interferences[10]. Once in Australia, the linguistic standardization process continued in the new environment, initially at a slower pace compared to those who remained in Italy, and occurred concurrently with learning English (Tosi, 1991: 338). The need to address

[8] As it is still the case in small Veneto villages, dialect is the language of the communication with family, friends and also with the town's public officers.

[9] This process has been termed "advergence" to distinguish it from a bilateral convergence (e.g. Auer, 2005).

[10] "RI is also the outcome of a process of divergence of geographical varieties from the national language; it results from a so-called "dialectization of (varieties of) Italian" (Berruto, 2005: 83). "Far from determining linguistic unification, advergence has caused an increasing differentiation across the national linguistic repertoire" (Cerruti, 2011: 12).

each other and to educate their children in a language that allowed them to share common cultural and emotional experiences across diverse Italian roots favoured the convergence towards shared forms (Gallina, 2011). Since the studies of Rando (e.g., 1973), Bettoni (1981, 1985, 1985b) and Bettoni, Rubino (1996), a great deal of attention has been devoted to the characteristics of such "community language" called Italo-Australian. Its linguistic base is multifaceted (Tosi, 1991), provided by a mixture of Italian dialects, regional varieties of Italian and the so-called *Italiano popolare*[11] (De Mauro, 1972; Cortelazzo, 1972), while its main feature is the importation of English lexical items within an Italian morphological frame.

The third language of their repertoire, Australian English, that was learned in adulthood as an L3 by immersion or, more seldom, at school, varies the most with respect to its role in the speakers' linguistic repertoire. Its acquisition depended on input type, availability and frequency, on the need to use it in daily circumstances, and on the presence and connections with speakers of compatible dialects and regional varieties of Italian.

1.3 The Italian language in Italy: Veneto Regional Italian and Veneto dialects

The variety of Italian spoken by Veneto immigrants (and by our informants as well) is Veneto Regional Italian (henceforth VRI). It presents specific phonetic realizations of Standard Italian (SI) consonants, characterized by a pronunciation that Canepari (1984) defines as "morbida" (*soft*). This means that consonants tend to be lenited, with a simpler articulation or shortened duration, or voiceless ones produced as voiced. An illustrative case is the pronunciation of /r/, which in Standard Italian is a trill or single tap, but in Veneto Italian is instead a flap [ɾ] or an approximant [ɹ].

According to Canepari (1984: 102), who offers a general picture of how Italian is spoken in the region, the VRI alveolar fricative /s/ and /z/ are apical and/or have a more retracted place of articulation than in SI; this endows them with a quality he dubbed "scibilante", i.e. they sound perceptually similar to the postalveolar fricatives [ʃ] and [ʒ] (see also Mioni, 2001: 157). The VRI postalveolar affricates [ʧ] and [ʤ], in turn, usually lose the lip protrusion that is present in SI (Canepari, 1984: 99), therefore reducing the length of the front resonating cavity.

The alveolar affricates [ts] and [dz] of SI tend to become fricatives in Veneto Italian due to their absence in the phonology of Veneto dialects. According to Canepari, when produced as affricates they are similar in place of articulation to SI. However, their production as alveolars may not be as homogeneous in the whole Veneto territory as Canepari seems to imply: based on data we are currently collecting in the province of Rovigo, in specific areas close to the Po river (Polesine), [ts] is produced as more advanced with respect to SI and is perceptually similar to the lamino denti-alveolar fricative present in the local dialect spoken in the same

[11] Cortelazzo (1972: 11) defines *Italiano popolare* as the type of Italian imperfectly acquired by a speaker whose L1 is the local dialect.

areas. In addition, in VRI, as in all Northern Italian varieties, dental affricates in word-initial position are produced as voiced and if geminated in word medial position as in SI [ˈpɛts:o] "pezzo" (*piece*), they are pronounced with a shorter duration (de-geminated).

The specific phonetic properties of VRI coronal obstruents originate in the properties of those consonants in the Veneto dialect. "Veneto", the dialect spoken in the Veneto region and named after it, is better described as a macro-dialect that shows intra-dialect differences among its five subsystems (Mioni, Trumper, 1977; Zamboni, 1974; 1988). Dental affricates are missing in four subsystems and are present only in the phonology of the most innovative dialect spoken in the area of Venice. In those four subsystems dental affricates are replaced by dental fricatives which, interestingly, differ in two of them: in the North-eastern system (trevigiano-feltrino-bellunese), centered on the towns of Belluno, Treviso and Feltre, and in the Central system (padovano-vicentino-polesano) centered on the towns of Padova, Vicenza and Rovigo (NeVen and CVen, respectively; Trumper, 1972; Mioni, Trumper, 1977; Zamboni, 1974; 1988).

In NeVen and CVen, words that in SI present /ts/ exhibit the (inter)dental fricative /θ/. However, in the southern part of the CVen system (province of Rovigo) the fricative that replaced the Old Italian /ts/ has further evolved to a somewhat more anterior lamino denti-alveolar fricative /s̪̻/, which is also present in the neighbouring dialect of Ferrara (Trumper, 1972; Zamboni, 1974; 1988). In both systems, the voiced counterpart [ð] also exists, that occurs as a lenited allophone of /d/ (Zamboni, 1988; Trumper, 1972)[12] together with a voiced dental approximant [ð̞] (Avesani et al., 2015). In Table 1 we present some examples.

Table 1 - *Correspondence of SI voiceless dental affricate in NeVen and CVen local dialects*

	Standard Italian	*NeVen dialect*	*CVen dialect*
clogs	[ˈtsɔk:oli]	[ˈθɔkoj]	[ˈs̪̻ɔkoj]
halter	[kaˈvets:a]	[kaˈʋeθa]	[kaˈʋes̪̻a]
one hundred	[ˈʧento]	[ˈθento]	[ˈs̪̻ento]
nail	[ˈkjɔdo]	[ˈʧɔð̞o]	[ˈʧɔð̞o]

In both local dialects, as well as in SI and VRI, coronal stops have a dental place of articulation[13].

[12] According to Trumper (1972: 25-27) and Zamboni (1988: 526-28) in dialectal systems that present apocopy (more in NeVen than in CVen) it is also possible to posit a morphophonemic contrast /d~ð/ as in [frˈeða] – [fret] vs [ˈmɛða] – [mɛθ].

[13] The classification of /t, d/ as dental is not fully agreed upon in the Italian phonetic literature. Most publications (9 out of 19 checked) indicate a dental place of articulation (e.g. Bertinetto, Loporcaro, 2005), some (4/19) indicate an alveolar place of articulation (e.g. Schmid, 1999), the remaining indicate both a dental and an alveolar place of articulation depending on speaker and following vocalic context (e.g. Tagliavini, 1963).

Comparing the phonetic properties and the distribution of coronal consonants in SI, local NeVen and CVen dialects and English, similarities and differences emerge. Stops are alveolar in English but dental in SI, VRI and in both NeVen and CVen dialects; moreover, they are aspirated in stressed syllable onset, but always unaspirated in SI and in Veneto dialects. The voiceless interdental fricative /θ/ is shared by English and by the NeVen dialect while SI lacks it and so does the CVen dialect. The voiced interdental fricative [ð] is a phone shared by English and by both local dialects while it does not occur in SI. Neither English nor SI nor the NeVen dialect has the lamino-denti-alveolar fricative [s̪] that is present in the CVen dialect of the Rovigo area. On the other hand, SI has dental affricates /ts, dz/ that do not occur in the two dialects nor in English and tend to be reduced to fricatives in VRI. Fricatives /s, z/ are alveolar in Australian English, but in Italian these are described as dental or (lamino-)alveolar according to several authors (Bertinetto, Loporcaro, 2005: 132; Mioni, 2001: 156). The voicelesss postalveolar fricative [ʃ] and voiced and voiceless postalveolar affricates [ʧ, ʤ] are shared by all systems. Table 2 summarises The crosslinguistic differences relevant for the present paper.

Considering the distribution of the coronals in the VRI productions of our speakers and by analysing their fine acoustic characteristics, we may be able to provide an index of the change in progress of L2-Italian in contact with L1-subregional dialect and/or with L3-English.

Table 2 - *Relevant differences in coronal obstruents*

SI	*English*	*NeVen dialect*	*CVen dialect*
[t̪, d̪, ts, dz]	[t, d, θ, ð]	[t̪, d̪, θ, ð]	[t̪, d̪, s̪, ð]

2. The present study: Italian by immigrants from two Veneto subregions

In light of the dynamics among the different languages mastered by Italian immigrants to Australia, our interest lies in examining acoustic and phonetic evidence for maintenance, attrition/drift or convergence of specific phonetic features of VRI in relation to Veneto dialect, both spoken as heritage languages, and English. Our informants are immigrants who originated from North-eastern Veneto (Belluno) and from the southern part of Central Veneto (Rovigo).

We consider three possible scenarios for their spoken Italian:

a) *Maintenance*. Due to the reduced amount of Italian usage in Australia, the VRI as spoken by immigrants from the two subregions may have frozen (crystallized) in the linguistic state it was at the time they left Italy. If so, their Italian could be expected to display the same phonetic and phonological characteristics attested in their subregion of Veneto. For example, we can expect that the speakers of both subregions will reduce dental affricates to fricatives and that [ts], if produced at all, will have a more advanced place of articulation in the speech of

CVen speakers, a feature induced by the contact with the lamino denti-alveolar fricative present in the local dialect (see § 1.3; Zamboni, 1974; 1988).

b) *Attrition.* Language attrition can involve not only deletion of structural elements, but also the drift of those elements toward the features of another, more dominant language (more frequently and widely used) with which it is in contact (Schmid, 2011). In Australia, Italian is in contact with both the immigrants' native language, their subregional dialect, and with their third language, English. If drift occurs in their Italian, it could take two possible directions:
    1. given the shortage of exposure to Standard Italian, the immigrants' Italian may drift toward dialect features, giving rise to new dialectized forms. For example, speakers of NeVen regional Italian could produce SI [ˈtsɔkːoli] (*clogs*) as [ˈθɔkoli] drifting toward the local dialectal form [ˈθɔkoj]. This is the process of de-standardisation (Auer, 2005: 25), which reflects increasing tolerance of regional features, and complements the opposing process of dialect-to-standard *convergence* (see point c below);
    2. given the dominance of English in Australia, Italian may drift toward English phonetic features in the direction of English either/both qualitatively (for example, Italian dental voiceless unaspirated stop consonants could become alveolar; or they could become more aspirated in relevant phonotactic contexts) and/or quantitatively (e.g., the voice onset time (VOT) of voiceless stops could become lengthened).

c) *Convergence*. Italian may be used as the shared language for communication between speakers whose dialects are too divergent for full mutual intelligibility. In this case, their different regional varieties of Italian could converge toward a new variety that abandons the most locally-marked forms, i.e., an Australian variety. Koinization is the process by which contact between different varieties of the same language creates a new, convergent variety of that language (Kerswill, 2002): a koiné (Siegel, 2001)[14]. Prototypic koinization involves two distinct processes: first, interactions contain a mixture of features from different language varieties; secondly, those features that are similar enough to be mutually intelligible are levelled, i.e., locally marked features are deleted. The emergent Australian variety of Italian would be an immigrant koiné, a linguistic variety born in a new area where speakers of different varieties have relocated (Siegel, 1985; 2001)[15]. Under this hypothesis, we expect either that a local variety of VRI includes phonetic features that are typical of other local varieties of RI, including other varieties of VRI; and/or that a local variety looses its most locally

[14] A stabilized contact variety that arises from the mixing and subsequent levelling of local features of different varieties of the same language, fostered by increased interaction or integration among the respective speakers (Siegel, 2001: 175).

[15] Siegel (1985, 2001) distinguishes between a *regional koiné*, which develops in the area historically occupied by the original varieties, and an *immigrant koiné*, which is born in a new area where speakers of different varieties have resettled.

marked forms of VRI (for example, CVen speakers could loose their lamino denti-alevolar pronunciation of [ts] and converge toward a dental pronunciation).

To evaluate these three possible paths for variation and change in the speech properties of an immigrant community's heritage language, the current paper examines the phonetic and acoustic properties of coronal obstruents in the Italian of first generation Italian-Australian immigrants from two dialectally differing subregions of Veneto. Specifically, we focus on the phonetic and acoustic properties of coronal obstruents in their spoken Italian ([t, d, ts, dz, s, z, ʃ, ʧ, ʤ]).

3. *Materials and methods*

3.1 Speakers

The speakers selected for the present study are four first generation Veneto immigrants from the Italian Roots in Australian Soil (IRIAS) corpus that we collected between 2011-2016 (http://irias.filefaustralia.org/; Galatà, Avesani, Best, Di Biase & Vayra, in preparation; Avesani et al., 2015, which was designed to investigate regionally-differing speech features in the dialect, Italian and English of first and second generation Italian-Australians from linguistically diverse regions of Italy. We will compare the properties of their Italian coronals with previous findings on coronals in dialect as spoken by the same four speakers (Avesani et al., 2015). In relation to the three scenarios described above, we will examine whether their coronals in Veneto Regional Italian (VRI) have maintained the fine phonetic features of their places of origin in Veneto which reflect the properties of coronals in their local substrate dialect, have drifted toward English, or have converged with other varieties of Italian through contact with the broader Italian community in Australia.

The speakers are two males and two females born and raised in two different linguistic subregions of the Veneto region: one male (GP) and one female (CZ) from the North-eastern area (NeVen) that includes the provinces and towns of Treviso, Belluno and Feltre; one male (JF) and one female (AM) from the southern part of the Central area (CVen) that includes the provinces and towns of Padua, Vicenza and Rovigo. CZ and GP were born in the NeVen province of Belluno. Both are native speakers of a Bellunese dialectal variety (BL). Speakers AM and JF are from the CVen province of Rovigo and their mother tongue is a Rodigino dialectal variety (RO). All speakers acquired the local dialect as L1, learned Italian as L2 when they entered elementary school in Italy, and learned English as L3 upon their arrival in Australia between 1929 and 1947, as adults or adolescents. None had been exposed to English before arriving in Australia. At the time of recording, their age range was 64-82 (see Table 3).

Table 3 - *Informants' sociolinguistic information: Age rec = age at time of recording; LOR = Length of Residence in Australia; AOA = age of arrival and onset of acquisition of English; RO = Rovigo dialect; BL = Belluno dialect; FI = formal instruction; SI = spontaneous immersion. All ages and time periods are shown in years*

ID	*Gender*	*Age rec.*	*AOA*	*LOR*	*L1 (dialect)*	*L2*	*L3*	*Highest education*	*Occupation*	*Years of education in Italian*
AM	F	64	14	50	RO	Italian	English FI	university	judge	8
JF	M	81	26	55	RO	Italian & Ferrarese dialect	English FI	primary school	storekeeper	5
CZ	F	74	17	57	BL	Italian	English SI	primary school	housewife	5
GP	M	82	29	53	BL	Italian	English SI	primary school	tradesman	5

The speakers are comparable for two sociolinguistic parameters (see Table 3): length of residence in Australia (LOR), which exceeds fifty years for all of them (range: 50-57 years); and age of acquisition of Italian as L2, which started at age 6 for all of them, when they entered primary school in Italy. They differ in age (females are younger than males) and age of arrival in Australia, which corresponds as well to their age of acquisition of English (AOA in Table 3). The females arrived in Australia 12 years on average before males during their adolescence, which gives them a longer period of acquisition and practice of English as L3. They also differ in the number of years of education they received in Italy, which gives them 5 years of education in Italian if they attended only primary school or 8 years if they attended also middle school.

3.2 Target coronals

The data selected for the present analysis were extracted for each speaker from the IRIAS corpus[16], in which a set of 46 target words were presented to the speaker in a picture naming task. The speakers spontaneously named the object depicted, described the target pictures in Italian and often commented at length. From the speech material collected in such (semi-)spontaneous condition a total of 619 tokens of coronal obstruents were selected. All speech samples were recorded at 96 kHz-24 bit mono in a quiet location at the participants' home or at MARCS

[16] The IRIAS corpus is a multilingual corpus of semi-spontaneous and spontaneous speech. Dialect, Italian and English productions, in that order, are elicited first in spontaneous conversations with an interviewer, then in a picture naming task in which a set of images are presented to the speaker on a computer screen. Objects depicted refer to words that contain a set of target coronal consonants that occur in the three languages. Readers are referred to Avesani et al. (2015) for details on the specific interview protocol adopted within the broader project.

Institute (Western Sydney University), using a MatLab recording tool (*SyncRec* developed by G. Tisato at the Institute of Cognitive Sciences and Technologies of the Italian National Research Council [ISTC-CNR], Padova).

The 46 target words analysed for the present study contain one or more target coronals in specific word positions: word-initially and word-medially, in intervocalic position or following a liquid or a nasal consonant. The amount of collected tokens is different for each speaker and varies from 96 to 191 occurrences[17]: this is an expected outcome of our eliciting procedure.

Our target sounds are the following set of coronal obstruents: [t, d, ts, dz, s, z, ʃ, ʧ, ʤ].

3.3 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 Italian recording. These included: a) segmentation and orthographic transcription in Italian at the sentence level; b) segmentation and narrow IPA transcription at the word level; c) segmentation and narrow IPA transcription at the phone level for the target consonants and for the preceding and/or following phonetic context (for affricate consonants we separately segmented and transcribed the closure and frication phases); d) coding of the target consonant for manner of articulation (stop, affricate or fricative) and position within the word (initial or medial). In order to ensure transcription accuracy, the segmentation and the coding done by one of the authors (FA) were checked by a second author (CA). Cases where there was disagreement were discussed to arrive at a consensus.

We limited this investigation to voiceless coronal fricatives, with the aim of identifying their acoustic realization in Italian and comparing them to the corresponding consonants on the same speakers' dialect speech. Therefore we used the same acoustic analysis and articulatory interpretation as in our previous report on voiceless coronal fricatives in their dialect (for full details see Avesani et al., 2015).

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) that was simultaneously accompanied by high-frequency 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).

A common way to acoustically analyse fricatives is to consider their power spectrum as a probability distribution, a technique that allows simultaneous capture of both local (spectral peak) and global (average spectral shape) features of a power spectrum (Jongman et al., 2000: 1253). Such an analysis turned out to be sensitive enough to index fine-grained articulatory differences such as those between the

[17] The number of tokens produced by each speaker is: AM = 191; JF = 149; CZ = 183; GP = 96. Of the four speakers, GP was the most parsimonious, and AM the most loquacious, during this task.

interdental versus the lamino-denti-alveolar fricatives described for the dialects of Belluno and Rovigo, respectively (see Avesani et al., 2015).

In a spectral moment analysis four measures are calculated: the spectral center of gravity (spectral mean or centroid frequency), standard deviation (or variance), skewness, and kurtosis[18]. These spectral moments (SMs) are also 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, and a negative value indicates that the left tail of the distribution extends further than the right. Skewness is the inverse of spectral tilt: positive skewness indicates negative tilt with a concentration of energy in the lower frequencies, while negative skewness indicates positive tilt, a predominance of energy in the higher frequencies. The fourth spectral moment, kurtosis (Kurt), indicates the peakedness of the distribution: a positive value indicates relatively high peakedness (the higher the value, the more peaked the distribution, with just one or a few relatively sharp peaks), while a negative value indicates a relatively flat distribution. Positive kurtosis thus suggests one or a few clear, well-resolved peaks, while negative kurtosis indicates a lack of clearly defined peaks.

Therefore, 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 relative to the length of the oral cavity (Li et al, 2009: 112). The lower the CoG value is, 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 indicates a concentration of energy in the lower frequencies, below the mean, suggesting a more posterior place of articulation. SDev and Kurt distinguish a compact and peaky spectral shape, respectively, from a diffuse and flat one. Thus, they help in differentiating fricatives based on their tongue posture, specifically apical vs. laminal tongue tip posture.

In order to compute the four spectral moments we adopted the time-averaging technique proposed by Shadle (2012)[19]. Adapting a PRAAT script by Di Canio (2013)[20], which implements Shadle's suggestion, prior to analysis we down-sampled

---

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

19 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).

20 http://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives.praat

all recordings to 48 kHz 16-bit and applied a 300 Hz low pass cut-off filter to remove any F0-related influence. Then CoG, SDev, Skew and Kurt were computed over the central 80% of the fricative segment's duration[21] using 5 DFTs with an analysis window set to 10ms. Duration, RMS intensity (dB), number of zero crossings, frequency and intensity of the spectral peak were also extracted for each target consonant.

4. *Results*

In the present section we present and comment the results of our analyses and will proceed as follows. First we will report the results of the distribution of the target consonants in the whole set of the 619 tokens analysed for the four speakers (§ 4.1). Stops are presented in § 4.1.1, affricates in § 4.1.2 and fricatives in § 4.1.3. In § 4.1.1 coronal stops as produced in VRI are compared to coronal stops as produced by the same speakers, and in the same recording session, in the local dialects. The analysis specifically focuses on the voiced coronal stop /d/ as both NeVen and CVen dialects display lenition of /d/ in intervocalic position. Therefore, a direct comparison of the /d/ allophones produced in VRI and in dialect provides a basis for verifying whether or not dialect phonetic details have transferred into any speaker's Italian.

After an interim summary (§ 4.2), in § 4.3 we present the results of the acoustical analysis of a selection of these tokens, focusing on the unvoiced coronal fricatives. In this set we include also the postalveolar affricate [ʧ], of which we will acoustically analyse only the fricative release phase ([ʃ]).

In the acoustical analysis of unvoiced fricatives (§ 4.3), we verify whether a three-way phonetic contrast exists, in VRI, in the coronal region as evidenced by the auditory analysis in § 4.1.3. We check whether lamino-alveolar, apico\retracted-alveolar and post-alveolar fricatives previously identified significantly differ in their spectral moment properties (§ 4.3.1). In § 4.3.2 we compare [s] and the fricative release phase of the post-alveolar affricate [(t)ʃ] in VRI and in the local dialects, as [s] and [(t)ʃ] are consonants shared by all varieties. If the fine phonetic details of the substrate dialectal consonants affect the pronunciation of same consonants in VRI as posited in § 2 in the *Attrition* account (specifically b1: attrition of Italian features, which are replaced by dialect features), then the acoustic features of the shared consonants should be comparable between the speaker's local dialect and their VRI.

4.1 Target consonants: Distributional analysis

4.1.1 Coronal stops

The total number of words presenting the voiceless stop /t/ amounts to 117 tokens occurring in 26 words (53 tokens produced by BL speakers, 64 by RO speakers). All were realized as expected, with no deviation from Standard Italian (SI): the coronal stop /t/ was always produced as a dental stop.

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

As for the voiced stop, we examined 67 occurrences of /d/, 9 tokens occurring word-initially and 58 occurring in intervocalic position, both within and across words. In all contexts /d/ is always realized as [d], as expected in SI.

In the local dialects, on the contrary, /d/ undergoes lenition in intervocalic context, as reported in the literature (e.g. Zamboni, 1988). To verify whether lenition of /d/ also occurs in the local dialect of our immigrant speakers, we analysed a set of words whose Italian counterparts present /d/ in intervocalic position (e.g., NeVen dialect [peˈð̞ɔʧo], SI [piˈdɔk:jo], *louse;* NeVen [ˈreð̞en], SI [ˈredine], *bridle*), including words from the picture naming task and from spontaneous production. We collected a total of 94 cases occurring within and across words. Globally, the allophones produced amount to 52 tokens of approximant [ð̞], 39 tokens of stop [d], 3 tokens of fricative [ð]. This shows that lenition of /d/ applies also in the dialect as spoken by our immigrant speakers. However, the lenited outputs of /d/ differ quantitatively among speakers. AM, the CVen female speaker, produced only [d], showing no lenition in intervocalic context. On the contrary, the NeVen male speaker, GP, did not produce any stop in intervocalic context, showing a free variation between a fricative [ð] and an approximant [ð̞] allophone skewed toward the approximant output ([ð]: 2/20 cases; [ð̞]: 18/20 cases). The remaining speakers show a free variation of [d] and [ð̞ / ð]: the NeVen female speaker CZ produces [d] and [ð̞] showing a preference for a non lenited allophone ([d]: 24/36; [ð̞]: 12/36); the CVen male speaker JF alternates between [d], [ð̞] and [ð] with a preference for the approximant allophone ([ð̞]: 22/29, [ð̞]: 6/29, [ð]: 2/29).

It is worth noting that while the literature attests a dental fricative as the only outcome of /d/ lenition in Veneto dialects, in our dialect data dental approximants were produced as output of /d/ lenition much more often than dental fricatives.

When speaking Regional Italian, however, none of the speakers produced any instance of [ð] or [ð̞]. The VRI data show no evidence of /d/ lenition and suggest that no transfer from the local dialect to Italian has taken place as far as coronal stops are concerned.

4.1.2 Coronal affricates

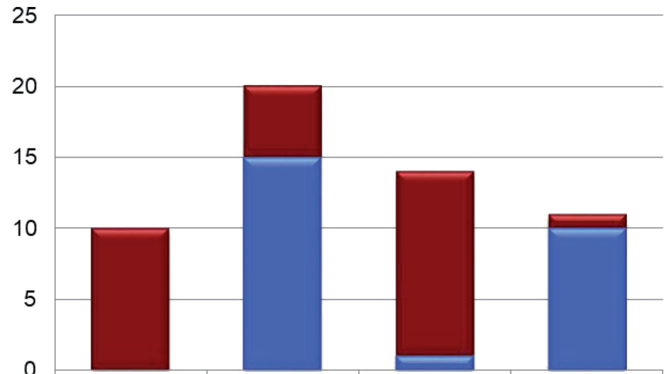
On the other hand, the analysis of how Italian coronal affricates were realized reveal noticeable evidence of transfer from dialect into Italian. The IRIAS corpus includes 7 Italian words with an instance of dental affricates (*zappa, zoccoli, zoppo, cavezza, cazzuola, fazzoletto, pezza*) and 11 words with one or two instances of post-alevolar affricates (*cenere, cento, cesta, ciliegia, cimice, cimitero, cipolla, damigiana, laccio, riccio, salice*). In word-initial position, as we have seen in § 1.3 dental affricates are produced as voiced (<*z*> in, e.g., *zappa (hoe), zoccoli (throwels*) despite the fact that the corresponding words in dialect show an unvoiced fricative in the same position. In the remaining words of the set, where the dental affricates appear in intervocalic position as geminates (<*zz*> as in *cavezza, pezza*), affricates are unvoiced in Veneto Regional Italian as in Standard Italian and are de-geminated[22]. In the word set of post-alveolar affricates, af-

[22] Northern Italian dialects lack consonant gemination. In RI as spoken in the northern regions of Italy either consonant gemination is absent or, if geminated consonants are produced, they are shorter

fricates that are orthographically represented as <*ce, ci*> are always unvoiced (e.g. [ʧ]) in all word positions, while those spelled with <*gi*> are produced as voiced [ʤ], which occurs in our Italian recordings of these speakers only in the words *ciliegia* and *damigiana*. Speakers spontaneously produced other instances of affricates in words not included in the list of target words. We obtained a total of 169 tokens: 52 with dental affricates and 117 with post-alveolar affricates (regardless of voicing status).

The dental affricates showed a strong tendency toward de-affrication and this was notably more frequent in the males than the females. Out of 55 expected cases of f dental affricates (29 voiced and 26 unvoiced) only 47% of them were realized as affricates (26/55), while the remaining 53% (29/52) were realized as fricatives. If we pool the number of affricates by the dialectal origin of the speakers, we obtain a balanced distribution: the two NeVen speakers were expected to produce 30 tokens of dental affricates and realized as such only 15 cases (50%); the two CVen speakers produced only 11 of the expected 25 cases as affricates (44%). However, pooling the data by gender, the picture reveals a highly unbalanced distribution: females were expected to produce 31 tokens of dental affricates and realized them as such 80,5% of the time (25/31) whereas only one affricate was produced by the males (4% of expected cases, e.g. 1/24). Examining the distribution by gender and by dialectal origin of the speakers (Figure 2), the NeVen male speaker GP never uttered any dental affricates, while the CVen male speaker JF uttered one. By comparison, the NeVen female produced a dental affricate in 15 of 20 expected cases (75%), while the CVen female speaker produced them in 10 of 11 expected cases (91%). Figure 1 summarizes such results.

Figure 1 - *Realization of expected affricates by speakers. Belluno speakers: GP (male) and CZ (female); Rovigo speakers: JF (male) and AM (female)*



	GP	CZ	JF	AM
fricatives	10	5	13	1
affricates	0	15	1	10

in duration with respect to the same consonants in SI, therefore reducing the phonetic difference between singletons and geminates. As de-gemination is a common feature in all northern Italian dialects and varieties of RI we do not take it as a specific case of transfer from Veneto dialect to Italian. We note however that this feature has been maintained also in the heritage VRI of our speakers.

Of the 29 cases of fricative output for expected dental affricates, all voiceless targets were produced as voiceless fricatives (12 cases). Voiced dental affricate targets produced as fricatives (17 cases) had voiced realizations except for one, which had a voiceless realization. It is interesting to note that the resulting voiced fricatives kept the same place of articulation as the fricated portion of the corresponding affricate, i.e., were realized as alveolar [z], while the voiceless fricative differed in place of articulation from the fricative portion of the dental affricate for two speakers. The male CVen speaker produced only two words that in Italian contain a voiceless dental affricate: one was produced as a dental affricate, the other was realized as a de-affricated lamino-denti-alveolar [s̻̪], which is the realization of the corresponding fricative of his dialect. However, the same fricative [s̻̪] also appears in all cases of de-affrication (6) by the NeVen female, whose dialect does not include such a lamino-denti-alveolar [s̻̪], providing further evidence of variability in the pronunciation of /s/ that may be attributed to a possible mixture of phonetic features from other Veneto linguistic varieties.

The post-alveolar affricates displayed much less de-affrication than did the dental affricates, and did so only for NeVen speakers. Out of 117 expected voiced and unvoiced affricates, 107 were indeed produced as such. Only 10 were realized as fricatives, maintaining the expected voicing (4 voiceless, 6 voiced post-alveolar fricatives). These fricatives were all produced by the NeVen speakers, more by the male (7/10, 4 voiceless, 3 voiced) than by the female (3/10, all voiced). Neither of the CVen speakers de-affricated any post-alveolar affricates.

4.1.3 Coronal fricatives

Our informants produced 183 cases of fricatives. In this set, the realizations that characterize the dialects of the NeVen and the CVen systems – respectively, the interdental [θ] and the lamino-denti-alveolar fricative [s̻̪] – are very few (9/183). The only token of [θ] in VRI was produced by the male of NeVen (GP), and it occurred in [kaˈθola], the word that GP used when speaking dialect. It is possible that he switched to dialect because he did not know the corresponding word in Italian ([kaˈts:wɔla], *trowel*): indeed, this is the speaker who acknowledged having scant practice of Italian. Similarly, the male of the CVen (JF) produced in VRI one token of the lamino denti-alveolar [s̻̪] that occurs in his dialect: his rendition of the SI word [kaˈvets:a] (*halter*) is [kaˈves̻̪a].

The remaining 7 cases of the lamino denti-alveolar fricative typical of CVen are unexpectedly produced by the female NeVen speaker (CZ) in words such as [ˈpɛs̻̪o] (*piece*), [fas̻̪oˈleto] (*handkerchief)*, [puˈpas̻̪o] (*puppet*), [ˈs̻̪apa] (*hoe)*, in which the fricative [s̻̪] substitutes the expected SI affricate [ts] in the first three words and the expected [dz] in the last one. Such a substitution is unexpected because that fricative does not occur in the phonological system of her BL dialect, and, as expected, she never used it uttering the corresponding words in dialect. In such words she used the interdental [θ] and the alveolar [s] instead: [ˈpɛθa], [faθoˈlet], [ˈsapa] (see Avesani et al., 2015).

The previous findings show evidence of a weak transfer of specific fricatives of the local dialect to the Italian spoken by the male Veneto speakers, and also show an interesting transfer of broader (not strictly local) dialectal features in the Italian speech of the female Central Veneto speaker CZ. Her use of fricatives that occur in the dialect of another Veneto subsystem (the North-Eastern one) show a mixing of features possibly induced by contact with other Veneto linguistic varieties.

Excluding the preceding 9 cases, of the remaining 174 cases, 28 are occurrences of voiced [z], 146 are voiceless alveolars. The narrow phonetic transcription of the voiceless fricative data reveals that 66% of total occurrences of VRI /s/ (96/146) are lamino-alveolar [s] as in Standard Italian, while in the remaining 33% of the cases (48/146) are apico-alvevolar with, possibly, a more back place of articulation (but not as far back as a post-alveolar [ʃ]). We chose to transcribe this variant as [s̠], using the IPA symbol for the alveolar fricative and the diacritic that indicates retraction. This type of fricative is not used as a substitute of SI /ts/ in words that have been de-affricated, but it alternates with [s] in words that in SI present /s/. [s̠] is produced by both NeVen speakers and by the male CVen speaker, while the female CVen speaker (AM, who professes to use dialect very little and Italian more often) uses only the SI allophone [s].

A set of Pearson χ^2 tests were conducted on the alveolar fricatives produced by 3 speakers (CZ, GP and JF, 108 tokens in total: [s] = 60, [s̠] = 48) in order to verify whether the choice of one over the other allophone depends on any of the following factors: position in the word (initial vs. non-initial[23]); stress of the syllable in which the fricative occurs (stressed vs unstressed); the following phonetic context (V vs C); the type of following vowel (front, central, back).

These three speakers produced comparable numbers of coronal fricatives (NeVen: CZ = 38, GP = 35; CVen: JF = 35), and the distribution of plain [s] versus retracted [s̠] did not differ significantly among them (χ^2 = 2.438; p = 0.295; number of tokens: CZ: [s] = 18, [s̠] = 20; GP: [s] = 21, [s̠] = 14; JF: [s] = 21, [s̠] = 14). The stress status of the syllable in which the fricatives occur had a non-significant effect on incidence of [s̠] (χ^2 = 0.337; p = 0.561), while their incidences are significantly affected by position in the word as well as by the following phonetic context. Occurrence of [s̠] was significantly higher in word-initial position than elsewhere (χ^2 =3.95; p = 0.047). There is also a higher probability that [s̠] precedes a vowel than a consonant (χ^2 = 7.896; p = 0.005), but does not differ by the type of following vowel (χ^2 = 1.483; p = 0.480).

4.1.4 Interim summary

The data presented so far show clear evidence of transfer from dialect to Italian for dental affricates. As this features is attested in VRI as spoken in Italy (Canepari, 1984), the data indicate that such feature of VRI is maintained in three out of four speakers. The data also show that de-affrication is not evenly distributed among the

[23] As only one case of [s̠] in word-final position is attested in our data, we did not take into consideration a "null" following phonetic context as a relevant level in the χ^2 tests.

speakers: males de-affricate more than females, independently from their language variety. Moreover, our analysis has shown that de-affrication applies also to post-alveolar affricates, though to a much lesser extent. As Canepari does not mention it and we do not have speech data yet of monolingual speakers currently living in the same areas in Veneto to compare with our heritage speakers, we cannot tell whether de-affrication of post-alveolar affricates is an innovative feature of heritage VRI induced by the contact with other varieties of RI or whether it is applied by our speakers to all affricates by analogy to de-affrication of dentals.

As for fricatives, three speakers (GP, CZ, JF) use two allophones of /s/: a "plain" SI lamino-alveolar [s] and an apical/retracted variant [s̠]. As Canepari (1984) claims that in VRI alveolar fricatives are apical or retracted, the additional presence of a lamino-alveolar [s] in their speech may indicate a step toward standardization of their Regional Italian. However, one speaker (CZ) also shows evidence of koinization, as the presence in her VRI of a coronal fricative typical of the local dialect of another Veneto sub-system indicates a mixing of phonetic features of different language varieties and may prelude to a subsequent levelling of the most marked forms (Siegel, 2001).

4.2 Acoustics of the Veneto speakers' Italian fricatives

4.2.1 Veneto Regional Italian

In the distributional analyses of VRI coronal fricatives presented above, we found that speakers varied in their productions; specifically, three of the four speakers alternated between a standard alveolar [s] and an alveolar with a relatively more posterior articulation that we labeled [s̠] (retracted). Excluding the female CVen speaker who only produced plain alveolar fricatives, we conducted χ^2 tests on the other three speakers, who did not differ in distribution of [s] and [s̠], and produced [s̠] significantly more often in word-initial and pre-vocalic than pre-consonantal positions.

Three of our four speakers alternated between plain and retracted alveolars, suggesting the presence of three phonetic categories in the coronal region for fricatives: lamino-alveolar, apico-/retracted alveolar and post-alveolar. This three-way partition of coronal fricative place of articulation differs from that displayed by both of the local dialects of the same speakers: the NeVen dialect has interdental, lamino-alveolar and postalveolar fricatives, while CVen dialect has lamino-denti-alveolar fricatives where NeVen has interdental ones.

In order to determine how the three fricatives differ acoustically in the speakers' VRI productions, we ran separate one-way Repeated Measure ANOVAs on each speaker for each of the four spectral moments measures with Type of Fricative ([s] vs. [s̠] vs. [(t)ʃ], i.e., the fricative portion of the affricate) as the independent factor. Details of these analyses are presented in Table 4, and mean spectral moment values in Table 5. For CVen speaker AM, who never uttered a retracted [s̠], the independent factor had only two levels: ([s] vs. [(t)ʃ]). Her [s] differed significantly from [(t)ʃ] on all spectral moments. A lower mean value of CoG and a higher value of Skew (see Table 5) indicate a longer front resonating cavity and suggest a more posterior place of articulation of her post-alveolar fricative; while a lower value of SDev indicates a less diffuse spectrum and suggests a non-laminal tongue posture for [(t)ʃ], relative to her alveolar [s].

The other three speakers all displayed a three-way contrast ([s] vs. [s̠] vs. [(t)ʃ]). For the CVen male speaker JF and the NeVen female speaker CZ, none of the spectral moments distinguished significantly among the three fricatives. Only the NeVen male speaker GP displayed differences among the three places of articulation, which were significant for three spectral moments and nearly significant for the fourth (Kurt). Tukey-Kramer HSD post-hoc tests on GP's data revealed that [s̠] failed to differ significantly from the fricative release of [tʃ] for any spectral moments, but that [s] differed significantly from both of the other fricatives on all spectral moments.

Table 4 - *F-values for one-way repeated measures ANOVA results for each speaker on each spectral moment of the target fricatives*

Contrast	*Language variety*	*Speaker*	*CoG*	*SDev*	*Skew*	*Kurt*
[s],[(t)ʃ]	RO	AM	$F(1, 36) = 47.441$ $p < 0.0001$	$F(1, 36) = 30.534$ $p < 0.0001$	$F(1, 36) = 39.309$ $p < 0.0001$	$F(1, 36) = 29.747$ $p < 0.0001$
[s], [s̠], [(t)ʃ]	RO	JF	n.s	n.s.	n.s.	n.s.
	BL	CZ	n.s.	n.s	n.s	n.s.
	BL	GP	$F(2, 33) = 9.82$ $p < 0.0005$	$F(2, 33) = 5.444$ $p = 0.0037$	$F(2, 33) = 3.988$ $p = 0.0281$	$F(2, 33) = 3.048$ $p = 0.061$

Table 5 - *Mean values of the spectral moments for each fricative by speaker*

Language variety	*Speaker*	*fricative*	*CoG*	*SDev*	*Skew*	*Kurt*
RO	AM	[s]	4800.59	2273.78	1.16	3.33
		--	--	--	--	--
		[ʃ]	3363.91	1542.28	2.59	4.12
RO	JF	[s]	4709.81	1803.04	2.23	10.08
		[s̠]	4763.27	1576.48	2.42	14.17
		[(t)ʃ]	4689.25	1640.00	1.77	8.24
BL	CZ	[s]	4167.49	1732.91	1.52	9.29
		[s̠]	4110.88	1658.73	1.83	11.31
		[(t)ʃ]	4167.49	1799.82	1.50	8.55
BL	GP	[s]	3849.00	1384.16	2.36	16.84
		[s̠]	3205.88	1044.45	4.10	38.31
		[(t)ʃ]	3413.42	1053.20	4.32	41.74

4.2.2 Italian and dialect fricatives compared

The [s] and [(t)ʃ] consonants are shared by Italian and by the local dialects of our speakers. The coronal fricatives that uniquely characterize the phonology of the NeVen and CVen dialects, i.e., the interdental [θ] and the lamino-denti-alveolar [s̪] respectively, are both absent from Italian. If the fine phonetic details of the substrate dialectal consonants percolate into VRI and affect the pronunciation of the shared consonants, as posited in the *Attrition* account (specifically b1: attrition of Italian features, which are replaced by dialect features) presented in our *Introduction,* then the acoustic features of the shared consonants should be comparable between the speaker's local dialect and their VRI.

To evaluate that hypothesis, we compared the spectral moments analysis of [s] and [(t)ʃ] in dialect reported in Avesani et al. (2015) to the corresponding analysis of the same consonants in Italian. Table 6 shows the articulatory interpretation of the results: if the consonants are not significantly different for a given SM, the phonetic symbols for the dialect and Italian fricatives are shown in curly brackets {x, y}; if they are significantly different, "x ⇒ y" indicates a significantly more advanced place of articulation in x than y, and "x > y" a significantly higher degree of laminality in x than y. Red bolded fonts indicate a change from dialect to Italian in the difference between the two consonants.

The comparison reveals that when speaking Italian the female speakers maintained the same features as in their local dialect: the NeVen speaker CZ did not distinguish [s] and [(t)ʃ] on any SM either in dialect or in VRI; the CVen speaker AM distinguished them in both languages on all SMs, producing a [ʃ] that was more retracted, i.e., articulated farther back in the vocal tract, and with a less laminal tongue shape. In contrast, the male speakers displayed different features in their fricatives when they speak dialect versus VRI, but in opposite directions. In Italian (VRI), the NeVen speaker GP produced a significantly more retracted and less laminal fricative in the post-alveolar affricate [(t)ʃ] than in [s], while in dialect he did not differentiate them on any SM. Conversely, the CVen speaker JF did differentiate these fricatives on all SMs in dialect, but produced both so variably in Italian that they showed no significant difference on any SM.

Table 6 - *Articulatory interpretation of spectral moments for each speaker. CoG ⇒: more advanced place; SDev >: more laminal; Skew ⇒: more advanced place; Kurt >: more laminal. Symbols within brackets show no significant difference. Red bolded fonts indicate a change from dialect to Italian in the difference between the two consonants*

	Dialect				*Regional Italian*			
Lang. variety	BL		RO		BL		RO	
Speaker	CZ	GP	AM	**JF**	CZ	GP	AM	JF
CoG	{[s],[ʃ]}	**{[s], [ʃ]}**	[s] ⇒ [ʃ]	**[s] ⇒ [ʃ]**	{[s],[ʃ]}	**[s] ⇒ [ʃ]**	[s] ⇒ [ʃ]	**{[s],[ʃ]}**
SDev	{[s], [ʃ]}	[s] > [ʃ]	[s] >[ʃ]	**[s] > [ʃ]**	{[s],[ʃ]}	[s] >[ʃ]	[s] >[ʃ]	**{[s],[ʃ]}**
Skew	{[s], [ʃ]}	**{[s], [ʃ]}**	[s] ⇒ [ʃ]	**[s] ⇒ [ʃ]**	{[s],[ʃ]}	**[s] ⇒ [ʃ]**	[s] ⇒ [ʃ]	**{[s],[ʃ]}**
Kurt	{[s], [ʃ]}	**{[s], [ʃ]}**	[s] > [ʃ]	**[s]> [ʃ]**	{[s],[ʃ]}	**[s]>[ʃ]**	[s] > [ʃ]	**{[s],[ʃ]}**

5. *Discussion and conclusions*

The data for these four speakers add to the socio-phonetic picture of a component of the Italian-Australian community who arrived in Australia more than 50 years ago, and provide new insights into how they have negotiated the languages that form their linguistic repertoire. In a previous study that focused on Veneto dialect of these speakers' native language (L1) (Avesani et al., 2015), it was observed that they had maintained the phonetic features that uniquely characterise their respective local dialects. Only one speaker, the female speaker AM from CVen, showed traces of initial attrition in the dialect-specific acoustic properties of the fricatives of her dialect. The authors interpreted this drift in her speech, which suggests a merger of the dialect's lamino-denti-alveolar into the alveolar fricative, as a result of her extensive contact with English rather than as an effect of contact with of Italian.

In the present study, we focused on the Veneto Regional Italian of the same four speakers, with the aim of ascertaining: 1) whether their VRI has crystallized at the stage it was at the time they left Italy; 2) whether it shows signs of attrition and, if so, whether it is due to replacement of features from the L1-dialect or from L3-English; or 3) whether their VRI shows signs of koinization due to the contact with other dialects and regional varieties of Italian.

The results on production of Italian coronal stops by these speakers indicate that at a phonetic level VRI has not drifted towards English for any of them, as their stop consonants have maintained the Italian dental place of articulation and have not shifted it back to become alveolars as in English. Nor has their VRI drifted toward dialect in any systematic way as none of the speakers lenited the Italian voiced dental stops in intervocalic position, even if some of them rarely used few fricatives of their local dialect.

The coronal affricates of their VRI, however, showed a strong tendency toward de-affrication as it has been attested in Veneto Italian (Canepari, 1984). As expected, de-affrication applies more to dental than to post-alveolar affricates in Italian, consistent with the fact that the former but not the latter affricates are missing in the phonemic inventory of both of the NeVen and CVen dialects. If de-affrication is the product of the long history of contact between local dialects and Italian in Italy, its presence in the VRI of our migrant speakers is compatible with the conservation of their regional Italian.

However, the amount of de-affricated consonants varies remarkably across the speakers. The difference cannot be attributed to their linguistic origin, as for a given Veneto regional variety (North-eastern and Central) males behave differently from females. Nor can it be easily attributed to a difference in their use of Italian, as the sociolinguistic questionnaires and the spontaneous interviews do not suggest a direct correlation between use and/or subjective proficiency of Italian and amount of de-affrication: for example, the Central Veneto female speaker who de-affricated only once (AM) is the one who admits using Italian only if compelled and only with interlocutors who do not speak English, while the male speaker (JF) of the same linguistic variety who, conversely, produced only one affricate, declares to speak Italian with many of his friends and to have a good mastery of it.

The variability reveals a strong gender effect: males speakers never (GP) or only once (JF) produced dental affricates and substituted them with coronal fricatives, while females produced affricates as such in 91% (AM) and in 75% of cases (CZ). Such a strong difference can be explained by Labov's (2001) second principle of linguistic variation, that states that women show a lower rate of stigmatized variants than men. De-affrication is among the few variants which are socially stigmatized in Veneto (Canepari, 1984: 101) and AM and CZ, who show a strong preference for using the standard form [ts], confirm that Labov's principle is at work also in the VRI of our heritage speakers. The two female speakers differ in their rates of de-affrication. Such variability could be traced back to the level of education in Italian they received while in Italy: AM, who almost always (91%) chose the prestige variant [ts], attended in Italy both primary and middle schools, while CZ, who chose it less often (75%), attended only primary school. It is possible that AM may have developed more metalinguistic awareness of variation in Italian than CZ.

As for their VRI fricatives, none of the speakers showed a drift toward their local dialect, as the interdental fricative of NeVen [θ] and the lamino-denti-alveolar [s̪] of CVen, amount to very few cases and appear in dialectal words used in place of the Italian ones. However, all speakers except the CVen female showed an alternation between a voiceless lamino-alveolar fricative [s] and a voiceless fricative [s̠] that sounds more retracted (but not as posterior as a postalveolar [ʃ]) and less laminal[24]. Statistical analyses showed that the retracted allophone occurred more often word-initially and pre-vocalically than non-initially and pre-consonantally. As Canepari (1984) suggested that alveolar fricatives in Veneto are apical and have a more retracted place of articulation than in Standard Italian, this result could indicate that three of our speakers maintain their VRI. Our data, though, show alternation with an alveolar that, as far as we know, has not been reported in the literature. If we assume that in Veneto all voiceless alveolar fricatives are retracted, the presence of a non-retracted alveolar allophone in alternation with a retracted one could be seen as the first step in a koinization process due to the contact with other varieties of Italian, a stage in which allophones of other varieties enter the system before a levelling can occur. However, as more specific data on fine phonetic features of alveolar fricatives in Veneto are still lacking, we cannot exclude the possibility that Veneto speakers in Veneto also alternate between the two allophones; in such a case, the alternation could be evidence of an ongoing process of dialect-to-Italian convergence that started in Italy and was maintained or even continued while in Australia.

More revealing hints of the dynamics of contact among the three languages mastered by these speakers come from the acoustic analysis of the fricatives and the fricative release of the affricate [tʃ]. The NeVen speakers CZ and GP and the male CVen speaker JF produced three phonetic categories in the coronal region: [s, s̠, (t)

[24] The alternation occurs in the set of words that in Standard Italian present a fricative. It does not apply to the fricative outcome of de-affrication.

ʃ]. But for two of them (NeVen, female CZ; CVen, male JF) the three categories are acoustically similar, as none of the Spectral Moments (SMs) significantly distinguished them due to the high acoustic variability of the exemplars within each category. Only for the male NeVen speaker GP were they acoustically different for any SM, but even here he only distinguishes [s] from [(t)ʃ] and [s̠], and does not distinguish the latter two from each other. For the fourth speaker (CVen, female AM), acoustic values indicate that [s] is significantly more anterior and more laminal than [(t)ʃ].

A comparison of alveolar and post-alveolar fricatives as produced by the speakers in their local dialect and in VRI revealed interesting variations not related to their linguistic origin. In speaking VRI, both females maintained the differentiation (AM) or the lack of it (CZ) between alveolars and postalveolars they showed when speaking dialect. Conversely, the male speakers did change their pronunciations of these fricatives between dialect and VRI: when speaking dialect, the NeVen speaker GP did not distinguish [s] from [(t)ʃ] by Center of Gravity, Skewness and Kurtosis, while in VRI he did distinguish them on all SMs; the CVen speaker JF who, on the contrary, did distinguish them in dialect for all SMs, failed to distinguish them when speaking VRI. Two different processes seem to be at work for the two males. For JF, a complete loss of acoustic differentiation between the fricatives that were clearly differentiated in the local dialect may indicate an ongoing process of convergence of his VRI toward the new variety of Italo-Australian. His Italian is in the initial stage of "mixing" in which JF, who declares to being in contact with speakers of many varieties of Italian besides Australian English, is exposed to many different exemplars of coronal fricatives and produces phones with such a high range of acoustic variation that the phonetic area of each category is included in the others.

For GP, a closer look at the SMs values suggests that the coronal fricatives of his VRI display a form of phonetic *attrition* by the coronals of his local dialect. Recall that in the NeVen phonological system /θ/ contrasts with /s/ and /(t)ʃ/[25], and that in his VRI this speaker produced [s], [s̠] as allophones of /s/ and [(t)ʃ] as the only allophone of /tʃ/. In dialect, the acoustic analysis shows that [θ] is the most anterior and laminal among his set of fricatives, and that [s] and [(t)ʃ] are distinct only by tongue posture, as indexed by a significant difference in SDev, which indicates a more apical tongue posture for [s], but they are not significantly different for place of articulation (as indexed by CoG and Skew). In VRI he still distinguishes [s] and [ʃ], but the distinction is based on significant differences in all SMs, showing a stronger degree of phonetic differentiation with respect to his dialect. The SDev mean values of the set of coronal fricatives he produced in dialect and VRI, respectively [θ], [s], [(t)ʃ] and [s], [s̠], [(t)ʃ], show that the degree of laminality in the phones of the two languages increases according to the following progression:

[25] In the present paper, as in Avesani et al. (2015), we analysed the post-alveolar fricative not per se but as the fricative release of the affricate. A palatal fricative /ʃ/ occurs in the phonological systems of Veneto's dialect and also in the corpus we analysed. However, words such as [ˈbiʃa] (*grass snake*) were excluded from our analysis due to an insufficient number of tokens.

Speaker GP	*Belluno dialect*					*Veneto Regional Italian*				
Degree of laminality as indexed by SDev	[θ]	>	[(t)ʃ]	>	[s]	[s]	>	[(t)ʃ]	<	[s̠]
	2555 (873)		1212 (321)		1126 (185)	1384 (246)		1053 (237)		1044 (257)

In the VRI set, one of the two allophones of /s/, [s], appears to have "replaced" the most laminal fricative ([θ]) of the dialect set, and the most apical fricative [s̠] took the place of the dialect [s].

In conclusion, our four Veneto speakers have shown specific paths of language change as revealed by the analysis of their spoken Italian. Individual differences in the accommodation of the languages of their multilingual repertoires emerged from the analysis of their coronal stops and obstruents: different speakers showed evidence of maintenance of some phonological and phonetic characteristics attested in their Regional Italian but also specific signs of ongoing koinization and phonetic attrition. The latter were possible to observe only through a fine-grained acoustic analysis of the frication noise, which has confirmed to be of great value for offering insights into the path of language change over time and distance.

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 as well as 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.

Bibliography

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.

Avesani, C., Galatà, V., Vayra, M., Best, C.T., Di Biase, B., Tordini, O. & Tisato, G. (2015). Italian roots in Australian soil: coronal obstruents in native dialect speech of Italian-Australians from two areas of Veneto. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*. Milano: Studi AISV, 61-86.

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.

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.

Bettoni, C. (1981). *Italian in North Queensland. Changes in the Speech of First and Second Generation Bilinguals*. Townsville: University of North Queensland Press.

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

Bettoni, C. (1985b). *Tra lingua dialetto e inglese: Il trilinguismo degli italiani in Australia*. Leichhardt, Australia: FILEF Italo-Australian Publications.

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

Cabrelli Amaro, J. (2012). L3 phonology. An understudied domain. In Cabrelli Amaro, J., Flynn, S. & Rothman, J. (Eds.), *Third language acquisition in adulthood*. Amsterdam/Philadelphia: John Benjamins, 33-60.

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

Canepari, L. (1984). *Lingua italiana nel Veneto*. Padova: CLESP.

Cavallaro, F. (2003). Italians in Australia: migration and profile. In *Altreitalie*, 26, 65-88.

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

Corazza, A., Grigoletti, M. & Pellegrini, E. (2012). *Australia solo andata. Un secolo di emigrazione veronese nella terra dei sogni*. Verona: Cierre Edizioni

Cortelazzo, M. (1972). *Avviamento critico allo studio della dialettologia italiana, Vol. III: Lineamenti di italiano popolare*. Pisa: Pacini.

De Leeuw, E., Schmid, M. & Mennen, I. (2010). The effect of contact on native language pronunciation in an L2 migrant setting. In *Bilingualism: Language and Cognition*, 13(1), 33-40.

De Mauro, T. (1972). *Storia linguistica dell'Italia unita*. Roma-Bari: Laterza.

De Mauro, T. (2014). *Storia linguistica dell'Italia repubblicana. Dal 1946 ai nostri giorni*. Roma-Bari: Laterza.

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.

Galatà, V., Avesani, C., Best, C., Di Biase, B. & Vayra, M. (in preparation). *The Italian Roots In Australian Soil multilingual speech corpus*.

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

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.

Kerswill, P. (2002). Koineization and accommodation. In Chambers, J.K. & Schilling-Estes, N. (Eds.), *The handbook of language variation and change*. Oxford: Blackwell, 669-702.

Labov, W. (2001). *Principles of Linguistic Change. Volume 2: Social factors*. Malden-Oxford: Blackwell.

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.

Major, R.C. (1992). Losing English as a first language. In *The Modern Language Journey*, 76, 190-208.

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

Mioni, A., Trumper, J. (1977). Per un'analisi del continuum linguistico veneto. In Simone, R., Ruggiero, G. (Eds.), *Aspetti sociolinguistici dell'Italia contemporanea*. Roma: Bulzoni, 329-372.

Montrul, S. (2013). Bilingualism and the Heritage Language Speaker. In Bhatia, T.K., Ritchie, W.C. (Eds.) *The Handbook of Bilingualism and Multilingualism*. Chichester: John Wiley & Sons, 168-188.

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, E.M. & Davydova, J. (Eds.), *Multilingualism and Language Diversity in Urban Areas. Acquisition, identities, space, education*. John Benjamins, 19-38.

Rando, G. (1973). Italiano parlato d'Australia. In *Il Veltro*, 17, 247-252.

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

Schmid, M.S. (2011). *Language Attrition*. Cambridge: Cambridge University Press, 11-17.

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.

Siegel, J. (1985). Koines and koineization. In *Language in Society*, 14(3), 357-378.

Siegel, J. (2001). Koine formation and creole genesis. In Smith, N., Veenstra, T. (Eds.), *Creolization and contact*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 175-197.

Tagliavini, C. (1963). *Elementi di fonetica generale*. Bologna: Patron.

Tosi, A. (1991). *L'italiano d'oltremare. La lingua delle comunità italiane nei paesi anglofoni/Italian overseas. The language of the Italian communities in the English-speaking world.* Firenze: Giunti.

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

Valdés, G. (2000). Introduction. In Anderson, N. (Ed.), *Spanish for native speakers. AATSP professional development series handbook for teachers K-16: Vol. 1*. New York: Harcourt College, 1-20.

Valdés, G. (2005). Bilingualism, Heritage Language Learners, and SLA Research: Opportunities Lost or Seized? In *The Modern Language Journal*, 89(3), 410-426.

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/O2103AISV000015

LORENZO CIAURELLI, DANIELA BELTRAMI

Spontaneous speech in patients with early-stage dementia and Mild Cognitive Impairment: The role of age of acquisition

The aim of this study is evaluating differences in the characteristics of words produced by patients with early-stage dementia(e-D) and Mild Cognitive Impairment (MCI) and healthy control (HC). We used two different corpora to obtain an age of acquisition (AoA) and word frequency values for words produced by subjects in two semi-spontaneous speech tasks. The results are in line with the tendencies found in previous studies for English and Italian (Silveri *et al.*, 2002; Forbes-McKay *et al.*, 2005; Rodríguez-Ferreiro *et al.*, 2009; Cuetos, Herrera & Ellis, 2010), that is the words produced by e-D patients have a lower age of acquisition value than those produced by MCI or HC. Concerning word frequency value, no significant difference was found between words produced by these different populations.

Key words: Early-stage dementia, Alzheimer's disease, Mild Cognitive Impairment, age of acquisition, spontaneous speech.

1. *Introduction*

Among the heterogeneity of symptoms related to dementia, language impairments are already present at the very early stage of cognitive decline (Taler, Phillips, 2008; Olney, Spina & Miller, 2017). These deficits vary along the course of the disease. Moreover, in the initial phase there is a prominent decline at lexical-semantic level, whereas phonological and syntactic abilities are relatively well preserved (Caramelli, Mansur & Nitrini, 1998). Because of their involvement in the preclinical phase of the disease, the deficits affecting the linguistic abilities can be used as clues for early diagnosis and dementia large-scale screenings. A number of studies, also based on the new sophisticated techniques from Natural Language Processing (NLP), have already demonstrated that linguistic features can be used for detecting and classifying dementia prodroms (Snowdon, Greiner & Markesbery, 2000; Chapman *et al.*, 2002; Jarrold *et al.*, 2010; Beltrami *et al.*, 2016).

The word retrieval impairment is one of the deficits affecting the language abilities in patients with dementia and it seems to follow the rule of "last-in, first-out" (i.e. the words acquired later are more vulnerable to cognitive decline). This process is also an effect of aging (Hodgson, Ellis, 1998), but it appears more severe in the cognitive decline (Holmes, Fitch & Ellis, 2006). For this reason, several studies have successfully used age of acquisition of words (i.e. the age at which a word is acquired) to assess the severity of cognitive decline (Silveri *et al.*, 2002; Forbes-McKay *et al.*, 2005; Rodríguez-Ferreiro *et al.*, 2009; Cuetos, Herrera & Ellis 2010). Furthermore, most of these studies have revealed that age of acquisition has an effect independent of word frequency (Morrison, Ellis, 1995; Räling, Schröder & Wartenburger, 2016). Such experiments usually rely on verbal fluency or picture naming tasks and use two different values of AoA: a) objective (directly from children) and b) subjective (by adult rating).

In this paper, we present a pilot study aimed to evaluate the role of AoA in the speech production of 48 subjects belonging to three different populations (early-stage dementia-e-D, Mild Cognitive Impairment – MCI and healthy control – HC). To this end, we analyzed the spontaneous speech of subjects in a semi-automatic way. Notwithstanding the spontaneous speech has a more intra- and inter-personal variability compared to structured evaluations like verbal fluency task, it allows a more naturalistic assessment of language abilities (Bucks *et al.*, 2000).

We expected that the subjects affected by cognitive decline would have produced a higher number of words with a lower value of AoA and that the more severe the impairment, the more evident the trend. Furthermore, we wanted to determine if the AoA value has an effect independent of word frequency.

2. *Data collection*

The sample was composed of 96 subjects (48 male and 48 female, age range 50-75, mean 63.3, SD 7.2): 48 healthy controls and 48 with cognitive decline previously evaluated with medical and neuropsychological assessment. The group with cognitive decline was composed of 32 with MCI and 16 with e-D. All subjects were native Italian speakers. Each subject underwent a neuropsychological screening composed of some of those tests considered the most reliable for the diagnosis of MCI or dementia (Grober *et al.*, 2008; Ismail, Raji & Shulman 2010): Mini Mental State Examination – MMSE, Montreal Cognitive Assessment – MoCA, General Practitioners assessment of Cognition – GPCog, Clock Drawing Test – CDT, Verbal fluency (phonemic and semantic) and the Paired Associate Learning (PAL, subtest of the Cambridge Neuropsychological Test Automated Battery – CANTAB).

The spontaneous productions of subjects were collected during the execution of three tasks, elicited by these input sentences:

a) "Describe this picture";
b) "Describe your typical working day";
c) "Describe the last dream you had or remember".

This data was collected, transcribed and POS-tagged under the OPLON project ("OPportunities for active and healthy LONgevity", Smart Cities and Community – DD 391/RIC, co-funded by Ministry of Education) (see Beltrami *et al.*, 2016 for a more detailed description of data collection procedures).

3. *Data analysis*

To assign an AoA value to the words produced by the subjects in the three different tasks we used two corpora available online:

- *AoAObj*. It contains 223 Italian nouns acquired by children up to 11 years old. The AoA value was objective, that is obtained directly from the children (Lotto, Surian & Job, 2010);
- *AoASubj*. It contains626 Italian nouns acquired by children up to 13 years old. The AoA value was subjective, that is obtained by adult rating (Barca *et al.*, 2002).

The two corpora were different in their composition[1]. They shared 79 tokens, that is 35% of the total of AoAObj and 12% of the total of AoASubj. Moreover, 79% of the lemmas of AoAObj and 99% of AoASubj were drawn from *Vocabolario di Base della Lingua Italiana* (De Mauro, 1980).

Table 1 - *Means number of the words and nouns produced by the subjects in each task and the nouns/words and matched-nouns/nouns ratio grouped by task, group and corpora*

Corpora	*Group*	*Task*	*Tot Words (mean)*	*Tot Nouns (mean)*	*Nouns / Tot Words (%)*	*Matched Nouns / Tot Nouns (%)*
AoAObj	CON	Picture	64.48	32.90	51.03	14.13
AoAObj	CON	Work	89.71	35.19	39.23	8.13
AoAObj	MCI	Picture	44.95	26.16	58.20	15.90
AoAObj	MCI	Work	67.00	26.26	39.20	9.82
AoAObj	e-D	Picture	29.50	16.63	56.36	18.83
AoAObj	e-D	Work	35.00	12.50	35.71	22.00
AoASubj	CON	Picture	64.48	32.90	51.03	37.05
AoASubj	CON	Work	89.71	35.19	39.23	29.89
AoASubj	MCI	Picture	44.95	26.16	58.20	41.06
AoASubj	MCI	Work	67.00	26.26	39.20	30.46
AoASubj	e-D	Picture	29.50	16.63	56.36	41.38
AoASubj	e-D	Work	35.00	12.50	35.71	36.56

After a preliminary analysis, the task "dream" was excluded from further analysis, due to low variation of lexicon.

We automatically matched the nouns contained in the corpora with the lemmas extracted from each subject production by using a script made in Python. The Table 1 shows the average number (in percentage) of the nouns mapped by this procedure,

[1] As regards to *AoAObj* corpus, in order to obtain an age of acquisition value for each noun, a set of 223 drawings were presented to a group of children aged 2-11 years and split in bands (6-months age bands for children up to 6 years old and 12-months band for children up to 11 years old). The AoA value assigned to a noun were the median value of the youngest age-level group able to reach the 75% of success in naming the drawings. The age of acquisition values contained in the *AoASubj* corpus were obtained from subjective adult ratings (on a 7-point scale) by presenting to the subjects 626 nouns printed in a booklet.

namely the nouns produced by the subjects for which we had an AoA value. The AoA and word frequency values were then given to the nouns thus obtained and for each subject and task we calculated the mean of the AoA and word frequency values. We grouped the data by task and then we performed a Kolmogorov-Smirnov non-parametric test to assess the statistical significance (p-value < 0.05) of the AoA and word frequency features.

4. *Results*

The results presented here are from a subset of 48 subjects (21 CON, 19 MCI, 8 e-D) balanced by sex and age (range 49-75, mean 62, SD 7.1). All subjects spoke Italian as their first language.

Figure 1 - *Distribution of the AoA values for the matched nouns of the task "work" using the AoAObj Corpus*

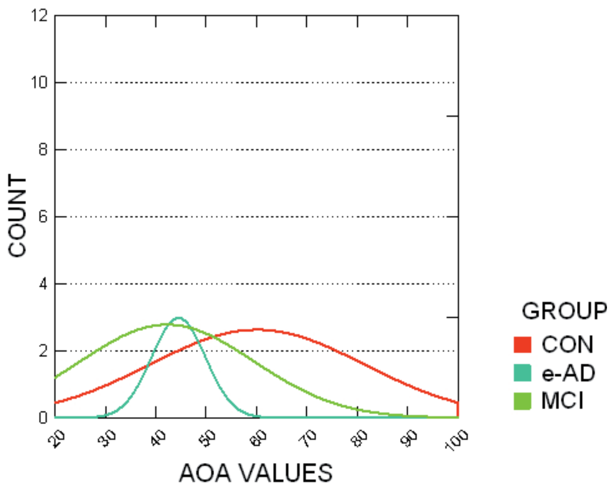


Figure 1 shows the distribution of AoA values assigned to the nouns matched using the AoAObj corpus in the task "work". The Kolmogorov-Smirnov nonparametric test was performed on AoA value with groups as a factor. A significant difference was found between the control group and the e-D group (d = 0.71; p-value = 0.042) in the task "work". A word frequency value was also assigned to the matched nouns and the Kolmogorov-Smirnov non-parametric test was also performed, but no significant difference was found among groups.

Figures 2 and 3 show the distribution of AoA values assigned to the nouns matched using the AoASubj corpus in the tasks "picture" and "work" respectively.

Figure 2 - *Distribution of the AoA values for the matched nouns of the task "picture" using the AoASubj Corpus*

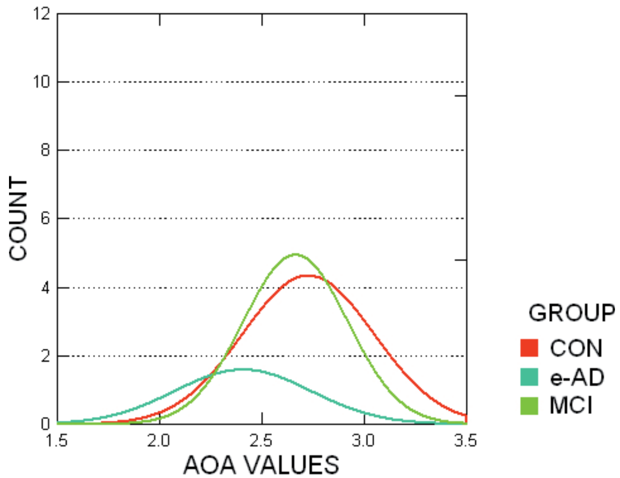
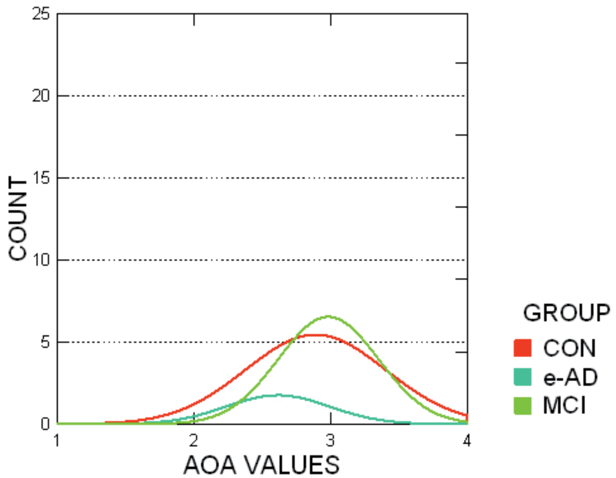


Figure 3 - *Distribution of the AoA values for the matched nouns of the task "work" using the AoASubj Corpus*



The Kolmogorov-Smirnov nonparametric test shows that AoA value can significantly differentiate the control group from e-D group (d = 0.57; p-value = 0.027) in the task "picture". Another significant difference was found between the MCI and e-D groups (d = 0.63; p-value = 0.048) in the task "work". Finally, a word frequency value was also assigned to the matched nouns and the Kolmogorov-Smirnov non-parametric test also performed, but no significant different was found among groups.

5. *Discussion and Conclusion*

This preliminary study seems to confirm for Italian the tendencies found in previous studies for English. In fact, using the AoAObj corpus we were able to differentiate the control group from the e-D group in the task "work" and, as expected, the control group had greater AoA values than the e-D group. A similar trend was found using the AoASubj corpus as the control group had greater AoA values than the e-D group in the task "picture" while the MCI group had greater AoA values than the e-D group in the task "work". No significant differences were found between MCI and control group in any tasks.

The choice of analyzing only the nouns is supported by the literature: it seems that the ability to refer to action (verbs) is relatively well preserved compared to object naming performance (nouns) (Williamson, Adair, Raymer & Hellman, 1998; Fung *et al.*, 2001).

In trying to interpret these results, we have to take into account some limitations of the experimental design. First, this kind of experiment generally relies on the verbal fluency task because spontaneous speech has more intra-personal variability. Moreover, the corpora used to match the nouns and to assign them an AoA value were too small (223 and 626 tokens respectively); indeed, the rate of mapped nouns was low. Finally, given that the years of education are significantly lower in the e-D group than in the CON and MCI groups, our results may have been confounded by this variable. This limit did not interfere with the neuropsychological results since most of the cognitive tests were adjusted for years and education according to the respective standardizations.

The two corpora differ in their composition (that is the percentage of nouns drawn from Vocabolario di Base della Lingua Italiana) and in the kind of AoA value assigned to the lemmas (objective vs. subjective). Comparing the two different corpora is out of the aim of this work. We can just assert, in line with the literature (Morrison, Chappel & Ellis, 1997; Lotto *et al.,* 2010), that the adult rating seems to be a reliable method to estimate the AoA value of acquired words. Due to these restrictions, this study must be considered as a pilot study aiming to assess the feasibility of utilizing available corpora to automate the analysis of a large sets of spontaneous speech samples. Nevertheless, some conclusions can be drawn.

The cognitive decline seems to be accompanied by loss of the words acquired later in life. In this process, the word frequency has no role and this could be due to the different origin of AoA and word frequency effects. In fact, AoA correlates highly with semantic variables (for example imageability or concreteness), but less with lexical variables such as word frequency (Morrison *et al.*, 1997; Räling *et al.*, 2016). So, it is likely that in the earliest phases of cognitive decline due to dementia the brain areas underlying the semantic process are more impaired than the areas responsible for lexical process (Taler, Phillips, 2008). Semantic impairment is also well documented in literature for MCI. So, the lack of statistical significance between MCI and control groups is more difficult to be accounted for. Given the heterogeneity in MCI populations and his fluctuating cognitive dysfunction (Feldam, Jacova, 2005), it is possible that analysis by group fails to pinpoint a common, subtle deficit. Another possible explanation may arise from the fact that the data on age of acquisition value drawn from spontaneous speech fails to capture the difference between intentional and automatic access to semantic memory. In fact, while it is well documented that even in the early-stage dementias like Alzheimer's disease or FTD both intentional and automatic access are impaired (Taler, Phillips, 2008; Olney, Spina & Miller, 2017), it seems that in Mild Cognitive Impairment only automatic access is inhibited (Duong *et al.*, 2006; Taler, Phillips, 2008).

A further study aimed to correlate type and duration of hesitation phenomena with the age of acquisition value of the nouns involved in the hesitations itself could enlighten the nature of deficit in semantic access in Mild Cognitive Impairment.

Bibliography

Barca, L., Burani, C. & Arduino, L.S. (2002). Word naming times and psycholinguistic norms for Italian nouns. In *Behavior Research Methods, Instruments, & Computers*, 34, 424-434.

Beltrami, D., Calzà, L., Gagliardi, G., Ghidoni, E., Marcello, N., Rossini Favretti, R. & Tamburini, F. (2016). Automatic identification of Mild Cognitive Impairment through the analysis of Italian spontaneous speech productions. In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J. & Piperidis, S. (Eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation*. Portorož, Slovenia, 2086-2093.

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.

Caramelli, P., Mansur, L. & Nitrini, R. (1998). Language and communication disorders in dementia of the Alzheimer type. In Stemmer, B., Whitaker, H.A. (Eds.), *Handbook of neurolinguistics*. San Diego: Academic, 463-473.

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.

Cuetos, F., Herrera, E. & Ellis, A.W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. In *Neuropsychologia*, 48, 3329-3334.

De Mauro, T. (1980). *Guida all'uso delle parole*. Roma: Editori Riuniti.

Duong, A., Whitehead, V., Hanratty, K. & Chertkow, H. (2006). The nature of lexico-semantic processing deficits in mild cognitive impairment. In *Neuropsychologia*, 44, 1928-1935.

Forbes-Mackay, K.E., Ellis, A.W., Shanks, M.F. & Venneri, A. (2005). The age of acquisition of word produced in a semantic fluency task is highly predictive of early Alzheimer's disease. In *Neuropsychologia*, 43, 1128-1137.

Fung, T.D., Chertkow, H., Murtha, S., Whatmough, C., Peloquin, L., Whitehead, V. & Templeman, F.D. (2001). The spectrum of category effects in object and action knowledge in dementia of the Alzheimer's type. In *Neuropsychology*, 15, 371-379.

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.

Hodgson, C., Ellis, A.W. (1998). Last in, first to go: Age of acquisition and naming in the elderly. In *Brain and Language*, 64, 146-163.

Holmes, S.J., Fitch, F.J. & Ellis, A.W. (2006). Age of acquisition affects object recognition and naming in patients with Alzheimer's disease. In *Journal Clinical and Experimental Neuropsychology*, 28, 1010-1022.

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.

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.), *Proceedings of Brain Informatics: International Conference*. Toronto, Canada, 299-307.

Lotto, L., Surian, L. & Job, R. (2010). Objective Age of Acquisition for 223 Italian Words: Norms and Effects on Picture Naming Speed. In *Behavior Research Methods*, 42, 123-133.

Morrison, C.M., Ellis, A.W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. In *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 116-133.

Morrison, C.M., Chappell, T.D. & Ellis, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. In *Quarterly Journal of Experimental Psychology*, 50A, 528-559.

Olney, N.T., Spina, S. & Miller, B.L. (2017). Frontotemporal Dementia. In *Neurologic Clinics*, 35, 339-374.

Räling, R., Schröder, A. & Wartenburger, I. (2016). The origin of age of acquisition and typicality effects: Semantic processing in aphasia and the ageing brain. In *Neuropsychologia*, 86, 80-92.

Rodríguez-Ferreiro, J., Davies, R., González-Nosti, M., Barbón, A. & Cuetos, F. (2009). Name agreement, frequency and age of acquisition, but not grammatical class, affect object and action naming in Spanish speaking participants with Alzheimer's disease. In *Journal of Neurolinguistics*, 22, 37-54.

Silveri, M.C., Cappa, A., Mariotti, P. & Puopolo, M. (2002). Naming in patients with Alzheimer's disease: influence of age of acquisition and categorical effects. In *Journal of Clinical and Experimental Neuropsychology*, 24, 755-764.

Snowdon, D.A., Greiner, L.H. & Markesbery, W.R. (2000). Linguistic Ability in Early Life and the Neuropathology of Alzheimer's Disease and Cerebrovascular Disease: Findings from the Nun Study. In *Annals of the New York Academy of Sciences*, 903, 34-38.

TALER, V., PHILLIPS, N.A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. In *Journal of Clinical and Experimental Neuropsychology*, 305(5), 501-556.

WILLIAMSON, D.J., ADAIR, J.C., RAYMER, A.M. & HELLMAN, K.M. (1998). Object and action naming in Alzheimer's disease. In *Cortex*, 34, 601-610.

DOI: 10.17469/O2103AISV000016

CATERINA PETRONE, ELISA SNEED GERMAN, SIMONA SCHIATTARELLA, GIOVANNA DE BELLIS, TIM MAHRT

Disturbi cognitivi come fonte di variazione nel dialogo: analisi preliminare nella Sclerosi Multipla

Participating in a dialogic conversation requires complex skills for interpersonal coordination. During turn-taking, for instance, conversational partners have to rapidly agree on who speaks next and when. This requires that they already start planning their utterances when listening to their interlocutors. This study investigates speech planning in French dialogues, by determining whether coordination mechanisms depend on speakers' cognitive abilities. We look at the effects of working memory and speed of processing in individuals with cognitive disorders and in particular with patients affected by multiple sclerosis. It is hypothesized that differences in speakers' cognitive capacities will result in different adjustments in gaps between patients with/without cognitive deficits. To test this, we have analysed the duration of gaps game in 24 dyadic interactions (36 participants). Preliminary results suggest that longer gaps indicate more preparation time for turn-taking for patients with cognitive deficits.

Key words: Multiple Sclerosis, cognitive deficits, working memory, turn taking.

1. *Introduzione*

La Sclerosi Multipla (SM) è una malattia degenerativa del sistema nervoso centrale che provoca la distruzione della guaina mielinica che ricopre le fibre nervose. Placche sclerotiche, che nell'aspetto possono essere paragonate a lesioni, si formano nelle zone interessate dalla demielinizzazione causando la compromissione della normale trasmissione degli impulsi nervosi. Eterogenei sono i disturbi causati da tale processo e non sempre presenti in tutte le persone che contraggono questa malattia, sulla cui eziologia ancora si discute (Brassington, Marsh, 1998; Bianconi, Poggioli, Morelli, Razzaboni & Pomelli, 2006). Le problematiche possono infatti interessare sia le capacità motorie che quelle cognitive.

L'indebolimento delle abilità cognitive riguarderebbe una percentuale di pazienti che oscilla tra il 40 e il 60-70%. La velocità nel processare le informazioni è la componente cognitiva più colpita dal danno neurologico (Langdon, 2011). In molti casi si riscontra anche una forte incidenza della malattia a danno della memoria di lavoro, vale a dire il sistema che permette di mantenere e manipolare le informazioni per brevi periodi di tempo (Baddeley, 2003). La compromissione di questo sistema è legata, come si diceva, al rallentamento nel processare le informazioni, rallentamento che risulterebbe più pronunciato soprattutto nelle prove in cui ai pazienti viene chiesto di portare a termine un compito in un breve lasso di tempo

(Langdon, 2011). Un danno è riscontrato anche nel funzionamento della memoria a lungo termine, esplicita ed episodica. Compromessa è inoltre l'attenzione sostenuta così come le funzioni esecutive, implicate nella capacità di pianificare, di risolvere problemi, di esprimere giudizi, ragionare e organizzare. L'identificazione dei disturbi cognitivi nella SM può essere difficile, perché questi deficit possono presentarsi in stadi precoci o avanzati della malattia e possono essere confusi con altri sintomi, come la stanchezza o la depressione (Guimarães, Sá, 2012; Kraemer, Herold, Uekermann, Kis, Wiltfang, Daum & Abdel-Hamid, 2013). I disturbi cognitivi hanno un impatto negativo sulla qualità di vita del paziente e delle persone che se ne prendono cura (Rao, Leo, Bernardin & Unverzagt, 1991; Labiano-Fontcuberta, Mitchell, Moreno-García & Benito-León, 2014).

Mentre i deficit cognitivi nella SM sono largamente conosciuti, pochi studi hanno investigato l'entità in cui essi possono colpire le abilità dei pazienti nel parlare. Un deterioramento nel funzionamento del sistema cognitivo ha necessariamente delle ripercussioni sul linguaggio dal momento che include disturbi della pianificazione, della presa di decisione, della memoria di lavoro, dell'attenzione e della velocità di trattamento dell'informazione. Il linguaggio, pur essendo prodotto naturalmente da ogni individuo quotidianamente, tuttavia è un'abilità cognitiva altamente sofisticata. Prendere parte a una conversazione richiede l'utilizzo di complesse competenze che consentono la coordinazione interpersonale. Quando gli interlocutori sono coinvolti in una conversazione devono essere in grado di pianificare velocemente quello che intendono dire. Essi scelgono il messaggio da trasmettere, la struttura sintattica, le parole, i segmenti (consonanti e vocali) e la prosodia delle loro frasi (Levelt, 1989). Inoltre, essi devono essere in grado di concordare rapidamente su chi parlerà e quando. In scambi conversazionali del tipo domanda-risposta le risposte sono date in un intervallo di tempo estremamente breve (Stivers *et al.*, 2009). Poiché i tempi di latenza implicati nella pianificazione della produzione del linguaggio possono essere molto più lunghi (Bögels, Magyari & Levinson, 2015), si suppone che i parlanti possano iniziare a pianificare le loro risposte mentre ascoltano i propri interlocutori. L'abilità dei pazienti di SM nel pianificare il proprio discorso è dunque cruciale perché la loro interazione verbale quotidiana abbia successo e quindi per preservare le loro relazioni sociali e occupazioni.

In questa sede verrà presentata in primo luogo una disamina degli studi sulla correlazione tra disturbi cognitivi e parlato dialogico nei pazienti SM rispetto a quanto avviene nei meccanismi dialogici normofasici (§ 1.1-§ 1.2). In seguito (§ 2), verrà illustrata l'ipotesi di ricerca alla base di questo lavoro, cui farà seguito la descrizione della metodologia e del corpus utilizzato (§ 3-§ 4). Successivamente si descriverà l'esperimento nelle sue varie fasi (§ 5) e verranno presentate le analisi dei dati dei test neuropsicologici (§ 6) e le analisi dei dati acustici (§ 7). Infine, saranno commentati e discussi i risultati ottenuti (§ 8-§ 9).

1.1 Disturbi cognitivi e parlato dialogico nei pazienti SM

Tra le problematiche riguardanti la sfera del linguaggio, molti studi attestano la presenza, nel 23-50% dei pazienti, di disturbi di disartria, sia nelle due forme, atassica e spastica, sia in una forma mista tra le due (Rosen, Goozée & Murdoch, 2008).

Pochi studi sono stati effettuati finora sull'impatto dei disturbi cognitivi sulla parola dei pazienti affetti da SM. Feenaughty, Tjaden, Benedict & Weinstock-Guttman (2013) e Rodgers, Tjaden, Feenaughty, Weinstock-Guttman & Benedict (2013) riportano sia nei compiti di lettura che nel parlato spontaneo che la velocità di articolazione è più lenta per pazienti anglofoni affetti da SM con una bassa capacità cognitiva rispetto agli adulti sani. Feenaughty *et al.* (2013) inoltre segnalano un maggior numero di pause silenziose e un minor numero di pause grammaticali[1] in pazienti con SM con bassa capacità cognitiva. La durata delle pause silenziose non era significativamente differente tra i gruppi. Rodgers *et al.* (2013) esaminano la relazione tra singoli predittori cognitivi e misure di velocità di eloquio. Nei loro studi è stata riscontrata una moderata associazione tra i punteggi compositi della velocità di trattamento dell'informazione (attraverso l'utilizzo di test come il PASAT e l'SDMT) e la velocità di articolazione, sia in compiti di lettura che nel parlato spontaneo, per i pazienti affetti da SM ma non per i casi controllo. L'associazione di misure di memoria e di velocità di eloquio era trascurabile nei compiti di lettura e limitata nel parlato spontaneo.

De Looze, Ghio, Moreau, Renié, Rico, Audoin, Viallet, Pelletier & Petrone (2017) hanno studiato l'effetto dei deficit cognitivi su diversi parametri prosodici di pianificazione in un compito di lettura. 45 soggetti francofoni, di cui 22 casi controllo e 23 pazienti nei primi stadi della SM recidivante-remittente, sono stati sottoposti a test neuropsicologici di valutazione di specifici processi cognitivi coinvolti nella pianificazione del parlato. I risultati mostrano che il parlato dei pazienti SM con problemi cognitivi è influenzato principalmente a livello temporale (velocità di eloquio, durata delle pause). Analisi di regressione ulteriori indicano che le misure della velocità sono correlate in maniera lineare con i punteggi della memoria di lavoro.

1.2 Meccanismi di coordinazione della parola in individui sani

Nella prospettiva pragmatica che considera il linguaggio come azione e interazione all'interno di un contesto, un ruolo centrale ha l'analisi del discorso come cooperazione tra gli interlocutori. L'avvicendamento dei turni di parola è l'elemento più caratteristico del parlato: esso, in condizioni normo-fasiche, consente ai parlanti "il migliore equilibrio possibile tra programmazione e produzione, da una parte, e recezione ed elaborazione, dall'altra" (Voghera, 2001: 76). Negli scambi conversazionali, l'ascoltatore riveste la stessa importanza del produttore per la costruzione del

---

[1] Una pausa silenziosa si realizza mediante un'interruzione più o meno prolungata del parlato. Essa può avere funzioni diverse, tra cui quella grammaticale: in questo caso, marca una cesura al termine di un enunciato grammaticalmente completo e può servire a mettere in rilievo elementi su cui il parlante vuole focalizzare l'attenzione dell'interlocutore. Quando la pausa silenziosa non assolve a questa funzione, essa può segnalare invece un'esitazione da parte del parlante (Duez, 1982).

discorso e del resto i ruoli cambiano continuamente. Uno degli aspetti rilevanti da analizzare in questo processo di scambio di ruoli tra produttore e ricettore del messaggio è proprio quello inerente alle modalità con cui avviene l'alternanza dei turni di parola in una conversazione: come gli interlocutori riescano più o meno a cooperare sincronizzando i propri interventi affinché la conversazione proceda felicemente.

Il modo in cui avviene l'avvicendamento dei turni è ciò che caratterizza la conversazione stessa. Sono state infatti messe in evidenza alcune regole che reggono l'avvicendamento dei turni: in una conversazione, la norma è che parli una persona alla volta e, sebbene le occorrenze in cui due o più parlanti parlino contemporaneamente siano frequenti, esse sono normalmente di breve durata. Inoltre, le transizioni tra un turno e l'altro possono avvenire senza vuoti, con accavallamenti, o con vuoti di durata variabile, normalmente breve (Sacks, Schegloff & Jefferson, 1974). La presenza di accavallamento tra i turni e di interruzioni è fortemente influenzata dal tipo di conversazione (più o meno formale), da fattori socioculturali propri degli interlocutori, dal contesto (Bazzanella, 1994). Per quanto riguarda il contesto, in una conversazione improntata alla risoluzione di un compito, come nel caso esaminato in questo studio, le interruzioni sono considerate del tutto normali e "supportive" (Bazzanella, 2005). Molto significativa è la presenza di pause, piene e vuote, sia all'interno dello stesso turno di parola, sia quando esse segnano il passaggio tra un turno e l'altro. All'interno dello stesso turno, la pausa ha un ruolo fondamentale, poiché, insieme ai fenomeni prosodici, contribuisce a chiarire la sintassi dell'enunciato, a dare particolare prominenza a una parte di esso, a fornire all'ascoltatore informazioni circa lo stato emotivo di chi parla. Pause prolungate all'interno dello stesso turno di parola possono poi indicare che il parlante ha bisogno di un tempo di pianificazione maggiore, mentre pause prolungate tra due turni possono essere dovute alla necessità di comprendere ciò che l'interlocutore ha detto e a pianificare una risposta. Recenti lavori volti ad analizzare i tempi di latenza che normalmente intercorrono tra un turno e l'altro hanno mostrato che, nello scambio dialogico, i processi di comprensione e quello di pianificazione in una certa misura si sovrappongono, occorrendo in maniera quasi parallela (Stivers *et al.*, 2009; Bögels *et al.*, 2015).

Accanto agli studi di (neuro-)pragmatica, alcune ricerche in psicolinguistica sui soggetti sani si sono incentrate sul ruolo delle capacità cognitive individuali nella pianificazione del parlato. Questi studi sono incentrati prettamente sul parlato monologico. I modelli tradizionali assumono che risorse cognitive come la memoria di lavoro hanno un impatto minimo sulla pianificazione, toccando per lo più i livelli alti di pianificazione, come la pianificazione del messaggio (Levelt, 1989). D'altra parte, recenti studi hanno indicato che la pianificazione del parlato (cioè, quanto in anticipo i locutori pianificano l'enunciato che stanno per realizzare) varia a secondo delle capacità cognitive individuali per diversi livelli di rappresentazione, e che l'impatto di tali risorse cognitive è molto più importante di quanto ipotizzato dai modelli tradizionali. Per esempio, Swets, Desmet, Hambrick & Ferreira (2007) hanno trovato che, in lettura, i parlanti creano più frontiere (o rotture) prosodiche se le loro capacità di memoria sono deboli perché probabilmente sono meno capaci di pianificare delle unità di parlato molto lunghe. Questo indicherebbe che la durata delle unità di pianificazione non è fissa ma flessibile, cioè variabile a seconda delle capacità cognitive individuali.

2. *Ipotesi di ricerca*

Nonostante i deficit cognitivi siano spesso riscontrati nei pazienti SM, è ancora poco chiaro in che misura tali disturbi influenzino la pianificazione del parlato spontaneo e in particolare i meccanismi di coordinazione dialogica. A partire da risultati nella letteratura in psicolinguistica (Swets *et al.*, 2007) e sulla SM (De Looze *et al.*, 2017), nel presente lavoro si è focalizzata l'attenzione sulla seguente ipotesi di ricerca che riguarda la durata complessiva delle transizioni silenziose nell'avvicendamento dei turni di parola. Si vuole verificare se nell'interazione dialogica i pazienti SM con deficit cognitivi (SM-DC) producano transizioni di durata complessivamente maggiore rispetto a quanto non si verifichi nell'interazione con i casi controllo (C) e con i pazienti senza deficit (SM-SDC). Infatti, ci si aspetta che il danno alla velocità di trattamento delle informazioni e alla memoria di lavoro comporti per i pazienti affetti da questi deficit un allungamento dei tempi di pianificazione del discorso che può essere verificato attraverso l'analisi dei tempi necessari alla presa di turno.

3. *Metodologia*

3.1 Partecipanti

Il nostro esperimento coinvolge tre tipologie di soggetti:

a) Gruppo di pazienti affetti da Sclerosi multipla (gruppo SM)

Il gruppo SM è formato da 12 persone che hanno una media di 47,3 anni di età, provenienti dal reparto di neurologia dell'*Hôpital des Pays d'Aix*. Tutti i pazienti affetti da SM sono in un periodo stazionario della loro malattia: ciò significa che non devono aver subito neppure una ricaduta (riacutizzazione o attacco) nell'arco di un mese precedente allo studio e di conseguenza non devono aver seguito un trattamento steroideo. Infatti, il trattamento d'elezione per le ricadute è la somministrazione di alte dosi di corticosteroidi, che hanno effetti immunomodulanti e antinfiammatori tali da ripristinare l'integrità della barriera ematoencefalica, ridurre il gonfiore e facilitare la ricostruzione della mielina, migliorando la conduzione dell'impulso nervoso (Rudick, Cohen, Weinstock-Guttman, Kinkel & Ransohoff, 1997).

Sono invece stati esclusi dall'esperimento i pazienti con: disartria; problemi uditivi; precedenti psichiatrici; precedenti di dipendenza da alcool o droghe; una terapia in corso con antidepressivi o sonniferi; altri disturbi neurologici. Tali criteri sono stati scelti per evitare l'effetto dei fattori appena elencati sui parametri prosodici e temporali.

Non è stato somministrato nessun test formale per la disartria. La disartria è stata valutata su base neurologica, in base alla diagnosi clinica (valutazione percettiva durante un colloquio orale) e in base all'autovalutazione dei pazienti. Un logopedista ha effettuato le registrazioni acustiche (dopo essere stato intensivamente addestrato da un esperto fonetista, tra gli autori). Lo stesso logopedista ha poi percettivamente giudicato la presenza di disartria in base a un compito di lettura somministrato prima del gioco di interazione. La disabilità fisica è stata valutata attraverso il punteggio ottenuto alla EDSS (*Expanded Disability Status Scale*; Kurtzke, 1983). Tutti i partecipanti sono di lingua madre francese.

b) Gruppo di controllo (C)

I soggetti che fanno parte del gruppo di controllo sono 12 madrelingua francesi dai 22 ai 71 anni, con un'età media 35 anni. Il gruppo è costituito da 10 donne e 2 uomini. I casi controllo sono stati abbinati ai pazienti SM secondo l'età, il genere e il livello di scolarizzazione (istruzione superiore). I casi controllo non hanno alcuna storia di malattia neurologica.

c) Gruppo di interlocutori (I)

Si compone di persone con esperienza più o meno solida nell'interazione con pazienti (logopedisti o psicologi). In particolare, 12 studenti di logopedia o psicologia clinica di età compresa tra i 20 e 24 anni, per una media di 22 anni, e di lingua madre francese. I soggetti hanno autocertificato di non presentare alcuna malattia neurodegenerativa, né alcun tipo di dislessia o di disturbo motorio.

Il reclutamento dei soggetti è stato approvato da un comitato etico (*Comité de Protection des Personnes*). Tutti i partecipanti hanno firmato un modulo di consenso e hanno compilato un foglio informativo, che ha permesso di conoscere meglio sia gli interlocutori (I) che i soggetti di controllo (C): età, sesso, livello di istruzione, luogo di nascita e di residenza, lingua madre. Per i pazienti poi è stata prodotta un'ulteriore documentazione comprendente alcune domande circa la loro malattia.

Le informazioni demografiche sono riassunte nella Tabella 1, dalla quale si può notare che i casi controllo sono tendenzialmente più giovani dei pazienti. Non sempre, infatti, è stato possibile reclutare soggetti che rientrassero nella medesima fascia d'età dei pazienti, così da ottenere un gruppo di riferimento perfettamente comparabile a quello in esame.

Tabella 1 - **Pazienti di SM recidivante remittente con deficit cognitivi (DC) e senza (SDC). Criteri di esclusione: terapia con antidepressivi, dislessia, disartria, abuso di droghe e/o alcol; precedenti di disturbi psichiatrici; deficit di udito. ** Casi controllo abbinati ai pazienti secondo il sesso, l'età e il livello di istruzione. *** Gruppo di interlocutori composto da logopedisti e psicologi*

	*SM-DC**	*SM- SDC**	*Controllo (C)***	*Interlocutori (I)****
N	6	6	12	12
Età	50.6 (6.3)	44 (11.4)	36.9 (16.1)	23.4 (3.4)
Sesso	5F/1M	5F/1M	10F/2M	10F/2M
EDSS[2]	5 (1.18)	3.2 (1.25)	--	--

[2] La scala EDSS (Expanded Disability Status Scale) è uno strumento ampiamente utilizzato negli studi clinici per misurare e valutare le caratteristiche dei pazienti affetti da sclerosi multipla. Il punteggio

3.2 Procedura sperimentale

L'esperimento è stato condotto principalmente nel locale del CNRS dell'ospedale di Aix-en-Provence (quando il soggetto è un paziente o un soggetto del gruppo di controllo) e talvolta nella camera anecoica del *Laboratoire Parole et Langage* (LPL) a Aix-en-Provence (quando il soggetto è anche un membro del LPL).

Sono state predisposte due distinte fasi. In primo luogo, tutti i partecipanti (pazienti SM, soggetti C e interlocutori I) hanno completato una serie di test neuropsicologici finalizzati a valutare la memoria di lavoro e la velocità nel trattamento delle informazioni. La durata della somministrazione dei test è di 20-25 minuti.

Successivamente, i partecipanti sono stati registrati durante un gioco di interazione di 10 minuti. La messa in atto della dinamica interazionale ha avuto lo scopo di valutare le capacità di pianificazione del discorso durante il dialogo.

4. *Corpus*

La prima tappa dell'esperimento consiste nell'aver sottoposto separatamente tutti i partecipanti (soggetti SM, soggetti C e interlocutori I) ai test neuropsicologici, selezionati grazie alla consulenza della neuropsicologa del reparto di neurologia dell'ospedale di Aix-en-Provence. Nel presente lavoro ci concentriamo su 3 test cognitivi, progettati ai fini di valutare la velocità di trattamento delle informazioni e la memoria dei partecipanti. In effetti, questi due parametri riscontrano un certo interesse ai fini dello studio poiché entrano in gioco al momento di pianificare il discorso (Swets *et al.*, 2007).

a) *PASAT-3 seconds* (*Paced Auditory Serial Addition Test*)

Il *PASAT-3 seconds* è un test composito che fa appello simultaneamente alla velocità di trattamento dell'informazione, all'attenzione e alla memoria di lavoro dei partecipanti. Questo test consiste in una serie di addizioni in sequenza per cui il soggetto deve usare la sua memoria di lavoro per trattenere un'informazione, trasformarla in funzione di ciò che viene detto subito dopo e quindi inibire il risultato di questa elaborazione al fine di concentrare rapidamente la sua attenzione sulle nuove informazioni che riceve di volta in volta. Queste operazioni devono essere ripetute sessanta volte a ritmo costante e sono dunque molto impegnative sul piano dell'attenzione. Infatti viene enunciata una serie di sessantuno cifre, una ogni tre secondi: il soggetto deve addizionare al primo numero il secondo e dare la propria risposta ad alta voce prima che gli venga fornito il numero successivo, da addizionare alla seconda cifra dell'operazione matematica, e così via. Per esempio: se sente 3, poi 4, dovrà dire 7.

EDSS totale viene determinato da due fattori: la capacità di deambulazione e i punteggi relativi a otto sistemi funzionali, variamente colpiti dalla malattia: sistema motorio, s. cerebellare, tronco encefalico, s. sensitivo, s. sfinterico, s. visivo, s. cerebrale e "altri". A ciascun sistema funzionale viene dato un punteggio di crescente gravità (da 1 a 5), con la sola eccezione dell'ultima categoria ("altri"), che serve unicamente a fornire informazioni su ulteriori problemi che possono insorgere differentemente da caso a caso (come, per esempio, la perdita di *capacità deambulatoria)*. Si veda Kurtzke (1983).

Se in seguito sente 6, dovrà dire 10 (cioè 6 + 4) e così a seguire. Per non complicare eccessivamente il processo e far sì che le addizioni vengano risolte rapidamente, le cifre utilizzate sono comprese tra 1 e 9.

Prima dell'inizio della prova, sono state riferite le istruzioni al soggetto, poi ci si è assicurati della loro corretta comprensione mediante un test di prova, che può essere ripetuto tre volte e che deve assolutamente essere stato superato prima di sottoporre il soggetto al test vero e proprio. Dal momento che la sequenza automatica di cifre non può essere interrotta, è stato chiesto al soggetto, qualora non fosse riuscito a dare la risposta nel tempo predisposto, di recuperare la successione di addizioni a partire dalle ultime cifre recepite. Le risposte sono state annotate in una griglia.

b) Sequenza lettere/cifre

Si tratta di un test per valutare unicamente la memoria di lavoro. Prevede l'enunciazione di una serie di tre sequenze di lettere e cifre, mescolate tra loro, che il soggetto ha il compito di ripetere, di volta in volta, disponendo sempre prima i numeri in ordine crescente e poi le lettere in ordine alfabetico. La difficoltà del test si accresce progressivamente: la prima sequenza è formata da una lettera e una cifra mentre l'ultima da quattro lettere e da quattro cifre, per un totale di sette sequenze, che costituiscono una serie. Il test si conclude nel momento in cui il soggetto sbaglia tre sequenze di una stessa serie. Il punteggio è 0 in caso di errore e 1 in caso di successo e il punteggio massimo è di 21.

c) Fluenza verbale

Si tratta di un test per la valutazione della memoria fonologica e semantica. Durante questa prova, è stato richiesto ai soggetti di produrre quante più parole possibili seguendo due tipi di criteri: uno di tipo fonetico (per es., parole che inizino con /p/) e uno di tipo semantico (per es., nomi di animali). Le parole devono avere radici differenti per essere valide. La prova ha una durata di due minuti e al termine di essa viene valutato il versante quantitativo di questa produzione. Non esiste un punteggio massimo raggiungibile.

5. *Il gioco del naufragio*

Il principale obiettivo di questo test è quello di ottenere la registrazione di un dialogo in interazione conversazionale da analizzare. I due partecipanti di ciascuna coppia sono stati fatti sedere alle due opposte estremità di una scrivania, al centro della quale è collocato un pannello che impedisce ai due di vedersi. Di fronte a ciascun soggetto è stato posizionato un monitor, sul quale appaiono, identiche, le istruzioni del gioco, che essi dovranno leggere e comprendere. Dopo essere stati invitati a rivolgere eventuali domande di chiarimento, inizia l'interazione, con l'obbligo tassativo per i soggetti di non parlare al di fuori del gioco. Il test richiede di calarsi in una situazione immaginaria: i due partecipanti devono fingere di essere due naufraghi, sopravvissuti su una scialuppa, in pieno oceano (Vaughan, Cullen, 2009). Essi hanno la possibilità di salvare quindici oggetti, rappresentati in una griglia che appare in modo identico su entrambi i loro monitor (Figura 1). Lo scopo dell'interazione è quello di classificare i quindici oggetti in ordine di importanza, realizzando un elenco in comune accordo. La classifica deve essere concordata unicamente mediante l'interazione verbale: il pannello che li separa, infatti, ha lo scopo di evitare qualsiasi forma di comunicazione non verbale, sia volontaria che non, come il linguaggio del corpo, le espressioni facciali, gli sguardi, la gestualità. Gli oggetti sul monitor non possono essere selezionati né spostati, per cui la classifica dovrà essere stilata facendo affidamento unicamente sulla memoria dei due soggetti.

Figura 1 - *Immagine che appare sui monitor dei partecipanti al gioco del naufragio*

6. *Analisi dei dati neuropsicologici*

Per ognuno dei tre gruppi (soggetti SM, soggetti C e interlocutori I) i punteggi dei test cognitivi ottenuti sono stati standardizzati, attraverso il calcolo della media e della deviazione standard. In particolare, la media dei tre test ha consentito di individuare all'interno del gruppo dei soggetti SM due distinti sottogruppi:
- sei pazienti dalle capacità cognitive compromesse, indicati con la sigla SM-DC;
- sei pazienti senza deficit cognitivi, indicati con la sigla SM-SDC.

7. *Analisi dei dati acustici*

7.1 Registrazioni

Le registrazioni acustiche sono state realizzate con microfoni professionali. Le sedute di registrazione sono state realizzate interamente grazie al software EVA2 (Ghio *et al.*, 2012). Le analisi acustiche sono state effettuate su PRAAT (Boersma, 2001).

7.2 Annotazioni ed estrazione automatica dei dati

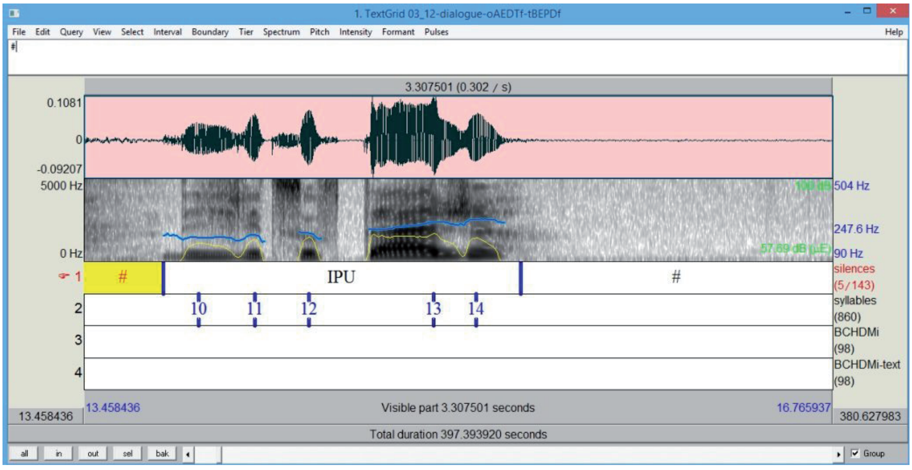
In una prima fase, uno script PRAAT, sviluppato appositamente per questo lavoro, è stato utilizzato per analizzare il parlato dei partecipanti. Ogni partecipante è stato registrato su una pista audio separata. Per ognuno di essi, sono state individuate in maniera automatica le unità interpausali e le pause silenziose a partire dalle analisi spettrografiche. Le informazioni sono state inserite automaticamente nei Textgrids corrispondenti.

Tutte le annotazioni relative alle frontiere acustiche ottenute automaticamente sono state corrette manualmente attraverso l'analisi combinata dello spettrogramma e della forma d'onda del segnale acustico. In particolare, sono stati presi in considerazione solo i silenzi di durata pari o superiore ai 100 ms. In questa fase di lavoro si è prestata grande attenzione alla correzione dell'individuazione automatica di PRAAT dei silenzi e delle unità interpausali: il sistema calcola infatti automaticamente come silenzi i tempi di ritenuta dei foni occlusivi, i foni fricativi e alcune realizzazioni del fono vibrante /r/ in fine di parola. Tali fenomeni sono stati accuratamente corretti e annotati come parlato. Inoltre fenomeni quali risate, colpi di tosse e rumori di fondo sono stati inclusi nei silenzi (dal momento che non costituiscono parlato) e opportunamente segnalati laddove il programma li segnava come unità interpausali.

In un secondo momento, un confronto dei Textgrids ottenuti per i due partecipanti di ciascuna coppia ha permesso di estrarre automaticamente la localizzazione e le durate delle pause silenziose tra turni di parola e le sovrapposizioni tra turni di parola.

In Figura 2 è riportato un esempio di analisi manuale effettuata con PRAAT.

Figura 2 - *L'immagine mostra un esempio di text grid analizzato e corretto secondo i criteri sopra esposti. I tiers utilizzati per le annotazioni sono quattro: nel primo sono stati individuati i segmenti di parlato (IPU, Inter-Pause Units) e i silenzi (#); nel secondo sono state numerate le sillabe (syllables); nel terzo sono stati annotati fenomeni quali Back Channel (BC), esitazioni (H), Marcatori Discorsivi (DM) e interruzioni (i); nell'ultimo tier è stato riportati per i BC e i DM il testo corrispondente al contenuto lessicale del fenomeno individuato (per esempio, nel caso dei DM: alors, donc, comment dire, etc.), mentre per H e i sono state segnalate le tipologie specifiche del macro-fenomeno individuato al tier superiore. Per H: filled pause (fp), false start (fs), initial syllable lengthening (isl), final syllable lengthening (fsl), babbling (b); per i: laugh (l)*



In una terza fase, è stato utilizzato un altro script PRAAT, anch'esso sviluppato ad hoc per estrarre le informazioni relative alle caratteristiche temporali delle interazioni conversazionali. Nel dettaglio, tali caratteristiche includono:

- il numero e la durata delle transizioni di parlato (transizioni silenziose e sovrapposizioni)
- il tempo totale di parlato e di silenzio (dato in percentuale: tempo di parlato / tempo totale di parlato nell'interazione *100; tempo di silenzio / tempo totale dell'interazione *100).

Tuttavia in questo articolo verrà trattato solo il fenomeno dei silenzi tra turni di parola (cui spesso ci si riferisce in letteratura come *gaps*).

7.3 Statistiche

Una serie di modelli lineari misti sono stati computati per modellizzare la relazione tra abilità cognitive e caratteristiche temporali del parlato. I modelli misti hanno il vantaggio di potersi facilmente applicare su dati disequilibrati, quali la variabilità inter-locutore e le misure ripetute. Come effetto fisso, il gruppo C/SM-SDC/SM-DC è stato incluso nel modello per testare l'impatto delle abilità cognitive sulle caratteristiche acustiche descritte nella sezione 7.2. Il gruppo SM-SDC è stato incluso come intercetta (livello di riferimento), a cui gli altri gruppi sono stati comparati. Ci aspettiamo che questo gruppo si comporti in modo simile a C ma differisca da SM-DC. Per questa analisi preliminare, i modelli sono caratterizzati da una struttura aleatoria semplice, con intercetta per soggetto e per coppia di locutori. La soglia di significatività è di 0.05.

8. *Risultati*

8.1 Test neurocognitivi

Il Grafico 1 mostra che i pazienti SM-DC al test PASAT hanno ottenuto risultati significativamente inferiori rispetto agli altri tre gruppi [$\beta = -37.5$, $SE = 6.8$, $t = -5.42$, $p < .05$]. Il gruppo dei pazienti SM-DC mostra una grande variabilità nei punteggi ottenuti a questo test, sebbene il valore della mediana, risultando molto vicino al valore minimo, riveli una media dei punteggi bassa. I punteggi ottenuti dagli altri tre gruppi sono molto al di sopra di quelli del gruppo SM-DC e allo stesso tempo simili tra loro ($p > .05$). Nella Tabella 2 sono riportati i valori medi dei punteggi ottenuti al test.

Grafico 1 - *Box plot dei punteggi al test PASAT. Sull'asse delle ascisse i diversi gruppi sono indicati con le iniziali delle rispettive denominazioni in inglese: I per Interlocutors (interlocutori), C per Control (casi controllo), MS-NCI per Multiple Sclerosis-Not Cognitive Impairments (in questo articolo, SM-SDC) e MS-CI per Multiple Sclerosis-Cognitive Impairments (in questo articolo, SM-DC)*

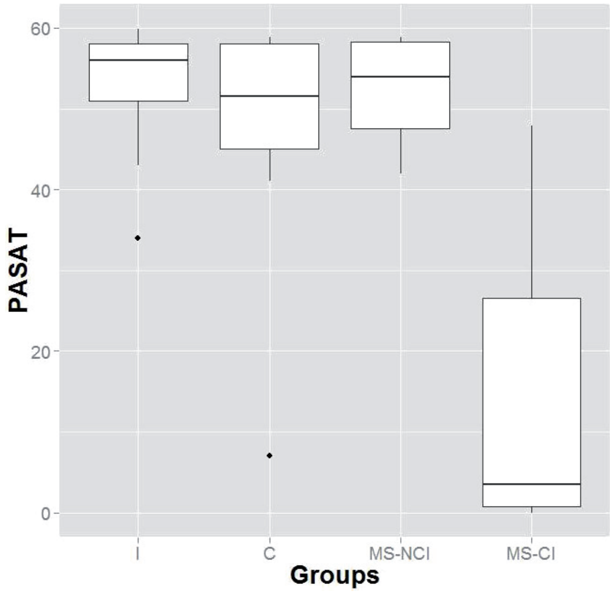


Tabella 2 - *Valori medi ed errori standard del test PASAT per ogni gruppo*

Gruppo di parlanti	*Valori medi PASAT*	*Errore standard PASAT*
1. I	52.9	1.6
2. C	48.2	4.1
3. SM-SDC	52.3	2.8
4. SM-DC	14.8	8.4

Per quanto riguarda il test di sequenza lettere/cifre, anche in questo caso, come si evince dal Grafico 2, i risultati ottenuti dal gruppo SM-DC sono significativamente più bassi (5.16) di quelli ottenuti dagli altri tre gruppi. Le medie dei punteggi ottenuti dai gruppi I (10.9) e C sono identiche (10.08). Il gruppo SM-SDC, in questo caso ottiene una media dei punteggi (8.1) inferiore rispetto ai gruppi I e C, nonostante si distacchi positivamente dal gruppo SM-DC. La analisi di regressione confermano una differenza significativa tra SM-DC e SM-SDC [$\beta = -3$, $SE = 1.4$, $t = -2.05$, $p < .05$], mentre la differenza tra SM-SDC e C non raggiunge la soglia di significatività. Nella Tabella 3 sono riportati i valori medi dei punteggi ottenuti al test.

Grafico 2 - *Box plot dei punteggi al test sequenza lettere/cifre (LN-seq). Sull'asse delle ascisse i diversi gruppi sono indicati con le iniziali delle rispettive denominazioni in inglese: I per Interlocutors (interlocutori), C per Control (casi controllo), MS-NCI per Multiple Sclerosis-Not Cognitive Impairments (in questo articolo, SM-SDC) e MS-CI per Multiple Sclerosis-Cognitive Impairments (in questo articolo, SM-DC)*

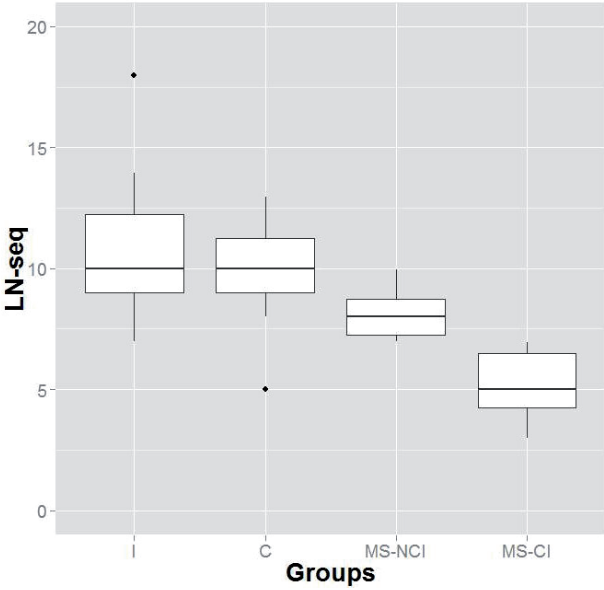


Tabella 3 - *Valori medi ed errori standard del test seq-L/C per ogni gruppo*

Gruppo di parlanti	*Valori medi test seq-L/C*	*Err standard test seq-L/C*
1. I	10.9	0.6
2. C	10.1	0.6
3. SM-SDC	8.1	0.4
4. SM-DC	5.1	0.6

Rispetto al test di fluenza verbale, come si evince dal Grafico 3, il gruppo SM-DC ottiene nuovamente una media dei punteggi ben al di sotto di quelle ottenute dagli altri tre gruppi. Il gruppo I ottiene i punteggi più elevati in assoluto; inoltre il gruppo C si caratterizza per una maggiore variabilità nella distribuzione dei punteggi ottenuti. Infine il gruppo SM-SDC ottiene punteggi inferiori a quelli dei gruppi I e C, con un valore della media dei punteggi leggermente più prossimo al valore minimo. Le analisi di regressione indicano una differenza significativa tra SM-DC e SM-SDC [$\beta = -16$, SE = 7.9, $t = -2.09$, $p < .05$] e una differenza significativa ma meno forte tra SM-SDC e C [$\beta = 14$, SE = 6.9, $t = 2.02$, $p = .04$]. Nella Tabella 4 sono riportati i valori medi dei punteggi ottenuti al test.

Grafico 3 - *Box plot dei punteggi al test di fluenza verbale. Sull'asse delle ascisse i diversi gruppi sono indicati con le iniziali delle rispettive denominazioni in inglese: I per Interlocutors (interlocutori), C per Control (casi controllo), MS-NCI per Multiple Sclerosis-Not Cognitive Impairments (in questo articolo, SM-SDC) e MS-CI per Multiple Sclerosis-Cognitive Impairments (in questo articolo, SM-DC)*

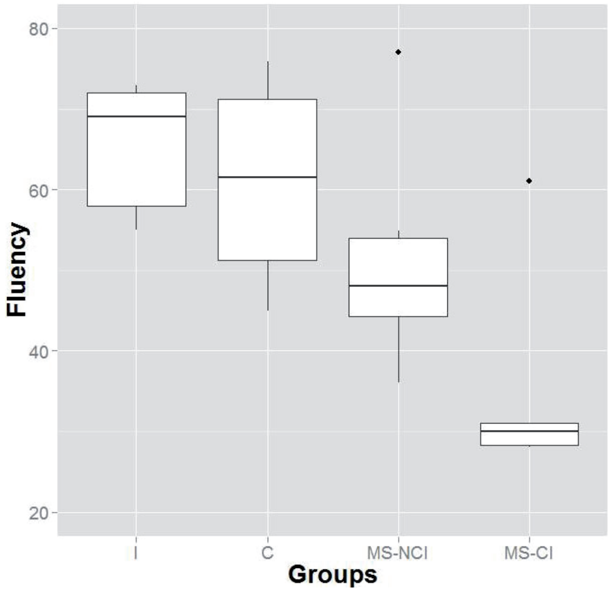


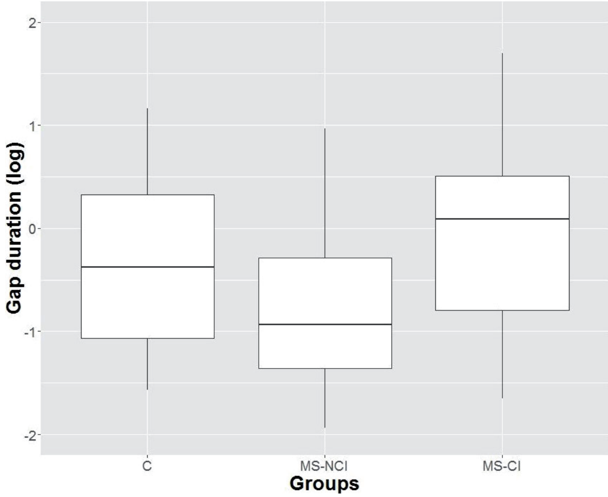
Tabella 4 - *Valori medi ed errori standard del test di fluenza verbale per ogni gruppo*

Gruppo di parlanti	*Valori medi test di fluenza verbale*	*Errore standard test di fluenza verbale*
1. I	70.8	2.7
2. C	65.3	4.1
3. SM-SDC	51.3	5.7
4. SM-DC	34.6	5.2

8.2 Gioco di interazione

Il Grafico 4, relativo alla durata delle transizioni silenziose nell'avvicendamento dei turni di parola, mostra che il gruppo SM-DC (nel grafico, MS-CI) è quello che complessivamente produce nell'interazione dialogica pause silenziose tra un turno e l'altro più lunghe. Nell'interazione col gruppo I il gruppo SM-SDC (nel grafico, MS-NCI) è quello che produce pause di transizione della durata più bassa anche rispetto al gruppo C. Le analisi statistiche confermano un contrasto significativo tra SM-DC e SM-SDC [$\beta = -0.85$, $SE = 0.23$, $t = -3.65$, $p < .05$], mentre la differenza tra SM-SDC e C non è significativa.

Grafico 4 - *Box plot della durata in log delle transizioni silenziose nell'avvicendamento dei turni. Sull'asse delle ascisse i diversi gruppi sono indicati con le iniziali delle rispettive denominazioni in inglese: C per Control (casi controllo), MS-NCI per Multiple Sclerosis-Not Cognitive Impairments (in questo articolo, SM-SDC) e MS-CI per Multiple Sclerosis-Cognitive Impairments (in questo articolo, SM-DC)*



9. *Discussione*

Dall'analisi dei test neuropsicologici emerge che tra i gruppi di soggetti privi di deficit cognitivi I è il più performante. Questo risultato non sorprende dal momento che I è il gruppo con età media più bassa e di conseguenza più reattivo dal punto di vista cognitivo. Al contrario, il gruppo C, costituito da soggetti di età media più elevata e quindi con capacità cognitive inferiori, ottiene mediamente punteggi più bassi rispetto al gruppo I. Il gruppo SM-SDC si colloca in una posizione intermedia tra i gruppi I e C, raggiungendo talvolta livelli pari a quelli del gruppo C, talvolta leggermente inferiori. SM-DC è il gruppo con peggiore *score* cognitivo in ognuno dei tre test.

Le analisi relative alla durata delle transizioni silenziose nell'avvicendamento dei turni di parola confermano la nostra ipotesi di partenza: il gruppo SM-DC è quello che complessivamente produce pause silenziose più lunghe tra un turno e l'altro. Nell'interazione col gruppo I, il gruppo SM-SDC è quello che produce pause di transizione più brevi anche rispetto al gruppo C, sebbene la differenza non sia significativa. Al contrario, le analisi statistiche confermano un contrasto significativo tra SM-DC e SM-SDC.

L'ipotesi si basava sull'esistenza di una correlazione tra deficit cognitivi causati dalla SM (*in primis* danno alla velocità nel processare informazioni e compromissione della memoria di lavoro) e la durata dei tempi di pianificazione nel parlato. Già studi precedenti (Swets *et al.*, 2007) hanno evidenziato come, nei compiti di lettura, i parlanti creino ai fini della pianificazione più frontiere (o rotture) prosodiche se le loro capacità di memoria sono deboli. Il presente studio contribuisce a rafforzare l'ipotesi secondo cui le capacità cognitive individuali influiscono in maniere decisiva sulla durata dei tempi di pianificazione.

Nelle conversazioni analizzate, quelle frutto dell'interazione tra il gruppo I e il gruppo SM-DC si caratterizzano per la maggiore durata media delle transizioni silenziose tra un turno e l'altro. Questo risultato conferma quello ottenuto da De Looze *et al.* (2017), in cui si evidenzia una correlazione tra deficit cognitivi nella SM e carenza nella pianificazione temporale nel parlato letto (velocità di eloquio).

I risultati qui presentati costituiscono solo un primo stadio nello studio della correlazione esistente tra deficit cognitivi nella SM e meccanismi di pianificazione del parlato. Nel futuro, il nostro studio si potrebbe avvantaggiare di un'analisi qualitativa dei corpora da incrociare con una modellizzazione ti tipo formale (e.g., interfaccia semantica/pragmatica) e statistica.

Una variabile da prendere in considerazione è il tipo di turno di parola. Per esempio, possiamo ipotizzare che, in sequenze domanda-risposta (attestate frequentamente nel nostro corpus), la durata della transizione silenziosa dipenda dal tipo di domanda. Si potrebbe ipotizzare che le domande aperte (wh-) implichino in generale una maggiore pianificazione della risposta rispetto alle domande polari. Quindi, la durata della transizione dovrebbe essere più lunga nelle domande aperte.

Un altro aspetto su cui attualmente si sta lavorando in riferimento allo stesso corpus riguarda la presenza di determinate categorie di *marcatori discorsivi* e la presenza di fenomeni di esitazione (Bazzanella, 1994). La funzione dei marcatori discorsivi dipende principalmente dal contesto in cui vengono utilizzati: il parlante può servirsene per segnalare strategie comunicative adottate nell'interagire col proprio partner, per esempio per manifestare la propria attenzione ed eventualmente il proprio accordo (sono i cosiddetti *back-channels*); ma possono essere utilizzati anche per strutturare meglio il discorso oppure per prendere tempo ai fini della pianificazione e gestire i turni di parola (si è soliti parlare di *pianificatori* e *riformulatori*; cfr. Bazzanella, 1994; Borreguero, 2015). Le nostre indagini attuali hanno pertanto lo scopo di analizzare le eventuali differenze che intercorrono tra soggetti normofasici e soggetti SM-DC nell'uso dei marcatori discorsivi.

In un'altra prospettiva di ricerca sul nostro corpus, può essere opportuno valutare come avviene la convergenza tra pazienti e terapeuti. La convergenza infatti è quel fenomeno per cui i parlanti che prendono parte ad una conversazione spontanea tendono ad adattare i propri comportamenti linguistici a quelli dei propri interlocutori, al fine di raggiungere un equilibrio. Dunque, una soddisfacente interazione verbale quotidiana dei pazienti SM dipende non solo dall'abilità di pianificare i turni di parola, ma anche dalle strategie di adattamento dei loro interlocutori. In tale direzione, sarebbe utile in futuro analizzare come avviene il fenomeno della convergenza tra pazienti e terapeuti proprio al fine di consentire a questi ultimi di migliorare la loro interazione con soggetti affetti da SM.

Nella prospettiva della ricerca applicata, studi come quello proposto in questa sede possono rivelarsi uno strumento prezioso ai fini della diagnosi dei deficit cognitivi conseguenti alla SM attraverso l'analisi del parlato conversazionale. Infatti attualmente per la valutazione e il monitoraggio dei deficit cognitivi vengono somministrati ai soggetti test di tipo neuropsicologico, che tuttavia spesso richiedono molto tempo e possono generare nel soggetto effetti controproducenti, come stress, ansia e frustrazione. Ad esempio, i risultati del *PASAT* sono notoriamente influenzati da fenomeni quali l'allenamento al test stesso, le capacità matematiche del soggetto e lo stress. Quindi sarebbe auspicabile lo sviluppo di ulteriori strumenti diagnostici ai fini di completare la valutazione corrente dei deficit cognitivi nella SM, come la classificazione automatica sulla base di caratteristiche proprie della produzione di parlato. Si potrebbe ottenere ciò attraverso la valutazione indiretta basata su un test interattivo, simile al "gioco del naufragio" utilizzato in questo studio, o in base alle prestazioni dei pazienti nelle attività quotidiane. I risultati ottenuti da questo studio preliminare dimostrano che l'analisi del discorso può costituire un metodo valido di diagnosi, poiché la capacità cognitiva può essere indirettamente valutata dalla riproduzione di un testo in lettura (De Looze *et al.*, 2017), così come nel parlato spontaneo delle interazioni quotidiane (per esempio, durante i colloqui tra medico e paziente). Ciò significa che la capacità cognitiva potrebbe anche essere monitorata a distanza, durante le attività di vita quotidiana tra un paziente e la persona che se ne prende cura.

Ringraziamenti

Questo lavoro è stato finanziato dalla *Agence Nationale de la Recherche* a Caterina Petrone per il progetto "Rappresentazione e pianificazione di prosodia" (ANR-14-CE30-0005-01). Si ringrazia inoltre l'Università degli Studi di Napoli Federico II per aver permesso, con i programmi *Erasmus Placement*, la realizzazione dei progetti di studio di Giovanna De Bellis e di Simona Schiattarella. Ringraziamenti vanno a Celine De Looze per la realizzazione degli *script* in PRAAT che hanno consentito di estrarre le informazioni relative alla caratteristiche temporali delle interazioni conversazionali dei dialoghi presi in esame. Un ringraziamento particolare infine va alla Prof.ssa Francesca M. Dovetto, docente di Glottologia e Linguistica presso l'Università degli Studi di Napoli Federico II, promotrice dei suddetti programmi di studio *Erasmus Placement* e relatrice delle tesi di Laurea Magistrale di Giovanna De Bellis e di Simona Schiattarella.

Riferimenti bibliografici

Baddeley, A.D. (2003). Working memory and language: An overview. In *Journal of Communication Disorders*, 36(3), 189-208.

Bazzanella, C. (1994). *Le facce del parlare: un approccio pragmatico all'italiano parlato.* Firenze: La Nuova Italia.

Bazzanella, C. (2005). Parlato dialogico e contesti di interazione. In Hölker, K., Maaß, C. (Eds.), *Aspetti dell'italiano parlato*. Münster: LIT Verlag, 1-22.

Bianconi, G., Poggioli, E., Morelli, E., Razzaboni, E. & Pomelli, D. (2006). Aspetti psicologici della sclerosi multipla. In *Giornale Italiano di Medicina, del Lavoro e di Ergonomia*, 28(1), 1.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. In *Glot International*, 5(9), 341-345.

Bögels, S., Magyari, L. & Levinson, S.C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. In *Scientific reports*, 5, 12881.

Brassington, J.C., Marsh, N.V. (1998). Neuropsychological aspects of multiple sclerosis. In *Neuropsychology Review*, 8(2), 43-77.

De looze, C., Ghio, A., Moreau, N., Renié, L., Rico, A., Audoin, B., Viallet, F., Pelletier, J. & Petrone, C. (2017). Effects of cognitive impairment on prosodic parameters of speech production planning in multiple sclerosis. In *Journal of Neuropsychology*. May 24. DOI: 10.1111/jnp.12127. [Epub ahead of print].

Duez, D. (1982). Silent and non-silent pauses in three speech styles. In *Language and Speech*, 25, 11-28.

Feenaughty, L., Tjaden, K., Benedict, R.H. & Weinstock-Guttman, B. (2013). Speech and pause characteristics in multiple sclerosis: A preliminary study of speakers with high and low neuropsychological test performance. In *Clinical Linguistics & Phonetics*, 27(2), 134-151.

Ghio, A., Pouchoulin, G., Teston, B., Pinto, S., Fredouille, C., De Looze, C., Robert, D., Viallet, F. & Giovanni, A. (2012). How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers? In *Speech Communication*, 54, 664-679.

Guimarães, J., Sá, M.J. (2012). Cognitive dysfunction in multiple sclerosis. In *Frontiers in Neurology*, 3, 74.

Kraemer, M., Herold, M., Uekermann, J., Kis, B., Wiltfang, J., Daum, I. & Abdel-Hamid, M. (2013). Theory of mind and empathy in patients at an early stage of relapsing remitting multiple sclerosis. In *Clinical Neurology and Neurosurgery*, 115(7), 1016-1022.

Kurtzke, J.F., (1983). Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). In *Neurology*, 33, 1444-1452.

Labiano-Fontcuberta, A., Mitchell, A.J., Moreno-García, S. & Benito-León, J. (2014). Cognitive impairment in patients with multiple sclerosis predicts worse caregiver's health-related quality of life. In *Multiple Sclerosis Journal*, 20(13), 1769-1779.

Langdon, D.W. (2011). Cognition in multiple sclerosis. In *Current Opinion in Neurology*, 24(3), 244-249.

Rao, S.M., Leo, G.J., Bernardin, L. & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis. In *Frequency, patterns, and prediction. Neurology*, 41(5), 685-691.

Rodgers, J.D., Tjaden, K., Feenaughty, L., Weinstock-Guttman, B. & Benedict, R.H. (2013). Influence of cognitive function on speech and articulation rate in multiple sclerosis. In *Journal of the International Neuropsychological Society*, 19, 173-180.

Rosen, K.M., Goozée, J.V. & Murdoch, B.E. (2008). Examining the effects of multiple sclerosis on speech production: Does phonetic structure matter? In *Journal of Communication Disorders*, 41(1), 49-69.

Rudick, R.A., Cohen, J.A., Weinstock-Guttman, B., Kinkel, R.P. & Ransohoff, R.M. (1997). Management of multiple sclerosis. In *The New England Journal of Medicine*, 337(22), 1604-1611.

Sacks, H., Schegloff, E.A. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. In *Language*, 50, 696-735.

Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J.P., Yoon, K.E. & Levinson, S.C. (2009). Universals and cultural variation in turn-taking in conversation. In *Proceedings of the National Academy of Sciences*, 106, 10587-10592.

Swets, B., Desmet, T., Hambrick, D.Z. & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. In *Journal of Experimental Psychology*, 136, 64-81.

Voghera, M.(2001). Teorie linguistiche e dati di parlato. In Albano Leoni, F., Stenta Krosbakken, E., Sornicola, R. & Stromboli, C. (Eds.), *Dati empirici e teorie linguistiche*. Roma: Bulzoni, 75-95.

Vaughan, B., Cullen, C. (2009). Emotional speech corpus creation, structure, distribution and re-use. In *Young Researchers Workshop in Speech Technology*. Dublin, Ireland.

DOI: 10.17469/O2103AISV000017

CLAUDIO ZMARICH, ELISA RACCANELLI, GIULIA RANCAN, CARLA LEONARDI

Instability of speech production as a marker of Childhood Apraxia of Speech (CAS): Segmental and acoustic evidence

According to ASHA (2007), CAS is "a neurological childhood disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits". The present study aims to verify whether the extreme variability (at segmental as well as acoustic level) in the production of the very same linguistic unit (word or syllable) is a marker of CAS. We engaged three CAS subjects in multiple productions of the same linguistic units using the TFPI, a new phonetic test not yet published. Then we analyzed the recorded signals by means of different metrics, chosen among the most sensitive ones to track speech variability, in order to assess the consistency and stability of CAS subjects production. This was compared to that of the control groups, i.e. lexical age peers (calculated with the Italian version of MacArthur Communicative Development Inventories or CDI), chronological age peers and adults. The results suggest that CAS speech is characterized by phonological inconsistency, i.e., multiple productions of the same target very often do not share the same phonological form, especially with the longest words. Moreover, the analysis of intra-syllabic CV anticipatory coarticulation suggests that CAS children have an immature speech motor control, similar to that of 2- or 3-years-old children.

Key words: Childhood Apraxia of Speech (CAS), Coarticulation.

1. *Introduction*

Childhood Apraxia of Speech (CAS) is a neurogenic Speech Sound Disorder (SSD; DSM-V, APA 2013) whose etiology and neurobiological correlates are still unclear. According to ASHA (2007), CAS is "a neurological childhood disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits", that is, a speech motor disorder whose core deficit involves the planning and/or programming of the spatiotemporal parameters of movement sequences (Terband, Maassen, 2010). From a segmental point of view, CAS subjects often produce a reduced phonetic inventory than expected for their chronological age, and the production is often characterized by inaccuracy in the realization of speech sequences, resulting in phonemes omissions, additions, substitutions and distortions that frequently make the speech unintelligible. Adding to this, segmental errors increase as the words get longer, and/or the phonemes are part of an accented syllable (Chilosi, Lorenzini, Cerri & Cipriani, 2014). The speed of articulation of CAS subjects, measured in number of syllables per second, is significantly slower than their peers (Chilosi, Lorenzini, Fiori, Graziosi, Rossi, Pasquariello, Cipriani & Cioni, 2015). CAS subjects may also present several markers of morphological and syntactic disorders. Syntactic structures suffer from the limitations of the vocabulary and CAS children tend to produce sentences of less than average length and with a simpler grammatical structure compared to their peers (Chilosi *et al.*, 2015). It has been hypothesized that the small vocabularies of children with CAS have an upstream impact on the cognitive-linguistic aspects of phonology (Velleman, 2011).

Considering the suspect of an underlying deficit in planning or programming speech movements, it is not surprising that coarticulation in CAS subjects has been thoroughly investigated in recent years, through both acoustic and kinematic analysis. Acoustic analysis of CAS speech production was performed by means of two different methodologies, mainly by L. Nijland and colleagues on the one hand and H. Sussman and colleagues on the other.

In a first experiment, Nijland, Maassen, van der Meulen, Gabreëls, Kraaimaat & Schreuder (2002) investigated the magnitude and variability of the anticipatory coarticulation exerted by the last vowel on the first vowel and on the consonant of the nonsense disyllables [əCV], where V was /a i u/ and C was /s x b d/. Results showed that CAS children were significantly more variable in the F2 trajectories than the age matched control children and the adults, and they exhibited less anticipatory coarticulation than their controls. In a second experiment (Nijland, Maassen, van der Meulen, Gabreëls, Kraaimaat & Schreuder, 2003), CAS children and control children produced high- and low-frequency of occurrence syllable utterances, in which the syllable structures were systematically manipulated, letting unchanged the phonemes sequence (-V1s#xV2 vs -V1#sxV2). Anticipatory coarticulation, using second formant trajectories, and durational structure were analyzed. This time the results showed stronger coarticulation in CAS children when compared to controls. Furthermore, at the prosodic level, CAS children, differently from controls, did not show any metrical contrast effects, i.e. they do not change the durational structure of V1 and /s/ according to changes in stress level and/or syllabic structure.

The study by Sussman, Marquardt & Doyle (2000) is exemplificative of the second methodology. The authors made use of the Locus Equation analysis of coarticulation, of which the first author is the main proponent (for a description of the method, see the procedural section of the present paper). The authors measured the degree of anticipatory coarticulation in CV monosyllabic words produced on imitation by a small group of CAS children and a group of age matched controls. Once compared to controls, CAS *k* values (the slopes of the Locus Equations indexing the degree of coarticulation) revealed a reduced distinctiveness in intra-syllabic coarticulatory extent across stop place categories (labial, dental, velars), and lower R2 and larger SE values, indicative of greater variability.

The purpose of the present study is to verify whether the extreme variability at the segmental (defined as phonological inconsistency) as well as at the acoustic level (measured as degree of anticipatory coarticulation) in multiple productions of the very same linguistic unit, respectively word or syllable, could be a marker of CAS, and, as to coarticulation, whether the CV syllables produced by CAS children are more or less coarticulated than those produced by their controls[1].

2. *Method*

2.1 Subjects

Three subjects were reported to be affected by CAS by clinicians (one of them being the last author of the present paper) (see Table 1).

The first one (GE) presented a lexical comprehension level slightly below the chronological age (PPVT, Stella, Pizzioli & Tressoldi, 2000), while the verbal production showed a serious deficit. Morphosyntax was seriously under the standard scores both for comprehension (TCGB, Chilosi, Cipriani, 1995; PVCL, Rustioni Metz Lancaster, 2007) and production (PVB, Caselli, Pasqualetti & Stefanini, 2007). The test scores for nonverbal oral and face movements showed that also this area was seriously affected, with more difficulty on performing praxias[2] on imitation than on request (Bearzotti e Fabbro, 2003). The Phonological Working Memory (PWM) could not go over the two elements (VAUMeLF test, Bertelli, Bilancia, 2008).

Table 1 - *Clinical characteristics of CAS subjects*

	GE	*RA*	*DO*
AGE (years.months)-SEX	11.0 - F	10.4 - F	8.3 - M
Lexical Production (CDI, Caselli *et al.*, 2007)	623 words (lexical age of slightly more than 36 months)	445 words (lexical age of 31 months)	617 words (lexical age of slightly more than 36 months)
Lexical Comprehension (PPVT, Stella *et al.*, 2000)	Slightly under chronological age	Slightly under chronological age	Under chronological age
Morphosyntax Production (CDI, Caselli *et al.*, 2007)	Seriously under chronological age	Seriously under chronological age	Seriously under chronological age

[1] Inconsistency is a clinical marker of CAS if it includes some important signs: i) inconsistent errors with groping (a silent research of the articulatory locus made by tongue, lips and jaw) for consonants and vowels during the production of syllables or words; ii) unpredictable articulatory production (i.e., consonant and vowels are performed in various ways both in different words and in the same word); iii) erratic errors, atypical and not recurrent phonological processes; iiii) efforts characterized by groping. Sometimes CAS speakers make repetitions or breaks that can be wrongly interpreted as stuttering. Interestingly, inconsistency has been observed in many languages such as Cantonese, Turkish, German, Portuguese.

[2] Praxia (from ancient Greek *prásso* «act») is an intentional action that requires the ability to plan, program and execute a sequence of movements to reach a purpose or an objective.

Morphosyntax Comprehension (Rustioni *et al.*, 2007)	Seriously under chronological age	Seriously under chronological age	Seriously under chronological age (TROG-2, Bishop 2009)
Nonverbal Oral & Speech Motor Control (Bearzotti, Fabbro, 2003)	Seriously affected (more on imitation)	Seriously affected (more on imitation)	Seriously affected (more on request)
Phonological Working Memory (VAUMeLF, Bertelli, Bilancia, 2008)	No more than two elements	Just 1 element	No more than two elements

The second subject (RA) presented a lexical comprehension level, according to PPVT, below the standard scores, while the lexical production showed a serious deficit. Morphosyntax was seriously below the standard scores both for comprehension (TCGB, PVCL) and production (PVB). The test scores for nonverbal oral and face movements (Bearzotti, Fabbro, 2003) revealed these skills to be seriously affected, with more difficulty when performing praxias on imitation than on request. The PWM scored 1 (very low). Articulatory diadochokinesis (Williams, Stackhouse, 1998) was also severely impaired.

The third subject (DO) presented lexical and morphosyntax comprehension levels under the standard scores (PPVT; TROG-2, Bishop 2009). The test scores for nonverbal oral and face movements (Bearzotti, Fabbro, 2003) evidenced a difficulty in performing praxias, more on request than on imitation. Articulatory diadochokinesis (Williams, Stackhouse, 1998) was also severely affected. The PWM was limited to two elements (VAUMeLF).

Control subjects were recruited according to three different criteria. A first group of children, raised in a monolingual context and developing in a typical way according to their parents, were recruited because they scored the same number of words at the Italian version of the MacArthur Communicative Development Inventories or CDI (Caselli *et al.*, 2007) as the CAS subjects (lexical age peers: ME, male, 3 years, 6 months, 633 words attested on CDI; NL, female, 2 years, 7 months, 445 words attested on CDI; DBN, male, 3 years, 2 months, 610 words attested on CDI).

Two other groups were considered for comparison: 3 typically developing children (according to parents' statements), which have the same chronological ages as the CAS subjects (chronological age peers: MA, female, 11 years, 0 months; FG, female, 10 years, 4 months; FS, male, 8 years, 3 months), and 3 adult subjects, all females because their F0 is more similar to children's F0 (HC, 28 years; MF, 23 years; PM 21 years), who declared not to have ever suffered for cognitive, speech or motor impairment and to speak Italian as their native language.

2.2 Procedure

The aim of the present study is to verify whether the extreme variability in the production of the very same linguistic unit (word or syllable) could be a marker of CAS. Consequently, we devised to engage the subjects in multiple productions of the same word, and then we analyzed the recorded signals by means of different metrics (both segmental and acoustic), chosen among the most sensitive in tracking variability, in order to compare the consistency/stability of the CAS subjects to that of their control subjects.

CAS subjects were administered the Italian version of CDI (Caselli *et al.*, 2007) which consists of a 680-words vocabulary production checklist, to be compiled by parents, which is designed for use with children in the 1;4-2;6-years range. It allowed us to calculate the lexical age, based on total words produced according to their parents (the same test, as already said, was administered to the younger control group). According to CDI, CAS subjects resulted to have a relatively restricted lexicon (GE: 623 words; RA: 464 words; DO: 617 words).

CAS subjects and control subjects filled (or the parents filled for them, according to the ages of the subjects) a custom-made, socio-linguistic questionnaire, which reported basic information about psycho-physiological and linguistic development, and they performed the Test Fonetico per la Prima Infanzia (TFPI), a new phonetic test not yet published (see Zmarich, Fava, Del Monego & Bonifacio, 2012). The TFPI contains a naming task divided into two subtests based on the age of the subjects (18-23 months; 24-47 months). Its construction took into account a Phonetic criterion, i.e. the consonant phonemes of Italian must be established in at least two different words for each position of the tested words, a Semantic/Frequency criterion, i.e. the words have to be high frequency concrete nouns, and a criterion of Gradualness in phonetic complexity, i.e. words must evidence an increase in complexity from the first to the second age group, by number and types of syllables. The TFPI form appropriate for subjects aged 25-47 months was chosen, which includes 78 figures displayed on a PC screen to be named by the subjects. Considered as a group, these words include all the 5 vowels of the Italian unstressed vowel system [i e a o u], and each vowel occur more than twice. This particular was critical in order to ensure variability in the vocalic segments, which is a pre-condition for the acoustic analysis of coarticulation by means of the Locus Equations method[3]. Crucially, in order to create opportunity for speech variability manifestation, the subjects had to produce the TFPI items three times in the same day (each time the order of items presentation was randomized) and the lexical productions were recorded at 44 kHz and 16 bit by means of an Edirol R-09 in a silent room.

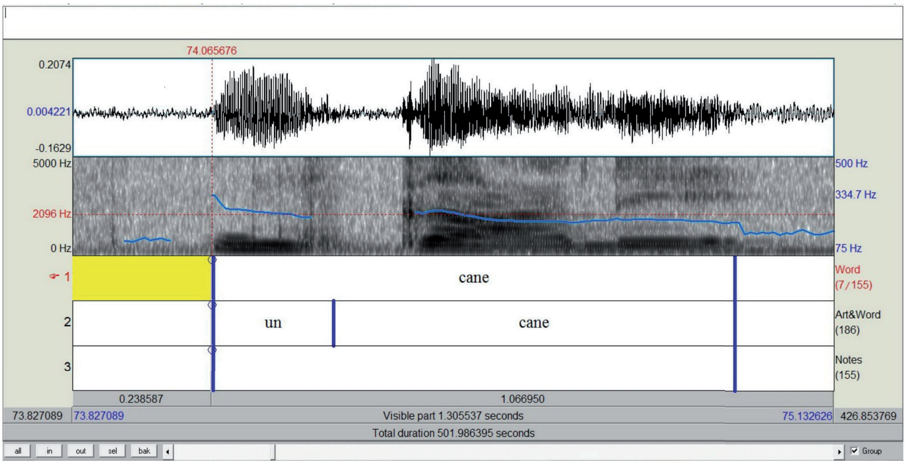
Phonetic transcription (in IPA) was always performed by the second or the third author working in an exclusive way on a single subject (i.e., each subject has been transcribed by the same transcriber). All the transcriptions were then checked

[3] For this reason, we made sure that all the subjects produced the syllables beginning with plosives with a statistically sufficient number of occurrences for at least the cardinal vowels [i a u].

by the first author, which took also the final decision as to the few disagreements. During the transcription process, the PRAAT software (http://www.fon.hum.uva.nl/praat/) was used as a practical help and also to segment and label the words produced by subjects. Only the fluent and intelligible productions of correct or incorrect words of TFPI were considered for the analysis.

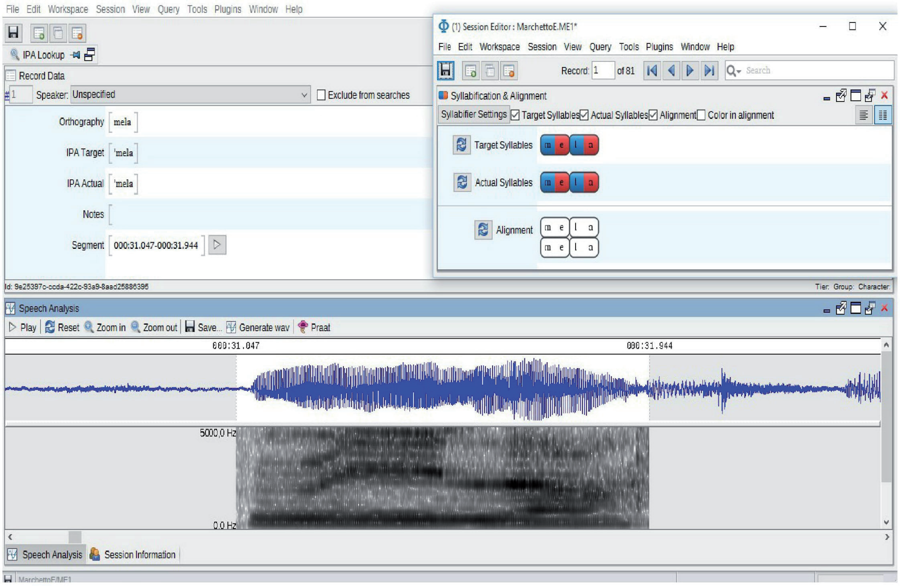
After that, words in the Word tier of PRAAT (see Figure 1) were exported as records to a PHON database (https://www.phon.ca/phontrac; Rose, Stoel-Gammon, 2015), in order to accomplish the phonological analysis made possible by the syllabification, segmental alignment, and matching between each lexical target (IPA Target) and the child's effective realization (IPA Actual, see Figure 2).

Figure 1 - *Praat display reporting the waveform of the lexical production segmented and annotated according to three tiers*



To this end, we exploited the interfacing between PRAAT and PHON to convert the content and the timing information, represented in the PRAAT Word tier by the orthographic word and by the initial and final border, respectively, into a record of Phon, by means of a custom script written by Vincenzo Galatà (CNR-ISTC). A PHON record is the basic element of a PHON database, and in this case is represented on the screen by an orthographic transcription of the target word (the word the child is attempting to), an IPA transcription of the target word, and an IPA transcription of the child attempt. Close to this section there is the acoustic signal of the word accompanied by the article, if present (corresponding to the Art/Word tier on PRAAT). In this way, a PHON database may contain a number of records corresponding to all the words produced by every single child. They appear already transcribed in IPA on the IPA Target tier thanks to an Italian phonetic dictionary incorporated into PHON. As a final step, the user of PHON will fill with IPA symbols the tier devoted to the effective child's production (IPA Actual). At this point data are ready to be analyzed through comparisons between target (adult/model) and actual (child/produced) form, and an algorithm that performs bestguess segmental alignments between corresponding target and actual IPA will help to classify the children's errors in one of the several types of phonological processes, or to produce a phonetic inventory, or the percentage calculation of consonants correctness, etc.

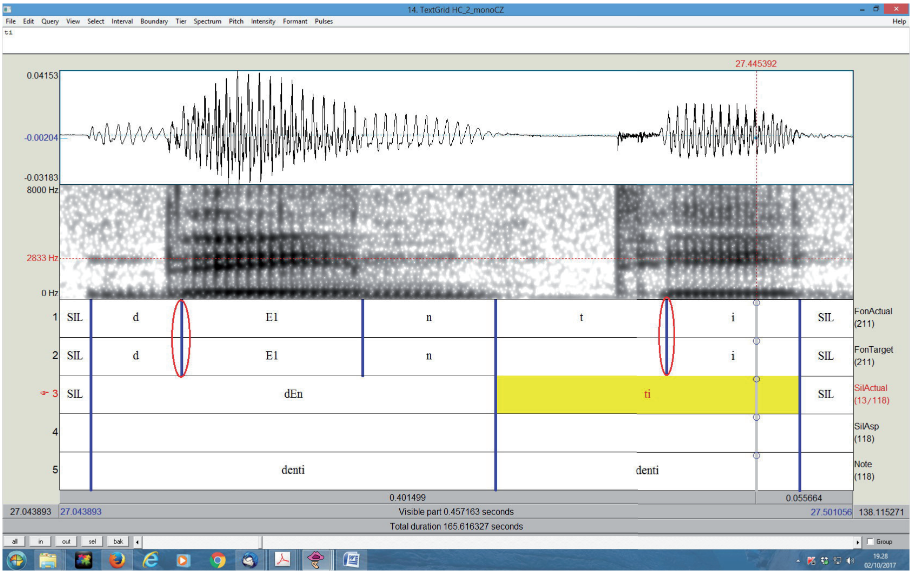
Figure 2 - *PHON display reporting the waveform of the lexical production (i.e. the record, at the bottom), three tiers containing the transcription of the lexical target (IPA Target) and the child's effective production (IPA Actual)*



Successively, any clear and intelligible C1V and C1VC syllables (C1 = plosive; V = any vowel) were looked for and further segmented and labeled at phone and syllable levels and processed by a PRAAT script (by V. Galatà) to obtain F2 values (Hz) on the first available oscillatory cycle after closure release (see "denti" in Figure 3) and at mid-vowel. A SilAsp tier was also included in order to report any aspirated stop release (more than 50 ms of aspiration after the release), which were later excluded from the coarticulation analysis for possible flaws. A second script calculated VOT, but we do not report the results of the VOT analysis here.

Since we would like to assess the degree of Consonant-Vowel intrasyllabic, anticipatory coarticulation, we calculated it by means of the Locus Equations method (Krull, 1988; Sussman, Marquardt & Doyle, 2000; see Sussman *et al.*, 1999, for an application to speech development), separately for bilabial, dental and velar place of articulation of the consonants.

Figure 3 - *PRAAT display: on the top, the waveform of a lexical production ("denti"), on the bottom, five tiers reporting the segmentations and labeling at different levels (word, syllables, segments). The boundaries used to extract F2 values at the C-V border are circled by a red line. The "SilAsp" tier is used to mark aspirated syllables, i.e. syllables beginning with a voiceless aspirated plosives (VOT greater than 50 ms)*



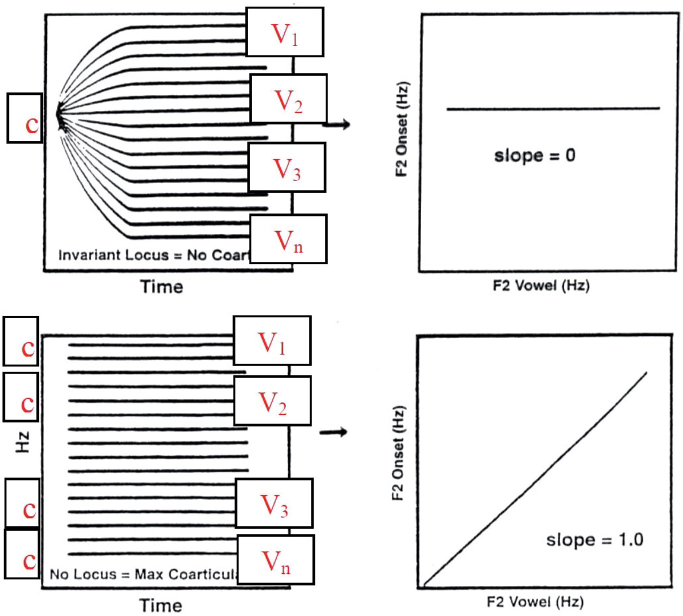
In the Locus Equations, the coefficient *k*, representing the slope of the regression line described by the equation: *F2 Onset* = *k* * *F2 Vowel* + *c*, indexes the degree of coarticulatory influence of V on C. The *k* coefficient and the *c* coefficient are calculated by means of a regression analysis of the dependent variable (the F2 consonant values) on the independent variable (the F2 vowel values). The value of *k* could vary between 0 (no coarticulation at all) and 1 (maximal coarticulation, Figure 4). It is interesting to note that F2 onset and F2 vowel, within a given place category, are consistently and robustly linearly correlated across diverse speakers (they are inherently normalized because they result from measures taken from the *same* syllable) and languages, and even under perturbation conditions as imposed by bite blocks. In addition, the particular linear function relating these two parameters is itself a function of place of articulation. Labials have been found to have the steepest regression functions, followed by velars, and then alveolars (for Italian, see Zmarich, Bortone, Vayra & Galatà, 2013).

3. *Results*

The first block of the results we present are drawn from Rancan (2015), which compared the segmental characteristics of the lexical productions of two CAS subjects (GE and RA) with those produced by the lexical age peers, on a number of phonetic-phonological measures (derived from the compilation of phonetic inventory and from the error analysis). We choose to match CAS subjects to lexical age peers because we felt that the alternative choice, involving a comparison with age matched normal subjects (from 8 to 11 years), would have been too "punitive" for CAS children.

The results from the acoustic analysis are drawn from Raccanelli (2016), which recruited another CAS subject (DO) and analyzed the lexical productions on a number of acoustic measures (VOT and degrees of intra-syllabic coarticulation). In the last work, CAS subjects were compared to lexical age peers, as before, but also to the chronological age peers and to the adults. In fact we felt that the previous cautions concerning the non-use of the age matched controls when using the phonetic test could be attenuated when a measure of speech motor functioning (coarticulation) was considered.

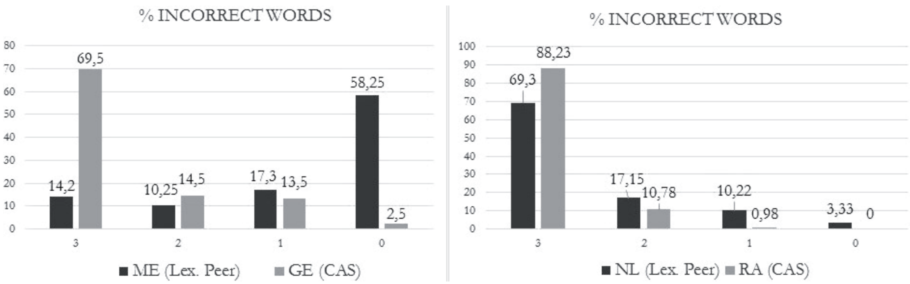
Figure 4 - *Hypothetical extremes of Locus Equation slopes (modified from Sussman et al., 1999). The top panel illustrates the F2 transition representing no coarticulation between the vowels and the consonants, and the Locus Equation slope of zero that would result from such a situation. The lower panel illustrates maximum coarticulation between vowels and consonants with no fixed consonantal locus and a resulting Locus Equation slope of 1.0*



3.1 Results of the segmental analysis

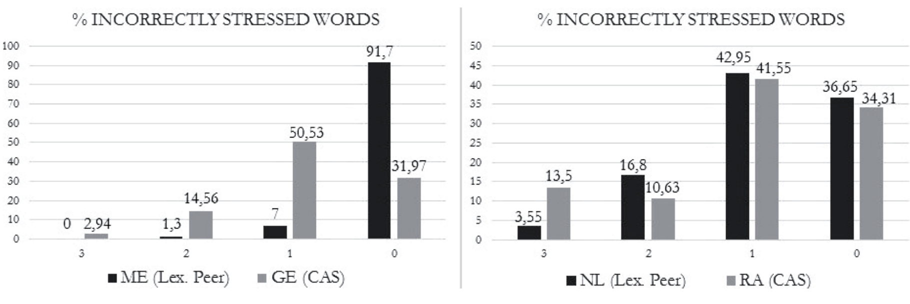
According to segmental analysis, the speech production of GE and RA sounds scarcely fluent and is affected by a wide range of error processes. In line with the literature findings, the subjects present a difficulty in combining phones in syllables and syllables in words. The percentage of errors increases with length and complexity of words. The subjects have an incomplete and atypical phonetic inventory, since they have not yet consistently acquired the consonants normally mastered since 24-32 months of age (Zmarich, Bonifacio, 2005). Speech is characterized by widespread abnormalities on different levels, in rhythm as well as in stress allocation. As far as consistency is concerned, we can select some analysis in order to demonstrate that multiple productions of the same speech targets are phonologically more inconsistent (from trial to trial) than those produced by the lexical age peers. Figure 5 shows the percentage distribution of incorrect words (due to segmental omissions, additions, substitutions and distortions), according to the number of words out of the totals, that were never mistaken (0), mistaken once out of three repetitions (1), mistaken twice (2), mistaken all the times (3). Although the difference is particularly evident for GE compared to the lexical peer ME, which produced incorrectly for three times the 69% of words, it is notable even for the other pair of subjects.

Figure 5 - *Percentage distribution of segmentally incorrect words according to the number of trials (3 = always incorrect; 0 = always correct)*



We know from the literature (i.e. Chilosi, Lorenzini, Cerri & Cipriani, 2014) that lexical stress is often displaced to another, normally unstressed, syllable. If we define a word as "incorrectly stressed" when we perceive that lexical stress is allocated on a syllable which is different from that one of the standard pronunciation, we found that one of the two CAS subjects (GE) stressed correctly all the repetitions of the lexical targets only the 31.97% out of all the attempts (Figure 6).

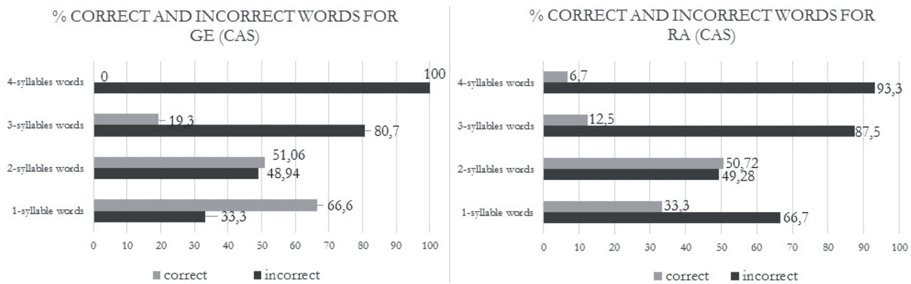
Figure 6 - *Percentage distribution of incorrectly stressed words according to the number of trials (3 = always incorrect; 0 = always correct)*



This low value contrasts with the 91% of total words always correctly stressed by the lexical peer (ME).

From the same literature we also know that longer words present greater difficulties for CAS subjects. Figure 7 shows the percentages of segmentally correct and incorrect words out of the total number of words having a given length in syllables (from one to four syllables). In this case, the total is represented by all the repetitions of the target words having the same length in syllables.

Figure 7 - *Percentage distribution of segmental correctness of the words, according to the number of constituting syllables, calculated over all the three trials*



The two CAS subjects perform similarly (but, again, GE is worse), generally making more and more mistakes as the words get longer[4]. However, since 4-syllable words have more segments than monosyllabic words, they are also more likely to contain more errors. In order to ascertain whether longer words are effectively more subject to errors, we calculated the number of incorrect syllables out of the number of the syllables in the lexical targets, by excluding errors relative to vocalism (because more prone to biases in the phonetic transcription), but including errors in syllabification (i.e. degemination, causing a CVC.CV word structure to become CV.CV). The results definitively excluded this possibility: as an example, the CAS subject GE, who was the worse, attempted to produce the 23 three-syllabic disyllabic target words three times each (for a total of 207 target syllables), and the resulting rate of incorrect/target syllables was 0.194 (almost one syllable incorrect out of five), which contrasts with the rate of 0.312 (almost one syllable incorrect out of three), when he attempted to produce the 50 disyllabic target words, three times each (for a total of 300 target syllables).

[4] There seems to be an exception: RA has more correct disyllabic than monosyllabic words. We can justify this behaviour by the current knowledge about the acquisition of speaking ability in CAS children: words and sentences tend to be acquired as whole units and the child can pronounce a whole word or even a sentence (such as "What time is it?") by the recruitment of an automatic mechanism. In this condition the articulation shows high rate fluency. On the contrary, when the child is trying to articulate intentionally rather than automatically, she/he may not be able to do it, even in the case of short syllables that frequently are codified as non-words. This situation influences the production by slowing down the fluency and/or leading to pronunciation errors. This phenomenon is called the automatic-voluntary dissociation. It is an erratic and highly variable behaviour, quite difficult to fully understand.

3.2 Results of the acoustic analysis

As already described in the procedure, the degree of anticipatory coarticulation of the vowel on the consonant, in all the CV and CVC syllables produced by the subjects in relation to the target words of the TFPI, has been calculated by means of Locus Equation method, separately for the bilabial, dental and velar place of articulation of the consonants. For this analysis, a third CAS subject was added, who was administered the phonetic test. This time, the control subjects were not limited to the lexical peers, but the chronological peers and the adults were also considered.

The experimental hypothesis consisted in verifying if, compared to controls: i) CAS subjects show similar degrees of coarticulation; ii) CAS subjects are able to differentiate the three places of articulation (i.e. are able to use a distinct degree of coarticulation for each place, as it is attested for Italian, see Petracco, Zmarich, 2006, and for other languages, see Sussman, McCaffrey & Matthews, 1991); iii) CAS subjects evidence more variability from trial to trial; iiii) CAS subjects evidence more variability in general, independently from trials.

The number of CV and CVC syllables of the intelligible lexical productions selected for each trial of each subject and categorized according to the articulatory place was around 20, ranging from a minimum of 4 (velars produced by the two CAS children GE and RA) to a maximum of 39 (dentals produced by an adult control subject, ME). A mixed ANOVA with 2 between factors (Subject status, Articulatory Place) and 1 within factor (Trials) has been performed on the dependent measure *k* coefficient. The levels of the Subject status factor were four: CAS subjects, lexical peers, chronological peers and adults. Since the statistical analysis did not reveal any significant interaction between the trials factor and the status factor, we excluded that CAS subjects were more variable from trial to trial than control subjects, and collapsed all the repetitions together.

We then run an ANOVA keeping only the factors Subject status and Articulatory Place as variables. Table 2 shows the number of syllables (or F2c-F2v pairs) on which the Locus Equations has been calculated. The values of the *k* coefficient of the equations (which indexes the degree of coarticulation) are reported on Figure 8, averaged across trials.

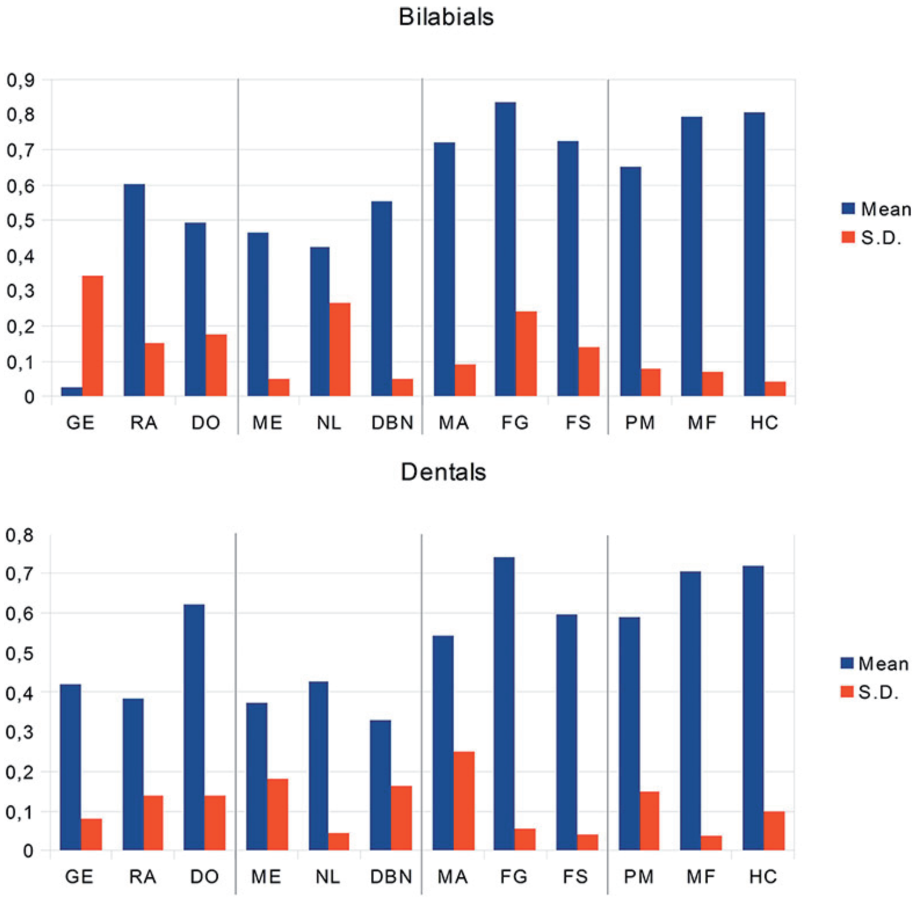
Table 2 - *Number of CV and CVC syllables selected for the coarticulation analysis in each subject, summed across the three trials*

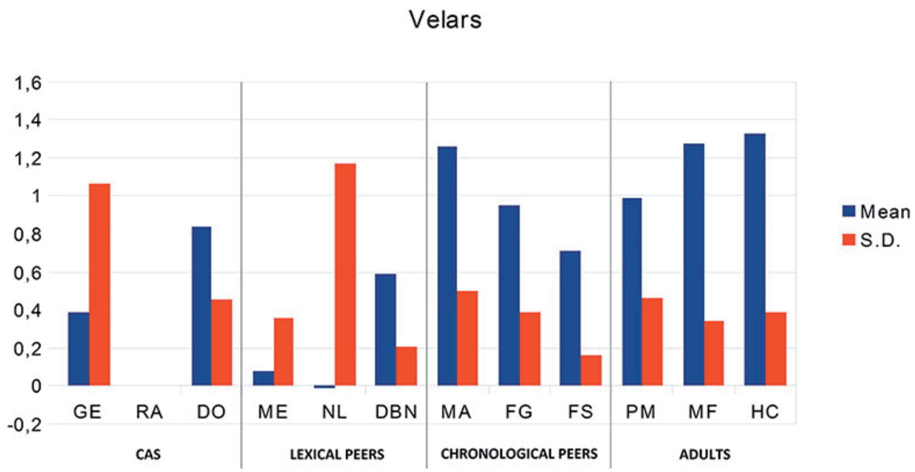
	CAS			*Lexical peers*			*Chronol. peers*			*Adults*		
	GE	RA	DO	ME	NL	DBN	MA	FG	FS	PM	MF	HC
BIL.	52	43	74	76	51	76	63	66	63	67	61	62
DEN.	73	59	79	114	78	107	69	79	79	85	71	79
VEL.	27	9	40	53	27	84	47	43	51	43	48	55

Results show that Subject status and Articulatory Place are statistically significant ($F_{(3,93)} = 13.166$, $p = 0.000$; $F_{(2,93)} = 3.419$, $p = 0.037$ respectively), as well as their interaction (F (6,93) = 2.337, p = 0.038). A post-hoc Tukey's Honestly-Significant-Difference Test on Subject status evidences that CAS and Lexical Peers have the lowest *k* values (i.e. less coarticulation: 0.486 and 0.359, respectively) and contrast significantly (p < .05) with Chronological Peers (0.786) and Adults (0.873).

Three ANOVAs, performed on the Subject status factor within each place of articulation (Bilabial, Dental, Velar) shows that CAS and Lexical Peers have the lowest *k* values and always contrast significantly (p < .05) with Chronological Peers and Adults. As a further analysis, we performed 4 mixed ANOVAs on *k* values within each Subject status condition, keeping Articulatory Place as a between factor and Trials as a within factor. Results show that only adults distinguish significantly the three places of articulation (F(2,6) = 15.730, p = 0.004).

Figure 8 - *The values of k, the coefficient of the Locus Equations indexing the degree of the anticipatory coarticulation of V on C, averaged across tree trials (S.D.: standard deviations)*





An ANOVA on Standard Deviations values resulting from the k values in the three trials does not evidence any significant difference, although CAS children always exhibit the higher SD values, indicating a trend towards more variability.

4. *Discussion*

According to the segmental analysis based on the phonetic transcription, the multiple productions (until three times) of the same speech target were phonologically more inconsistent (from trial to trial) for the two CAS subjects rather than their lexical peers, and the percentage of incorrect words increased with length and complexity of words. There were, however, exceptions: for instance, the CAS subject RA showed more correct disyllabic than monosyllabic words (see above, Footnote 4).

To sum up, CAS subject speech was characterized by widespread abnormalities at different levels, in phonological segments as well as in stress allocation. These results agree with the existing literature (e.g. Chilosi *et al.*, 2014). Whereas the performances of the two CAS subjects were quite similar to one another, the control subject ME performed much better than the control subject NL. This discrepancy might raise a question about the criteria of selection of control subjects; we will keep it as a warning for the future. However, the alternative choice of matching the subjects for the segmental analysis of the words produced by means of a naming task (i.e. the phonetic test) on the basis of the chronological age would have been even worse.

As to the analysis of intra-syllabic CV anticipatory coarticulation, a third CAS subject was added, and he was administered the phonetic test. All the CV and CVC syllables in the intelligible lexical productions elicited through the phonetic test were analysed. The three CAS subjects and the lexical age peers obtained low coefficients of coarticulation and they were significantly different from the chronological age peers and the adults, suggesting that CAS children could have an immature speech motor control (like that of 2- or 3-years-old normally developing children). Anyway, we did not find the CAS children to be significantly different from controls, as to intra-individual variability among multiple repetitions of the same lexical target, as it was initially hypothesized.

The finding of weaker coarticulation in CAS children as compared to age-matched controls (but similar to lexical peers aged from two to three years) is difficult to interpret. We still do not know if the CAS children are different from age-matched controls. If they are, then a question can be raised, about whether the developmental meaning of the difference is abnormality or delay. In other words, we need to know both the characteristics of normal speech development and the characteristics of a normal mature speech system to be able to reliably recognize immaturity. Only a few studies, using different methodologies as summarized in the introduction, have compared directly the coarticulation of CAS subjects to that of normally developing children and normal adults, and they arrived at different conclusions. Weaker coarticulation in CAS subjects with respect to age-matched peers was found by Nijland *et al.* (2002), but the coarticulation degree of the age-matched peers was stronger than in the adults. In a later experiment (Nijland *et al.*, 2003), the coarticulation pattern was found to be stronger in CAS children than in the age-matched controls, and both of them were stronger than the adult coarticulation pattern. Since the authors related greater immaturity to stronger coarticulation pattern as a possible consequence of a more global and undifferentiated planning and programming of the syllables (Nittrouer, Studdert-Kennedy & McGowan, 1989), this result could qualify CAS subjects as immature. According to this interpretation, the low degree of coarticulation exhibited by the CAS subjects of our experiment would not qualify them as immature.

The pattern of weaker coarticulation that we found in CAS children was also found by Sussman *et al.* (2000), and their discussion could bring a new light into the question. Sussman *et al.* (2000) start from the established fact that in adult speakers the regression lines, interpolating the x,y coordinates of F2 values of each syllable, have a typical slope, described by the coefficient *k* of the Locus Equation, which associates to the place of articulation of the consonant. In normal condition, the value of k is maximal for velars, intermediate for bilabials and minimal for dentals. This ordering was confirmed several times by independent researchers and has been confirmed also for Italian (Petracco, Zmarich, 2007; Zmarich *et al.* 2013). According to Sussman et al (2000: 309) “LE slopes directly encode extent of anticipatory coarticulation by capturing the strength of the vowel’s influence on the preceding stop consonant”. The CAS children of Sussman *et al.* (2000) experiment, differently than age-matched controls, did not show this clear ordering. In the words of Sussman et al (2000: 309), they “showed a reduced distinctiveness in coarticulatory extents across stop place categories as indexed by LE slopes”. The authors advance the suggestion that this lack of contrastiveness contributes to the lower intelligibility of the speech of CAS speakers. Interestingly, low *k* values and lack of contrastiveness across stop place categories were also found as a result of our experiment in typical developing children much younger than CAS, matched to CAS subjects on the basis of the same number of words attested in CDI (the so called lexical peers). This similarity could suggest that CAS children could suffer from an immature speech motor control, which is characteristic of 2- or 3-years-old children (Raccanelli, 2016). This hypothesis could find some confirmation in the fact that even at a more advanced age between 42 and 47 months, a group of typically developing children was found to have lower *k* values and lower distinctiveness across stop place categories than adults (Zmarich *et al.*, 2013). Possibly, the same reasons invoked by Zmarich *et al.* (2013) to interpret their results could find an application here: in the case of bilabials, the occlusion (being anatomically not binding for tongue) provides the adult speaker a maximal temporal overlap of the tongue and lip gestures. The degree of anticipatory coarticulation is maximal since at the time of the stop release the back of the tongue is already in position for the vowel (coarticulation as co-production). The lower *k* values of CAS subjects are possibly due to the fact that the tongue is still in motion at the moment of release, indicating that these children still have to learn to synchronize the tongue occlusion movement with the release. As to dentals, CAS children possibly still need to learn to differentiate and coordinate the tip and the back of the tongue, which can be moved quite independently one from the other even though they belong to the same organ. In the case of velars the biomechanical constraints are maximal (C and V are articulated with the back of the tongue), and in adult speech the acoustic effect highlights a very high degree of coarticulation as a reciprocal adaptation. Possibly in CAS children the consonant does not change the place of articulation as a function of the vowel's place, forcing the articulators to an extra effort to reach the required positions to produce the vowel quickly.

As a final note, some caution about the interpretation of these results could come from the absence of a further analysis regarding speaking rate, which could be an important determinant of coarticulation patterns (Agwuele, Sussman & Lindblom, 2008), especially when CAS subjects, whose speech rate is notoriously slow and laborious, are considered. Our next effort will be then that of trying to relate coarticulatory patterns to syllable duration.

Bibliography

American Speech Language Hearing Association (2007). *Childhood Apraxia of Speech: Ad Hoc Committee on Apraxia of Speech in Children*.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5).* Washington, DC: American Psychiatric Pub.

Agwele, A., Sussman, H.M. & Lindblom, B. (2008). The Effect of Speaking Rate on Consonant Vowel Coarticulation. In *Phonetica*, 65, 194-209.

Bearzotti, F., Fabbro, F. (2003). Test per la valutazione delle Prassie orofacciali nel bambino, Protocollo B e C. In *Giornale di Neuropsichiatria dell'Età Evolutiva*, 23, 406-417.

Bertelli, B., Bilancia, G. (2008). *VAUMeLF: Batteria per la valutazione dell'attenzione uditiva e della memoria di lavoro fonologica nell'età evolutiva*. Firenze: Giunti O.S.

Bishop, D.V.M. (2009). *TROG-2: test for reception of grammar, version 2.* Firenze: Giunti O.S.

Caselli, M.C., Pasqualetti, P. & Stefanini, S. (2007). *Parole e frasi nel "Primo Vocabolario del Bambino"*. Milano: Franco Angeli.

Chilosi, A.M., Cipriani, P. (1995). *TCGB Test di comprensione grammaticale per bambini*. Edizioni del Cerro.

Chilosi, A.M., Lorenzini, I., Cerri, B. & Cipriani, P. (2014). Disprassia Verbale Evolutiva: inquadramento clinico e diagnosi differenziale con il disturbo fonologico. In Marotta, L., Caselli, M.C., (Eds.), *Disturbi del linguaggio, caratteristiche, valutazione, trattamento*. Trento: Erickson, 145-161.

Chilosi, A.M., Lorenzini, I., Fiori, S., Graziosi, V., Rossi, G., Pasquariello, R., Cipriani, P. & Cioni, G. (2015). Behavioral and neurobiological correlates of childhood apraxia of speech in Italian children. In *Brain and Language*, 150, 177-185.

Grigos, M.I., Moss, A. & Lu, Y. (2015). Oral Articulatory Control in Childhood Apraxia of Speech. In *Journal of Speech, Language, and Hearing Research*, 58, 1103-1118.

Krull, D. (1988). Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops. In *Phonetic Experimental Research at the Institute of Linguistics*, Stockholm University, 7, 66-70.

Maassen, B., Nijland, L. & Terband, H. (2010). Developmental models of Childhood Apraxia of Speech. In Maassen, B., Van Lieshout, P.H.H.M (Eds.), *Speech motor control: New developments in basic and applied research*. Oxford: Oxford University Press, 243-258.

Nijland, L., Maassen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F.W. & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. In *Clinical Linguistics & Phonetics*, 16, 461-483.

Nijland, L., Maassen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F.W. & Schreuder, R. (2003). Planning of syllables in children with developmental apraxia of speech. In *Clinical Linguistics & Phonetics*, 17, 1-24.

Nittrouer, S., Studdert Kennedy, M. & McGowan, R.S. (1989). The Emergence of Phonetic Segments Evidence from the Spectral Structure of Fricative-Vowel Syllables Spoken by Children and Adults. In *Journal of Speech, Language, and Hearing Research*, 32, 120-132.

Petracco, A., Zmarich, C. (2006). La quantificazione della coarticolazione nello sviluppo fonetico. In Giordani, V., Bruseghini, V., Cosi, P. (Eds.), *Scienze vocali e del linguaggio. Metodologie di valutazione e risorse linguistiche.* Torriana (RN): EDK Editore srl, 135-150.

Raccanelli, E. (2016). Un'analisi acustica della coarticolazione e del VOT in bambini con Disprassia Verbale Evolutiva (DVE). Tesi di Laurea in Logopedia, Università di Padova.

Rancan, G. (2015). Applicazione di un nuovo Test Fonetico a soggetti con Disprassia Verbale Evolutiva e a soggetti con sviluppo tipico di età tra i 36 e i 48 mesi. Tesi di laurea in Logopedia, Università di Padova.

Rose, Y., Stoel-Gammon, C. (2015). Using PhonBank and Phon in studies of phonological development and disorders. In *Clinical Linguistics & Phonetics*, 29, 686-700.

Rustioni Metz Lancaster, D. (2007). *Prove per la valutazione della comprensione linguistica*. Firenze: Giunti O.S.

Stella, G., Pizzioli, C. & Tressoldi, P.E. (2000). *PPVT Peabody – Test di vocabolario recettivo.* Torino: Omega.

Sussman, H.M., McCaffrey, H.A. & Matthews, S.A. (1991). An investigation of locus equations as a source of relational invariance for place categorization. In *The Journal of the Acoustical Society of America*, 90, 1309-1325.

Sussman, H.M., Duder, C., Dalston, E. & Cacciatore, A. (1999). An Acoustic Analysis of the Development of CV Coarticulation. A Case Study. In *Journal of Speech, Language, and Hearing Research*, 42, 1080-1096.

Sussman, H.M., Marquard, T.P. & Doyle, J. (2000). An Acoustic Analysis of Phonemic Integrity and Contrastiveness in Developmental Apraxia of Speech. In *Journal of Medical Speech-Language Pathology*, 8, 301-313.

Terband, H., Maassen, B. (2010). Speech Motor Development in Childhood Apraxia of Speech: Generating Testable Hypoteses by Neurocomputational Modeling. In *Folia Foniatrica & Logopaedica*, 62, 134-142.

Terband, H., Maassen, B., van Lieshout, P. & Nijland, L. (2010). Stability and composition of functional synergies for speech movements in children with developmental speech disorders. In *Journal of Communication Disorders*, 44, 59-74.

Velleman, S.L. (2011). Lexical and phonological development in children with childhood apraxia of speech-a commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. In *Journal of Child Language*, 38, 82-86.

Williams, P., Stackhouse, J. (1998). Diadochokinetic skills: Normal and atypical performance in children aged 3-5 years. In *International Journal of Language & Communication Disorders*, 33, 481-486.

Zmarich, C., Bonifacio, S. (2005). Phonetic inventories in Italian children aged 18-27 months: a longitudinal study. In *Proceedings of INTERSPEECH 2005 – Eurospeech, 9th European Conference on Speech Communication and Technology.* NewYork (NY): ISCA, 757-760.

Zmarich, C., Fava, I., del Monego, G. & Bonifacio, S. (2012). Verso un "Test Fonetico per la Prima Infanzia". In Falcone, M., Paoloni, A. (Eds.), *La voce nelle applicazioni*. Roma: Bulzoni, 51-66.

Zmarich, C., Bortone, E., Vayra, M. & Galatà, V. (2013). La coarticolazione e il VOT nello sviluppo fonetico: studio sperimentale su bambini dai 42 ai 47 mesi d'età. In Galatà, V. (Ed.), *Multimodalità e multilingualità: la sfida più avanzata della comunicazione orale.* Roma: Bulzoni, 475-493.

PARTE IV

BALBUZIE E TEORIA FONETICA

DOI: 10.17469/O2103AISV000018

CLAUDIO ZMARICH

Stuttering and phonetic theory: An introduction

Stuttering has been traditionally defined by making reference to the auditory detection and qualitative assessment of the dysfluencies (some of them being abnormal for number, type, duration and position) and their distributional patterns have been explained by invoking the same dysfunctions of the linguistic representations and processes which generate lapsus and disfluencies in non-stuttering speakers. However, fluency is multidimensional: not only the fluent speech is (relatively) devoid of discontinuities, but it is also produced with a regular rhythmic beat, at fast rate and without an excessive physical and mental effort. As a phonetician, two main questions about stuttering are worthwhile. The first is: "Why would Phonetics be so important in the study of stuttering?" A possible answer could be that Phonetics is at the convergence of different scientific realms, and for such a role it holds a privileged key for unifying and simplifying the understanding of the multidimensional aspects of stuttering (made of sociocultural, psychological, physiological and genetic factors). In order to accomplish a causal function in stuttering, each of these variables must at the end interact with the motor control processes of the speech apparatus, traditionally studied by Phonetics. As to the second question "Why should a phonetician be interested in stuttering?" we could answer that phoneticians could feel a potential attraction towards a speech disorder that selectively affects fluency, in individuals that are judged to be healthy and normally endowed with reference to cognitive and emotional aspects. At the same time, the speech aspects under investigation in stuttering are at the heart of a number of theories of speech production, for what it is about the conceptualization of the time dimension and of the speech variability.

Key words: disfluencies, stuttering, phonetics, psycholinguistics.

1. *Introduction*

Within the realm of Phonetic Sciences, over the last thirty years, an informal community of researchers begun to take shape and grow around a series of conferences known with the general name of "Laboratory Phonology" (the first of them was organized and the proceedings edited by Kingston, Beckman, 1990). This community was later celebrated by three of their most representative founders in a programmatic paper (Pierrehumbert, Beckman & Ladd, 2000), which I will refer to, because it makes clear how phoneticians and people working on stuttering may benefit the ones from the others. What is shared by Laboratory Phonology is the belief that Phonology (a label covering much of the Phonetics, in the authors' point of view) is one of the natural sciences, and that everything in language, including language-specific features and sociolinguistic variation, is part of the natural world. This attitude is contrary to the mind-body dualism, and believes that categorization underlying the phonological constructs is ultimately based on physical non-linearity grounded in the systems of speech production, speech perception and in the acoustic medium, which, acting together through speech variability, can lead to the emergence of the abstract categories. For this reason, researchers who recognize themselves in Laboratory Phonology's approach make use of laboratory methods to discover and explain the phonetic form of language. According to Lindblom (1995), Phonetics is in a privileged position compared to the other domains of Linguistics (e.g. Syntax) to develop this program. It can invoke a type of knowledge that is relevant to language but which has been acquired independently of it, such as information on general mechanisms of hearing and motor control (i.e., it uses facts and principles whose empirical motivation is independent from the data to be explained). Hence we can see a first difference with those speculating on language "starting from syntax", who believe that languages are constructed in arbitrary and unnatural ways. A second fundamental difference is that, contrary to what is claimed by those who speculate on language starting from syntax, these physiological mechanisms would not be modular, that is, specific to the language of *Homo Sapiens*, as they also serve other physiological functions, evolved naturally from mammals. Starting from the assumption that speech forms are the means the languages provide to make interpersonal communication possible and are therefore public actions (and not mental categories; Fowler, 2014), a theory of Phonology should then explain the properties of these actions; a theory of speech production, how these actions are accomplished; a theory of speech perception, how these actions are perceived. Finally, since humans are subject to natural evolution, the theoretical construction have to explain how these actions could emerge through interpersonal communication exchanges in the child, and how these actions could be reduced or distorted in speech and language disorders (as it happens in stuttering).

The research activity underlying the elaboration of these theories must be based on the cooperation of scientists of different extraction, on a common vocabulary, on the existence of auxiliary theories (such as those relating to the functioning of the instruments used), and on a mathematical formulation, specifically of continuous type (as opposed to discrete). In this perspective, clinical disorders should be considered as an opportunity for basic scientific research: "When basic science research is integrated with knowledge about clinical disorders, the disorders are viewed as natural experiments and opportunities to observe factors that are not normally available for experimental manipulation in humans, and/or that can not be viewed across a sufficient range of settings under normal conditions" (Bernstein, Weismer, 2000: 225).

Stuttering has been defined as a "disorder in the rhythm of speech, in which the individual knows precisely what he wishes to say, but at the time is unable to say it because of an involuntary repetition, prolongation or cessation of a sound" (International Classification of Diseases and Related Health Problems, ICD-9, World Health Organization, 1977: 202). We prefer this somehow dated definition to the current ICD 10 (2007) definition[1], because it marks explicitly the peculiar nature of stuttering disfluencies, their being "involuntary" and their being perceived by People Who Stutter (henceforth PWS) as a "loss of controls" on the articulators (Perkins, 1990; Perkins, Kent & Curlee, 1991). This single characteristic would suffice to suggest the reference to a motor disorder rather than a language disorder (Perkins, 1990).

Previously, we declared our proximity to Laboratory Phonology, but our inquiry on noble ancestors would be incomplete without quoting Clinical Linguistics, which in Crystal's words "concerns the application of linguistic science to the study of communicative disabilities, as they meet in clinical situations" (Crystal, 1981: 31). According to Ball, Kent (1987), in their preface to the first issue of the *Clinical Linguistics and Phonetics* journal, it covers the application of analytical linguistic/phonetic techniques to clinical problems, or the demonstration of how clinical data could contribute to theoretical issues in Linguistics/Phonetics. Following these reasoning, as a phonetician, two main questions about stuttering are worthwhile. The first is: "Why would Phonetics be so important in the study of stuttering?" A possible answer could be that Phonetics is a borderline discipline, both in the sense that it has a theoretical as well as an applicative character, and in the sense that it is at the convergence of different scientific realms, such as Communication Engineering, Physical Acoustics, Psychology, Anatomy, Physiology, Linguistics, Applied Linguistics, Computer Science and Poetry (Ladefoged, 1988). As a borderline discipline, it holds a privileged key for unifying and simplifying the understanding of stuttering that presents itself as a multidimensional phenomenon, in which sociocultural, psychological, physiological and genetic factors are involved. In fact, one can say that, in order to accomplish a causal function in stuttering, each of those variables must at the end interact with the motor control processes of the speech apparatus, whose defective functioning can be considered the proximal cause of stuttering (Smith, Kelly, 1997).

The second, and specular, question is: "Why should phoneticians be interested in stuttering?" Once again, we could answer that phoneticians could feel attraction towards a speech disorder that selectively affects fluency, leaving essentially intact the syntactic and grammatical structures in individuals that are judged to be healthy and normally endowed with reference to cognitive and emotional aspects (remember the ICD 9 definition which suggests a speech motor, not a language,

[1] F98.5: Other behavioral and emotional disorders with onset usually occurring in childhood and adolescence: "Speech that is characterized by frequent repetition or prolongation of sounds or syllables or words, or by frequent hesitations or pauses that disrupt the rhythmic flow of speech. Minor dysrhythmias of this type are quite common as a transient phase in early childhood, or as a minor but persistent speech feature in later childhood and adult life. They should be classified as a disorder only if their severity is such as markedly to disturb the fluency of speech. There may be associated movements of the face and/or other parts of the body that coincide in time with the repetitions, prolongations, or pauses in speech flow. Stuttering should be differentiated from cluttering (see below) and from tics. In some cases, there may be an associated developmental disorder of speech or language, in which case this should be separately coded under F80."

problem). At the same time, the speech aspects under investigation in stuttering are at the heart of a number of theories of speech production (Weismer, Tjaden & Kent, 1995). In fact, these theories can be affiliated to two great families on the basis of their solutions to the problem of speech timing: the Extrinsic and Intrinsic timing theories (Fowler, 1980), also named, in the terminology of Weismer *et al.* (1995), Translational and Gestural theories, respectively. The first ones postulate the existence of a timer, possibly not specific to the speech mechanism, which puts in sequence a series of discrete and timeless units (i.e. columns of distinctive features); the second ones, bring the timing organization back to the general dynamic property of the articulatory system. The validity of these theories could – at the end – be proved by their power in explaining timing phenomena that are characteristics of the motor speech disorders, like stuttering: articulatory slowness, abnormal scaling (in magnitude) of the articulatory gestures, variability of speech production (across repetitions), abnormal degree of coarticulation (see Weismer *et al.*, 1995; Kent, 1997; Van Lieshout, Goldstein, 2008).

Turning to the ICD approach to speech disorders like stuttering, it is too restricted and has been criticized because its most central assumption is that an underlying clinical entity or medical condition is responsible for stuttering. As a consequence, the goal of Medicine is to intervene on such alterations, which are generally represented by the objective symptoms, i.e. the stuttering disfluencies, and they must be treated independently from any other associated behavioural, psychological and social factors.

These factors are all taken into consideration by another type of broad-based classification system, the International Classification of Functioning, Disability and Health (ICF, WHO, 2001), which starts from assumptions that are radically different from the ICD's ones. ICF could offer an alternative classification for stuttering because emphasis is placed on the fact that disorders like stuttering involve more than just observable behaviours. Specifically, the speaker's experience of stuttering can involve negative emotional, behavioural, and cognitive reactions (both in the speaker and in the communication partner(s)), as well as significant limitations in the speaker's ability to participate in daily activities and a negative impact on the speaker's overall quality of life (Yaruss, Quesal, 2004). According to the ICF's definition, if there is a known sensory, neurological or craniofacial impairment (e.g. hearing loss, cerebral palsy, cleft lip and/or palate), the speech impairment is classified at the Body Structure level. If there is no known cause, the speech impairment is classified at the Body Function level.

The ICD and ICF classification systems are well representative of the philosophical division that, in the word of Tetnowski, Scaler Scott (2010) "is driving both research and clinical interests." [...] This issue "is the dilemma regarding the impact of Behaviourism [as represented by ICD] versus Social Constructivism [as represented by ICF]".

Returning to the ICD classification system (World Health Organization's ICD-10, 2007), this is one of the two broad aetiological-based classification systems (or nosographic systems) in speech pathology, the other being the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013). They provide a classification system for more than communication disorders[2].

Beyond these, specific classification systems exist for specific clusters of speech disorders: as to stuttering, Yairi (2007) and Seery, Watkins, Mangelsdorf & Shigeto (2007) enlist a great number of them, and we invite the reader to refer to these authors in order to know the assumptions of these systems in more detail. For sake of simplicity, I think that they could be easily redistributed under the same three main labels used by Waring and Knight (2013) about the Speech Sound Disorders (SSD): aetiologically based models, linguistic-descriptive models, and psycholinguistic models.

Aetiologically based models are atheoretical and start from a position of pathology rather than normality. Their underpinning premise is that an unvarying relationship exists between an identifiable genetic anomaly and a specific type of speech behaviour. This kind of model is also known as the model of the Mayo Clinic (Darley, Aronson & Brown, 1975), and has been recently questioned by Weismer (2006), which challenged the legitimacy and usefulness of a direct link between medical diagnosis and linguistic symptoms, and, in clinical practice, the reinforcement of non-linguistic oral functions in order to improve the linguistic ones.

The descriptive-linguistic approach to stuttering aims at classifying the subgroups of PWS according to the influence of speech sounds on the specific locations and types of stuttering (see the reviews of Zmarich, 1991; Zmarich, 2012; Zmarich, 2015). The approach is developmental (because of differences between children up to 7-8 years and older subjects) and it relies on the identification and description of the differences between PWS's speech compared to PWNS peers. Nonetheless, the descriptive-linguistic approach is unsuitable when it affords to data based on phonetic transcriptions delicate questions about the nature of disfluencies: as Smith (1999: 27) once highlighted, "static units of disfluency counts, for example, part-word repetitions or sound prolongations, are convenient fictions [...]" but "stuttering is not a series of 'stutter events' [...]", because "stuttering is a dynamic disorder". Any explanation based on transcriptional data is also undermined by the opacity introduced by the relative distance between the more or less central cause of the pathology and the more or less distal periphery in which the acoustic or perceptive events are measured. Moreover, the phonetic-phonological theories based on perceptive or acoustic targets are not suitable for explaining motor events of intrinsically dynamic nature.

This traditional view identifies the dysfluent loci in the stuttered utterance and explains these distributional patterns by invoking the same dysfunctions of the mental representations and processes which generate lapsus and disfluencies in normal speakers (Wingate, 1988; Zmarich, 1991). Apart from the methodological error of attributing causal relationships to the statistical associations of two events

[2] The difference between ICD and DSM rests on the wider scope of the first, which applies to all kind of diseases, with respect to the second, which applies only to mental disorders.

(the disfluency occurrence and the linguistic structure affected by the disfluency), the main defect of this interpretation rests on the exclusive attention to the disfluencies, which render the speech "discontinuous". But fluency is a multidimensional concept (see Lickley, this issue): not only the fluent speech is devoid of discontinuities, but it is also produced with a regular rhythmic beat, in rapid rate and without excessive physical and mental effort (Starkweather, 1987). We know that exist stutterers without disfluencies: they are affected by "covert/subperceptual stuttering" (Bloodstein, Bernstein Ratner, 2008) and often perceive in speaking excessive levels of muscular effort and "cognitive tension" that can pass unobserved to the eye and the ear of the clinician, because they are undetectable without laboratory instrumentation (like those used in experimental phonetics). From my point of view, Phonetics is in a privileged position for measuring the fluency reductions at the level of rhythm (Harrington, 1988) and rate (Andrade, Cervone & Sassi, 2003) and in a good position to measure them at the level of physical (Ingham, Warner, Byrd & Cotton, 2006) and mental effort (Panico, Healey, 2009). As a last criticism, research has demonstrated that, at least in adult stutterers, we cannot easily distinguish between signs of learned coping strategies (the secondary behaviours of stuttering) and signs of an underlying disorder (i.e. the primary behaviours of stuttering). In conclusion, we can not avoid to do the phonetic transcription of PWS's speech, but we have to take this as a starting point, not as an ending point, in order to do phonetic research on stuttering.

The psycholinguistic processing approach employs models of speech processing in children in order to explain 'how' speech impairment arises. Warin, Knight (2013: 34) described the psycholinguistic processing approach to SSD as a bridge between aetiological classification and linguistic descriptions: "Psycholinguistic speech processing models vary considerably in their complexity; however, the application to individuals is the same: a series of hypotheses are developed and systematically tested to find where the breakdown(s) [in the flow of information] is occurring".

Psycholinguistic and phonetic studies have a long tradition in stuttering research, and the advances in knowledge permitted by them have not lost any significance even in recent years when there has been an explosion of genetic and neurophysiologic studies (Bloodstein, Bernstein Ratner 2008). Genetic research established that the predisposition to stuttering is genetically transmitted, although the responsible mechanism at this level is still unknown (for a review, Kraft, Yairi, 2012). A number of brain structural and functional anomalies have been evidenced in adult stutterers (Etchell, Civier, Ballard & Sowman, 2017; Busan, Battaglini & Sommer, 2017; see also the contribution of Busan, this issue). Since similar findings have been documented also in children (Weber-Fox, Wray & Arnold, 2013; Chang, Zhu, Choo & Angstadt, 2015) both developmental stuttering and adult stuttering seem to rely on possibly shared dysfunctional cerebral mechanisms. This don't exclude complete recovery in early childhood, probably due to high neural plasticity (favourable factors are young age and short time interval from the onset, see Ludlow, Hoit, Kent, Ramig, Shrivastav, Strand, Yorkston & Sapienza, 2008).

However important they may be, these results risk to be incomplete and difficult to interpret: on the one hand, genetic studies face with distal causes at such a "molecular" level that they can not at present provide an explanation of the "proximal" causes of the stuttering behavior. Brain-imaging research aiming at testing a certain kind of linguistic process must be guided by psycholinguistic and phonetic hypotheses (Indefrey, Levelt, 2004; Indefrey, 2007), independently formulated. The model of speech production (and perception) most widely adopted by scholars is that of P.I.M. Levelt and colleagues (Levelt, 1989; Levelt, Roelofs & Mejer, 1999; Cholin, Levelt, 2009). The model was elaborated by considering evidences coming from different research fields: (1) speech hesitations (Mahl, 1956; Johnson, 1961; Goldman Eisler, 1968); spontaneous or provoked speech errors (Fromkin, 1973); self-repairs, (Levelt, 1983); (2) temporal reactions (RT) in simple tasks involving descriptions (Oldfield, Wingfield, 1965), or RT task complicated with priming (Lupker, 1979), and with different types of SOA (stimulus onset asynchronies, cf. Schriefers, Meyer & Levelt, 1990); (3) articulatory behaviors, inferred from acoustic analysis (Kent, Kim, 2003) and/or achieved directly through kinematic analysis (Gracco, 1992).

A psycholinguistic model like that of Levelt and coll. consists in a flow diagram that identifies the processing units, explicitly mentions the linguistic information and highlights the stages and processes of information processing that take place (e.g. activation, selection, monitoring, correction, etc.). Information processing proceeds in a unidirectional (from top to bottom levels), and incremental way (i.e., a lower level process is activated as soon as it receives an initial part from the higher stage).

Recently, two psycholinguistic theoretical hypotheses have been advanced which incorporate the view that stuttering is a dynamical and multidimensional phenomenon (as put forward by Smith *et al.*, 1997; Smith, 2016). The first one, *the Packman and Attanasio3-factor causal model of moments of stuttering* (Packman, 2012), states that (1) a deficit in the neural processing underpinning speech production renders the speech production system unstable and prone to perturbation, (2) the perturbation is triggered by some inherent features of speech (like stress or linguistic complexity) that increase the motoric task demands on that system, and (3) it is modulated by intrinsic factors (like physiological arousal) which determine the triggering threshold. The second one is the *Variable Release Threshold hypothesis of stuttering* (Brocklehurst, Lickley & Corley, 2013). It takes the best from two previous models (*Anticipatory Struggle Hypothesis*, Bloodstein, 1975) and *EXPLAN revised model* (Howell, 2003), whereby "the anticipation of upcoming difficulty leads to the setting of an excessively high threshold for the release of speech plan".

Regarding the part about the articulatory preparation and execution stages, which is not very detailed in the model of Levelt, you have to look for elsewhere, but luckily not too far. The model of speech motor control that have received most attention over recent years was developed by F.H. Guenther and colleagues, GODIVA

(Bohland, Bullock & Guenther, 2010), as a specialized derivation from the so-called Hybrid Motor Control or also State Feedback Control models (Hickok, Houde & Rong, 2011; Tian, Poeppel, 2012). It has been recently used in order to account for disfluent production in stuttering (Civier, Bullock, Max & Guenther, 2013), by simulating the consequences over the time course of blood flow caused by deficits in the basal ganglia (excessive levels of dopamine) and in white matter (low density), in a cerebral region below the left precentral gyrus.

Before finishing this excursus, I would like to come back again to the problem of distinguishing between the direct manifestations of stuttering and the reactions of the subjects to it. Regarding this point, it could be very important to study the affected subjects before they develop coping reactions, that is, we must study stuttering in young preschool children. In fact, stuttering could be defined as a typically childhood disease, because it begins between 16 and 66 months of age and less than 5% of PWS begin to stutter after they pass the 5th year of age (Yairi, Ambrose, 2005). This is the period of the greatest and fastest development in anatomo-physiological structures and functions, and in linguistic, cognitive and motor abilities as well. Its incidence is around 10% of all the children but its prevalence is only around 1%, due to the overwhelming probability of spontaneous recovery (around 90%, Yairi, Ambrose, 2013). Anyway, if spontaneous recovery does not happen within four years from the stuttering onset, that child is very likely "destined" to persistence. Maybe the key for solving the mystery of stuttering resides in the understanding why 9 children out of 10 which start to stutter, later recover spontaneously.

We know that non stuttering children are more disfluent when they attempt to produce new syntactic structures (Colburn, Mysak, 1982; Wijnen, 1990). According to Rispoli (Rispoli, 2003; Rispoli, Hadley & Holt, 2008), increases in Mean Length of Utterance parallel increases in revisions, and increases in utterance length parallel increases in "stallings" (i.e., part-word repetitions and/or prolongations). After the children are four years-old, only children who stutter continue to produce stallings consistently (Wagovich, Hall & Clifford, 2009). Moreover, research has established that, as to linguistic structures or psycholinguistic skills, stuttering children are not different from nonstutterers, in the average, but lower and upper performers are over-represented (Seery *et al.*, 2007). There are more frequent dissociations in stuttering children than nonstuttering ones among language capacity (often higher than normal), and articulatoy skills (often lower than normal, Coulter, Anderson & Conture, 2009).

As to childhood stuttering, a recent review (Sasisekaran, 2014) suggested a relationship between stuttering and phonology (excluding Phonetics, considered by the author as a low-level motor production of sounds) in 3 areas:

1. effects of phonological complexity on the location (loci) of stutter events;
2. outcomes of standardized test measures in children who do and do not stutter;
3. studies of phonological encoding in children and adults who stutter.

The results from the loci studies, according to Sasisekeran (2014: 95), "offer some support for the role of phonological complexity in the occurrence of stuttering. Studies of performance in standardized tests of phonology have not identified differences between children who do and do not stutter. Studies of phonological encoding have been equivocal in reporting differences between children and adults who stutter and those who do not stutter".

We can integrate the results from this review on phonological factors with the experimental findings on the articulatory skills of preschool children: in the words of the authors, "they provide new evidence that preschool children diagnosed as stuttering lag their typically developing peers in maturation of speech motor control processes" (Smith, Goffman, Sasisekaran & Weber-Fox, 2012: 344; see also Walsh, Mettel & Smith, 2015; Smith, 2016).

We can conclude this introduction on stuttering and phonetic theories by presenting the Speech Motor Skills (SMS) model of van Lieshout and colleagues (see, among others, Namasivayam, van Lieshout, 2011): according to them, stutterers lay at the lower end of a hypothetical non-pathological continuum that characterizes speech motor skills. Disfluencies reflect errors in motor control, but stuttering is not a motor disorder (such as dysarthria or dyspraxia), but reflects an "innate" limitation of the verbal motor control system (clumsiness). Clumsiness emerges when programming and performing complex motor tasks in the presence of emotional, motor, cognitive and linguistic influences, and/or when demands are increasing for both accuracy and rapidity of movement. Most of the time these resources are crashed by dual-task processes. Not surprisingly, the linguistic condition that causes a worsening of stuttering is the "ecological" communication exchange (i.e. a conversation) where the PWS must simultaneously handle feedback at different levels and plan at the cognitive and syntactic level while performing in real time phonological planning and articulatory execution processes. Phonological coding can be considered a sort of dual-task process, because it requires simultaneous planning of subsequent language units during the articulatory execution of previous units (in addition to handling various feedback), PWS are limited in motor skills: they receive poor benefits from practice and are not able to generalize acquisitions to similar tasks, and/or to maintain them over time.

2. *The organization of the special session on stuttering and phonetic theory*

The organization of the individual contributions at this special session will proceed from more general and more "peripheral" (in the sense of the analysis of stutterers' speech based on the "auditory perceptions" of the hearers), to the most recent instrumental researches, passing from acoustics and kinematics to arrive to neurophysiology. Robin Lickley opens the special session by laying the foundations for every analysis of stuttered speech, that is the analysis of disfluencies: blocks, prolongations and repetitions, which are the hallmark of the disorder (see the ICD 9 definition at beginnings; see also Lickley, 2015). Lickley shows that "while typical disfluencies are mostly influenced by cognitive issues in the planning of speech and only rarely by motor control issues, stuttered disfluencies result from a break down in the coordination of the complex motor commands necessary for successful articulation" (from the slide presentation). In other words, drawing from his presentation: "Typical disfluency is due to problems with planning, lexical access, word finding, errors, while stuttering is due to neurological problems".

As reported above, in recent years, the focus on the motor aspects of stuttering led to considering disfluencies as only one of the many ways speech could become disfluent. Considerations like these led speech researchers to concentrate on the perceptually fluent speech of stutterers, on the belief that PWS' speech could be abnormal even when the person is not openly stuttering. Acoustic and kinematic analysis, often associated to brain imaging or to electrophysiological techniques, have been used more and more, and speech motor control theories have become the preferred theoretical frame for most of the scientists. Moreover, for the reasons I explained above, researchers focussed on the affected subjects before they could possibly develop coping reactions, that is, in the preschool years. The contribution by Giovanna Lenoci illustrates a particular application of this methodological recommendation, such as the acoustic and kinematic analysis of coarticulation, which has been so much in the focus of research on stuttering that it was included in more than one definition: "[...] the difficulty is not manifested in the articulatory postures essential to that sound, but instead in moving on the succeeding one(s) (Wingate, 1964)"; "the lack of anticipatory coarticulation is probably the primary elements in the core behaviours of stuttering" (Stromsta, 1986). The degree of intra-syllabic anticipatory coarticulation also maintains a great potential as early predictor of stuttering persistence (Subramanian, Yairi & Amir, 2003), and Lenoci presents the first results of the CNR longitudinal project on early predictive indexes of persistent stuttering in early infancy. Forty pre-school children at high risk to develop stuttering due to familiar antecedents were tested when they were 2-years old for a number of linguistic, cognitive and physiological aspects. Those who developed stuttering were later followed longitudinally in order to ascertain the predictive power for stuttering persistence of some phonetic indexes, degrees of anticipatory coarticulation included (Zmarich, Bernardini, Lenoci, Natarelli & Pisciotta, in press). Although the predictive value of this last variable is currently still not clear, the authors found significant differences between peer-aged controls and all the PWS, without distinction between recovered or persistent sub-groups. Lenoci then presents some preliminary insights about her new project about the use of Ultrasound Tongue Imaging (UTI) for investigating anticipatory coarticulation and, more generally, speech (in)stability in primary school-aged PWS and PWNS. This focus on childhood helps to remember that all of us, as adult fluent speakers of one or more language, attained this unique capacity through a long process of acquisition and learning, during the pre-school years. This is the period of the greatest and fastest development in anatomo-physiological structures and functions, and in linguistic, cognitive and motor abilities as well, and Phonetic Sciences could tell us a lot about the possible ways this acquisition process could go wrong and produce stuttering.

Pierpaolo Busan will conclude the special session by illustrating his innovative research on PWS' neurophysiology, and he will show how the most recent models of speech motor control, as the State Feedback Control models (see above), have been applied to stuttering, which has been interpreted as a defect in sensory-motor integration. He states that "stuttering may result from speech components that are not properly synchronized, also because of time pressure; it may be the result of a series of errors that are present before, after, or during word execution; it could be influenced from load of cognitive processing as well as from phonological complexity; the anticipation of upcoming difficulties may cause the setting of higher thresholds for the release of speech motor plans" (from the slide presentation). His own work based on Transcranial Magnetic Stimulation (TMS) paired with EEG recordings is illustrated and the results support specific speech models of motor control in stuttering. The neural mechanism causing stuttering is a deficit in motor-to-sensory transformation: both dopamine dysfunction and white matter impairment may cause stuttering during speech motor control (see Civier *et al.*, 2013 and its simulation of stuttering starting from the GODIVA model). Even in this case, "findings may be useful for new treatment solutions for stuttering, ranging from neuromodulation to neurofeedback".

Bibliography

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5).

Andrade, C.R.F.D., Cervone, L.M. & Sassi, F.C. (2003). Relationship between the stuttering severity index and speech rate. In *Sao Paulo Medical Journal*, 121(2), 81-84.

Ball, M., Kent, R.D. (1987). Editorial. In *Clinical Linguistics & Phonetics*, 1, 1-5.

Bernstein, L.E., Weismer, G. (2000). Basic science at the intersection of speech science and communication disorders. In *Journal of Phonetics*, 28(3), 225-232.

Bloodstein, O. (1975). Stuttering as tension and fragmentation. In Eisenson, J. (Ed.), *Stuttering: A second symposium*. New York: Harper & Row, 1-96.

Bloodstein, O., Bernstein Ratner, N. (2008). *A Handbook on Stuttering*. New York (NY): Thomson Delmar Learning.

Bohland, J.W., Bullock, D. & Guenther, F.H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. In *Journal of Cognitive Neuroscience*, 22(7), 1504-1529.

Brocklehurst, P.H., Lickley, R.J. & Corley, M. (2013). Revisiting Bloodstein's Anticipatory Struggle Hypothesis from a psycholinguistic perspective: A Variable Release Threshold hypothesis of stuttering. In *Journal of Communication Disorders*, 46(3), 217-237.

Busan, P., Battaglini, P.P. & Sommer, M. (2017). Transcranial magnetic stimulation in developmental stuttering: Relations with previous neurophysiological research and future perspectives. In *Clinical Neurophysiology*, 128, 952-964.

Chang, S.E., Zhu, D.C., Choo, A.L. & Angstadt, M. (2015). White matter neuroanatomical differences in young children who stutter. In *Brain*, 138(3), 694-711.

Cholin, J., Levelt, W.J. (2009). Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level. In *Language and Cognitive Processes*, 24(5), 662-684.

Civier, O., Bullock, D., Max, L. & Guenther, F.H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. In *Brain and Language*, 126(3), 263-278.

Colburn, N., Mysak E.D. (1982). Development disfluency and emerging grammar. I. Disfluency characteristic in early syntactic utterances. In *Journal of Speech and Hearing Research*, 25, 414-420.

Coulter, C.E., Anderson, J.D. & Conture, E.G. (2009). Childhood stuttering and dissociations across linguistic domains: A replication and extension. In *Journal of Fluency Disorders*, 34, 257-278.

Crystal, D. (1981). *Clinical Linguistics*. Vienna &New York: Springer.

Darley, F.L., Aroson, A.E. & Brown, J.R. (1975). *Audio seminars in speech pothology: Motor speech disorders*. Philadelphia: WB Saunders.

Etchell, A.C., Civier, O., Ballard, K. & Sowman, P.F. (2017). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. In *Journal of Fluency Disorders*. Mar 12 DOI: 10.1016/j.jfludis.2017.03.007 [Epub ahead of print].

Fowler, C.A. (1980). Coarticulation and theories of extrinsic timing. In *Journal of Phonetics*, 8(1), 113-133.

Fowler, C.A. (2014). Talking as doing: Language forms and public language. In *New ideas in psychology*, 32, 174-182.

Fromkin, V.A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Goldman Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Gracco, V.L. (1992). Analysis of speech movements: practical considerations and clinical application. In *Haskins Laboratories, Status Report on Speech Research*, SR-109/110, 45-58.

Harrington, J. (1988). Stuttering, delayed auditory feedback, and linguistic rhythm. In *Journal of Speech and Hearing Research*, 31(1), 36-47.

Hickok, G., Houde, J. & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. In *Neuron*, 69(3), 407-422.

Howell, P. (2003). Is a perceptual monitor needed to explain how speech errors are repaired? In *Gothenburg Papers in Theoretical Linguistics*, 90, 31-34.

Indefrey, P. (2007). Brain-imaging studies of language production. In Gaskell, M.G. (Ed.), *The Oxford Handbook of Psycholinguistics*. Oxford (UK): Oxford University Press, 547-564.

Indefrey, P., Levelt, W.J.M (2004). The spatial and temporal signatures of word production components. In *Cognition*, 92, 101-144.

Ingham, R.J., Warner, A., Byrd, A. & Cotton, J. (2006). Speech effort measurement and stuttering: Investigating the chorus reading effect. In *Journal of Speech, Language, and Hearing Research*, 49(3), 660-670.

Johnson, W. (1961). Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and non stutterers. In *Journal of Speech and Hearing Disorders*, 7, 1-20.

Kent, R.D. (1997). Gestural phonology: Basic concepts and applications in speech-language pathology. In Ball, M.J., Kent, R.D. (Eds.), *The new phonologies: Developments in clinical linguistics*. San Diego: Singular Publishing Group, 247-268.

Kent, R.D., Kim, Y.J. (2003). Toward an acoustic typology of motor speech disorders. In *Clinical Linguistics & Phonetics*, 17(6), 427-445.

Kingston, J., Beckman, M.E. (Eds.), (1990). *Papers in Laboratory Phonology, 1, Between the Grammar and Physics of Speech*, Cambridge: Cambridge University Press.

Kraft, S.J., Yairi, E. (2012). Genetic Bases of Stuttering: The State of the Art, 2011. In *Folia Phoniatrica et Logopaedica*, 64, 34-47.

Ladefoged, P. (1988). A view of phonetics. *UCLA Working Papers in Phonetics*, 70, 41.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. In *Cognition*, 14(1), 41-104.

Levelt, W.J.M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: Mit Press.

Levelt, W.J.M., Roelofs, A. & Mejer, A.S. (1999). A theory of lexical access in speech production. In *Behavioral and Brain Sciences*, 22, 1-75.

Lickley, R.J. (2015). Fluency and Disfluency. In Redford, M. (Ed.), *The Handbook of Speech Production*. Wiley-Blackwell, 445-469.

Lindblom, B. (1995). A view of the future of phonetics. In Elenius, K., Branderud, P. (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, vol. 4. Stockholm, Sweden, 462-469.

Ludlow, C.L., Hoit, J., Kent, R., Ramig, L.O., Shrivastav, R., Strand, E., Yorkston K. & Sapienza, C.M. (2008). Translating principles of neural plasticity into research on speech motor control recovery and rehabilitation. In *Journal of Speech, Language, and Hearing Research*, 51(1), 240-258.

Lupker, S.J. (1979). The semantic nature of response competition in the picture-word interference task. In *Memory & Cognition*, 7(6), 485-495.

Mahl, G.F. (1956). Disturbances and silences in the patient's speech in psychotherapy. In *Journal of Abnormal Society Psychology*, 3, 1-15.

Namasivayam, A., van Lieshout, P.H.H.M. (2011). Speech Motor Skill and Stuttering. In *Journal of Motor Behavior*, 43, 477-489.

Oldfield, R.C., Wingfield, A. (1965). Response latencies in naming objects. In *Quarterly Journal of Experimental Psychology*, 17(4), 273-281.

Packman, A. (2012). Theory and therapy in stuttering: A complex relationship. In *Journal of Fluency Disorders*, 37(4), 225-233.

Panico, J., Healey, E.C. (2009). Influence of text type, topic familiarity, and stuttering frequency on listener recall, comprehension, and mental effort. In *Journal of Speech, Language, and Hearing Research*, 52(2), 534-546.

Perkins, W.H. (1990). What is stuttering? In *Journal of Speech and Hearing Disorders*, 55(3), 370-382.

Perkins, W.H., Kent, R.D. & Curlee, R.F. (1991). A theory of neuropsycholinguistic function in stuttering. In *Journal of Speech, Language, and Hearing Research*, 34(4), 734-752.

Pierrehumbert, J., Beckman, M.E. & Ladd, D.R. (2000). Conceptual foundations of phonology as a laboratory science. In Burton-Roberts, N., Carr, P. & Docherty, G. (Eds.), *Phonogical knowledge: conceptual and empirical issues*. Oxford: Oxford University Press, 273-304.

Rispoli, M. (2003). Changes in the nature of sentence production during the period of grammatical development. In *Journal of Speech, Language, and Hearing Research*, 46, 818-830.

Rispoli, M., Hadley, P. & Holt, J. (2008). Stalls and revisions: A developmental perspective on sentence production. In *Journal of Speech, Language, and Hearing Research*, 51, 953-966.

Sasisekaran, J. (2014). Exploring the link between stuttering and phonology: A review and implications for treatment. In *Seminars in speech and language*, 35, 95-113.

Schriefers, H., Meyer, A.S. & Levelt, W.J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. In *Journal of Memory and Language*, 29(1), 86-102.

Seery, C.H., Watkins, R.V., Mangelsdorf, S.C. & Shigeto, A. (2007). Subtyping stuttering II: Contributions from language and temperament. In *Journal of Fluency Disorders*, 32, 197-221.

Smith, A. (1999). Stuttering: a unified approach to a multifactorial, dynamic disorder. In Bernstein Ratner, N., Healey, N.E.C. (Eds.), *Stuttering research and practice*: *Bridging the Gap*. Mahwah, NJ: Lawrence Erlbaum Associates, 27-44.

Smith, A. (2016). A multifactorial neurodevelopmental approach to stuttering: (1) Language and motor factors and (2) Pathways to persistence and recovery. In Tomaiuoli, D. (Ed.), *Proceedings of the 2nd international Conference on Stuttering*. Trento: Erickson, 31-40.

Smith, A., Kelly, E. (1997). Stuttering: a dynamic, multifactorial model. In Curlee, R.F., Siegel, G.M. (Eds.), *Nature and treatment of stuttering. New directions*. Boston: Allyn & Bacon, 218-235.

Smith, A., Goffman, L., Sasisekaran, J. & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: evidence from behavioral and kinematic indices of nonword repetition performance. In *Journal of Fluency Disorders*, 37(4), 344-358.

Starkweather, C.W. (1987). *Fluency and Stuttering*. Englewood Cliffs, NJ: Prentice Hall.

Stromsta, C. (1986). *Elements of stuttering*. Atsmorts Publishing.

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(1), 59-75.

Tetnowski, J.A., Scaler Scott, K. (2010). Fluency and Fluency Disorders. In Damico, J.S., Müller, N. & Ball, M.J. (Eds.), *The Handbook of Language and Speech Disorders*. Wiley-Blackwell, 431-454.

Tian, X., Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. In *Frontiers in Human Neuroscience*, 6, 1-11.

Van Lieshout, P.H., Goldstein, L.M. (2008). Articulatory Phonology and Speech Impairment. In Ball, M.J., Perkins, M.R., Müller, N. & Howard, S. (Eds.), *The handbook of clinical linguistics*. Oxford: Blackwell, 467-479.

Wagovich, S.A., Hall, N.E. & Clifford, B.A. (2009). Speech disruptions in relation to language growth in children who stutter: An exploratory study. In *Journal of Fluency Disorders*, 34, 242-256.

Walsh, B., Mettel, K.M. & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. In *Journal of Neurodevelopmental Disorders*, 7(1), 27.

Waring, R., Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. In *International Journal of Language & Communication Disorders*, 48(1), 25-40.

Weber-Fox, C., Wray, A.H. & Arnold, H. (2013). Early childhood stuttering and electrophysiological indices of language processing. In *Journal of Fluency Disorders*, 38(2), 206-221.

Weismer, G. (2006). Philosophy of research in motor speech disorders. In *Clinical Linguistics & Phonetic*s, 20(5), 315-349.

Weismer, G., Tjaden, K. & Kent, R.D. (1995). Can articulatory behavior in motor speech disorders be accounted for by theories of normal speech production? In *Journal of Phonetics*, 23(1), 149-164.

Wijnen, F. (1990). The development of sentence planning. In *Journal of Child Language*, 17, 651-75.

Wingate, M.E. (1964). A standard definition of stuttering. In *Journal of Speech and Hearing Disorders*, 29(4), 484-489.

Wingate, M.E. (1988). *The Structure of Stuttering: A Psycholinguistic Analysis*. Springer Science & Business Media.

World Health Organization (1977). *International statistical classification of diseases, injuries, and causes of death*. (ICD-9), Geneva.

World Health Organization (2001). *International Classification of Functioning, Disability and Health (ICF)*, Geneva.

World Health Organization (2007). *International Statistical Classification of Diseases and Related Health Problems (ICD)*, 10th edition. http://apps.who.int/classifications/apps/icd/icd10online/.

Yairi, E. (2007). Subtyping stuttering I: A review. In *Journal of Fluency Disorders*, 32(3), 165-196.

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.

Zmarich, C. (1991). Una revisione critica degli studi 'linguistici' sulla balbuzie. In *Acta Phoniatrica Latina*, XIII, 4, 495-514.

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. (2015). Le teorie psicolinguistiche e fonetiche della balbuzie. In Tomaiuoli, D. (Ed.), *Balbuzie: fondamenti, valutazione e trattamento*. Trento: Erickson, 177-210.

Zmarich, C., Bernardini, S., Lenoci, G., Natarelli, G. & Pisciotta, C. (in press). Could the frequencies of stuttering-like-dysfluencies predict persistent stuttering in children who have just started to stutter? In Savy, R., De Meo, A. & Alfano, I. (Eds.), *La fonetica sperimentale nell'insegnamento e nell'apprendimento delle lingue straniere*.

DOI: 10.17469/O2103AISV000019

ROBIN LICKLEY

Disfluency in typical and stuttered speech

This paper discusses what happens when things go wrong in the planning and execution of running speech, comparing disfluency in typical speech with pathological disfluency in stuttering. Spontaneous speech by typical speakers is rarely completely fluent. There are several reasons why fluency can break down in typical speech. Various studies suggest that we produce disfluencies at a rate of around 6 per 100 fluent words, so a significant proportion of our utterances are disfluent in some way. Stuttering can halt the flow of speech at a much higher rate than typical disfluency. While persons who stutter are also prone to the same kinds of disfluency as typical speakers, their impairment results in the production of other forms of disfluency that are both quantitatively and qualitatively different from typical forms. In this paper, I give an overview of the causes of disfluency in both typical and stuttered speech and relate these causes to their articulatory and phonetic realisations. I show how typical and stuttered disfluencies differ in both their cause and their realisations.

Key words: fluency, disfluency, stuttering.

1. *Introduction*

This work discusses phenomena that occur in everyday speech, both for typical speakers and for speakers who stutter: Disfluencies. We begin with an operational definition of fluency, before examining how planning difficulties can lead to a failure to maintain a perfectly smooth flow. Types and causes of typical disfluencies are described before we move on to consider disfluencies that occur in stuttered speech.

2. *Fluency*

Before we begin to discuss disfluency in speech, it is important to define what we mean by fluency. Essentially, the word fluency refers to the flow of speech. Flow involves regular, continuous motion over time. If speech is fluent, then the sounds that we hear are perceived as flowing smoothly and without unexpected breaks.

Some breaks in the flow are to be expected in extended periods of speech and these do not necessarily appear disfluent. Speakers need time to breathe in from time to time, after all.

Breath pauses tend to fall at structural boundaries. So, in fluent speech, pauses in the flow are expected at sentence and phrase boundaries. The pattern holds more for read speech than for spontaneous speech (Wang, Green, Nip, Kent & Kent, 2010) and in faster speech, breath pauses become less regularly associated with structural units (Grosjean, Collins, 1979). But within phrases, words typically follow on from each other continuously. Sounds in the speech stream influence each other and coarticulate, rather than simply concatenating like typed letters. Fluency between sentences and phrases may entail pauses between structural units, but fluency within phrases involves continuity between words, with gestural overlap between adjacent phonemes.

The production of speech demands a complex interaction of processes on many levels. Fluency at the level of speech output depends on smooth functioning within and between all of those levels, from the conceptual planning of what message to convey and how to express it, through formulation, involving lexical selection and syntax building and phonological encoding, all the way to the creation of a motor plan to suit the message and its articulation via the muscle systems that allow us to speak. Thus, it is possible to consider fluency at multiple levels of production prior to the spoken output. If conceptual planning is straightforward, and the speaker can quickly decide what in general terms to say and how to make it appropriate given the current discourse, then there is no need for hesitation before formulation begins, and conceptualisation can be said to be fluent. At the formulation stage, if the lemmas required to express the planned utterance are easily accessed, if construction of the required syntactic structure, the assignment of appropriate morphology and the phonological encoding and prosodic patterns present no problems, then a prearticulatory plan can be constructed fluently. If motor commands to the multiple muscle systems involved in articulation can be synchronised precisely, without a hitch, then a fluent utterance can be produced.

In reading out loud, and in carefully rehearsed performances, the speaker avoids the need for much of the rapid planning behind the act of speaking, and it is relatively rare for fluency to break down. In spontaneous speech, though, the complexity of the task means that it is rare for a speaker to continue absolutely fluently for long.

3. *Typical Disfluency*

If there is a breakdown in fluency, then we can say that the resultant speech is disfluent. Disfluencies occur frequently in typical spontaneous speech, at a rate of around 6 per 100 words (Bortfeld, Leon, Bloom, Schober & Brennan, 2001; Eklund, 2004; Fox Tree, 1995; Shriberg, 1994). They occur at a higher rate in longer utterances (Oviatt, 1995; Shriberg, 1994) and in more complex utterances (Lickley, 2001; Shriberg, 1994). Individuals vary considerably in the rate at which they produce disfluencies, but it is difficult to find a speaker who is never disfluent.

The word 'disfluency' is defined in several different ways in the research literature and there seems to be no consensus on what phenomena it includes, so it is important to begin this piece with our own operational definition. Our definition of fluency refers to the flow of speech, so disfluency involves a break in that flow, when the speaker stops for a moment in a place or for a length of time not predicted by typical fluent production.

We will now describe what phenomena fall under this definition, with the help of some examples, before discussing how these disfluencies result from problems at various points in the production processes.

From a formal point of view, the major subtypes of disfluency are (1) those that involve simply a halt in the production process, (2) those that involve repetition of part of the utterance and (3) those that involve alteration of part of the utterance. We can refer to these as hesitations, repetitions and repairs.

The simplest forms of disfluencies are hesitations. Hesitations may be realized as silent pauses, prolongations, filled pauses and repetitions, as well as combinations of these. The simplest form of hesitation is a silent pause. Though it may be the simplest form, it is also one of the hardest to define. One problem is that it is difficult to define a minimum duration for what should constitute a disfluent silence. Another is that silences occur in fluent speech, between turns in dialogue (though speakers often overlap), between sentences and phrases (though by no means all of them), and within words, where a short silence can occur during the closure phase of a stop consonant. Distinguishing between fluent and disfluent pauses is far from easy.

On the duration issue, many researchers have accepted 250 ms as a minimum for within-sentence hesitation pauses, following Goldman-Eisler's assertion that only pauses longer than this should be included in accounts of the cognitive processes involved in hesitation (Goldman-Eisler, 1958a; 1958b; 1961). In Goldman-Eisler's view, shorter silences were more likely a consequence of typical fluent production processes. This rather arbitrary value possibly stemmed from the limitations of the technology available at the time for analyzing speech. Since rate of speech varies not only between speakers on their average rate, but also within speakers, depending on discourse context and within single utterances, and in slower speech one would expect longer silences, it should be clear that it is not even desirable to set a single threshold for a hesitation silence. Since Goldman-Eisler, of course, the topic has been addressed by many authors. Among these, Butcher (1981) examined the effect of prosodic context on the perception of pauses in German. Listeners regarded a 220 ms pause as excessive when it occurred between tone groups; but a pause of only 80 ms was regarded as excessive when it occurred within a tone group. How long a pause has to be to be judged as disfluent depends on context, so an absolute value for a hesitation is hard to justify.

On distinguishing fluent from disfluent pauses, we need to consider structural issues at the syntactic and prosodic levels. There is not a straightforward relationship between prosodic and syntactic structure and Ferreira (1993; 2007) suggests that the durations of fluent pauses are licensed by prosodic structure, while hesitation pauses, most often caused by planning issues, are more closely related to syntactic structure, since it is at points related to structural planning and lexical access that speakers need to find time.

In classifying disfluencies within speech corpora, given that there are problems both in defining pause durations and in distinguishing fluent from disfluent pauses, it seems safest, where possible, to employ subjective judgement, supported by inter-judge agreement, in classifying pauses in running speech, a system used, for example, by Nakatani and Hirschberg (1994) and Eklund (2004).

Aside from these issues with defining silent pause, as intimated previously, it is unclear whether we should distinguish silent from breath pauses. A major issue here is that in analyzing recordings of speech it may not always be possible to judge accurately when a pause includes a breath. But this is a reflection of life beyond the speech laboratory, where such acoustic detail may easily be drowned out by background noise.

Speakers also use prolongation of syllables in otherwise fluent speech as a means of pausing. For the same reasons as those given for silent pauses, prolongations are hard to classify in a rule-based way, since syllables vary in duration in fluent speech according to context. So, again, subjective judgement is often needed. Prolongation may accompany other signs of pausing, preceding a silent pause, for example (Bell, Jurafsky, Fosler-Lussier, Girand, Gregory & Gildea, 2003). On their own, prolonged syllables can give the listener a clear impression that the speaker has halted momentarily, even in the absence of any acoustically measureable silence (Duez, 1993).

By convention, when a speaker is expected to say something, they will say something, if only to let others know that they're about to say something more. A long silence is open to misinterpretation (Does the person realize that it is their turn to speak? Do they have nothing to say?). So a vocal sign of hesitation makes sense from the point of view of continuity within a dialogue. The most recognizable sign of vocalized pausing is what is usually referred to as the filled pause. In English, filled pauses can be represented orthographically as "um" or "uh" (amongst many other variants). The written form disguises the fact that the phonetic realization varies greatly between accents of English. Being syllabic and relatively long, filled pauses are more easily recognised than silent pauses and prolongations, so one can report their frequency in speech corpora with greater confidence. For example, for the 64 speakers in the HCRC Map Task Corpus (Anderson, Bader, Bard, Boyle, Doherty, Garrod, Isard, Kowtko, McAllister, Miller, Sotillo, Thompson & Weinert, 1991, mostly Scottish-accented English) we find an overall average rate of 1.3 filled pauses per 100 words, with individual speakers ranging from 0.18 to 6.66 per 100 words (Lickley, 2015), while reports of other corpora in American English vary between 2.6 (Bortfeld *et al.*, 2001) and 1.6-2.2 (Shriberg, 1994, for 3 different corpora). The differences in frequency between corpora are likely due to task complexity, rather than dialect differences. In languages other than English, different frequencies have been reported. For example, Eklund (2004) reports a rate of 3.6 per 100 words for a Swedish corpus and Maekawa (2004) reports 7.23 per 100 words for a Japanese corpus. It is hard to make valid cross-corpus comparisons and to conclude anything from such differences, when not only the demands of the speaking tasks but also the definitions of 'filled pause' may vary between studies. Japanese filled pauses, for example, include lexical fillers, loosely transcribed as "ano" and "eeto" (Watanabe, 2009), while many studies of 'filled pause' also include prolonged syllables and repetitions under the definition.

It was suggested above that repetitions form a disfluency subtype of their own. This is true in that they are different in form from others, but a case can be made for some repetitions functioning as hesitations and others as repairs. When speakers pause mid-utterance they have the option of continuing from where they left off or restarting the phrase that they paused, thus repeating a word or two. Disfluent repetitions in English tend to be on unstressed function words (Clark, Wasow, 1998; Fox, Jasperson, 1995; Lickley, 1994; Shriberg, 1994). Lickley (1994) found that 96% of disfluent repetitions were on function words (very often the definite article, "the").

When discussing repetitions, we also need to distinguish fluent from disfluent cases. Occasionally, two instances of a word need to be produced together for an utterance to convey the intended meaning: telephone numbers often contain repeated digits, for example. People regularly repeat intensifiers for rhetorical effect (e.g., "that was a very, very strange election result"). Where repetitions form part of a fluent utterance, they will fall into the prosodic pattern expected of the utterance: In saying a telephone number out loud, there is a characteristic intonation pattern, which would be the same whether or not a digit is repeated. Disfluent repetitions are often accompanied by pause of another kind, like silence and prolongation, and they typically repeat the pitch level as well as the word. So, it is relatively easy to distinguish fluent from disfluent cases of repetition, on the basis of prosodic information.

The third major type of disfluency, Repair, can take several forms. Repairs take place when something has gone awry in the production process and the speaker needs to adjust what they have said in order to produce a corrected version of the utterance. This entails a certain amount of backtracking. Most relatively recent corpus-based accounts of disfluency include the repair subtypes Substitution, Insertion and Deletion, as well as combinations of these, and complex cases (Eklund, 2004; Heeman, 1997; Lickley, 1998; Shriberg, 1994).

In Substitutions, an erroneous string is replaced by a correction. The error may be at one of a number of different levels. In this example, the non-word aphosia is produced via a phonological error – anticipation of the stressed vowel from the following word – which is quickly repaired: "Then you can calculate the aphosia – aphasia quotient". In the following example, the speaker selects the wrong lexical item and replaces it after backtracking to the verb: "Go along the road and turn left – turn right at the traffic lights".

In Insertions, the speaker adds something to what they have already said, usually to be more specific. Here, the speaker chooses to add the modifier sharp to make the instruction clearer: "After you reach the post office, take a right – a sharp right turn".

In deletions, the speaker abandons the utterance altogether and immediately begins a new utterance, without apparently altering or adding to the original utterance: "Go along the road and turn – have you visited Edinburgh before?" In this case, the speaker changes plan, deciding to verify that the listener possesses relevant background knowledge before issuing further instructions.

As mentioned above, repetitions can also sometimes viewed as a manifestation of repair. This follows from the view that errors may be detected in the production process prior to being articulated, covertly in other words. If the monitoring and repairing processes are rapid enough, there may be no overt production of the error. But since it takes a certain amount of time to make the correction, the speaker may repeat the onset of the phrase containing the repaired error. Levelt (1983) introduced the notion of covert error and repair, and since then researchers have included repetitions amongst the phenomena that signal covert repair (The Covert Repair Hypothesis: Postma, Kolk, 1993). There is some evidence that there are acoustic differences between types of repetitions, making it possible to distinguish repetitions that are associated with covert repair from others (Plauché, Shriberg, 1999).

Having introduced the major forms of disfluencies that are typically found in spontaneous speech, we now need to discuss in more detail why they occur.

Preparing to speak can take time. As speakers progress through the complex set of processes that are required to produce an utterance, there are many points at which problems can occur and it can become necessary to buy some extra time.

In planning the overall message, the complexity and the length of the message can affect the likelihood of hesitation or repair. Giving instructions on how to fix the brakes on a bicycle involves mentioning more tools, more mechanical parts and more actions than simply inflating a tyre, so keeping track in memory of all the elements required to give a successful string of instructions for the former task is far more demanding, cognitively, than issuing the simpler set of instructions in the latter. Longer and more complex utterances are more prone to hesitation and repair-inducing errors.

Also at the level of overall planning, decision making can take time. For example, if a speaker needs to respond to a question by retrieving information from long term memory, then planning the response will take longer when the information is harder to retrieve (as demonstrated in responses to quiz questions: Brennan, Williams, 1995; Smith, Clark, 1993). When it is hard to plan a response, speakers will find time by using hesitation disfluencies.

Planning at the conceptual level is not only about decision making and information retrieval: It can also involve error detection and correction. During planning, a speaker may decide on a goal, begin to execute the plan and then realize that the plan is not ideal and decide to make alterations. This will entail some kind of repair, possibly correcting a factual error, or possibly modifying the output to be more accurate or specific, as suggested by Levelt's Appropriateness Repairs (Levelt, 1983). In the classification scheme outlined above, a change of plan can take the form of an insertion repair, where a slight modification is required, a substitution repair, where an alternative is preferred, or even a deletion, if the entire plan is to be ditched and a new one commenced.

Once a speaker has the overall plan ready and is accessing the lexicon, there are more challenges that can result in disfluency. Words that we learned at an earlier age and words that we use more frequently can be accessed from the lexicon more quickly (e.g., Jescheniak, Levelt, 1994; Morrison, Ellis & Quinlan, 1992; Oldfield, Wingfield, 1965). Late-learned, less frequent words take longer to access and are therefore more likely to be preceded by hesitations.

Similarly, words with lower name agreement (more competing lexical items for the same concept) attract more hesitations (Hartsuiker, Notebaert, 2010). Physical characteristics of the words themselves can also affect their accessibility and, one can assume, their likelihood of being preceded by hesitations. Additionally, in experimental settings longer words have been found to elicit longer reaction times (e.g., Bates, D'Amico, Jacobsen, Székely, Andonova, Devescovi, Herron, Lu, Pechmann, Pléh, Wicha, Federmeier, Gerdjikova, Gutierrez, Hung, Hsu, Iyer, Kohnert, Mehotcheva, Orozco-Figueroa, Tzeng & Tzeng, 2003; Severens, Lommel, Ratinckx & Hartsuiker, 2005). So, several characteristics of lexical items can affect the likelihood that they will be preceded by hesitation.

Context can also influence hesitation before words. The context of the current discourse has an impact on the probability that a speaker will have trouble accessing a word and therefore need to hesitate. As far back as the 1950s, it was show that the probability of a word given the local context was inversely related to the probability of a hesitation preceding that word (Goldman-Eisler, 1958a; Lounsbury, 1954). Another contextual influence is that of interference with lexical access caused by the close proximity of phonologically similar words. In an experiment with running speech, words were more likely to be preceded by hesitation disfluencies when they were adjacent to phonologically similar words (e.g., "hand the hammer" Jaeger, Furth & Hilliard, 2012).

So, in addition to the intrinsic properties of lexical items, the context in which they occur can influence hesitation.

Errors in lexical access are relatively rare in typical speech, though they do occur, mostly with semantically related words, as in the example above, repeated here: "Go along the road and turn left – turn right at the traffic lights". When they do occur, they are usually repaired. With one word being substituted for another, such a repair would be classified as a substitution repair.

Once words have been accessed, phonological anticipation, perseveration and exchange errors can come about as a result of confusion in working memory. Look at the following example produced by a BBC radio reporter: "We still have this ban on cabbage – uh – on cabin baggage". Here, the words "cabin" and "baggage" had been accessed and were ready to articulate. The speaker had no plan to say "cabbage" and the semantic associations of the word were presumably not active. But confusion in working memory between the two target words caused the blend to be produced. The substitution repair followed after a brief hesitation.

Hesitation and repair disfluencies, do, of course, occur together. When an error occurs and a repair is needed, this demands some processing time. So some repairs are accompanied by hesitation phenomena (as in the cabbage/baggage example above). However, it is not the case that filled pauses regularly mark repair sites. Studies of disfluencies in speech corpora suggest that a small minority (6-10%) of repetition and repair disfluencies are accompanied by filled pauses at the interruption point (Lickley, 1994; Nakatani, Hirshberg, 1994; Shriberg, 1994), and it is often the case that repair follows the interruption point so rapidly that there is no gap at all between interruption and repair (Blackmer, Mitton, 1991).

This brief overview of disfluency in typical speech is intended to give the reader a summary of the main forms that typical disfluencies take and an explanation of why they occur. For a more complete discussion, see Lickley (2015) and for a discussion of some phonetic aspects, see Shriberg (2001).

4. *Disfluency in stuttering*

Now, we turn to disfluencies in stuttering. The discussion will focus on the speech of adults who stutter. As proficient speakers of their native language, we should start by stating that there is no reason in principle for adults who stutter (AWS) not to produce exactly the same repertoire of typical disfluencies as anyone else. Is it simply that stuttering is at the high end of a continuum of typical disfluency with AWS producing far more frequent disfluencies than typical speakers? The answer is clearly 'No'. Stuttering is different. Not only do stuttered disfluencies have different physical and phonetic realizations from typical disfluencies, but their causes are also quite different. So, in addition to an expectation that AWS should be prone to the same types of disfluency as typical speakers, they produce additional 'stuttering-like disfluencies'.

Stuttering-like disfluencies typically occur more frequently than typical disfluencies. Wingate (2002) suggests an average of around 10 stuttering events per 100 words as a reasonable estimate of an average, taken from a range of studies. But speakers vary widely in the frequency, with rates of 50% reported in severe cases, while some cases self-report as stuttering, while apparently fluent. The frequency of stuttering also varies considerably within speaker: Some AWS report being fairly fluent on some occasions and very disfluent on others. While typical disfluency is usually rated by frequency alone, stuttering severity can also be measured by the duration of instances of stuttering. It is estimated that instances of stutters average around 1 second in duration, while rarely extending beyond 5 seconds (Bloodstein, 1944, 1987, cited in Guitar (2006)).

Disfluencies in stuttering are usually characterized as consisting of three types, according to popular academic text books and official diagnostic documents: Repetitions, Blocks and Prolongations. This is an unsatisfactory characterization, for a number of reasons, explained below.

One issue is that we need more detail than simply 'repetitions, blocks and prolongations', because the description is incomplete. We discussed repetitions and prolongations in some detail above, with reference to typical speech. But it is crucial to understand that in stuttering, these repetitions and prolongations are different. A key characteristic of stuttered disfluencies is that stuttering is usually accompanied by physical tension. This factor is ignored in that simple three-way description, yet, it is one major factor that clearly differentiates stuttered from typical disfluency. Tension is focused on the muscle systems involved in speech articulation. Alongside the tension is the fact that stuttered disfluencies usually last longer than typical disfluencies. Whereas typical disfluencies rarely take more than a fraction of a second, stuttered disfluencies can involve the speaker getting stuck on a sound for more than one second, and sometimes several seconds. And finally, a stuttering episode may contain a mixture of blocking, prolongation and repetition.

A second issue is that the simple three-way description implies that there may be three different causes for stuttering. After all, we've described a number of different causative factors behind the various typical disfluencies. But in the case of stuttering, it seems most plausible that the three types are articulatory realizations of the same underlying difficulty: An inability to progress from one sound to the next. To some extent the realizations of this inability to progress are conditioned by the nature of the units of speech (phonemes and syllables) that are involved.

A stuttered repetition will usually involve a sound that can be repeated easily, like an oral stop consonant or cluster (electr-tr-tr-tronically), or a whole syllable (au-au-au-australia) or short word (I I I work in an office). Such repetitions are typically very rapid and may consist of just one repeated token or (more often) several repeated tokens.

A stuttered block will usually involve a long, tense articulatory closure on a stop consonant (perhaps representing a more severe case than a repetition: elect::::ronically). But blocks also occur on the glottal closure before a word-onset vowel ([ʔ]::::oil).

A stuttered prolongation usually involves a sound that can be prolonged easily, like a continuant (sssssssssseven). In considering prolongations, it is most evident that the repetition-block-prolongation distinction should be seen as realizations of the same basic phenomenon: If a speaker becomes stuck on a [t] sound, for example, and cannot release it into the following vowel, then one might class this as a 'block'; but from a functional point of view it is no different from a speaker being unable to make the transition from a [s] to a following vowel and prolonging the [s].

In all cases, if the repetition, block or prolongation leads to a successful release, then there is likely to be an intense burst, because of the extended build-up of subglottal pressure. It is also important to note that stutters mostly occur on stressed syllables, or on utterance-initial syllables (Natke, Grosser, Sandrieser & Kalveram, 2002).

This somewhat simplistic description belies the fact that close analysis and description of episodes of stuttering is usually quite problematic. For example, a repetition may vary in quality, with intermittent vocalization (t-t-t-tuh-tuh-t-t-t-take); A block may become a partially-blocked repetition (b:::::ba-b:::::ba- b::: back); a tense prolongation of /s/ may vary in intensity, giving the impression of the rapid repetition of the sound. If a successful release is not achieved, the speaker may repeat the process, with further attempts.

The picture is further complicated by the production of secondary speech behaviours. An adult with many years of experience of stuttering will typically have learned speech behaviours that are intended to enable them to get through a stutter successfully. Amongst these are deliberately slowing down or prolonging sounds, soft articulatory contacts, resembling slurring of speech, using well-practiced fluent words or phrases intended to launch them into a fluent stream of speech (e.g., "So", "I mean", "well", "what I mean is", "Ok" ...), use of monotone or unusual intonation, or intentionally repeating sounds or words in a slow and controlled manner (voluntary stuttering). All of these features make the description of stuttered speech more challenging.

In addition to the speech-based characteristics of stuttering, many AWS present with secondary behaviours, such as eye-rolling and blinking, head movements, tongue thrusts, facial grimaces, movements of arms or hands.

In general, there is a lot of interspeaker variation in the presentation of stuttering, and perhaps because of that, in addition to the complexity outlined above, a lack of detailed phonetic description of stuttered speech.

What can be said is that for the most part, stuttering-like disfluencies are different in form from typical disfluencies. While words and syllable onsets are repeated in typical speech, it is relatively rare for a word or sound to be repeated more than once. While there is prolongation in typical speech, the duration of prolongations is relatively short, where it can extend over a second in duration in stuttering, and it is more common at the ends of words in typical speech, rather than at the syllable onsets that are affected by stuttering. Blocking is not considered by researchers examining typical disfluency. There is no evidence of physical tension associated with typical disfluency, but this is a defining characteristic of stuttering-like disfluency.

Earlier, we discussed how typical disfluencies can be explained with reference to the various levels of processing involved in speech production. Repetitions and prolongations are found frequently in typical speech, and we have attributed them to problems at the conceptualization and formulation stages, where the speaker needs to buy time. But in stuttering, speakers report that when they get stuck on a word, they know exactly what they want to say, so there is no problem with conceptualization or formulation. Although there is some evidence that word-finding difficulties may also play a role (Hubbard, Prins, 1994; Prins, Main & Wampler, 1997), the main problem lies in executing the plan for articulation. While it is tempting to say that a speaker who stutters on a given sound has a problem with that sound, this is a mischaracterization. The problem lies in the transition from the repeated, blocked or prolonged sound to the next sound, in synchronizing the change in articulation and/or voicing between one sound and the next. This view forms the basis of Wingate's fault-line hypothesis (Wingate, 1988), which suggests that in stuttering a speaker suffers a delay in the syllable rhyme following production of the onset. A similar view is also modelled in the EXPLAN Theory (Howell, Au-Yeung, 2002), which posits that stuttering may occur when there is a misalignment between a plan and execution such that a syllable is initiated before the plan for the rest of the syllable is ready.

Beyond the stuttered disfluencies themselves, there are other factors related to stuttering that can potentially result in disfluencies that may be seen as more typical. Adults who stutter are very aware that they are likely to stutter in certain situations and on certain words and sounds, and this creates apprehension (Blood, Blood, Tellis & Gabel, 2001; Kelso, 1997; Neiman, Rubin, 1991).

In his Anticipatory Struggle Hypothesis, Bloodstein (1975) viewed the anticipation of difficulty in speaking as a cause of the muscular tension and then the fragmentation of speech that is characteristic of stuttering. Although there is little support for theories of stuttering that place anxiety about speech at the centre, there is some evidence that anticipation of trouble in speech can increase the likelihood that an AWS will stutter on a word (Brocklehurst, Lickley & Corley, 2012). It has been demonstrated that AWS are more sensitive than are typical speakers to minor disfluencies in the speech of other people, as well as in their own speech (Lickley, Hartsuiker, Corley, Russell & Nelson, 2005). But often, AWS plan ahead to avoid the possibility of stuttering. In some cases, this may entail avoiding speaking altogether. In other cases, an AWS may make an alteration to their plan, either at the conceptual level, by deciding to change the message that they were originally going to produce, or at the lexical level, by finding an alternative word, to avoid one that they anticipate to present difficulties. These processes can take time, and may result in hesitation or reformulation of an utterance, creating disfluency, though potentially avoiding stuttering. On the other hand, some AWS are so adept at circumlocution that they appear fluent, although they are still well aware of the underlying problem: This is usually referred to as covert stuttering (Murphy, Quesal & Gulker, 2007). It is not clear whether a person with covert stuttering can be so controlled in their speech that they avoid typical disfluencies, too, though it is likely that they would present with an elevated rate of hesitations.

5. *Conclusion*

It is clear, then, that stuttered disfluencies differ greatly in form from typical disfluencies. Although similar terms have been used to describe them (repetitions, prolongations, silent pauses or blocks), they differ in quantity as well as in quality. In more severe cases, stutters occur at a far greater frequency than typical disfluencies. The more prominent typical disfluencies are relatively short (just over 400 ms for the filled pause "um", Lickley, 2015), whereas stutters can average around 1 second in duration and extend to 5 seconds or more. Repetitions in stuttering involve multiple repeats of a sound or syllable, whereas it is relatively rare for a typical disfluent repetition to consist of more than one repeat. Similarly, stuttered prolongations are typically significantly longer than those in typical speech. Stuttering-like blocks are extremely rare in typical speech, though brief silent pauses are not uncommon. Finally, stuttering-like disfluencies are typically accompanied by muscular tension, which is hardly ever present in typical disfluency.

The fact that stuttering-like disfluencies are different from typical disfluencies in quality and quantity does not necessarily rule out the possibility that they are just extreme cases of certain types of otherwise typical hesitation disfluencies. However, it is also clear that stuttered disfluencies differ greatly from typical disfluencies in their provenance. In most cases, typical disfluencies are responses to difficulties in planning and formulation of speech. In stuttering, the primary symptoms result from difficulties in the coordination of motor programmes for the execution of existing plans. Further evidence for this come from the observation that in cases where the planning of the message and the words to be used are already supplied – in reading out loud – typical speakers are only rarely disfluent, whereas many AWS stutter at a higher rate than in spontaneous speech.

We began with an operational definition of fluency. We then discussed how between typical speakers and speakers who stutter the speech of the two populations is characterized by different types of disfluency. Some researchers object to the term disfluency being used to describe features of speech that is entirely typical. In this sense, for some adults who stutter, a therapy goal might be 'typical fluency' insofar as that alternative definition of 'fluency' might include typical disfluency.

Bibliography

Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S. & Weinert, R. (1991). The HCRC Map Task Corpus. In *Language and Speech*, 34(4), 351-366.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C.C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A. & Tzeng, O. (2003). Timed picture naming in seven languages. In *Psychonomic Bulletin & Review*, 10(2), 344-380.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M. & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. In *The Journal of the Acoustical Society of America*, 113, 1001-1024.

Blackmer, E.R., Mitton, J.L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. In *Cognition*, 39, 173-194.

Blood, G.W., Blood, I.M., Tellis, G. & Gabel, R. (2001). Communication apprehension and self-perceived communication competence in adolescents who stutter. In *Journal of Fluency Disorders*, 26(3), 161-178.

Bloodstein, O. (1975). Stuttering as tension and fragmentation. In Eisenson, J. (Ed.), *Stuttering: A second symposium*. New York: Harper & Row, 1-96.

Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F. & Brennan, S.E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. In *Language and Speech*, 44(2), 123-147.

Brennan, S.E., Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. In *Journal of Memory and Language*, 34(3), 383-398.

Brocklehurst, P.H., Lickley, R.J. & Corley, M. (2012). The influence of anticipation of word misrecognition on the likelihood of stuttering. In *Journal of Communication Disorders*, 45(3), 147-160.

Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communicative functions. PhD Dissertation, Christian Albrechts Universität zu Kiel.

Clark, H.H., Wasow, T. (1998). Repeating words in spontaneous speech. In *Cognitive Psychology*, 37(3), 201-242.

Duez, D. (1993). Acoustic correlates of subjective pauses. In *Journal of Psycholinguistic Research*, 22(1), 21-39.

Eklund, R. (2004). Disfluency in Swedish human-human and human-machine travel booking dialogues. PhD Dissertation, University of Linköping.

Ferreira, F. (1993). Creation of prosody during sentence production. In *Psychological Review*, 100, 233-253.

Ferreira, F. (2007). Prosody and performance in language production. In *Language and Cognitive Processes*, 22(8), 1151-1177.

Fox Tree, J.E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. In *Journal of Memory and Language*, 34, 709-738.

Fox, B.A., Jasperson, R. (1995). A syntactic exploration of repair in English conversation. In Davis, P.W. (Ed.), *Descriptive and theoretical modes in the alternative linguistics*. Amsterdam: John Benjamins, 77-134.

Goldman-Eisler, F. (1958a). The predictability of words in context and the length of pauses in speech. In *Language and Speech*, 1(3), 226-231.

Goldman-Eisler, F. (1958b). Speech production and the predictability of words in context. In *Quarterly Journal of Experimental Psychology*, 10(2), 96-106.

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. In *Language and Speech*, 4(1), 18-26.

Grosjean, F., Collins, M. (1979). Breathing, pausing and reading. In *Phonetica*, 36(2), 98-114.

Guitar, B. (2006). *Stuttering: An integrated approach to its nature and treatment* (3rd ed.). Philadelphia: Lippincott Williams and Wilkins.

Hartsuiker, R.J., Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. In *Experimental Psychology (Formerly Zeitschrift Für Experimentelle Psychologie)*, 57(3), 169-177.

Heeman, P.A. (1997). Speech repairs, intonational boundaries and discours markers: Modeling speakers' utterances in spoken dialog. PhD Dissertation, University of Rochester.

Howell, P., Au-Yeung, J. (2002). The EXPLAN theory of fluency control applied to the diagnosis of stuttering. In *Amsterdam Studies in the Theory and History of Linguistic Science Series*, 4, 75-94.

Hubbard, C.P., Prins, D. (1994). Word familiarity, syllabic stress pattern, and stuttering. In *Journal of Speech, Language, and Hearing Research*, 37(3), 564-571.

JAEGER, T.F., FURTH, K. & HILLIARD, C. (2012). Phonological overlap affects lexical selection during sentence production. In *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1439-1449.

JESCHENIAK, J.D., LEVELT, W.J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. In *Journal of Experimental Psychology-Learning Memory and Cognition*, 20(4), 824-843.

KELSO, K. (1997). The relationship between communication apprehension, communication competence, and locus of control in stutterers and nonstutterers. In *Journal of Fluency Disorders*, 2(22), 153.

LEVELT, W.J.M. (1983). Monitoring and self-repair in speech. In *Cognition*, 14, 41-104.

LICKLEY, R.J. (1994). Detecting disfluency in spontaneous speech. PhD Dissertation, University of Edinburgh.

LICKLEY, R.J. (1998). *HCRC disfluency coding manual.* HCRC Technical Report No. TR-100. University of Edinburgh.

LICKLEY, R.J. (2001). Dialogue moves and disfluency rates. In *Proceedings of DiSS: ISCA Tutorial and Research Workshop on Disfluency in spontaneous speech*, Edinburgh, UK.

LICKLEY, R.J. (2015). Fluency and disfluency. In REDFORD, M. (Ed.), *The handbook of speech production.* Wiley-Blackwell, 445-469.

LICKLEY, R.J., HARTSUIKER, R.J., CORLEY, M., RUSSELL, M. & NELSON, R. (2005). Judgment of disfluency in people who stutter and people who do not stutter: Results from magnitude estimation. In *Language and Speech*, 48(3), 299-312.

LOUNSBURY, F.G. (1954). Transitional probability, linguistic structure and systems of habit-family hierarchies. In OSGOOD, C.E., SEBEOK, T.A. (Eds.), *Psycholinguistics: A survey of theory and research problems*. Baltimore: Williams and Wilkins, 93-101.

MAEKAWA, K. (2004). Design, compliation and some preliminary analyses of the corpus of spontaneous Japanese. In MAEKAWA, K.Y.K. (Ed.), *Spontaneous speech: Data and analysis*. Tokyo: The National Institute for Japanese Language, 87-108.

MORRISON, C.M., ELLIS, A.W. & QUINLAN, P.T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. In *Memory & Cognition*, 20(6), 705-714.

MURPHY, B., QUESAL, R.W. & GULKER, H. (2007). Covert stuttering. In *Perspectives in Fluency and Fluency Disorders*, 17(2), 4-9.

NAKATANI, C.H., HIRSCHBERG, J. (1994). A corpus-based study of repair cues in spontaneous speech. In *The Journal of the Acoustical Society of America*, 95, 1603-1616.

NATKE, U., GROSSER, J., SANDRIESER, P. & KALVERAM, K.T. (2002). The duration component of the stress effect in stuttering. In *Journal of Fluency Disorders*, 27(4), 305-318.

NEIMAN, G.S., RUBIN, R.B. (1991). Changes in communication apprehension, satisfaction, and competence in foreign dialect and stuttering clients. In *Journal of Communication Disorders*, 24, 353-366.

OLDFIELD, R.C., WINGFIELD, A. (1965). Response latencies in naming objects. In *Quarterly Journal of Experimental Psychology*, 17(4), 273-281.

OVIATT, S. (1995). Predicting spoken disfluencies during human-computer interaction. In *Computer Speech and Language*, 9, 19-35.

Plauché, M., Shriberg, E. (1999). Data-driven subclassification of disfluent repetitions based on prosodic features. In *Proceedings of the International Congress of Phonetic Sciences*. San Francisco, USA, 1513-1516.

Postma, A., Kolk, H. (1993). The Covert Repair Hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. In *Journal of Speech and Hearing Research*, 36, 472-487.

Prins, D., Main, V. & Wampler, S. (1997). Lexicalization in adults who stutter. In *Journal of Speech, Language, and Hearing Research*, 40 (2), 373-384.

Severens, E., Lommel, S.V., Ratinckx, E. & Hartsuiker, R.J. (2005). Timed picture naming norms for 590 pictures in Dutch. In *Acta Psychologica*, 119 (2), 159-187.

Shriberg, E. (1994). Preliminaries to a theory of speech disfluencies. PhD Dissertation, University of California.

Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. In *Journal of the International Phonetic Association*, 31(1), 153-169.

Smith, V.L., Clark, H.H. (1993). On the course of answering questions. In *Journal of Memory and Language*, 32(1), 25-38.

Wang, Y.T., Green, J.R., Nip, I.S., Kent, R.D. & Kent, J.F. (2010). Breath group analysis for reading and spontaneous speech in healthy adults. In *Folia Phoniatrica Et Logopaedica*, 62(6), 297-302.

Watanabe, M. (2009). *Features and roles of filled pauses in speech communication: A corpus-based study of spontaneous speech*. Tokyo: Hituzi Syobo Publishing.

Wingate, M.E. (1988). *The structure of stuttering*. New York: Springer.

Wingate, M.E. (2002). *Foundations of stuttering*. San Diego, CA: Academic Press.

DOI: 10.17469/O2103AISV000020

GIOVANNA LENOCI

Anticipatory coarticulation in the speech of people who stutter

Stuttering is a fluency disorder that manifests itself through frequent interruptions in the smooth flow of speech. The disfluencies characterising stuttering strongly suggest the presence of breakdowns in the precisely timed and coordinated articulatory movements required for fluent speech. For these reasons, coarticulation has been one of the most studied aspects in relation to stuttering. Studies investigating coarticulation in speakers with impaired speech production have obvious importance for advancing our understanding of the disorder itself, which ultimately has implications for the diagnosis and treatment of speech impairments. The purpose of this chapter is to summarize the most interesting experimental results emerged from the acoustic and articulatory study of lingual coarticulation in the speech of people who stutter. The last section of this chapter is devoted to present an ultrasound tongue imaging study developed with a group of Italian stuttering children and a matched control group. Preliminary results suggest that lingual coarticulation in the fluent speech of children who stutter presents some differences compared to normally fluent children.

Key words: stuttering, speech motor control, coarticulation, second formant transitions, Locus Equations, Ultrasound Tongue Imaging.

1. *Introduction*

In everyday life, humans accomplish a great number of different and often complex motor activities in an automatic and apparently effortless trend.

The literature on motor control provides various definitions of 'motor skill', such as: "skill consists in the ability to bring about some end result with maximum certainty and minimum outlay of energy, or time and energy" (Guthrie, 1952: 17); "a skilled response is highly organized, both spatially and temporally. The central problem for skill learning is how such organization or pattering comes about" (in Kelso, 1997). Therefore, a 'skilled' motor activity requires practice, a high spatial and temporal organization, and effectiveness in terms of energy and time costs. This general description of motor skill fits very well with fluent speech production mechanisms in the adult speech.

The motor skills required to produce speech are among the most sophisticated learned by humans: it requires a remarkably complex combination of 'linguistic' and motor processes. Prior to the articulation of the speech sounds, it involves rapid interactions of processes related to the planning and formulation of the intended utterance[1]. From a motor control perspective, a large number of muscles are activated to move respiratory, phonatory and articulatory structures; the nervous system must generate sets of commands that must be coordinated in time and space for the appropriate sequences of muscle activation to occur, and for speech to be produced fluently.

As all motor skills, speech movements are not innate: they require practice for an extended period of time (up to adolescence) before they reach a certain level of skill, as in adults' speech (Smith, Goffman, 1998; Smith, Zelaznik, 2004). Learning to speak is a task of immense complexity: children must learn to control their breathing sufficiently to produce the subglottal pressure necessary for speech; they have to learn that consonants and vowels tend to alternate with one another although consonants can occur in clusters. Furthermore, children will learn to produce supra-segmental constraints of intonation and timing that govern phrases while, within words, children must learn not only the order of syllables and segments, but also the details of their temporal control (Hawkins, 1984).

As for other motor skills, speech movements are organized into coordinative structures[2] (Mac Neilage, Davis, 2000): they are characterized by a high spatial and temporal organization; they are very effective because they accomplish their goals (the acoustic targets) with a minimum of effort, 'if possible' (Lindblom, 1990) and in a relatively fast pace. Speech gestures are also naturally flexible to change, in fact they adapt themselves to their linguistic context. In short, producing speech is indeed a motor skill: the articulatory organization of the adult's speech is symptomatic of a high degree of motor ability, because the speech sounds are co-produced rapidly, efficiently and with a high degree of accuracy through the phenomenon of coarticulation.

Coarticulation is a term used to describe the ubiquitous overlapping of the articulatory movements associated with separate sound segments. One of the consequences of coarticulation is therefore that speech sounds vary according to the context in which they are produced, and to the nature of sounds which precede or follow them. Coarticulation effects are often described in terms of the direction and the extent of influence. Right-to-left or anticipatory coarticulation occurs when a speech sound is influenced by a following sound, while – if a sound shows influence of a preceding sound – this is called carry-over or perseverative (left-to-right) coarticulation. Carry-over effects are often attributed to inherent kinematic characteristics of the speech organs. Anticipatory coarticulatory effects are generally regarded to be a characteristic of a skilled speech behavior. At a cognitive level, anticipatory movements are the evidence of a universal tendency for the brain to 'scan ahead' of

[1] For example, according to Levelt, Roelofs & Meje (1999), after the conceptual preparation of the linguistic message, word generation proceeds through lexical selection, morphological and phonological encoding, phonetic encoding, and articulation itself (i.e., mapping abstract intended linguistic structure to dynamic sequences of movements).

[2] "Highly evolved task-specific ensembles of neuromuscular and skeletal components constrained to act as a single unit" (Kelso, 1998: 205).

time (Lashley, 1951), and it is suggested that such anticipation may be disrupted in many types of speech disorder, affecting normal speech motor control, such as stuttering (Hardcastle, Tjaden, 2008).

1.1 Stuttering as a limitation in the speech motor skills

Stuttering is a speech motor disorder that typically arises in the first childhood: for the persistent stuttering group, the mean age at onset is 35.14 months (Ambrose, Yairi, Loucks, Seery & Throneburg, 2015). According to recent epidemiological estimates, stuttering is characterized by a high rate of spontaneous recovery: around 90% of stuttering children recover without therapeutic treatment by the fourth year from the onset of the disorder (Yairi, Ambrose, 2013). The phenomenon of recovery gives rise to important questions pertaining the differences between persistent and recovered stuttering[3] and prognosis. For this reason, recent research on stuttering focuses on the possibilities to discriminate, as soon as possible, children who will recover spontaneously from children who will become persistent, and consequently to determine if they exhibit different speech and/or non-speech characteristics even before the different developmental processes separate them (i.e., recovered vs persistent stuttering). Early prediction of the eventual course of the disorder will allow clinicians to make informed decisions about selective treatment strategies, for example, reserving immediate clinical intervention to children showing high chances of chronicity. This approach could increase the chance to recover from stuttering symptoms.

Developmental stuttering is characterized by disruptions in the production of speech sounds, also called Stuttering-Like Disfluencies (Yairi, Ambrose, 2005): monosyllabic-word repetitions, part-word repetitions, silent and audible sounds prolongations are the hallmark characteristics of the disorder.

Recent accounts of stuttering agree in defining it as a multifactorial disorder: many variables – such as language, motor, cognitive, emotional and genetic factors – are supposed to interact in complex ways in the development of the disorder and in the overt breakdowns in speech motor control that are perceived as stuttering-like disfluencies. The importance of the interactions of two of these factors, namely language and speech motor processes, is supported by a number of experimental findings according to which, for example, increases in utterances length and syntactic complexity are associated with the increased occurrence of stuttering-like disfluencies in children and adult who stutter (Buhr, Zebrowski, 2009; MacPherson, Smith, 2013). It is well known that stuttering is associated with reduced motor speech performance, and it is also true that moments of linguistic complexity tend to highly correlate with motor speech complexity. In studies where group differences emerged (Kleinow, Smith, 2000), people who stutter (henceforth, PWS) were found to show poorer performances on the timing and coordination of motor events compared to

[3] Separating the two sub-groups, should increase precision of experiments in various aspects of the disorder and provide evidence-based data to re-consider traditional view of stuttering as a unitary disorder (Subramanian, Yairi &Amir, 2003)

normally fluent speakers for a given speech task. In other words, even if most theories of the causes of stuttering postulate that many factors are involved in producing these motor breakdowns, it is clear that abnormal speech motor output is an essential component of stuttering (Olander, Smith & Zelaznik 2010).

Van Lieshout and co-workers (Van Lieshout 1995; Van Lieshout, Hulstijn & Peters 2004) proposed the Speech Motor Skills approach to explain motor aspects in PWS: speech production is a motor skill similar to any other (fine) motor skill and stuttering may arise from limitation in speech motor skills. PWS are located more toward the lower end of a presumed normal speech motor skill continuum, with people who do not stutter (henceforth PWNS) distributed across the more skilled end. From this perspective, stuttering is not viewed as a motor disorder such as dysarthria or dyspraxia, but rather as a reflection of an innate limitation of speech motor control system to prepare and perform complex motor actions in the presence of cognitive, linguistic, emotional and speech motor influences (Namasivayam, Van Lieshout, 2011).

Thus, in PWS the speech motor control system has been argued to be the critical weak link in the chain of events that lead to the production of speech. Therefore, stuttering-like disfluencies are the direct manifestation of failures of the speech motor system to address the appropriate command signals that drive the muscles involved in speech production.

Coarticulation has been one of the most studied motor aspects in relation to stuttering speech: it is a crucial mechanism for fluency, referring to the neuromuscular organization that mediates the complex and precise movements involved in the speech production. Stuttering has been hypothesized to stem from breakdown in coarticulation or difficulty in transitioning between sounds. Past research on stuttering tried to investigate coarticulatory processes in the stuttering speech using different instrumental techniques. Findings from studies investigating CV coarticulation in PWS are equivocal, especially because they vary in terms of the age group of interest, the methodology or the measures used to infer coarticulation, and the speech samples – which in the case of stuttering means whether perceptually fluent or dysfluent tokens were of interest. However, these results seem to confirm that the lingual coarticulation that accompanies a stuttering-like disfluency clearly differs from the coarticulation that characterizes normal fluency[4].

Due to space reasons, this paper will present only experimental results obtained from the acoustic and ultrasound tongue imaging data collected from the literature on the fluent speech of PWS. The rationale for studying coarticulatory patterns in the fluent productions of PWS is that differences, however subtle, between the perceptually fluent speech of PWS and normally fluent speakers may provide insight into the (disordered) speech motor control strategies of the former group. Furthermore, the more general implication is to establish that PWS are speaking abnormally even when they are not stuttering at all.

[4] The bulk of studies focusing on dysfluent utterances of PWS suggests atypical or absent F2 transitions (Harrington, 1987; Yaruss, Conture, 1993). Thus, coarticulation in dysfluencies of PWS appears to differ from normally fluent speech, at least as inferred from F2 transition characteristics.

2. *Acoustic measurements of coarticulation*

Many researchers have used relatively indirect techniques, such as acoustic analysis mainly because of the practical difficulties associated with articulatory tracking techniques.

However, acoustic analysis is still a valuable tool for exploring general contextual effects that occur as a result of coarticulatory processes. For example, second formant transitions (F2) allows for inferences to be made concerning lingual position and movement during speech production.

One method of estimating formant transition characteristics involves the use of visual criteria to determine the onset and the offset of the formant transitions, i.e., the first glottal pulse in the target vowel and the point in which a maximally steady-state is visually identified in the vowel formants on a wideband spectrograms. In this way, the entire formant transition can be measured in terms of duration[5] and frequency extent[6] (i.e., amplitude). The characteristics of these formant transitions (i.e., duration and extent) provide us the possibility to calculate the slope or 'trajectory' for the transition, and this measure allows to assess the articulatory gestures underlying coarticulation. Due to the temporal and positional aspects involved in the coarticulation of speech sounds, the slope coefficient can be regarded as 'an ordinal index of the rate of change in vocal tract geometry' (Weismer, 1991). The rate of frequency change in F2 transitions, or the speed with which formant frequencies changed during the transition, was estimated by calculating the absolute value of the extent of F2 transition (Hz) divided by the duration of the transition (msec). This measure is believed to approximate the speed with which speech articulators move from one location to the next (Yaruss, Conture, 1993). Accordingly, a large slope coefficient would reflect considerable positional and temporal movement of the tongue body inside the oral cavity following consonant release (Weismer, 1991).

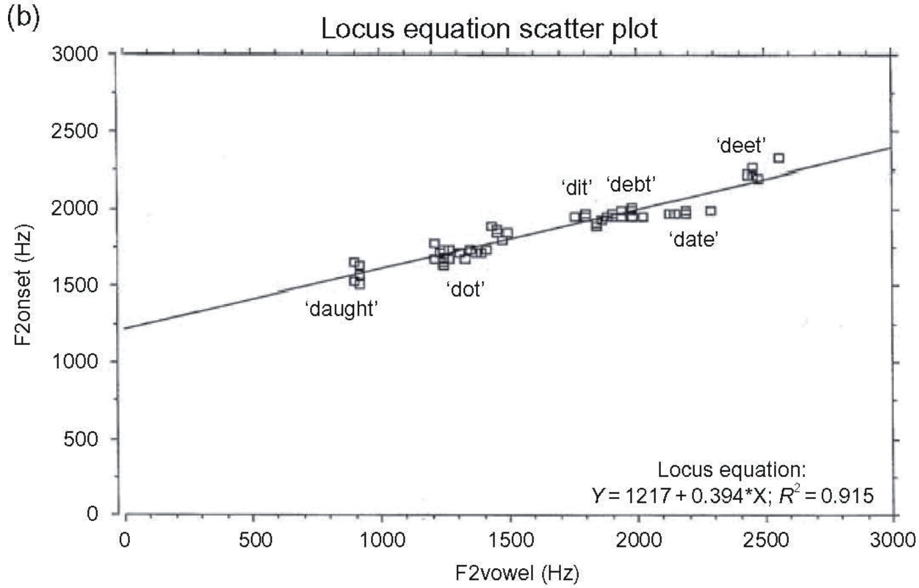
Another possibility to assess anticipatory CV coarticulation is Locus Equations metric (henceforth, LE). LEs are linear regressions of the frequency of the F2 transition sampled at its onset (on the first glottal pulse of the vowel) on the frequency of F2 when measured in the vowel nucleus (at a so-called steady-state location) (Lindblom, 1963). These frequency values are plotted for a single consonant produced with a wide range of following vowels: $F2_{onsets}$ are plotted along the y-axis and $F2_{midpoints}$ along the x-axis (cfr. Figure 1). For a given stop place category, data coordinates have been consistently shown to tightly cluster in a positively correlated distribution (Sussman, Hoemeke & McCaffrey, 1992; Sussman, Hoemeke & Ahmed, 1993), and the slope

[5] The duration of the formant transition can be estimated by calculating the difference in time (in msec) between the onset of the F2 transition at the beginning of formant movement from the initial sound, and the offset of the F2 transition at the beginning of the steady-state portion of the following sound (e.g., the vowel in CV utterances). This measure is believed to assess the amount of time the articulators spend moving from one position to another during the transition from one sound to the following one (Yaruss, Conture 1993).

[6] This measure can be estimated by calculating differences between the offset and the onset center frequencies of the transition, and is believed to approximate the overall movement of the articulators during the transition (Yaruss, Conture 1993).

of the linear regression line is said to be linked to the degree of coarticulation (Krull, 1988). The LE slope was shown to vary between the extremes of k = 0 and k = 1.0: flatter regression slopes indicated limited anticipatory coarticulation, as the C onset remains constant across varying vowel contexts, while steeper LE slopes indicated increasing extents of anticipatory coarticulation, with maximal coarticulation having $F2_{onsets}$ varying as a direct and linear function of the following vowel.

Figure 1 - *A representative locus equation scatter plot for 50 [dVt] tokens produced across 10 vowel contexts. This scatter plot is fit with a linear regression line, of the form F2onset = k*F2vowel + c, where k and c are respectively the slope and y-intercept of the line (adapted from Sussman et al., 2010: 3)*



The next section will summarise the main acoustic findings concerning the study of anticipatory coarticulation in the fluent speech of PWS.

The discovery of abnormal coarticulatory patterns in the perceptively fluent productions of PWS would corroborate the idea of a limitation in the speech motor system, which is instable and prone to interferences. In fact, to overcome interferences and achieve fluency, it is a common opinion that PWS have to adopt different articulatory strategies to accomplish the planned acoustic target (Kleinow, Smith, 2000; Van Lieshout *et al.*, 2004).

3. *Analyses of F2 transitions in the stuttering speech*

Literature on coarticulation in the stuttering speech has varied with regard to the analysis methods and speech samples; therefore, it is not surprising that a consistent picture has yet to emerge as to whether coarticulation is deviant for PWS (Hardcastle, Tjaden, 2008).

Robb, Blomgren (1997), for example, analyze CV lingual coarticulation in the speech of 5 adults who stutter (henceforth, AWS) and 5 adults who do not (AWNS). The hypothesis tested was that the perceptively fluent speech of the first group differs from AWNS in the slope of F2 transitions as a result of abnormal lingual coarticulation behavior. Acoustic results show that the F2 transitions in the stuttering group were characterized by greater frequency extents compared to the control group. These results show that the regulation of transitioning from different vocal tract configurations were handled differently: the steeper slopes found in the fluent speech of AWS would indicate greater and quicker movement of the tongue body within the oral cavity and hence a lower degree of coarticulation compared to non-stuttering speakers.

A more recent study on Farsi-speaking AWS (Dheqan, Yadegari, Blomgren & Scherer, 2016) aims to validate previous results: authors suggest that a limitation in the generalization of acoustic studies on stuttering is due to the fact that most of them have been conducted on English speakers. For these reasons, the same acoustic parameter (F2 transition slope[7]) and the same stimuli used by Robb & Blomgren were used to infer anticipatory coarticulation in the fluent speech of their subjects. Other F2 transitions features were investigated to compare articulatory dynamics: the overall frequency extent, the overall duration of the transition and the speech rate. The participants of the study were 10 Iranian AWS and 10 AWNS. The findings revealed significant differences in a number of measures between stuttering and non-stuttering speakers. Compared to control group, perceptually fluent utterances of PWS were characterized by greater F2 frequency extents during transitions, longer F2 transition which took more time to reach vowel steady state and slower speaking rate. Concerning the transition slope, results showed no differences in the overall F2 slope between the two groups of speakers but, when the comparison window was constrained to the beginning of the transition (i.e., on the first 30 ms), differences were identified. Stuttering speakers, in fact, exhibited greater absolute initial F2 slope compared to control group. Overall, these findings corroborate previous results found in the literature on stuttering (Robb, Blomgren, 1997), and suggest that PWS adopt different articulatory strategies to reach fluency, compared to normally fluent speakers. The longer F2 transitions (and the lower speaking rate) suggest that PWS need a lengthening of time to complete an articulatory movement to an intended target; the greater overall F2 transition frequency extent (or change) suggests that stuttering speakers had more lingual movement (or displacement) during the transition compared to control group. Authors conclude that there are both spatial and temporal features of F2 transitions in stuttering speakers that need to be better understood.

Sussman *et al.* (2010) focus on anticipatory coarticulation using LE metric in the fluent and disfluent productions of stop + vowel sequences in 8 PWS and 8 PWNS. Linear regression functions were performed separately for the fluent and disfluent utterances. Results show that PWS slope values, both for fluent and for dysfluent slope

[7] F2 transition slopes were calculated using the fixed-time point method (Nearey, Shammass, 1987): the offset formant frequencies were specified at the distinct time-points of 30 ms and 60 ms from the onset of the transition (corresponding to the first glottal pulse of the vowel following the C).

tokens, fall within the normative ranges. Despite this, PWS show more variation in repeated productions of the target stimuli, as demonstrated by Standard Error Estimate values collected from LE analysis. Authors suggest that this argues against 'any serious deficits in the motor planning/execution of stop + vowel anticipatory coarticulation in PWS' (Sussman *et al.*, 2010: 12). It is important to underline that, in this study, the more severe stutterers showed, both for fluent and for dysfluent utterances, the most distant slope values from the control group speakers. This means a lower degree of coarticulation compared to control group and to moderate/mild stutterers.

Therefore, the picture emerging from the study of a possible inefficiency in the stutterers' ability to plan and execute the proper degree of anticipatory coarticulation is still controversial.

A further step towards a deeper understanding of speech dynamics responsible for stuttering is to assess the speech of children close to the onset of stuttering[8]. Studying the speech of children who stutter (henceforth, CWS) allows to discriminate the core characteristics of the disorder from possible articulatory strategies that AWS could adopt to overcome interruptions in the smooth flow of speech.

Chang, Ohde & Conture (2002), for example, assess anticipatory coarticulation and Formant Transition Rate (FTR) on the fluent speech of 14 CWS and 14 non-stuttering children who do not stutter (henceforth, CWNS). Only fluent utterances were selected for the acoustic analyses. The degree of anticipatory coarticulation was evaluated through the LE metric, while the speed (the velocity) at which the tongue moves from one position in the oral cavity to another was assessed with FTR measurement. Results show that while CWS and CWNS did not differ in terms of degree of coarticulation (as measured by the LEs slope and y-intercept), they did differ in FTR measures. CWS differentiated FTR less than CWNS for place of articulation. This result indicates that CWS were slower than CWNS in executing anticipatory movement of articulators for the vowel during consonantal production. Therefore, authors suggest that there is a kinematic-based difference (speed of movement) between stuttering and control groups, rather than a direct coarticulation-based difference.

Another important rational for studying CWS's speech is that much of the actual research on developmental stuttering is focusing at the identification of clinical predictors of chronicity. One of the phonetic indexes proposed as a potential marker of chronicity was F2 transition. Stromsta (1965), for example, report an early longitudinal study in which F2 transitions were analyzed in 63 children identified by their parents as having stuttering. Disfluent segments were analyzed and results reveal that speech disfluencies characterized by abnormal formant transitions and abnormal terminations of phonation were found in the speech of those children whose stuttering became persistent, while children exhibiting normal formant tran-

[8] "coarticulatory behaviours of adults who do and do not stutter are not readily generalizable to the coarticulatory behaviours 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).

sitions recovered. Unfortunately, details about this study are scarce and is difficult to assess the generality of its results.

Some years later, Yaruss, Conture (1993) investigate F2 transition differences among 7 young children considered at 'low-risk' of persistence and 6 children regarded as 'high-risk' for chronic stuttering. Several measures of F2 transition, such as the duration, the extent and the transition rate of F2 in sound/syllable repetitions were made in comparing the extra disfluent segment with the fluent segment (e.g., b-but; a-and). The authors show that "children who stutter do produce missing (25-29%) or atypical (10-16%) formant transitions during the first iteration of their sound/syllable repetition" (p. 893), but the presence of abnormal F2 transitions was not sufficient to differentiate the two groups. A critical issue of the study is that the validity of the classification of a child as 'high' or 'low' risk of chronicity was not verified through longitudinal observations.

A more recent study, realized by Subramanian *et al.* (2003), investigate the predictive value of F2 transitions as an early marker of chronic stuttering in children close to the onset of the disorder. Twenty CWS and 10 CWNS were audio recorded during the initial visit, when the eventual classification of a child as 'persistent' or 'recovered' was unknown. The final status of each child was evaluated after a minimum of 36 month post-onset of the disorder.

The acoustic parameters considered were the frequency change and the duration of F2 transitions; to allow the comparison with the control group, only fluent utterances were selected for the analysis.

Results suggest that "the frequency dimension of the formant transition, rather than the time dimension, is the most significant contributor to the differences between stuttering and non-stuttering children as well as between the two stuttering subgroups" (Subramanian *et al.*, 2003: 70). As we can see, in fact, in Table 1, mean values for the duration measures did not differ between the three groups, while more interesting findings emerged for the frequency change measure.

Table 1 - *Means and standard deviations (in parentheses) of the three groups for frequency change (Hz) and duration (ms) (from Subramanian et al., 2003: 68)*

Measure	*Persistent*	*Recovered*	*Controls*
Duration	58.75 (22.72)	59.07 (19.61)	66.14 (37.12)
Frequency change	395.78 (196.72)	583.99 (229.77)	502.35 (232.10)

The persistent group shows a smaller frequency change compared to recovered and control groups. This result may be interpreted to reflect restricted spatial movement of the tongue from one position to another. It seems that, to reach fluency, children with persistent stuttering have to undershoot their articulatory target, showing a higher degree of coarticulation. Authors conclude that F2 transitions could be a good marker of persistent stuttering in children close to the onset of the disorder, because results showed smaller frequency changes (reduced amplitude of the articulatory gestures) for those children whose stuttering become chronic, compared to children whose stuttering recovered.

We will focus now on an Italian project named 'Phonetic predictive indexes of chronic stuttering in preschool age children'[9]. The aim of this project is to test the prognostic value of a set of clinical predictors proposed as good markers of persistent stuttering: the Disfluencies Profile (Zmarich, Bernardini, Lenoci, Natarelli, Pisciotta, *to appear*) and the speech-associated attitudes of preschool and kindergarten children, as measured by kiddyCAT (Vanrykeghem, Brutten, 2007). Adding to these more clinical markers, the degree of CV coarticulation was also investigated.

The 13 CWS enrolled in the study were audio-recorded from the stuttering onset stage to 16-22 month post-onset. The validation of the prognosis (persistent vs recovered stuttering) was evaluated through structured telephone interviews with the parents after 3 and 4 years from the onset. A group of 26 CWNS, matched for age and sex, served as control group.

In Lenoci (2015) acoustic data are presented for a group of 5 stuttering children (3 persistent vs 2 recovered stutters), and results were compared with those of 26 non-stuttering children, selected in order to be matched for sex and age with the former group. The author collected spontaneous speech recordings at 3 different main stages: onset of the disorder, second and third semester post-onset. The last two stages were used to investigate the prognostic value of F2 transitions to discriminate persistent vs recovered stutterers. Only fluent CV and CVC sequences, containing bilabial, alveolar and velar stops in any vowel context were selected for the acoustic analyses. The metric used to investigate anticipatory coarticulation was Locus Equations. Slope values indexing anticipatory coarticulation were used for comparison analysis among the three groups (persistent stutterers, recovered stutterers and control group). Results show that for the bilabial and alveolar place of articulation, 2 of the children who later developed persistent stuttering presented lower slope values compared to the other two groups. This tendency has been observed through the three stages analysed altogether (from the onset of stuttering to the third semester post-onset). Children who later recover present slope values very similar to control group's ones for the three places of articulation. Velar slope values show less differences between the three groups, even if the more severe stuttering children present a lower degree of coarticulation compared to control group. These preliminary results suggest that speech gestures are poorly coordinated and performed with higher amplitude in the speech of children with chronic stuttering. Even if the small sample of subjects does not allow for prognostic inferences, these preliminary results corroborate previous findings from the pertinent literature (Robb, Blomgren, 1993), according to which PWS have a lower degree of coarticulation compared to control group. These results can also be interpreted with the proposal made by Van Lieshout *et al.* (2004), according to which PWS show a stronger reliance on the kinesthetic feedback compared to normal speakers, in

[9] This is a longitudinal project granted to Claudio Zmarich by the CNR in 2008 (Lenoci, Allegri, Bernardini, Chiari, Crivelli, Dadamo, de Biase, Galatà, V., Pisciotta, Polesel, Stanchina, Stocco, Vayra, Zmarich, 2012).

order to maintain stability in the speech motor control. To increase the feedback gain, probably, CWS need to increase the range of articulatory movements and this means to perform speech gestures with a low degree of coarticulation.

Despite all, acoustic results regarding coarticulation in the speech of CWS seem to be still controversial even if they provide the rationale for further investigations. For this reason, in the next section we will focus on more direct measures of speech dynamics, such as Ultrasound Imaging of the Tongue[10].

4. *Ultrasound Tongue Imaging studies on coarticulation in the stuttering speech*

Ultrasound Tongue Imaging (henceforth, UTI) has been increasingly used in speech sciences research over the last few decades and it offers a direct representation of tongue movements in speech. It is a safe and non-invasive articulatory technique, providing information about the shape and the position of the main articulator involved in the production of consonants and vowels, the tongue. Such direct measures allow to get a deeper insight into lingual articulation dynamics that are not revealed through acoustic measures. Furthermore, particular attention has been devoted to the potential power of UTI as a biofeedback tool for modifying atypical articulations in speech disordered speakers (Cleland, Scobbie & Zharkova, 2016).

So far, two ultrasound studies were realized on the speech of adults who stutter (Heyde, Scobbie, Lickley & Drake, 2015; Frisch, Maxfield & Belmont, 2016). The first study aims at corroborate the hypothesis that PWS struggle not when initiating the absolute syllable initial consonant but when transitioning from the C into the following vowel (Wingate, 1988). Fluent productions of CV syllables (C = /k/; V = /a, i, ə/) from 3 PWS and 3 PWNS were analysed for duration and peak velocity relative to articulatory movement towards (onset) and away (offset) from the consonantal closure.

Measures of displacement and velocity were collected at the point of maximum displacement of the tongue surface and results show that while the two groups had comparable onset behaviours, they do differ in offset peak velocity. PWS, in fact, displayed lower velocity at reaching the vowel target compared to controls. According to authors, results confirm the hypothesis according to which coarticulation from a sound to the succeeding one is impaired in the stuttering speech and that this could be an indicator for an underlying motor control impairment.

Frisch *et al.* (2016)'s study was specifically devoted to investigate anticipatory coarticulation in the fluent speech of 23 AWS and 23 AWNS. The study examines the coarticulation of anticipatory velar-vowel sequences in the adjustment of velar closure location for /k/[11] depending on the following vowel context (the nine Standard American English vowels /i e æ ʌ ɝ ɑ ɔ o u/). For the purposes of the study,

[10] The advantage of ultrasound over acoustic analysis is that the last one provides only indirect evidence of articulatory movements, posing some issues (for example in the detection of formants) especially for children acoustic analysis.

[11] "the target phoneme in this study was /k/, which is known to have relatively large variation in production across contexts" (p. 288).

authors selected for each velar-vowel production the UTI frames displaying maximum velar closure for subsequent analysis. Coarticulation was determined through curve-to-curve distance[12] comparison between tongue contours across the variety of all vowel contexts for each speaker following Zharkova, Hewlett & Hardcastle (2012). Differences among tokens (18 monosyllabic CV or CVC sequences) were measured, and average measures of coarticulation obtained from each speaker were then used in statistical analysis to compare overall patterns of coarticulation.

Results show that curve-to-curve distances between front and back vowel contexts did not differ for PWS and PWNS. This means that the two groups show identical patterns of coarticulation. In line with previous findings from literature (Sussman *et al.*, 2010; Smith, Sadagopan, Walsh & Weber-Fox 2010), PWS of this study were found to be more variable than PWNS in the production of the same articulatory target (i.e., the same velar-vowel sequence). Authors conclude that PWS's stuttering-like disfluencies are not attributable to immature motor planning as measured by anticipatory coarticulation (Frisch *et al.*, 2016).

4.1 An UTI study of lingual coarticulation in the speech of Italian children who stutter

An ongoing project, developed in the Laboratorio di Linguistica of Scuola Normale Superiore, focuses on studying the speech of CWS. The aim of the project is to assess the presence of differential articulatory patterns between a group of Italian stuttering children and a matched control group. The motor aspects under investigation are those underlying the anticipatory coarticulation and the stability of movements through multiple repetitions of the same item.

So far, 10 school age children (age range from 6 to 12 years) were recruited for the study: 5 CWS[13] (3 males and 2 females) and 5 CWNS, balanced for age and sex. Subjects were recorded by means of an ultrasound system (*Mindray UTI system-30 Hz*) with a microconvex probe (*Mindary probe 65EC10EA*); the UTI images were synchronized with the audio through a synchronisation unit (*Synch Bright-up unit*). In a child-friendly set up, participants were seated comfortably in a chair and the ultrasound probe was held under the chin with the head stabilization unit (*Articulate stabilisation headset*, Articulate Instruments ltd) for stabilizing the position of the transducer with respect to the head. The experimental task consisted in the production of /CV/ sequences with the consonant C corresponding to the bilabial stop /b/, alveolar /d/ and velar /g/ and the vowel V corresponding to the high front /i/, low /a/ and high back /u/. The three cardinal vowels allowed for testing diverging tongue positions. The target syllables were embedded in dysillabic pseudo-words of the type /ˈCVba/. Bilabials following the

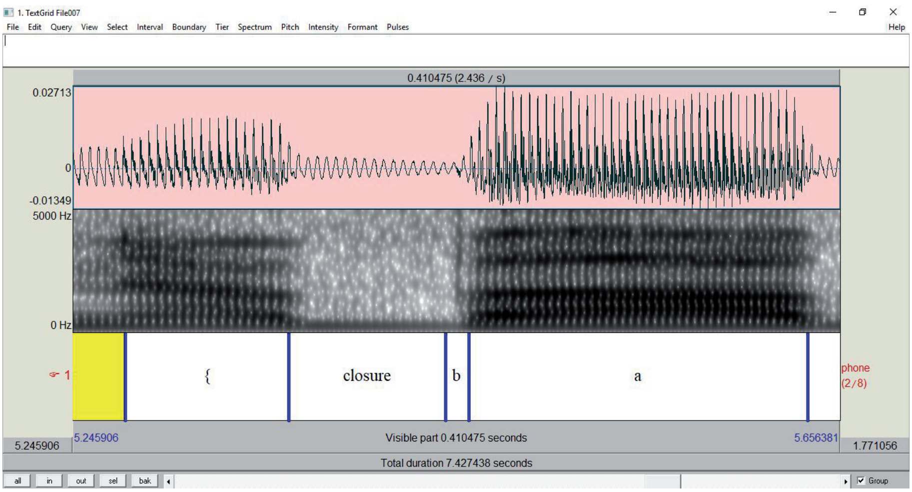
[12] "Mean distance in midsagittal tongue surface outline between tokens of the same phoneme across two different environments was taken as a measure of the phoneme's susceptibility to environment influence" (Zharkova, Hewlett, 2009: 3).

[13] Parents of stuttering children reported a history of developmental stuttering with no other language or hearing disorders. The presence of stuttering was evaluated from a speech pathologist and all stuttering children had begun therapeutic treatment in the last few years. Parents of the typically fluent group reported no history of speech, language and hearing disorders.

target syllables were used to eliminate the influence of additional lingual coarticulation within pseudo-words. Twelve repetitions of each CV sequence, embedded in short carrier phrases (ex. 'La gattina DUba salirà' the little cat DUba will go up), were collected in random order, for a total of 108 utterances for each participant. Here we will present only preliminary data for the coarticulation during the fluent speech[14] of one CWS and one CWNS. To achieve the goal we transposed measures of LE to the articulatory domain in line with previous recent studies (Noiray, Menard & Iskarous, 2013). The adaptation of LE to the articulatory domain was conducted on the lingual data simultaneously recorded with the acoustic speech signal. Instead of F2 transitions, we used the horizontal position of the highest point of the tongue at the mid-point of the consonantal closure (dependent variable) and at the mid-point of the following vowel (independent variable)[15]. We chose the mid-point of the consonant closure as dependent variable to ensure that what we were measuring was indeed 'coarticulation', as opposed to being part of the 'transition' from the C to the V articulation.

First, acoustic data were phonetically segmented and labelled using PRAAT. Three intervals were selected on the spectrograms of each CV sequences (see Figure 2): the consonantal closure (from the offset of the preceding vowel to the burst of the stop); the VOT of the C (from the burst to the first glottal pulse of the following vowel); the V-target (from the first glottal pulse to the offset of the vowel: see Figure 1).

Figure 2 - *Example of a PRAAT window, with the waveform, the spectrogram and the text-grid containing the segmentation and labelling of a [ba] sequence*



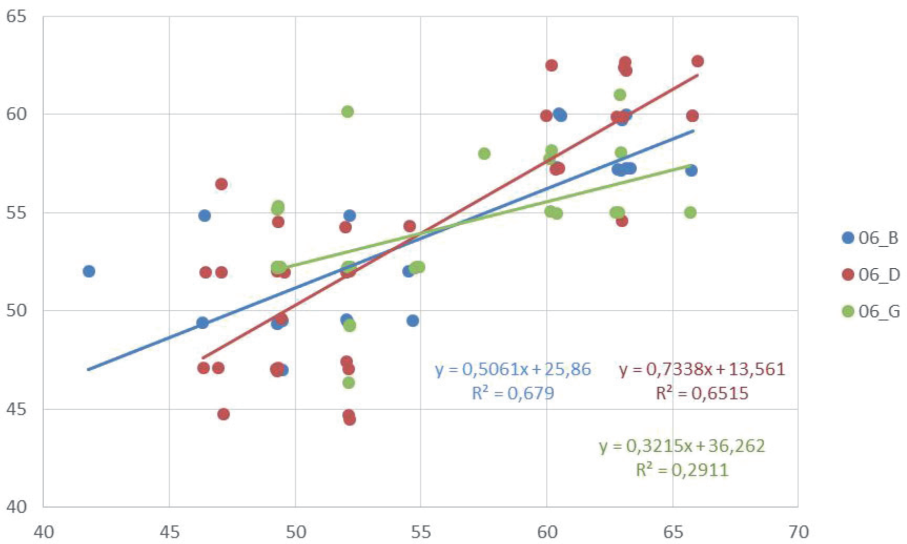
[14] CWS's speech was judged as perceptually fluent on the basis of audible acoustic data. The absence of syllable repetitions, blocks and sound prolongations (cfr. Stuttering-like disfluencies) was used as criteria to judge the speech as perceptually fluent.

[15] Since several studies have associated the frontness of the body of the tongue in the vocal tract to F2, Iskarous, Fowler & Whalen (2010) measured, with EMMA, the horizontal position of the tongue body at the release of consonant and in the middle of the vowel, in order to investigate the articulatory origins of LE. Authors found the same linear relations present in the acoustic domain.

The annotations were imported in the software used for the articulatory analysis (*Articulate Assistant Advanced* - Articulate Instruments ltd) for the selection of the ultrasound relevant frames within each C-V intervals. A semi-automatic tongue contour splining was performed in the acoustic interval spanning from the beginning of the consonantal closure to the end of the following vowel.

As already described, relevant ultrasound frames were selected for each CV sequence at two time points: the midpoint of the consonantal closure and the midpoint of the vowel. Each tongue curve (spline) was converted in terms of x,y coordinates and the x value corresponding to the highest y point on each curve was selected for subsequent linear regression analysis. Linear regression functions were calculated for each place of articulation and for each child[16]. Figures below display the regression lines obtained from the fluent utterances of one CWNS (Figure 3) and one CWS (Figure 4).

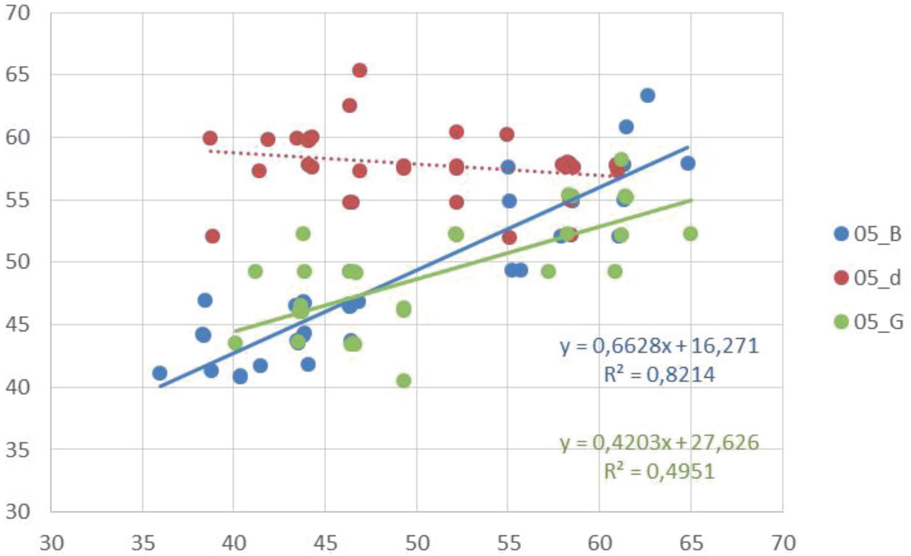
Figure 3 - *Linear regressions between consonant-closure midpoint and vowel midpoint for one child who do not stutter. Slopes are b: 0.50, d: 0.73, g: 0.29 respectively for the bilabial (blue), alveolar (red) and velar (green) place of articulation*



For CWNS we can observe that, for each place of articulation, data points tightly cluster across the regression lines. This means that, in line with previous literature, the linearity of LE originates in linearity in articulation between the horizontal position of the tongue dorsum in the consonant and to the horizontal position of the tongue dorsum in the vowel (Iskarous *et al.*, 2010).

[16] The experimental procedure and the statistical analyses were developed with the technical support of Irene Ricci, from the Laboratorio di Linguistica of Scuola Normale Superiore.

Figure 4 - *Linear regressions between consonant-closure midpoint and vowel midpoint for one stuttering child. Slopes are b: 0.66, g: 0.42 respectively for the bilabial (blue), alveolar (red) and velar (green) place of articulation*



As for the CWS, we can observe that for the bilabial and velar place of articulations this CWS shows slightly higher slope values compared to the CWNS. For the alveolar consonants instead, the regression line did not fit the data (red dots). This means that for the CWS there is no interaction between the position of tongue dorsum at the mid-point of the consonantal closure and the midpoint of the following vowel. Consequently, in terms of LEs metric, there is no anticipatory effect in the production of /dV/ sequences. Interestingly enough, a linearity emerged when we used a different time point as dependent variable: the consonantal offset[17] instead of the mid-point of the closure. According to Zharkova *et al.* (2012), a possibility to measure the extent to which a speech sound varies systematically according to the identity of the following one, and to compare the effects between speakers, is to compare the size of any coarticulatory effect at a selected time point. The result obtained in our study could mean that anticipation of the upcoming V gesture starts later in the CWS compared to CWNS. Another aspect to underline is that even though a coarticulatory effect emerged for the stuttering child, the slope value was considerably lower compared to the normally fluent child. These preliminary results suggest that, even if no significant differences emerge for the degree of coarticulation of /bV/ and /gV/ sequences between the two children, some differences emerged when the synergistic use of different part of the same articulator (the tip and the body of the tongue) is required. Our opinion is that, this result can be inter-

[17] We selected as dependent variable the horizontal position of the tongue on the UTI frame included between the burst and first glottal pulse of the vowel.

preted as a less mature motor control system for the CWS who need more time to reach the articulatory target for the C and, consequently, cannot anticipate earlier the vowel gesture.

5. *Conclusion*

A still controversial picture emerges from this review on the acoustic and articulatory studies investigating coarticulation on the speech of adults and children who stutter. The ambiguous results may be due to the different age samples, speech samples and methodologies used across studies here reviewed. Overall, most studies have shown that PWS differ in respect to coarticulatory dynamics compared to PWNS, even during their perceptually fluent speech. Sometimes, however, PWS and PWNS seem to be very similar, showing only extremely subtle differences, and these results corroborate Van Lieshout *et al.* (2004)'s motor skill perspective, according to which there is an individual variability along a continuum for *all speakers*, and PWS stay in the low end of that continuum. Anyway, further investigations especially on CWS are needed in order to both assess the real manifestations of the disorder, and to equip clinicians with the instruments for a better rehabilitation. Ultrasound Tongue Imaging analysis is a valuable tool for both these purposes. Preliminary data on two Italian children show, for example, that some differences can characterize the fluent speech of the stuttering child: for the alveolar stops, for example, we observed a lower degree of coarticulation, which emerges later compared to the control peer. The lower degree of coarticulation suggests that for the stuttering child, the alveolar consonants adapt to the vowels less than the control child. A possible interpretation is that the CWS have not learned how to achieve the articulatory differentiation between two parts of the same articulator, the tip and the body of the tongue, and how to coordinate them in the proper way, as the normally fluent peer does. Furthermore, this articulatory result corroborates previous acoustic results obtained for a group of pre-school Italian children who stutter (Lenoci, 2015) and it is an aspect that need to be investigate further.

Bibliography

Ambrose, N.G., Yairi, E., Loucks, T.M., Seery, C.H. & Throneburg, R. (2015). Relation of motor, linguistic and temperament factors in epidemiologic subtypes of persistent and recovered stuttering: Initial findings. In *Journal of Fluency Disorders*, 45, 12-26.

Buhr, A.P., Zebrowski, P.M. (2009). Sentence position and syntactic complexity of stuttering in early childhood: A longitudinal study. In *Journal of Fluency Disorders*, 34, 155-172.

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.

Cleland, J., Scobbie, J. & Zharkova, N. (2016). Insights from ultrasound: Enhancing our understanding of clinical phonetics. In *Clinical Linguistics & Phonetics*, 30, 171-175.

Dehqan, A., Yadegari, F., Blomgren, M. & Scherer, R.C. (2016). Formant transitions in the fluent speech of Farsi-speaking people who stutter. In *Journal of Fluency Disorders*, 48, 1-15.

Frisch, S.A., Maxfield, N. & Belmont, A. (2016). Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter. In *Clinical Linguistics & Phonetics*, 30, 277-291.

Guthrie, E.R. (1952). *The psychology of learning.* New York: Harper and Row.

Hardcastle, B., Tjaden, K. (2008). Coarticulation and Speech Impairment. In Ball, M.J., Perkins, M.R., Muller, N. & Howard, S. (Eds.), *The Handbook of Clinical Linguistics*. Malden: Blackwell Publishing Ltd., 506-524.

Harrington, J. (1987). Coarticulation and stuttering: an acoustic and electropalatographic study. In Peters, H.F.M., Hulstijn, W. (Eds.), *Speech Motor Dynamics in Stuttering.* Wien: Springer Verlag, 381-392.

Hawkins, S. (1984). On the Development of Motor Control in Speech: Evidence from Studies of Temporal Coordination. In Lass, N.J. (Ed.), *Speech and Language: Advances in basic research and practice.* New York: Academic Press, 317-374.

Heyde, C.J., Scobbie, J.M., Lickley, R. & Drake, E.K.E. (2015). How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound. In *Clinical Linguistics & Phonetics*, 30, 292-312.

Howell, P., Vause, L. (1986). Acoustic analysis and perception of vowels in stuttered speech. In *Journal of the Acoustical Society of America*, 79, 1571-1579.

Iskarous, I., Fowler, C. & Whalen, D. (2010). Locus equations are an acoustic expression of articulator synergy. In *Journal of the Acoustical Society of America*, 128, 2021-2032.

Kelso, J.A.S. (1997). *Dynamic patterns. The self-organization of brain and behavior.* Cambridge, MA: MIT Press.

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.

Krull, D. (1988). Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops. In *Phonetic Experimental Research at the Institute of Linguistics*. Stockholm University: PERILUS X, 87-101.

Lashley, K.S. (1951). The problem of serial order in behaviour. In Jeffress, L.A. (Ed.), *Cerebral Mechanisms in Behavior.* New York: Wiley, 112-36.

Lenoci, G. (2015). Balbuzie evolutiva e indici fonetici predittivi di cronicità. Tesi di Dottorato, Scuola Normale Superiore, Pisa.

Lenoci, G., Allegri, S., Bernardini, S., Chiari, F., Crivelli, N., Dadamo, V., de Biase, M., Galatà, V., Pisciotta, C., Polesel, L., Stanchina, S., Stocco, D., Vayra, M. & Zmarich, C. (2012). Il progetto di ricerca longitudinale "indici fonetici predittivi di balbuzie cronica in età prescolare: primi risultati". In Paoloni, A., Falcone M. (Eds.), *La voce nelle applicazioni*, Roma. Bulzoni Editore: Roma, 19-34.

Levelt, W.J.M., Roelofs, A. & Mejer, A.S. (1999). A theory of lexical access in speech production. In *Behavioral and Brain Sciences*, 22, 1-75.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W., Marchal, A. (Eds.), *Speech production and speech modeling.* Amsterdam: Kluwer Academic Publishers.

Lindblom, B. (1963). *On vowel reduction, Report No. 29*. Stockholm: The Royal Institute of Technology, Speech Transmission Laboratory.

MacNeilage, P.F., Davis, B.L. (2000). The Motor Core of Speech: A Comparison of Serial Organization Patterns in Infants and Languages. In *Child Development*, 71, 153-163.

MacPherson, M.K., Smith, A. (2013). Influences of Sentence Length and Syntactic Complexity on the Speech Motor Control of Children Who Stutter. In *Journal of Speech, Language, and Hearing Research*, 56, 89-102.

Namasivayam, A.K., Van Lieshout, P. (2011). Speech Motor Skill and Stuttering. In *Journal of Motor Behavior*, 43, 477-489.

Nearey, T., Shammass, S. (1987). Formant transitions as partly distinctive invariant properties in the identification of voiced stops. In *Canadian Acoustics*, 15, 17-24.

Noiray, A., Menard, L. & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. In *Journal of Acoustical Society of America*, 133, 444-452.

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 stutterers and nonstutterers. In *Journal of Fluency Disorders*, 22, 1-16.

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., Sadagopan, N., Walsh, B. & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. In *Journal of Fluency Disorders*, 35, 1-18.

Smith, A., Zelaznik, H.N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. In *Developmental Psychobiology*, 45, 22-33.

Stromsta, C. (1965). A spectrographic study of disfluencies labeled as stuttering by parents. In *De Therapia Vocis et Loquellae*, 1, 317-320.

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.

Sussman, H.M., Byrd, C.T. & Guitar, B. (2010). The integrity of anticipatory coarticulation in fluent and non-fluent tokens of adult who stutter. In *Clinical Linguistics & Phonetics*, 25, 169-186.

Sussman, H.M., Hoemeke, K.A. & Ahmed, F. (1993). A cross-linguistic investigation of locus equations as a relationally invariant descriptor for place of articulation. In *Journal of the Acoustical Society of America*, 94, 1256-1268.

Sussman, H.M., Hoemeke, K.A. & McCaffrey, H.A. (1992). Locus Equations as an index of coarticulation for place of Articulation Distinctions in Children. In *Journal of Speech and Hearing Research*, 35, 769-781.

Van Lieshout, P.H.H.M. (1995). *Motor planning and articulation in fluent speech of stutterers and nonstutterers*. Nijmegen: The Netherlands University Press Nijmegen.

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 stutter: A motor skill approach. In Maassen, B., Kent, R., Peters, H.F.M., Van Lieshout, P.H.H.M & Hulstijn, W. (Eds.) *Speech motor control in normal and disordered speech*. Oxford: Oxford University Press, 313-355.

Vanryckeghem, M., Brutten, G.J. (2007). *The KiddyCAT: communication attitude test for preschool and kindergarten children who stutter*. San Diego: Plural Publishing.

Weismer, G. (1991). Assessment of articulatory timing. In Cooper, J. (Ed.), *Assessment of speech and voice production: Research and clinical applications*. Bethesda, Maryland: National Institute on Deafness and other Communication Disorders.

Wingate, M.E. (1988). *The structure of stuttering: A psycholinguistic analysis*. New York: Springer Verlag.

Yairi, E., Ambrose, N. (2005). *Early Childhood Stuttering: for clinicians by clinicians*. Austin, TX: Pro-Ed.

Yairi, E., Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. In *Journal of Fluency Disorders*, 38, 66-87.

Yaruss, J., Conture, E.G. (1993). F2 transition during sound/syllable repetitions of children who stutter and predictions of stuttering chronicity. In *Journal of Speech and Hearing Research*, 36, 883-896.

Zharkova, N., Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. In *Journal of Phonetics*, 37, 248-256.

Zharkova, N., Hewlett, N. & Hardcastle, W.J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. In *Journal of the International Phonetic Association*, 42, 193-208.

Zmarich, C., Bernardini, S., Lenoci, G., Natarelli, G. & Pisciotta, C. (in press). Could the frequencies of stuttering-like-dysfluencies predict persistent stuttering in children who have just started to stutter? In Savy R., De Meo, A. & Alfano I. (Eds.), *La fonetica sperimentale nell'insegnamento e nell'apprendimento delle lingue straniere*.

DOI: 10.17469/O2103AISV000021

GIOVANNI DEL BEN, PIERPAOLO BUSAN

The neural correlates of developmental stuttering: A brief overview of the literature

Developmental stuttering (DS) is a disruption of the rhythmic flow of speech, and its aetiology is still obscure. Neuroimaging/neurophysiological techniques have been used to study, the neural system of people with DS highlighting the presence of widespread structural/functional abnormalities, especially in the motor system. Reduced white matter integrity and altered functioning of the basal ganglia-thalamo-cortical circuit play a key role in DS. Evidence from transcranial magnetic stimulation suggests the presence of an altered interplay between excitatory and inhibitory signals, especially in the left motor cortices; findings of neurophysiological indexes obtained from non-speech related muscles, support the theory that stuttering is the overt symptom of a more general motor disorder. Further investigations need to be conducted to better elucidate the neural basis of this disorder, in order to find better rehabilitative solutions.

Key words: Stuttering, connectivity, motor cortex, basal ganglia, transcranial magnetic stimulation.

1. *What is stuttering?*

Stuttering is a particular condition in which the normal rhythmic flow of speech is disrupted by frequent pauses, blocks, hesitations, repetitions of syllables, words, and sounds. It may result in a highly negative impact on the quality of life and daily activities of people who stutter, affecting not only their spoken communication attitudes but also their emotional stability and mental health status (Craig, Blumgart & Tran, 2009). Stuttering may be associated with lower social interaction capacities, educational and occupational disadvantages, self-imposed isolation and elevated levels of social anxiety (see Craig, Tran, 2014; Iverach, Rapee, 2014). Developmental stuttering (DS) is the most common form of the disturbances and includes all cases with gradual onset in childhood that are not the result of an acquired brain damage (Costa, Kroll, 2000); disfluencies occur predominantly at the beginning of words and phrases (Bloodstein, 1995) and they are characterized by possible adaptation phenomena (Craig-McQuaide, Akram, Zrinzo & Tripoliti, 2014). The overt symptoms of stuttering are also usually accompanied by spasms and associated movements of various muscular districts (especially facial muscles), that may initially help people with DS to overcame the disfluencies (Riva-Posse, Busto-Marolt, Schteinschnaider, Martinez-Echenique, Cammarota & Merello, 2008). DS is a relatively frequent disorder in childhood, especially in males. In many cases, it recovers either naturally or as result of a specific clinical/behavioural treatment (see Yairi, Ambrose, 2013). However, due to the complex nature of speech production the exact aetiology of DS has not yet been fully understood: multifactorial components, as well as genetic predispositions, seem to play an important role in the pathophysiology of this complex disorder (Neef, Anwander & Friederici, 2015a). Epidemiological twin studies, adoption studies, family aggregation studies and different sex ratios suggest a possible influence of genetic components in stuttering (see Yairi, Ambrose, 2013). For example, mutations in proteins involved in the lysosomal enzyme-targeting pathway (Kang, Riazuddin, Mundorff, Krasnewich, Friedman, Mullikin & Drayna, 2010) have been identified in people with DS: protein trafficking plays a key role in the biogenesis and maintenance of myelin sheaths and thus the white matter abnormalities described in DS (e.g. Sommer, Koch, Paulus, Weiller & Buchel, 2002; Watkins, Smith, Davis & Howell, 2008; see below) could be related to those mutations (see also Buchel, Watkins, 2010). In this regard, a mouse model genetically engineered to carry one of these mutations has been created: with respect to the littermate wild type controls, mutant puppies show stuttering-like behaviours, emitting fewer vocalizations with longer pauses between them (Barnes, Wozniak, Gutierrez, Han, Drayna & Holy, 2016). Over the years, scientists from different fields, ranging from psychology to linguistics, from biomechanics to neuroscience, have proposed different theories to describe the causal role of DS (Neef *et al.*, 2015a). One of the older and most influential theories of DS pathophysiology suggests that stuttering may occur as a result of an incomplete dominance of the speech and motor centres of the left hemisphere over the right homologue areas (Travis, 1978). In this context, it has been widely shown that stuttering improves during various "fluency-inducing conditions", such as singing, choral speech and external rhythmic cues (e.g. speaking with a metronome); these empirical observations suggest that the disorder has its origin in the central nervous system, likely at a speech motor planning level, rather than in the peripheral nervous system or in abnormalities of the vocal apparatus (see Craig-McQuaide *et al.*, 2014). In fact, the occurrence of stuttering during direct intra-operative electrical stimulation of brain regions such as supplementary motor area (Penfield, Welch, 1951) and/or thalamus (Ojemann, Ward, 1971) indicates the implication of a cortico-subcortical circuit in the disorder. This is further supported by the evidence of stuttering occurrence in fluent speakers after the stimulation of the left frontal aslant tract (Kemerdere, Champfleur, Deverdun, Cochereau, Moritz-Gasser, Herbert & Duffau, 2016) which connects the pars opercularis of the inferior frontal gyrus and the anterior supplementary motor/pre-supplementary motor areas (see Catani, Dell'Acqua, Vergani, Malik, Hodge, Roy, Valabregue & Thiebaut de Schotten, 2012). All this evidence sustains also the suggestion that stuttering may be related to a series of disturbances that involve also the auditory-motor integration level (useful for speech), especially in adults (e.g. Daliri, Max, 2015). As a consequence, the most relevant vision includes, at the moment, the proposal that stuttering should be more properly viewed as a disconnection syndrome (see below and Sommer *et al.*, 2002), also characterized by the aberrant functioning of the basal ganglia system (see below and Alm, 2004). In this context, in the present work, we will briefly review the available neurophysiological evidence in DS, trying to combine it with one of the more recent computational models of DS (Civier, Bullock, Max & Guenther, 2013), in order to propose new suggestions for research, and toward more effective rehabilitative solutions.

2. *Neuroimaging correlates of developmental stuttering*

Numerous non-invasive brain imaging studies have provided advanced structural and functional descriptions of the neural system of people with DS, and they have highly contributed to define some neural markers of DS. Techniques such as functional magnetic resonance, diffusion tensor imaging and positron emission tomography, were able to individuate widespread white matter abnormalities and metabolism alterations in DS: reduced levels of white matter are present bilaterally in many brain regions, such as in the posterior inferior frontal gyrus, in precentral gyrus, in ventral premotor cortex, in the cerebral peduncles (see Watkins *et al.*, 2008), in the arcuate fasciculus, in the left angular gyrus, in the left corticospinal tract and in the left corticobulbar tract (see Connally, Ward, Howell & Watkins, 2014). In this context, especially during speech production, people with DS show lower neural activity in the left ventral premotor cortex, the left and antero-medial Heschl's gyrus, left and right sensorimotor cortex and in the rolandic opercular cortex (Watkins *et al.*, 2008). Interestingly, reduced white matter integrity is present in the left rolandic operculum, immediately below the motor representations of tongue, larynx and pharynx (see also Sommer *et al.*, 2002), while a prominent increase in white matter volume is present in various regions of the right hemisphere, likely as the result of compensatory mechanisms, in the superior temporal gyrus, the inferior frontal gyrus including the pars opercularis, and in the sensorimotor areas including hand and mouth motor representations (Jäncke, Hänggi & Steinmetz, 2004). An increased mean diffusivity is also present, bilaterally, in the frontal aslant tract (Kronfeld-Duenias, Amir, Ezrati-Vinacour, Civier & Ben-Shachar, 2016). Other regions characterized by white matter abnormalities include chorona radiata, left superior longitudinal fasciculus (Chang, Zhu, Choo & Angstadt, 2015), corpus callosum and thalamo-cortical circuits (Choo, Kraft, Olivero, Ambrose, Sharma, Chang & Loucks, 2011). Similarly, in the adult population of DS decreased/increased volumes of grey matter may be evident in basal ganglia, cerebellum, inferior frontal gyrus, middle temporal gyri, pre- and post-central gyri and superior temporal gyri, in both hemispheres (see Beal, Gracco, Lafaille & Denil, 2007; Lu, Chen, Ning, Ding, Guo, Peng, Yang, Li & Lin, 2010; Song, Peng, Jin, Yao, Ning, Guo & Zhang, 2007). The major part of available data has been obtained from the adult population, while a reduced amount of studies have been conducted in children with DS. Also in this population differences in neural networks with respect to fluent speakers have been highlighted: a decrease in gray matter volume is present, in DS, in the left and right inferior frontal gyri, left anterior cingulate gyrus, right temporal regions and, bilaterally, in the supplementary motor area (Chang, Erickson, Ambrose, Hsegawa-Johnson & Ludlow, 2008). A reduction is evident also in the left putamen (Beal, Gracco, Brettschneider, Kroll & Denil, 2013). Conversely, grey matter volume is increased, in the right hemisphere, in the middle frontal gyrus, in the post-central gyrus, in the superior temporal gyrus, in the inferior parietal lobule and in the rolandic operculum in DS children (Beal *et al.*, 2013). Reduced white matter integrity in the tract underlying the left rolandic opercular region is also present in DS children but not associated with higher white matter volumes in right hemisphere speech regions, suggesting that the anatomical increase present in adults may be the result of compensatory mechanisms (see Chang *et al.*, 2008). In fact, in DS there is an overactivation of the right hemisphere motor systems, including primary motor cortex, supplementary motor area, superior lateral premotor regions and cerebellum (Fox, Ingham, Ingham, Hirsch, Downs, Martin, Jerabek, Glass & Lancaster, 1996). Moreover, a greater activity is present in the left and right midbrain, at the level of the substantia nigra, in the subthalamic nucleus, peduncolo-pontine nucleus and red nucleus, as well as in the left and right posterior lobe of the cerebellum (Watkins *et al.*, 2008). Close to the right anterior insula, a systematic activation of the right frontal operculum (the right hemisphere homologue of Broca's area), may be evident in DS, especially during speech tasks (Preibisch, Neumann, Raab, Euler, von Gudenberg, Lanfermann & Giraud, 2003). The negative correlation between neural activation of this brain region and stuttering severity helps to exclude a direct causal role of this cortical area in DS but favours a vision that claims a compensatory activity of right frontal operculum in the disturbance. These findings sustain the idea of a compensatory role of the right hemisphere in DS (see Kell, Neumann, von Kriegstein, Poserienske, von Gudenberg & Euler, 2009) that may develop during a life of stuttering (see also Ingham, Grafton, Bothe & Ingham, 2012). Metabolism abnormalities are also present in DS: glucose hypometabolism has been highlighted in the neural system of people with DS in Wernicke's area, Broca's area, medial cerebellum, superior frontal cortices, ventral posterior cingulated cortex, frontal orbital cortex, anterior prefrontal cortex and angular gyrus; moreover in DS, the left caudate nucleus is nearly 50% less active both during stuttering and fluency-enhanced conditions (Wu, Maguire, Riley, Fallon, Lacasse, Chin, Klein, Tang, Cadwell & Lottenberg, 1995). The reduced glucose uptake seems in part related to an altered dopamine metabolism which may be present in people with DS: an increased dopamine uptake activity is present in DS in the left caudate tail, and in the right ventro-medial prefrontal cortex, which is an area functionally connected to the supplementary motor area. Other regions of enhanced uptake activity include the amygdala, the left insular cortex, the right deep orbital cortex, left insular cortex and the left pulvinar (Wu, Maguire, Riley, Lee, Keator, Tang, Fallon & Najafi, 1997). This is indirectly supported by the evidence of fluency enhancements after the administration of dopamine D2 antagonists such as haloperidol (Murray, Kelly, Campbell & Stefanik, 1977), risperidone (Maguire, Riley, Franklin & Gottshalk, 2000) and olanzapine (Maguire, Riley, Franklin, Maguire, Nguyen & Brojeni, 2004). Similarly, paroxetine, a selective serotonin reuptake inhibitor, is also effective in the management of stuttering symptoms (Busan, Battaglini, Borelli, Evaristo, Monti & Pelamatti, 2009) probably via a serotonin mediated and indirect anti-dopaminergic mechanism (Schreiber, Pick, 1997). As a consequence, it is evident that one of the main neural mechanisms related to DS may rely on the possible dysfunction of the basal ganglia system. For this reason, in the following section, we will try to focus our attention on possible dysfunctional cortico-basal-thalamo-cortical mechanisms in DS.

3. *Cortico-basal-thalamo-cortical networks in stuttering*

Stuttering shares a series of characteristics with various basal ganglia-related disorders such as Parkinson's Disease, attention deficit and hyperactivity disorders, Tourette's Syndrome and focal dystonia. In this context, acquired neurogenic stuttering often occurs after lesions of basal nuclei (see Craig-McQuaide *et al.*, 2014). An abnormal activity of basal ganglia (see the previous section but see also Alm, 2004), along with a consequent impairment of the cortico-basal-thalamo-cortical network that is mainly able to reach supplementary motor area complex, seems to play a key role in DS. In fact, fluent speech production is a highly demanding motor task that requires the punctual motor planning and execution of articulated movements through the integration of excitatory and inhibitory neural signals useful for the correct coordination of the muscles of speech apparatus. Supplementary motor area is involved in planning and execution of voluntary movements as well as in word production: anterior pre-supplementary motor area may have a role in lexical selection process, while its posterior portion may have a role in linear sequence encoding; finally, the "proper" supplementary motor area is fundamental in articulation of motor output (Alario, Chainay, Lehericy & Cohen, 2006). Basal ganglia are strongly involved in neural activity related for example to motor control of voluntary movements, learning, cognitive and limbic functions (Graybiel, 2000). An anomalous activation of basal ganglia in DS is often reported (e.g. Watkins *et al.*, 2008; Lu *et al.*, 2010): stuttering severity positively correlates with bilateral caudate nucleus activity and negatively correlates with left substantia nigra activity (Giraud, Neumann, Bachoud-Levi, von Gudenberg, Euler, Lanfermann & Preibisch, 2008). Weaker connectivity is present in DS when considering regions of the posterior middle temporal gyrus and the putamen, whereas a stronger connectivity is present from putamen to the thalamus and from this latter region to temporal cortices and supplementary motor area, as well as between them (Lu *et al.*, 2010). As a consequence, the understanding of DS neurophysiology may take advantage from the utilization of techniques that have been already extensively used in other basal ganglia related motor disorders (e.g Parkinson's Disease, dystonia and Tourette's Syndrome), such as transcranial magnetic stimulation (TMS).

4. *Electro/magneto-neurophysiological correlates of developmental stuttering*

The aforementioned findings, mainly obtained by using neuroimaging techniques such as functional magnetic resonance and positron emission tomography, have shed light on different aspects of DS neurophysiopathology. Similarly, non-invasive brain stimulation tools such as transcranial magnetic stimulation (TMS) have been employed to investigate the functioning of the motor system in adults with DS. TMS provides useful information on the excitability of motor cortex, cortico-spinal and cortico-bulbar physiology and on the role of the intracortical networks in the modulation of the final motor output (see Kobayashi, Pascual-Leone, 2003). Only few authors have employed, at the moment, TMS in DS, often concentrating on non-speech related muscles (see Sommer, Wischer, Tergau & Paulus, 2003; Alm, Karlsonn, Sundberg & Axelson, 2013), probably due to the challenging methods required to record motor evoked potentials (MEPs) directly from the speech apparatus (see D'Ausilio, Jarmolowska, Busan, Bufalari & Craighero, 2011). Early findings from TMS highlighted that indexes of intracortical inhibition (ICI) and facilitation (ICF), recorded from right hand muscles when stimulating only the left motor cortex, are normal in DS, but an increased resting and active motor threshold in the left motor cortex is evident, suggesting that a dysfunction at a cortico-spinal level is present (Sommer *et al.*, 2003). In this regard, evidence from recruitment curves suggests that left hand cortical excitability is lower in DS probably due to a reduced number of cortical projecting neurons or due to a reduced strength of the cortico-spinal pathway (Busan, D'Ausilio, Borelli, Monti, Pelamatti, Pizzolato & Fadiga, 2013). The same study also highlighted that cortical silent period duration is normal in bilateral hand motor cortex in DS supporting the evidence that, when compared to fluent speakers, no differences are present in terms of intracortical inhibition in people with DS. In every case, a negative correlation between silent period duration and stuttering severity was also evident in the right hemisphere of stuttering males. Fluent speakers usually show lower motor thresholds (i.e. increased excitability) in the left hemisphere, while in DS the pattern is usually reversed: motor thresholds tend to be higher in the left hemisphere, in comparison to their own right and to the left hemisphere of fluent speakers (Alm *et al.*, 2013). Hand motor cortex in DS seems also characterized by the absence of an aberrant interhemispheric inhibition (IHI) and ipsilateral cortical silent period duration (Sommer, Knappmayer, Hunter, Gudenberg, Neef & Paulus, 2009). On the other hand, the chronic administration of paroxetine decreases TMS-evoked silent period (i.e. an index of intracortical inhibition) duration registered from right hand muscles and reduces DS associated spasms and movements (Busan *et al.*, 2009). In fluent speakers, a sub-threshold repetitive TMS (rTMS; 1 Hz for 20 min) over the left dorsolateral premotor (dPM) cortex during auditory paced finger tapping tasks, prolongs ipsilateral hand asynchrony, while right stimulation is ineffective; in DS the pattern is reversed: rTMS over the right dPM cortex increases contralateral asynchrony but no effects were present after left dPM stimulation (Neef, Jung, Rothkegel, Pollok, von Gudenberg, Paulus & Sommer, 2011a). This evidence suggests an altered control of timed non-speech movements in DS (Neef *et al.*, 2011a). On the other hand, the stimulation of cortical representations of primary motor cortex representations of tongue muscle performed during no concurrent speech tasks has pointed out the presence of alterations in motor intracortical networks. More specifically, different asymmetries are present in terms of motor thresholds: in fluent speakers, tongue motor cortex excitability is increased in the left hemisphere while in DS left motor cortex excitability is decreased and right is increased (Barwood, Murdoch, Gozee & Riek, 2013; Busan, Del Ben, Bernardini, Natarelli, Bencich, Monti, Manganotti & Battaglini, 2016). Neef, Paulus, Neef, von Gudenberg & Sommer (2011b) evaluated a series of neurophysiological indexes where the main outcome is the presence of a bilaterally reduced intracortical facilitation and a reduced short term intracortical inhibition in the right hemisphere: these findings suggest the presence of alterations in intracortical modulation of inhibitory and facilitatory circuits underlying tongue motor representations in DS. More recently, Busan *et al.* (2016) have further investigated cortico-bulbar excitability and intracortical inhibition in DS, mainly focusing on neurophysiological indexes not previously evaluated: in adults with DS, silent period threshold of the left hemisphere is higher in comparison to their own right; moreover, silent period duration is prolonged in the left hemisphere compared to the left hemisphere of fluent speakers. The pathophysiological mechanism underlying enhanced intracortical inhibition in stuttering is not clear, but a possible explanation of this pattern of findings can be the presence of an imbalance between excitatory and inhibitory inputs to the motor cortex, probably in relation with an abnormal activity of inhibiting interneurons, influencing the final level of excitability of motor cortex. The prolonged cortical silent period duration can be the result of a decrease in excitation modulated by afferent pathways to motor cortex as a result of widespread white matter abnormalities already described in the DS neural system (see Watkins *et al.*, 2008; Connally *et al.*, 2014), favouring a prolonged GABA-mediated inhibition on pyramidal cells. Moreover, stuttering severity positively correlates with silent period durations of right hand muscles, and negatively with left hand muscles: the associations of higher stuttering severity with higher intracortical inhibition in the left hemisphere and lower intracortical inhibition in the right one, also in cortical areas that are not directly involved in speech muscle control, support the idea that stuttering may be only the overt symptom of a more general motor disorder (Busan *et al.*, 2013; Busan *et al.*, 2016). Finally, TMS applied during speech tasks highlighted that, in fluent speakers, a conspicuous increase of motor cortex excitability (facilitation) is present in the left hemisphere tongue motor cortex during a speech transition phase, but not in DS (Neef, Hoang, Neef, Paulus & Sommer, 2015b). Thus, it is evident that TMS studies led to further highlight the presence of an altered functioning of the motor system in DS; however, it would be interesting to investigate this aberrant modulation of excitatory and inhibitory networks also in different populations of people who stutter, and in particular in children, in order to elucidate if such neurophysiological abnormalities are present since stuttering onset or they may be the result of compensatory mechanisms. Also electroencephalography (EEG) and magnetoencephalography (MEG) have further highlighted neural differences between DS and fluent speakers. Altered oscillations in the beta frequency band (about 13-30Hz), which are associated with motor activity, are often reported in DS: hyperactivity in the cortical beta band may be present in adults during aloud reading, but it is reduced in delayed auditory feedback conditions (Rastatter, Stuast & Kalinowski, 1998). On the contrary, reduced beta band activity may be present in children (Özge, Toros & Cömelekoglu, 2004). Again, it has been proposed that this hyperactivation in adults likely reflects a compensatory mechanism for hypoactivity in beta oscillations, starting from basal ganglia (Etchell, Johnson & Sowman, 2014). Intrahemispheric alterations in resting state are mainly present for high frequencies band (beta and gamma -i.e. > 30 Hz-): functional connectivity for high frequencies oscillations is mainly decreased, in DS, between Broca's area and right motor cortex, between right premotor cortex and left and right pars opercularis and right motor cortex, between left premotor area and Broca's area (beta), and between left motor and premotor area and Broca's area (gamma) (Joos, De Ridder, Boey & Vanneste, 2014). Finally, MEG showed in DS a relevant suppression of beta rhythms during the preparation stage of overt speech production and a consequent higher synchronization in mouth motor cortex, bilaterally (Mersov, Jobst, Cheyne & De Nil, 2016). Moreover, before stuttering occurs, the left inferior frontal and orbitofrontal cortices are less active, while an extra-activation may be present in the homologous right hemisphere regions and, bilaterally, in sensorimotor and auditory cortical regions (Sowman, Crain, Harrison & Johnson, 2012).

5. *A computational model of stuttering*

In light of the above reported evidence, functional and structural abnormalities of DS may be also verified by using computational models. In this view, a recently implemented "stuttering" version of the neuro-computational speech production model GODIVA (Gradient Order Directions Into Velocities of Articulators) (see for descriptions Bohland, Bullock & Guenther, 2010; Civier *et al.*, 2013) has been employed to test the main hypothesis of a causal disruption in DS (i.e. the basal ganglia dysfunction hypothesis and the white matter disruption hypothesis). Interestingly, in comparison with the normal performance of the "healthy" model, the GODIVA model with basal ganglia dysfunction (i.e. elevated levels of dopamine in this region) reads out the motor program for the word initial syllable with a significant neural delay. Differently, in the simulation of white matter fibre impairment, the motor program for the word second syllable is readout with delay. As a consequence, this computational simulation of DS seems to support both hypothesis and may suggest different neural substrates for different DS symptoms (e.g. blocks vs. repetitions): high levels of dopamine and basal ganglia dysfunctions are associated with stuttering occurrence especially in the first syllable of the word/sentence, whereas the white matter hypothesis may be associated with stuttering occurrence mainly in the following part of the word/utterance (Civier *et al.*, 2013). This may open the discussion to the possibility that different subgroups of stuttering may exist.

6. *Different stuttering subgroups from a neural point of view?*

From a structural and functional point of view, neuroimaging and neurophysiological studies have shown that neural differences may exist also among the stuttering population: for example, between children with DS and those who recovered, as well as between children and adults with DS (see Chang *et al.*, 2008). Interestingly, different neurophysiological profiles are present also between males and females with stuttering (e.g. Busan *et al.*, 2013; Ingham, Fox, Ingham, Xiong, Zamarripa, Hardies & Lancaster, 2004). Moreover, findings of the previously reported studies are often difficult to reproduce and sometimes discordant, especially in adults, which undergo a series of modifications and adaptations in their neural system, likely to overcome stuttering. It has been hypothesized that the population of adults with DS can be divided into subgroups also from a clinical/behavioural point of view (see Alm, 2004 for a review): for example, one subgroup may be characterized by individuals with a genetic predisposition to DS, while others may be composed of people who suffered from early neural injuries (Poulos, Webster, 1991). Another possible classification is based on the level of secondary concomitants and anxiety (e.g. higher anxiety and a higher incidence of attention deficit and hyperactivity vs. lower anxiety and higher familiar history of DS) (Alm, Risberg, 2007) Finally, evidence of different responses to pharmacological treatment may suggest a possible and further subdivision.

7. *Conclusions and future perspectives in stuttering research*

The present work is a very brief (and partial) overview on DS neurophysiology, and it has been proposed based on the currently available literature. It is evident that the exact aetiology of DS is not completely clear; however, clinical observations along with neuroimaging and neurophysiological studies have provided evidence that an abnormal functioning of the brain, and especially of the motor system, is present. The main features of DS seem to include alterations in brain regions useful to prepare, execute and control motor acts; in particular, a widespread reduced white matter integrity, an abnormal functioning of the cortico-basal-thalamo-cortical circuit, a strong activation of right hemisphere during speech, and an altered balance between excitatory and inhibitory neural signals in motor cortex have been highlighted. It is still not clear if these abnormalities are specific features of DS or the result of compensatory mechanisms due to a lifetime stuttering. In this regard, subgroups of people with DS may share different neurophysiological profiles. The future research in neurophysiology of DS should attempt to answer to every remaining question by using different techniques and different approaches (see Busan, Battaglini & Sommer, 2017), in order to define more focused and effective treatments ranging from pharmacological to neuromodulatory (see Chesters, Watkins & Möttönen, 2017) and behavioural ones (Ingham, Ingham, Euler & Neumann, *forthcoming*).

Bibliography

Alario, F.X., Chainay, H., Lehericy, S. & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. In *Brain Research*, 1076, 129-143.

Alm, P.A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. In *Journal of Communication Disorders*, 37, 325-369.

Alm, P.A., Karlsson, R., Sundberg, M. & Axelson, H.W. (2013). Hemispheric lateralization of motor thresholds in relation to stuttering. In *PLoS ONE*, 8, e76824.

Alm, P.A., Risberg, J. (2007). Stuttering in adults: the acoustic startle response, temperamental traits, and biological factors. In *Journal of Communication Disorders*, 40, 1-41.

Barnes, T.D., Wozniak, D.F, Gutierrez, J., Han, T.U., Drayna, D. & Holy, T.E. (2016). A mutation associated with stuttering alters mouse pup ultrasonic vocalizations. In *Current Biology*, 26, 1009-1018.

Barwood, C.H.S., Murdocj, B.E., Gozee, J.V. & Riek, S. (2013). Investigating the neural basis of stuttering using transcranial magnetic stimulation: preliminary case discussions. In *Speech, Language and Hearing*, 16, 18-27.

Beal, D.S., Gracco, V.L., Lafaille, S.J. & Denil, L.F. (2007). Voxel-based morphometry of auditory and speech-related cortex in stutterers. In *Neuroreport*, 18, 1257-1260.

Beal, D.S., Gracco, V.L., Brettschneider, J., Kroll, R.M. & Denil, L.F. (2013). A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. In *Cortex*, 49, 2151-2161.

Bloodstein, O. (1995). *A handbook on stuttering*. San Diego: Singular Publishing Group.

Bohland, J.W., Bullock, D. & Guenther, F.H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. In *Journal of Cognitive Neuroscience*, 22, 1504-1529.

Buchel, C., Watkins, K.E. (2010). Genetic susceptibility to persistent stuttering. In *New England Journal of Medicine*, 362, 2226-2227.

Busan, P., Battaglini, P.P., Borelli, M., Evarisot, P., Monti, F. & Pelamatti, G. (2009). Investigating the efficacy of paroxetine in developmental stuttering. In *Clinical Neuropharmacology*, 32, 183-188.

Busan, P., Battaglini, P.P. & Sommer, M. (2017). Transcranial magnetic stimulation in developmental stuttering: Relations with previous neurophysiological research and future perspectives. In *Clinical Neurophysiology*, 128, 952-964.

Busan, P., D'Ausilio, A., Borelli, M., Monti., F., Pelamatti, G., Pizzolato, G. & Fadiga, L. (2013). Motor excitability evaluation in developmental stuttering: A transcranial magnetic stimulation study. In *Cortex*, 49, 781-792.

Busan, P., Del Ben, G., Bernardini, S., Natarelli, G., Bencich, M., Monti, F., Manganotti, P. & Battaglini, P.P. (2016). Altered modulation of silent period in tongue motor cortex of persistent developmental stuttering in relation to stuttering severity. In *PLoS ONE*, 11, e0163959.

Catani, M., Dell'Acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., Valabregue, R. & Thiebaut De Schotten, M. (2012). Short frontal lobe connections of the human brain. In *Cortex*, 48, 273-291.

Chang, S.E., Erickson, K.I., Ambrose, N.G., Hasegawa-Johnson, M.A. & Ludlow, C.L. (2008). Brain anatomy differences in childhood stuttering. In *Neuroimage*, 39, 1333-1344.

Chang, S.E., Zhu, D.C, Choo, A.L. & Angstadt, M. (2015). White matter neuroanatomical differences in young children who stutter. In *Brain*, 138, 694-711.

Chesters, J., Watkins, K.E. & Möttönen, R. (2017). Investigating the feasibility of using transcranial direct current stimulation to enhance fluency in people who stutter. In *Brain and Language*, 164, 68-76.

Choo, A.L., Kraft, S.J., Olivero, W., Ambrose, N.G., Sharma, H., Chang, S.E. & Loucks, T.M. (2011). Corpus callosum differences associated with persistent stuttering in adults. In *Journal of Communication Disorders*, 44, 470-477.

Civier, O., Bullock, D., Max, L. & Guenther, F.H. (2013). Computational modelling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. In *Brain and Language*, 126, 263-278.

Connally, E.L., Ward, D., Howell, P. & Watkins, K.E. (2014). Disrupted white matter in language and motor tracts in developmental stuttering. In *Brain and Language*, 131, 25-35.

Costa, D., Kroll, R. (2000). Stuttering: An update for physicians. In *Canadian Medical Association Journal*, 162, 1849-1855.

Craig, A., Blumgart, E. & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. In *Journal of Fluency Disorders*, 34, 61-71.

Craig, A., Tran, Y. (2014). Trait and social anxiety in adults with chronic stuttering: Conclusions following meta-analysis. In *Journal of Fluency Disorders*, 40, 35-43.

Craig-McQuaide, A., Akram, H., Zrinzo, L. & Tripoliti, E. (2014). A review of brain circuitries involved in stuttering. In *Frontiers in Human Neuroscience*, 8, 884.

D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I. & Craighero, L. (2011). Tongue corticospinal modulation during attended verbal stimuli: Priming and coarticulation effects. In *Neuropsychologia*, 49, 3670-3676.

Daliri, A., Max, L. (2015). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. In *Brain and Language*, 150, 37-44.

Etchell, A.C., Johnson, B.W. & Sowman, P.F. (2014). Beta oscillations, timing, and stuttering. In *Frontiers in Human Neuroscience*, 8, 1036.

Fox, P.T., Ingham, R.J., Ingham, J.C., Hirsch, T.B., Downs, J.H., Martin, C., Jerabek, P., Glass, T. & Lancaster, J.L. (1996). A PET study of the neural systems of stuttering. In *Nature*, 382, 158-161.

Giraud, A.L., Neumann, K., Bachoud-Levi, A.C., Von Gudenberg, A.W., Euler, H.A., Lanfermann, H. & Preibisch, C. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. In *Brain and Language*, 104, 190-199.

Graybiel, A.M. (2000). The basal ganglia. In *Current Biology*, 10, R509-R511.

Ingham, R.J., Fox, P.T., Ingham, J.C., Xiong, J., Zamarripa, F., Hardies, L.J. & Lancaster, J.L. (2004). Brain correlates of stuttering and syllable production: gender comparison and replication. In *Journal of Speech, Language, and Hearing Research*, 47, 321-341.

Ingham, R.J., Grafton, S.T., Bothe, A.K. & Ingham, J.C. (2012). Brain activity in adults who stutter: similarities across speaking tasks and correlations with stuttering frequency and speaking rate. In *Brain and Language*, 122, 11-24.

Ingham, R.J., Ingham, J.C., Euler, H.A. & Neumann, K. (forthcoming). Stuttering treatment and brain research in adults: A still unfolding relationship. In *Journal of Fluency Disorders*.

Iverach, L., Rapee, R.M. (2014) Social anxiety disorder and stuttering: Current status and future directions. In *Journal of Fluency Disorders*, 40, 69-82.

Jäncke, L., Hänggi, J. & Steinmetx, H. (2004) Morphological brain differences between adult stutterers and non-stutterers. In *BMC Neurology*, 4, 23.

Joss, K., De Ridder, D., Boey, R.A. & Vanneste, S. (2014). Functional connectivity changes in adults with developmental stuttering: A preliminary study using quantitative electroencephalography. In *Frontiers in Human Neuroscience*, 8, 783.

Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J.C. & Drayna, D. (2010). Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. In *New England Journal of Medicine*, 362, 677-685.

Kell, C.A., Neumann, K., Von Kreigstein, K., Poserienske, K., Von Gudenberg, A.W. & Euler, H. (2009). How the brain repairs stuttering. In *Brain*, 132, 2747-2760.

Kemerdere, R., De Champfleur, N.M., Deverdun, J., Cochereau, J., Mortiz-Gasser, S., Herbet, G. & Duffau, H. (2016). Role of the left frontal aslant tract in stuttering: A brain stimulation and tractographic study. In *Journal of Neurology*, 263, 157-167.

Kobayashi, M., Pascual-Leone, A. (2003). Transcranial magnetic stimulation in neurology. In *Lancet Neurology*, 2, 145-156.

Kronfeld-Duenias, V., Amir, O., Ezrati-Vinacour, R., Civier, O. & Ben-Shachar, M. (2016). The frontal aslant tract underlies speech fluency in persistent developmental stuttering. In *Brain Structure and Function*, 221, 365-381.

Lu, C., Chen, C., Ning, N., Ding, G., Guo, T., Peng, D., Yang, Y., Li, K. & Lin, C. (2010). The neural substrates for atypical planning and execution of word production in stuttering. In *Experimental Neurology*, 221, 146-156.

Maguire, G.A., Riley, G.D., Franklin, D.L. & Gottschalk, L.A. (2000). Risperidone for the treatment of stuttering. In *Journal of Clinical Psychopharmacology*, 20, 479-482.

Maguire, G.A., Riley, G.D., Franklin, D.L., Maguire, M.E., Nguyen, C.T. & Brojeni, P.H. (2004). Olanzapine in the treatment of developmental stuttering: A double-blind, placebo-controlled trial. In *Annals of Clinical Psychiatry*, 16, 63-67.

Mersov, A.M., Jobst, C., Cheyne, D.O. & De Nil, L. (2016). Sensorimotor oscillations prior to speech onset reflect altered motor networks in adults who stutter. In *Frontiers in Human Neuroscience*, 10, 443.

Murray, T.J., Kelly, P., Campbell, L. & Stefanik, K. (1977). Haloperidol in the treatment of stuttering. In *The British Journal of Psychiatry*, 130, 370-373.

Neef, N.E., Anwarder, A. & Friederici, A.D. (2015). The Neurobiological Grounding of Persistent Stuttering: from Structure to Function. In *Current Neurology and Neuroscience Report*, 15, 63.

NEEF, N.E., HOANG, T.N., NEEF, A., PAULUS, W. & SOOMER, 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.

NEEF, N.E., JUNG, K., ROTHKEGEL, H., POLLOK, B., VON GUDENBERG, A.W., PAULUS, W. & SOMMER, M. (2011). Right-shift for non-speech motor processing in adults who stutter. In *Cortex*, 47, 945-954.

NEEF, N.E., PAULUS, W., NEEF, A., VON GUDENBERG, A.W. & SOMMER, M. (2011). Reduced intracortical inhibition and facilitation in the primary motor tongue representation of adults who stutter. In *Clinical Neurophysiology*, 122, 1802-1811.

OJEMANN, G.A., WARD, A.A. JR. (1971). Speech representation in ventrolateral thalamus. In *Brain*, 4, 669-680.

ÖZGE, A., TOROS, F. & CÖMELEKOGLU, U. (2004). The role of hemispheral asymmetry and regional activity of quantitative EEG in children with stuttering. In *Children Psychiatry and Human Development*, 34, 269-280.

PENFIELD, W., WELCH, K. (1951). The supplementary motor area of the cerebral cortex: A clinical and experimental study. In *AMA Archives of Neurology and Psychiatry*, 66, 289-317.

POULOS, M.G., WEBSTER, W.G. (1991). Family history as a basis for subgrouping people who stutter. In *Journal of Speech and Hearing Research*, 34, 5-10.

PREIBISCH, C., NEUMANN, K., RAAB, P., EULER, H.A., VON GUDENBERG, A.W., LANFERMANN, H. & GIRAUD, A.L. (2003). Evidence for compensation for stuttering by the right frontal operculum. In *Neuroimage*, 20, 1356-1364.

RASTATTER, M., STUAST, A. & KALINOWSKI, J. (1998). Quantitative electroencephalogram of posterior cortical areas of fluent and stuttering participants during reading with normal and altered auditory feedback. In *Perceptual and Motor Skills*, 87, 623-633.

RIVA-POSSE, P., BUSTO-MAROLT, L., SCHTEINSCHNAIDER, A., MARTINEZ-ECHENIQUE, L., CAMMAROTA, A. & MERELLO, L. (2008). Phenomenology of abnormal movements in stuttering. In *Parkinsonism and Related Disorders*, 14, 415-419.

SCHREIBER, S., PICK, C.G. (1997). Paroxetine for secondary stuttering: Further interaction of serotonin and dopamine. In *Journal of Nervous and Mental Disease*, 185, 465-467.

SOMMER, M., KNAPPMEYER, K., HUNTER, E.J., GUDENBERG, A.W., NEEF, N. & PAULUS, W. (2009). Normal interhemispheric inhibition in persistent developmental stuttering. In *Movement Disorders*, 24, 769-773.

SOMMER, M., KOCH, M.A., PAULUS, W., WEILLER, C. & BUCHEL, C. (2002). Disconnection of speech relevant brain areas in persistent developmental stuttering. In *Lancet*, 360, 380-383.

SOMMER, M., WISCHER, S., TERGAU, F. & PAULUS, W. (2003). Normal intracortical excitability in developmental stuttering. In *Movement Disorders*, 18, 826-830.

SONG, L.P., PENG, D.L., JIN, Z., YAO, L., NING, N., GUO, X.J. & ZHANG, T. (2007). Gray matter abnormalities in developmental stuttering determined with voxel-based morphometry. In *National Medical Journal of China*, 87, 2884-2888.

SOWMAN, P.F., CRAIN, S., HARRISON, E. & JOHNSON, B.W. (2012). Reduced activation of left orbitofrontal cortex precedes blocked vocalization: A magnetoencephalographic study. In *Journal of Fluency Disorders*, 37, 359-365.

Travis, L.E. (1978). The cerebral dominance theory of stuttering: 1931-1978. In *Journal of Speech and Hearing Disorders*, 43, 278-281.

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, 50-59.

Wu, J.C., Maguire, G., Riley, G., Fallon, J., Lacasse, L., Chin, S., Klein, E., Tang, C., Cadwell, S. & Lottenberg, S. (1995). A positron emission tomography [18F]deoxyglucose study of developmental stuttering. In *Neuroreport*, 6, 501-505.

Wu, J.C., Maguire, G., Riley, G., Lee, A., Keator, D., Tang, C., Fallon, J. & Najafi, A. (1997). Increased dopamine activity associated with stuttering. In *Neuroreport*, 8, 767-770.

Yairi, E., Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. In *Journal of Fluency Disorders*, 38, 66-87.

PARTE V

LAVORI IN CORSO

DOI: 10.17469/O2103AISV000022

FRANCESCO AVANZINI, PIERO COSI, ROLANDO FÜSTÖS,
ANDREA SANDI

When fantasy meets science: An attempt to recreate the voice of Ötzi the "Iceman"

Ötzi the Iceman's, the mummy found some years ago on the top of the Similaun mountain, lived 5300 years ago, and obviously, we cannot say which language, which phonemes, or even which sounds could belong to him. In this work, still in its initial stage, a possible reconstruction of an "approximation" of Ötzi the Iceman's voice, recreating the timbre or colour of his stone-age possible vowels will be described.

Key words: Ötzi, Vocal Tract, CT Scan.

1. *Introduction*

Ötzi (*the Iceman*) was found by two German hikers in 1991, frozen and mummified in Schnalstal glacier, Ötztal Alps, near Hauslabjoch in South Tyrol on the border between Austria and Italy.

Ötzi and his artefacts have been exhibited at the South Tyrol Museum of Archaeology in Bolzano, Italy since 1998 (see Figure 1).

Radiocarbon dating gives ages between 5300 and 5200 years, placing it in the Age of copper, the moment of transition between the Neolithic and Bronze Age. So Ötzi is an ancient mummified specimen of homo sapiens and is Europe's oldest known natural mummy, providing researchers with an interesting picture describing what life was like around 3300 BC.

There are particular difficulties working with a 5300-year-old, especially conserved body – by far the oldest mummified person ever found; those of ancient Egypt are at least 1,000 years younger. Nonetheless, during these last years, scientists have discovered many interesting things about the mummy. The discovery made by X-rays and CT scans that Ötzi had an arrowhead lodged in his left shoulder led researchers think that he died of blood loss from the wound (Jha, 2007). Further research found bruises and cuts to the hands, wrists and chest, and cerebral trauma suggested a blow to the head. At present, researchers believe that death could have been also caused by a blow to the head but it is difficult to say if this was due to a fall, or from being struck with a rock by another person (Carrol, 2002).

Figure 1 - *Ötzi (the iceman), the mummy kept in the Archaeological Museum of Alto Adige of Bolzano*

Ötzi exhibits the oldest preserved tattoos in the world. These 61 tattoos are concentrated in areas where bones X-ray examination showed "age-conditioned or strain-induced degeneration", thus led to hypothesise that these tattoos may have been related to pain relief treatments similar to acupuncture (Singh, Ernst, 2008), and not to spiritual or magical meanings. If so, this is at least 2000 years before their previously known earliest use in China (c. 1000 BC).

Modern DNA sequence analysis (Palmer, 2012) revealed that Ötzi had brown eyes, blood type 0, was lactose intolerant, and was likely to suffer heart disease. He was closely related to modern Corsicans and Sardinians and he was also the first known case of a person infected by the Lyme disease bacterium. He is middle-aged.

Summarizing in other words, thanks to him, experts were able to give us a quite complete picture of his past, and, by reconstructing his voice, we hope to gain more insight into what humans might have sounded like back then.

Obviously, we cannot say that we could reconstruct Ötzi's original voice because we miss too many crucial information from the mummy. However, with two reconstructed but true measurements, the length of both the vocal tract and the vocal cords, we have been able to recreate with a "high degree of fantasy", a hypothesized approximation of the mummy's vocal tract structure and consequently a hypothesized approximation of its voice.

2. *Vocal tract reconstruction*

We had to face several challenges as we worked to reconstruct the 5300-year-old mummy's vocal tract configuration.

We had to deal with Ötzi's position, which makes the CT scanning quite difficult. Ötzi's arm is covering his neck and throat, and, for our project, this is the worst position you can imagine. Moreover, the hyoid bone, or tongue-bone, was partially absorbed and dislocated.

Considering the scan of similar human body as a reference, and with the help of a special dedicated software, available at SINTAC Biomedical Engineering in Padova[1], we were able to virtually move Ötzi's arm from his position, reposition his skull in the erect position, reconstruct his vertebrae, from the first one (C1) closest to the skull to the first thoracic vertebra (T1), and reconstruct and reposition the hyoid bone.

At last, as illustrated in Figure 2, we ended up with a complete model of the vocal tract (Vocal Tract Lenght = 127.3mm, Vocal Chord Length = 20.47), but still we underline again that we were missing important data such as the tension and density of the vocal cords or the thickness and composition of the soft tissues that take an important role in the human voice.

MRI (magnetic resonance imaging) scans would have helped us getting more insights, but this technology could not be used because of the condition of Ötzi's mummified body. Thus, we had only to rely on mathematical models and a software that simulates the way the vocal tract works.

Taking into consideration that, despite its short slender body, Ötzi's had a rather large head, it was possible to hypothesize that his voice could probably had a fundamental frequency between 100 Hz and 150 Hz, in line with today's average male (voices).

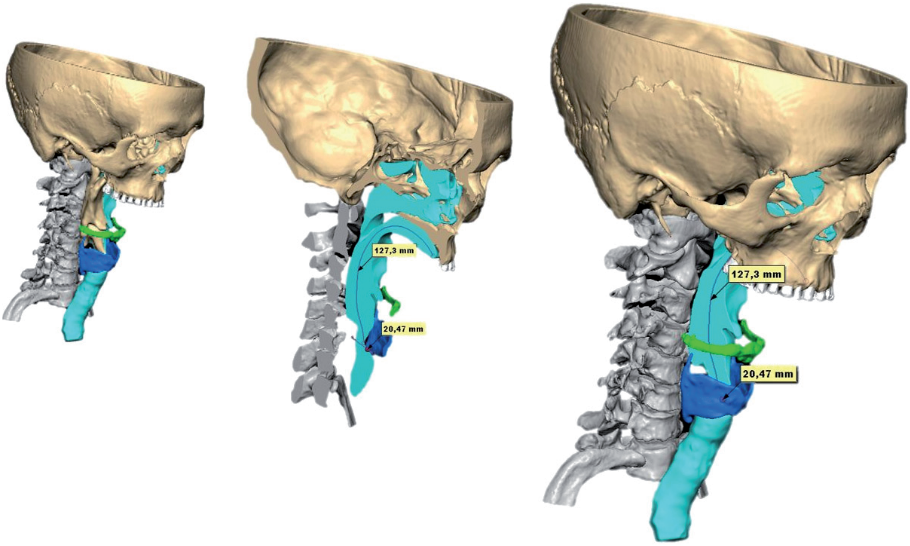
With only these simple measurements and this hypothesized f0 finally, with the help of a specialized software developed by Peter Birkholz of the Institute of Acoustics and Speech Communication of the Technische Universität of Dresden (Birkholz, Lehnert & Neuschaefer-Rube, 2009), we let Ötzi speak with few vowel like sounds and few simple words/sentences.

3. *From vocal tract to speech sounds*

Acoustic tube models of speech production have been heavily studied (Chiba, Kajiyama, 1941; Dunn, 1950; Stevens, Kasowski & Fant, 1953; Fant, 1959, 1960; Kelly, Lochbaum, 1962; Flanagan, Rabiner, 1973; Tousignant, Lefevre & Lecours, 1979). In these studies, it was shown that from a given tube shape, the resonance frequencies could be obtained, and the inverse problem of determining a unique tube shape from resonance characteristics has also received considerable attention in many other studies (Schroeder, Mermelstein, 1965; Mermelstein, 1967; Schroeder, 1967; Heinz, 1967; Sondhi, Gopinath, 1971).

[1] SINTAC Biomedical Engineering, Padova, http://www.sintac.it/.

Figure 2 - *Ötzi's vocal tract reconstructed model: Vocal Tract Lenght = 127.3mm, Vocal Chord Length = 20.47*



From Linear Prediction of Speech Model theory (Markel, Gray, 1976), it is well known that the vocal tract could be modeled as a succession of elementary cylindrical tubes, and the corresponding area function can be represented by reflection coefficients between adjacent sections of the tube. This LP model allows the parameters of these acoustic tube models to be estimated directly from the acoustical speech waveform.

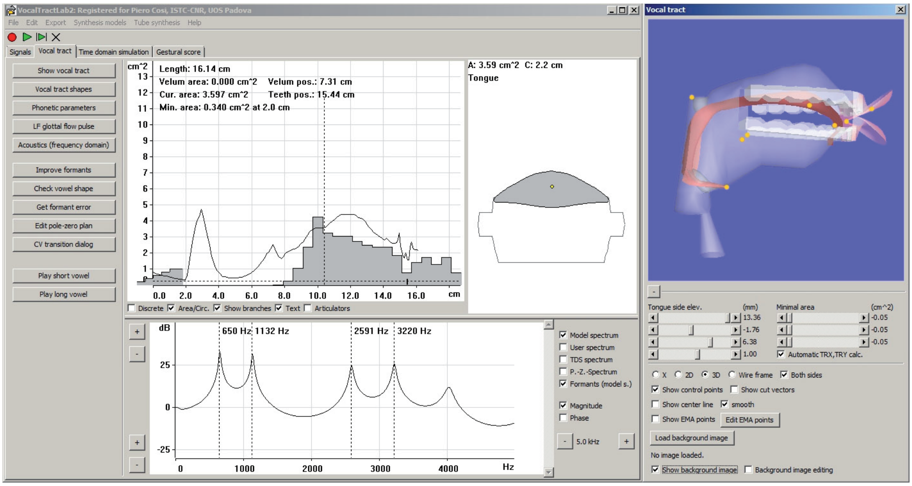
The first attempt at directly computing an acoustic tube model of the vocal tract from the speech waveform was due to Atal (Atal, 1970) who demonstrated that the formant frequencies and bandwidths are sufficient to uniquely determine the areas of an acoustic tube having a specified number of sections. He also demonstrated that a transfer function with M poles is always realizable as the transfer function of an acoustic tube consisting of M cylindrical sections of equal length (Atal, Hanauer, 1971). Thus, a unique discrete tube shape can be reconstructed from a given-order polynomial transfer function. Wakita (1972) showed that the same acoustic tube model is equivalently represented from the inverse filter A(z) obtained by linear prediction of the acoustical speech waveforms. He also demonstrated the important experimental result that if the speech is properly pre-emphasized, and if boundary conditions of the acoustic tube are properly chosen, then very reasonable vocal tract shapes can be directly estimated using the autocorrelation method of linear prediction.

All these findings are the basis for the VocalTractLab (Birkholz *et al.*, 2009) open source software[2].

[2] http://www.vocaltractlab.de/.

The central element of VocalTractLab is a three-dimensional model of the human vocal tract, which represents the surfaces of the articulators and the vocal tract walls. An interactive visualization of the model is shown in Figure 3.

Figure 3 - *Visualization of a screenshot of the VocalTractLab software*



The shape and/or position of the articulators is defined by a number of vocal tract parameters, which can be changed interactively by dragging some control points (see the dots in the vocal tract picture in Figure 3). The pictures on the right side of Figure 3 display the cross section through the vocal tract at a selected position along the centerline (top), and the vocal tract area function (bottom). At the bottom of the screen is a graph of the volume velocity transfer function.

All possible articulators parameters are easily set by the user but, for Ötzi, we could have only two of them: the length of the vocal tract and the length of the vocal chords. The voiced sound corresponding to the adjusted vocal tract shape can be thus mathematically reconstructed, by using the above-described knowledge, and played back to the user.

4. *Final considerations*

We would like again to repeat that, obviously, we could not say that we can reconstruct Ötzi's original voice because we miss too many crucial information from the mummy. However, with two true measurements, the length of both the vocal tract and the vocal cords, we have been able to recreate with a "high degree of fantasy", a hypothesized approximation of the mummy's vocal tract structure and consequently a hypothesized approximation of its voice.

Even knowing perfectly the Ötzi's respiratory capacity, his vocal chords dimensions and his vocal tract perfect configuration measurements, still, there will be too many unknown factors, such as tissue characteristics to cite one, that make impossible a true reconstruction of its voice.

With all these limitations in mind, we wanted only to explore with the help of a lot of fantasy, but also "a certain degree" of science what could have been the old voice of Ötzi.

Acknowledgements

We would like to strongly acknowledge EURAC (the Institute for Mummies and the Iceman), and the South Tyrol Museum of Archaeology of Bolzano, for the courtesy to make us working on their Ötzi's CT scannings, and Peter Birkholz of the Institute of Acoustics and Speech Communication of the Technische Universität of Dresden, for his kindness and patience, in explaining us the functioning of the VocalTractLab open source software, and for letting us use it, even for the complete speech synthesis stage.

Bibliography

Atal, B.S. (1970). Determination of the vocal tract shape directly from the speech wave. In *Journal of the Acoustical Society of America*, 47, 65(A).

Atal, B.S., Hanauer, S.L. (1971). Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. In *Journal of the Acoustical Society of America*, 50, 637-655.

Birkholz, P., Lehnert, B. & Neuschaefer-Rube, C. (2009). VocalTractLab – Ein neues Softwaretool für die artikulatorische Sprachsynthese in der Lehre. In *Proceedings of the 26th Jahrestagung der DGPP*. Leipzig, Germany, 209-211.

Carroll, R. (2002). How Oetzi the Iceman was stabbed in the back and lost his fight for life. *The Guardian*. https://www.theguardian.com/world/2002/mar/21/humanities.research1.

Chiba, T., Kajiyama, M. (1941). *The vowel, its nature and structure*. Tokyo: Tokyo Kaiseikan Publisher Company.

Dunn, H.K. (1950). The calculation of vowel resonances, and an electrical vocal tract. In *Journal of the Acoustical Society of America*, 22, 740-753.

Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to Swedish. Stockholm: Ericsson.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton De Gruyter.

Flanagan, L., Rabiner, L.R. (1973). *Speech Synthesis.* Stroudsburg, Pennsylvania: Dowden, Hutchinson & Ross.

Jha, A. (2007). Iceman bled to death, scientists say. In *The Guardian*, https://www.theguardian.com/science/2007/jun/07/archaeology.internationalnews.

Heinz, M. (1967). Perturbation functions for the determination of vocal tract area functions from vocal tract eigenvalues. In *Speech Transmission Laboratory Quarterly Progress and Status Report*, 8, 1-14.

KELLY J.R., L., LOCHBAUM, C. (1962). Speech synthesis. In *Proceedings of the Stockholm Speech Communication Seminar*. Reprinted in FLANAGAN, J.L., RABINER, L.R. (Eds.) (1973), *Speech Synthesis*. Stroudsburg, Pennsylvania: Dowden, Hutchinson & Ross, 127-130.

MARKEL, J.D., GRAY, A.H. JR. (1976). *Linear prediction of speech*. Berlin-New York: Springer-Verlag.

MERMELSTEIN, P. (1967). Determination of the vocal-tract shape from measured formant frequencies. In *Journal of the Acoustical Society of America*, 41, 1283-1294.

PALMER, J. (2012). *Oetzi the Iceman's nuclear genome gives new insights*. BBC News: Science & Environment. http://www.bbc.com/news/science-environment-17191398.

SCHROEDER, M.R (1967). Determination of the Geometry of the Human Vocal Tract by Acoustic Measurements. In *Journal of the Acoustical Society of the America*, 41, 1002-1010.

SCHROEDER, M.R., MERMELSTEIN, P. (1965). Determination of smoothed cross-sectional area function of the vocal tract from formant frequencies. In *The Journal of the Acoustical Society of America*, 37, 1186.

SINGH, S., ERNST, E. (2008). *Aghi, pozioni e massaggi. La verità sulla medicina alternativa*. Milano: Rizzoli.

SONDHI, M.M., GOPINATH, B. (1971). Determination of vocal tract shape from impulse response at lips. In *Journal of the Acoustical Society of America*, 49, 1867-1873.

STEVENS, K.N., KASOWSKI, S. & FANT, C.G.M. (1953). An electrical analog of the vocal tract. In *Journal of the Acoustical Society of America*, 25, 734-742.

TOUSIGNANT, B., LEFEVRE, J-P. & LECOURS, M. (1979). Speech synthesis from vocal tract area function acoustical measurements. In *Proceedings of the International Conference on Acoustical Speech and Signal Processes*. Washington, D.C.: USA.

WAKITA, H. (1972). *Estimation of the vocal tract shape by optimal inverse filtering and acoustic/articulatory conversion methods (Tech. Rep. No9)*. Santa Barbara, California: Speech Communication Research Laboratory.

DOI: 10.17469/O2103AISV000023

SONIA CENCESCHI, ROBERTO TEDESCO, LICIA SBATTELLA

Verso il riconoscimento automatico della prosodia

This paper presents our approach to automatic recognition of prosodic forms. In particular, we present: CALLIOPE, a multi-dimensional model aiming at categorizing all prosodic forms; SI-CALLIOPE, a sub-space for which we defined a corpus of recorded prosodic forms; and the psychoacoustic experiment we are currently planning for investigating main acoustic behaviours and features involved into the discrimination of prosodic forms. The results of the experiment will be useful for defining the acoustic/textual features to rely on for automatic recognition of prosodic forms. For that reason, we are also defining a classifier, based on Neural Nets. This study is part of the LYV project, which focuses on improving prosodic expressiveness skills of Italian speakers with autism and other cognitive disabilities.

Key words: prosody, human-computer interaction, paralinguistics, Neural Networks.

1. *Introduzione*

In questo articolo presentiamo una proposta di approccio al riconoscimento automatico di forme prosodiche. In particolare, verranno descritti:
- CALLIOPE: un modello multidimensionale che ha lo scopo di catalogare tutte le possibili forme prosodiche pronunciabili da un parlante generico.
- SI-CALLIOPE: un sottospazio di CALLIOPE, per il quale definiamo un corpus di registrazioni audio.
- Gli esperimenti psicoacustici necessari per indagarne i principali comportamenti acustici.
- Il modello di un classificatore, basato su Reti Neurali, per riconoscere automaticamente alcune forme prosodiche.

Questo studio è parte del progetto LYV[1] (Lend Your Voice), incentrato sul miglioramento delle capacità prosodiche ed espressive di parlanti italiani con disabilità cognitive, tramite l'utilizzo della tecnologia e in contesti complessi (Sbattella, 2007).

2. *Il modello CALLIOPE*

CALLIOPE è uno spazio multidimensionale, dove ogni dimensione rappresenta un fattore che influenza l'interpretazione della singola Unità Informativa (Cresti, 2000). Ogni dimensione *i* è rappresentata da una variabile qualitativa l_i, che assume

[1] LYV è un progetto Polisocial award 2016-2017, http://www.polisocial.polimi.it.

valore all'interno di un insieme F_i; ogni UI è quindi associata a un punto dello spazio. Più formalmente, una generica UI è associata ad una tupla T(UI) composta di dodici etichette:

$$T(UI) = (l_1, l_2, ..., l_{12}) : l_i \in F_i\ , 1 \leq i \leq 12$$

dove l_i è la *i*-esima etichetta scelta nel corrispondente insieme F_i. Diverse UI possono essere associate alla stessa T(UI), ma ogni UI è univocamente associata ad una e una sola T(UI).

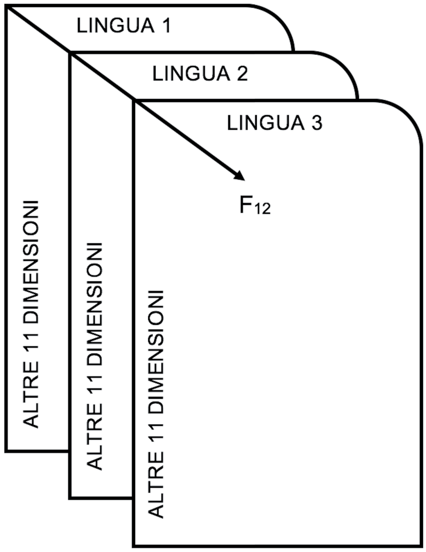
Le dimensioni di CALLIOPE sono divise *Dialogic Dimensions* (caratteristiche correlate al contesto comunicativo) e *Background Dimensions* (caratteristiche che esistono a prescindere dalla presenza di interazione tra individui).

Le Dialogic Dimensions, da F_1 a F_9, sono: *Struttura*, *Modalità Linguistica*, *Focus Intonativo*, *Forma Retorica*, *Stato Motivazionale*, *Speech Mood*, *Spontaneità, Forme di Punteggiatura,* ed *Emozioni*. La *Struttura* comprende le costruzioni verbali (Prieto, Borràs-Comes & Roseano, 2010-2014) composte da singoli elementi grammaticali e linguistici (un esempio è la forma della domanda diretta). La *Modalità* rappresenta l'intenzione comunicativa scelta dal parlante per ottenere un particolare effetto sull'interlocutore (Kratzer, 2012). Il *Focus Intonativo* cataloga le diverse motivazioni per cui una parola può essere enfatizzata rispetto alle altre all'interno del dialogo (Ouyang, Kaiser, 2012), mentre la *Forma Retorica* comprende solo le figure retoriche che influenzano la componente sonora della UI. Lo *Stato Motivazionale* (Liotti, Monticelli, 2008) indica come il suono delle UI è in relazione con il ruolo del parlante nel contesto del dialogo: la stessa frase pronunciata da un giudice o da una madre può "suonare" evidentemente differente a causa del diverso rapporto con l'interlocutore. Il *Mood* si riferisce all'intensità applicabile alle UI in un dato contesto: la prosodia di una stessa UI cambia integralmente se sussurrata o urlata. La *Spontaneità* è relativa al "livello di improvvisazione" che il parlante può, o deve, adottare in una data situazione (Nencioni, 1983): parlato letto, recitato, spontaneo assumono prosodie differenti per una stessa UI. Le *Forme di Punteggiatura* sono qui da considerarsi come un indice della presenza di pause all'interno di una UI e sono indicate, in particolare, dalla presenza di virgole. Talvolta la virgola è utilizzata come separatore di UIs, ma in questo caso consideriamo solamente un suo utilizzo interno alla singola unità intonativa. Le *Emozioni*, infine, sono un'evidente causa di modifica della prosodia di una UI (Cowie, Cornelius, 2003; Tomkins, 1984). CALLIOPE propone, per ognuno degli insiemi da F_1 a F_9, un numero finito di etichette.

Le Background Dimensions sono F_{10}, F_{11} e F_{12}: *Capacità Espressive Soggettive*, *Contesto Sociale* e *Linguaggio, Dialetto o Variazione Linguistica Locale*. La *Capacità Espressiva* può influenzare una UI modificando il modello necessario per una corretta trasmissione del messaggio. *Contesto Sociale*, e soprattutto *Linguaggio, Dialetto o Variazione Linguistica Locale*, modificano in modo non banale la prosodia dei parlanti. CALLIOPE propone, per ognuno degli insiemi F_{10}, F_{11} e F_{12}, alcune etichette che però non intendono essere esaustive.

Ciascuna etichetta in F_{12} è da immaginarsi come un *sottospazio*[2]: un livello all'interno del quale attribuire le restanti dimensioni (Figura 1). Uno di tali sottospazi è l'Italiano Standard, che sarà utilizzato nel seguito.

Figura 1 - *Diversi linguaggi*



3. *SI-CALLIOPE e validazione*

La validazione dell'intero modello CALLIOPE è molto complessa. Abbiamo quindi deciso di focalizzarci su un sotto-spazio, denominato SI-CALLIOPE (*Standard Italian* CALLIOPE). Tale modello è utile per gli scopi del progetto LYV, che l'utilizzerà l'Italiano Standard (F_{12}) come descritto da Canepari (1986), prendendo come modello la voce di parlanti normodotati (F_{10}) simulando situazioni prosodiche tipiche della quotidianità (F_{11}). L'analisi è stata poi ulteriormente ristretta a frasi non retoriche e con sospensione (F_4) e parlato recitato (F_7).

3.1 Il Corpus Prosodico: un sottospazio di SI-CALLIOPE

Grazie ai volontari di Libro Parlato Onlus[3] e ad alcuni attori professionisti, abbiamo registrato un corpus contenente una lista di frasi e pseudo-frasi. Quando l'interpretazione non era chiara, abbiamo fornito ai parlanti dei suggerimenti o una descrizione del contesto. Le UI sono così suddivise:

- F_1, Struttura:

[2] A rigore, ogni dimensione di CALLIOPE definisce un insieme di sottospazi. Tuttavia, poiché si immagina che tipicamente il modello sarà applicato a un linguaggio alla volta, visualizzare lo spazio di CALLIOPE come indicato in Figura 2 permette di comprenderne meglio la struttura senza perdita di generalità.

[3] Libro Parlato Onlus, Centro Internazionale del Libro Parlato (CILP), Feltre, Italy, www.libroparlato.org.

1. Dichiarativa
2. Interrogativa ad 1 unità tonale
3. Interrogativa a 2 o più unità tonali
4. Interrogativa disgiuntiva
5. Domanda eco
6. Esclamativa
7. Vocativo

– F_6, Speech mood:
8. Sospirato
9. Urlato
10. Standard

– F_3, Focus intonativo
11. Focus contrastivo

– F_4, Forma retorica
12. Sospensione

– F_8, Forma di punteggiatura
13. Lista

I parlanti sono 14, 7 uomini e 7 donne, con età compresa tra 33 e 48 anni. Ogni parlante ha registrato circa 1 ora di parlato, 278 UI (139 con significato, 139 pseudo-frasi) per un totale di 1946 frasi con significato e 1946 pseudo-frasi. I file audio sono stati registrati, con differenti modalità e microfoni, in formato WAV (44.1 kHz), per ottenere un modello quanto più possibile indipendente dal mezzo tecnico utilizzato.

3.2 Generazione delle pseudo-frasi

La fase sperimentale richiede la generazione e la registrazione di pseudo-frasi: frasi composte da parole senza significato, ma che rispettano la fonotassi della lingua Italiana e che quindi "suonano" come italiane. Siamo partiti dal corpus di parole italiane CoLFIS[4], dal quale sono state rimosse le parole contenenti caratteri nell'insieme {'w', 'y', 'j', 'k', 'x'}, e le parole contenenti segni diacritici diversi dall'accento grave o acuto. Le parole così rimaste sono state divise in sillabe tramite Hyphenator 0.5.1, un modulo Python che sfrutta la sillabazione del dizionario fornito da OpenOffice. Quindi, è stato addestrato un trigramma di sillabe che codifica un'approssimazione della fonotassi italiana. Il trigramma utilizzato dal nostro generatore è definito come la seguente distribuzione di probabilità condizionata:

$$P\,(s_i \mid s_{i\text{-}1}, s_{i\text{-}2}) : s_i \in \mathrm{S}$$

dove s_i è la sillaba i-esima della parola da trasformare in pseudo-parola, e S è l'insieme delle sillabe ricavate dall'analisi del corpus CoLFIS.

[4] Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS), http://linguistica.sns.it/CoLFIS/Home.htm.

L'algoritmo parte da una frase italiana e genera, per ogni parola più lunga di 3 caratteri, una pseudo-frase con lo stesso numero di sillabe. In particolare, partendo da una parola composta da *n* sillabe, l'algoritmo sceglie *n* sillabe dall'insieme S, tramite *n* campionamenti pseudo-casuali della distribuzione di probabilità del trigramma[5], e le accosta consecutivamente, costruendo la corrispondente pseudo-parola. Per migliorare la leggibilità delle frasi ottenute, le parole più corte di 4 caratteri (per lo più articoli, preposizioni e alcune forme del verbo "essere") non sono modificate e vengono semplicemente copiate nella pseudo-frase. L'algoritmo applica quindi alle pseudo-parole un insieme di regole di concordanza che migliorano la leggibilità globale, poi ulteriormente raffinata tramite un ultimo controllo manuale.

4. *Esperimento psicoacustico*

Il test per l'esperimento psicoacustico è disponibile su web, all'indirizzo http://calliope.deib.polimi.it. Lo scopo principale dell'esperimento è comprendere quali informazioni utilizza il nostro cervello per la comprensione della componente prosodica:

– Quali forme prosodiche sono riconosciute grazie alla mera componente acustica del parlato?
– Quali forme prosodiche sono riconosciute grazie alla componente acustica e alla fonotassi?
– Quali forme prosodiche necessitano anche del significato della frase?

Il test è composto da 36 quesiti, suddivisi in tre sezioni, per ognuno dei quali un audio è estratto, con tecniche di randomizzazione, dal corpus. L'ascoltatore deve ascoltare e riconoscere la forma prosodica. Vi sono quindi:

– 13 frasi con significato
– 13 pseudo-frasi
– 10 frasi ridotte al solo inviluppo del pitch[6]

La terza sezione manca dei tre quesiti riguardanti la percezione dell'intensità, perché essa è impossibile da rilevare utilizzando il solo pitch. Ogni quesito comprende tre audio e l'ascoltatore deve segnalare, per ognuno di essi, se riconosce o no la forma prosodica suggerita dal quesito stesso (per esempio "Quali tra gli audio seguenti è una domanda diretta?"). Per ogni quesito, il sistema inserisce m_{yes} audio che l'ascoltatore dovrebbe riconoscere e $m_{no} = 3-m_{yes}$ audio che l'ascoltatore dovrebbe scartare, dove $m_{yes} \in \{1,2,3\}$ è scelto a caso.

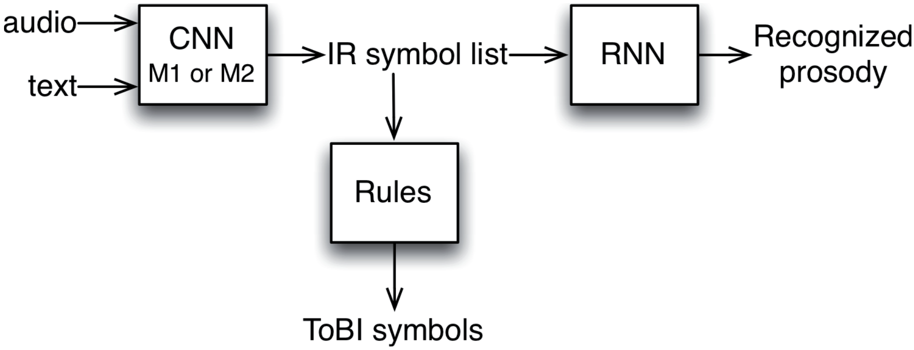
[5] Ciò significa che la probabilità di scegliere una cerca sillaba, in posizione *i*-esima, dipende dalle sillabe in precedenza scelte per le posizioni $i-1$ e $i-2$. È questa dipendenza che permette di approssimare la fonotassi.

[6] Per la generazione di questa versione, abbiamo utilizzato la funzione Hum del programma Praat, che genera un suono simile a una vocale (suona come una sorta di "a"), secondo l'andamento voluto del pitch.

5. *Il classificatore basato su Reti Neurali*

L'architettura generale del riconoscitore è mostrata in Figura 2.

Figura 2 - *Schema di massima del riconoscitore*



Una Convolutionary Neural Network (CNN) sarà utilizzata per trovare pattern prosodici, che saranno trasformati in una sequenza di simboli. Questa rappresentazione intermedia (Intermediate Representation – IR) è utile perché più astratta e informativa rispetto ai dati audio reali in input. Stiamo al momento definendo questi simboli, che potranno essere un'integrazione del sistema di taggatura ToBI (Beckman, Hirschberg & Shattuck-Hufnagel, 2005), ma in una versione più raffinata che possa descrivere in modo più approfondito il modello prosodico.

Una Recurrent Neural network (RNN) riconoscerà la sequenza di simboli classificandoli come una delle forme prosodiche scelte. Dalla sequenza di simboli IR, un modello a regole genererà le corrispondenti etichette ToBI. La CNN è in realtà composta da due modelli separati: il più semplice, M1, considera solo feature audio e ha lo scopo di riconoscere forme prosodiche dove le caratteristiche sonore sono condizione sufficiente per il riconoscimento. Il secondo, M2, necessita anche di feature testuali e ha lo scopo di riconoscere un più largo insieme di forme prosodiche. M1 è indicato per situazioni in cui il testo non è disponibile o l'allineamento testo-audio non è possibile. In entrambi i casi non possono essere utilizzate feature testuali.

Sia la CNN che la RNN saranno addestrate utilizzando il corpus audio derivato da SI-CALLIOPE. Il modello che genererà i simboli ToBI a partire da quelli IR sarà composto da regole definite manualmente e non necessiterà di una fase di addestramento.

6. *Conclusioni*

Questo lavoro è una proposta che potrà essere utilizzata per il riconoscimento automatico della prosodia vocale in diversi campi. Nonostante si sia deciso di partire da una serie di frasi utili nell'ambito del progetto LYV, nulla vieta che la ricerca possa essere ampliata, seguendo la stessa metodologia, ad ulteriori forme prosodiche e applicata ad ambiti di ricerca. Potrebbe risultare utile, per esempio, come modello di ricerche inerenti altre lingue, o come guida per pianificare ulteriori studi sperimentali. Il corpus può essere ampliato con facilità, aggiungendo nuovi esempi di prosodie e creando le relative pseudo-word con il software che abbiamo scritto.

Riferimenti bibliografici

Beckman, M.E., Hirschberg, J. & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Jun, S.A. (Ed.), *Prosodic typology – The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 9-54.

Canepari, L. (1986). *Italiano standard e pronunce regionali*. Padova: CLEUP.

Cowie, R., Cornelius, R.R. (2003). Describing the emotional states that are expressed in speech. In *Speech Communication*, 40, 5-32.

Cresti, E. (2000). *Corpus di italiano parlato*. Firenze: Accademia della Crusca.

Kratzer, A. (2012). *Modals and Conditionals: new and revised perspectives Oxford Studies in Theoretical Linguistics*. Oxford: Oxford University Press.

Liotti, G., Monticelli, F. (2008). *I sistemi motivazionali nel dialogo clinico*. Milano: Raffaello Cortina editore.

Nencioni, G. (1983). *Di scritto e di parlato*. Bologna: Zanichelli.

Ouyang, I.C., Kaiser, E. (2012). Focus-marking in a tone language: prosodic cues in Mandarin Chinese. In *Proceedings of the Linguistic Society of America, Extended abstracts of the annual meeting*, 3, 1-5.

Prieto, P., Borràs-Comes, J. & Roseano, P. (2010). *Interactive atlas of Romance intonation*. http://prosodia.upf.edu/iari/.

Sbattella, L. (2007). Le défi de la complexité: harmoniser et composer. In *La formatività del musicale*. Milano: Esagramma.

Tomkins, S.S. (1984). *Affect theory*. In Scherer, K.R., Ekman, P. (Eds.), *Approaches to emotion*. Hillsdale: Lawrence Erlbaum Associates.

DOI: 10.17469/O2103AISV000024

ANDREA CALVO, PIERO COSI, GIULIO PACI, VALENTINA PASIAN,
GIACOMO SOMMAVILLA, FABIO TESSER

Is voice personalization suitable and useful for ALS patients?

This study arises from the collaboration between MIVOQ, the ALS association CRESLA, and CNR, and it is still in an early preparatory and development phase. The aim of this experimental study is that of verifying the impact on the quality of life of patients and of his family members who use a speech synthesizer able to replicate the vocal timbre of the user's voice. The specific objectives of this work include understanding how "Your Digital Voice" is suitable and useful for ALS patients.

Key words: ALS, Voice Personalization, TTS.

1. *Introduction*

This work will be focused on voice preservation for patients with an illness that will inhibit them to speak (for ex. ALS Amyotrophic Lateral Syndrome/Lou Gehrig's disease).

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease characterized by a progressive impairment of motor neurons, which causes the gradual loss of volunteer muscles function. 25-30% of the patients with ALS presents dysarthria as initial or predominant sign early in the disease phases. Dysarthria come as a first symptom in patients with ALS and it is 8-times more frequent compared to dysphagia (swallowing difficulty). Moreover, dysarthria affects up to 70% of patients with spinal onset in late phases (Chiò *et al*., 2017).

These deficit symptoms may not be evident until about 80% of the motor neurons have been damaged. Moreover, the interval between the onset of the articulators deficit and the diagnosis can vary from 33 months before the ALS diagnosis up to 60 months after the diagnosis.

Dysarthria occurs in more than 80% of the ALS patients and can produce a considerable disability, in a more precocious way in those with bulbar onset that can get an anarthria after few months. The loss of the communicative possibilities prevents these patients to participate in many activities and lead to social isolation. Dysarthria therefore significantly reduces the quality of life of people affected by ALS.

The intervention in support of such symptoms is speech therapy, but a direct intervention only to orality is not recommended as it may result in patient fatigue and prevent effective communication. In these cases, intervenes the AAC (Augmentative Alternative Communication) which is a field of clinical practice which has as its objective the compensation of a disability (temporary or permanent) of expressive language. They try to recreate the ideal conditions so that the subjects have the opportunity to effectively and efficiently communicate, translating the content of their thoughts in a series of intelligible signs for their interlocutors. Often the use of communication aids allows to reach those goals.

2. *Personalized TTS for ALS patients*

The personalized speech synthesis (hereinafter referred as PTTS) allows to capture the timbric and prosodic characteristics of a specific user, and to make them available in the form of a voice for text-to-speech synthesis.

To generate a custom TTS voice, users must record the sentences with their own voice in the early stage of the disease, and convert it to a digital format through a TTS software. In this study, it will be used the MIVOQ technology, which allows to create a TTS voice with a hundred sentences. For reference, these sentences are recorded by users without disabilities in about 30 minutes.

MIVOQ PTTS implements two Speech Technology innovative techniques, known in the field of research as Statistical Parametric Speech Synthesis (SPSS) (Zen, Tokuda & Black, 2009) and Speaker Adaptation (Yamagishi, Kobayashi, Nakano, Ogata & Isogai, 2009). They have several advantages and among them:

- a customized synthetic voice can be created with few speech samples;
- TTS is "flexible" and "creative"[1].

This is the reason why we adopt for this aim "Your Digital Voice", a text-to-speech (TTS) service, developed by MIVOQ s.r.l., that allows fast and easy customization of synthetic voices. In order to get his/her own digital voice, the user has to record just a few dozens of sentences, without the need of a professional audio setup. Those users, who subsequently have lost the ability to speak, can still use their voice, in digital form, with the help of "Your Digital Voice" and a Speech Generating Device (SGD). The added value of this innovative technology is that it requires very few sentences to be recorded by the user, it is multilingual, it can be used in complete autonomy even by non-expert users, and it also allows for a more creative use of TTS (emotive speech, focus on specific words, etc.). The current prototype of the MIVOQ "Your Digital Voice" service consists of a web application where users can record sentences and automatically obtain a personalized TTS voice model. The voice model possesses the vocal characteristics (timbre, prosody) of the author of the recordings and can be used to convert any text into speech.

[1] Mivoq is developing a creative and flexible technology for TTS: it will be possible to create synthetic speech which incorporates emotions (sad, happy,...), audio effects (childish voice, monster voice,...) and vocalizations (laughter, cry, fillers like "uh" or "um",...).

3. *Goals*

The aim of this experimental study is that of verifying the impact on the quality of life of patients and of his family members who use a speech synthesizer able to replicate the vocal timbre of the user's voice. The specific objectives of this work include understanding how "Your Digital Voice" is suitable and useful for ALS patients.

Also, it is important to understand users' needs more deeply. In fact, we need to better understand also how much TTS personalization can improve the life of patients. Experiments should be set up that could clarify these issues. Therefore, the experimentation with patient is a priority and, to do so, it is necessary to integrate MIVOQ "Your Digital Voice" system into Speech Generating Devices currently used by ALS patients.

Moreover, this study aims at subjectively evaluate the TTS / PTTS. Firstly, comparing the custom synthetic voice of a user, from the point of view of quality and intelligibility, with standard synthetic voices already present in Speech Generating Devices; and, subsequently, by assessing how the custom voice resembles the patient's real voice.

4. *Methodology*

Patients diagnosed with ALS with the absence of dysarthria, but with a likely development of the symptom and patients with initial dysarthria (ALSFRS_R value of language ≥ 3) will be recruited, while patients with frontotemporal dementia diagnosis will be excluded from recruiting. A few dozens of sentences produced by these patients will be recorded and sampled.

Later, the newly created personalized TTS voice will be used as soon as the patient will require a support to communication (possible uses will be via smartphones, dynamic communicator, eye tracker etc. that is medium and high-tech aids).

After a trial period of about 2 months, the subjects will undergo a number of questionnaires that assess the perceived patients' and caregivers' quality of life:

- McGill Scale;
- Quality of Life during Serious Illness-Family Carers – QOLLTI- Fv2;
- Needs and effective communication – CETI-M;
- Psychosocial Impact of Assistive Devices Scale – PIADS;
- Quebec Evaluation of User Satisfaction with Assistive Technology – QUEST 2.0;
- an ad hoc created questionnaire.

Bibliography

Chiò, A., Mora, G., Moglia, C., Manera, U., Canosa, A., Cammarosano, S., Ilardi, A., Bertuzzo, D., Bersano, E., Cugnasco, P., Grassano, M., Pisano, F., Mazzini, L. & Calvo, A. (2017). Secular Trends of Amyotrophic Lateral Sclerosis: The Piemonte and Valle d'Aosta Register. In JAMA Neurology, 74(9), 1097-1104.

Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K. & Isogai, J. (2009). Analysis of Speaker Adaptation Algorithms for HMM-Based Speech Synthesis and a Constrained SMAPLR Adaptation Algorithm. In *IEEE Transactions on Audio, Speech, and Language Processing*, 17(1), 66-83.

Zen, H., Tokuda, K. & Black, A.W. (2009). Statistical parametric speech synthesis. In *Speech Communication*, 51, 1039-1064.

Autori

Giulia Angonese – Università di Padova
angonese.giulia@gmail.com

Fabio Ardolino – Università di Pisa

Francesco Avanzini – ENT Department, General Hospital, Bolzano
francesco.avanzini@asbz.it

Cinzia Avesani – Istituto di Scienze e Tecnologie della Cognizione del CNR (ISTC-CNR), Padova
cinzia.avesani@pd.istc.cnr.it

Daniela Beltrami – Arcispedale S. Maria Nuova, Reggio Emilia
beltrami.daniela@yahoo.it

Catherine T. Best – MARCS Institute for Brain, Behaviour and Development, Western Sydney University and Haskins Laboratories, New Haven CT

Paolo Bravi – Conservatorio di Musica "G. P. Palestrina", Cagliari
pa.bravi@tiscali.it

Pierpaolo Busan – IRCCS Fondazione Ospedale San Camillo, Venezia
pbusan@units.it

Andrea Calvo – Dipartimento di Neuroscienze "Rita Levi Montalcini", Università degli Studi di Torino, CRESLA – Centro Regionale Esperto per la Sclerosi Laterale Amiotrofica, A.O.U. Città della Salute e della Scienza di Torino
andrea.calvo@unito.it

Violetta Cataldo – Università degli Studi di Salerno
violetta.cataldo@live.it

Chiara Celata – Scuola Normale Superiore, Pisa
chiara.celata@sns.it

Sonia Cenceschi – Dipartimento di Elettronica, Informazione e Bioingengeria (DEIB), Politecnico di Milano
sonia.cenceschi@polimi.it

Cheng Chen – Scuola Normale Superiore, Pisa
cheng.chen@sns.it

Lorenzo Ciaurelli – Università di Roma "La Sapienza"
lorenzociaurelli@gmail.com

Claudia Roberta Combei – Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa
roberta.combei@fileli.unipi.it

Piero Cosi – Istituto di Scienze e Tecnologie della Cognizione del CNR (ISTC-CNR), Padova
piero.cosi@pd.istc.cnr.it

Giovanna De Bellis – Università degli Studi di Napoli Federico II
giodebellis@libero.it

Giovanni Del Ben – B.R.A.I.N. Center for Neuroscience, Department of Life Sciences, University of Trieste
giovanni.delben@studenti.units.it

Bruno Di Biase – School of Humanities & Communication Arts, Western Sydney University

John H. Esling – Department of Linguistics, University of Victoria, Canada
esling@uvic.ca

Susanne Fuchs – Laboratory Phonology Group, Leibniz-Centre General Linguistics (ZAS), Berlin
fuchs@leibniz-zas.de

Rolando Füstös – ENT Department, General Hospital, Bolzano
rolando.fustos@sabes.it

Vincenzo Galatà – Libera Università di Bolzano & Istituto di Scienze e Tecnologie della Cognizione, CNR, Padova
vincenzo.galata@pd.istc.cnr.it

Davide Garassino – Romanisches Seminar, Universität Zürich
davide.garassino@gmail.com

Barbara Gili Fivela – CRIL – DREAM, Università del Salento, Lecce
barbara.gili@unisalento.it

Glenda Gurrado – Dipartimento di Lettere Lingue e Arti. Italianistica e culture comparate, Università degli Studi di Bari
glendagurrado@libero.it

Massimiliano M. Iraci – CRIL – DREAM, Università del Salento, Lecce
massimiliano.m.iraci@gmail.com

Giovanna Lenoci – Scuola Normale Superiore, Pisa
giovanna.lenoci@sns.it

Carla Leonardi – Università di Padova & SCIAF – Servizio di Neuroriabilitazione Età Evolutiva Padova, AULSS 6 Euganea
carla.leonardi@unipd.it

Robin Lickley – Speech and Hearing Sciences, Queen Margaret University, Edinburgh
RLickley@qmu.ac.uk

Michele Loporcaro – Romanisches Seminar, Universität Zürich
loporcar@rom.uzh.ch

Tim Mahrt – Laboratoire Parole et Langage, CNRS, Aix-en-Provence
timmahrt@gmail.com

Chiara Meluzzi – Scuola Normale Superiore Pisa, Università di Pavia
chiara.meluzzi@sns.it

Daniela Mereu – Università di Bergamo – Dipartimento di Lingue, letterature e culture straniere
mereudaniela@hotmail.it

Nicholas Nese – Università di Pavia
nicholas.nese@libero.it

Riccardo Orrico – Università degli Studi di Salerno
rorrico@unisa.it

Giulio Paci – MIVOQ, Bassano del Grappa
giulio.paci@mivoq.it

Valentina Pasian – CRESLA – Centro Regionale Esperto per la Sclerosi Laterale Amiotrofica, A.O.U. Città della Salute e della Scienza di Torino
valentina.pasian1@gmail.com

Caterina Petrone – Laboratoire Parole et Langage, CNRS, Aix-en-Provence
caterina.petrone@univ-amu.fr

Duccio Piccardi – Dipartimento di Filologia, letteratura e linguistica. Sezione di linguistica – Università di Pisa
duccio.piccardi@fileli.unipi.it

Elisa Raccanelli – Università di Padova
elisa.raccanelli@yahoo.it

Giulia Rancan – Università di Padova
rancangiulia90@gmail.com

Irene Ricci –Laboratorio SMART-ELiTe – Scuola Normale Superiore, Pisa
irene.ricci@sns.it

Andrea Sandi – SINTAC Biomedical Engineering, Padova
andrea.sandi@sintac.it

Renata Savy – Università degli Studi di Salerno
rsavy@unisa.it

Licia Sbattella – Dipartimento di Elettronica, Informazione e Bioingengeria (DEIB), Politecnico di Milano, Fondazione Sequeri Esagramma Onlus, Milano
licia.sbattella@polimi.it

Simona Schiattarella – Università degli Studi di Napoli Federico II
simona.schiattarella@yahoo.it

Stephan Schmid – Phonetisches Laboratorium, Universität Zürich
stephan.schmid@uzh.ch

Elisa Sneed German – Laboratoire Parole et Langage, CNRS, Aix-en-Provence
elisa.sneed@lpl-aix.fr

Giacomo Sommavilla – MIVOQ, Bassano del Grappa
giacomo.sommavilla@mivoq.it

Patrizia Sorianello – Dipartimento di Lettere Lingue e Arti. Italianistica e culture comparate, Università degli Studi di Bari
patrizia.sorianello@uniba.it

Roberto Tedesco – MultiChancePoliTeam, Politecnico di Milano
roberto.tedesco@polimi.it

Fabio Tesser – MIVOQ, Bassano del Grappa
fabio.tesser@mivoq.it

Ottavia Tordini – Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa
ottavia.tordini@fileli.unipi.it

Mario Vayra – Dipartimento di Filologia Classica e Italianistica, Università di Bologna, ISTC-CNR, Padova

Claudio Zmarich – Istituto di Scienze e Tecnologie della Cognizione del CNR (ISTC-CNR), Padova
claudio.zmarich@pd.istc.cnr.it

Revisori del volume

Iolanda Alfano (Università degli Studi di Salerno)
Cinzia Avesani (ISTC-CNR, Padova)
Martin J. Ball (Bangor University)
Pier Marco Bertinetto (Scuola Normale Superiore, Pisa)
Chiara Bertini (Scuola Normale Superiore, Pisa)
Maria Paola Bissiri (Scuola Normale Superiore, Pisa)
Silvia Calamai (Università degli Studi di Siena)
Gianluca Campana (Università degli Studi di Padova)
Chiara Celata (Scuola Normale Superiore, Pisa)
Lidia Costamagna (Università per Stranieri di Perugia)
Claudia Crocco (Ghent University)
Franco Cutugno (Università di Napoli Federico II)
Esther de Leeuw (Queen Mary University of London)
Olga Dmitrieva (Purdue University, IN)
Susanne Fuchs (Leibniz-Centre General Linguistics – ZAS, Berlin)
Vincenzo Galatà (Libera Università di Bolzano & ISTC-CNR, Padova)
Barbara Gili Fivela (Università del Salento, Lecce)
Mirko Grimaldi (Università del Salento, Lecce)
Giovanna Lenoci (Scuola Normale Superiore, Pisa)
Robin Lickley (Queen Margaret University, Edinburgh)
Paolo Mairano (Università degli Studi di Torino)
Giovanna Marotta (Università di Pisa)
Chiara Meluzzi (Scuola Normale Superiore, Pisa)
Giulia Natarelli (Università degli Studi di Padova)
Rosalba Nodari (Scuola Normale Superiore, Pisa)
Antonio Origlia (Università di Napoli Federico II)
Riccardo Orrico (Università degli Studi di Salerno)
Daniel Recasens (Universitat Autònoma de Barcelona)
Irene Ricci (Scuola Normale Superiore, Pisa)
Maja Rogić (University of Split)
Luciano Romito (Università della Calabria)
Stephan Schmid (University of Zurich)
Patrizia Sorianello (Università degli Studi di Bari Aldo Moro)
Patrycja Strycharczuk (University of Manchester)
Fabio Tamburini (Università di Bologna)
Claudio Zmarich (ISTC-CNR, Padova)

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.

Chiara Bertini è responsabile tecnico del Laboratorio ELiTe della Scuola Normale Superiore di Pisa, dove si occupa dell'elaborazione dei dati raccolti in studi linguistici sperimentali in fonetica e neurolinguistica.

Chiara Celata è ricercatore in Linguistica presso la Scuola Normale Superiore di Pisa. Si occupa dei meccanismi fisiologici e cognitivi che stanno alla base della produzione del parlato, e della pronuncia dell'italiano da parte di adulti mono- e bilingui.

Giovanna Lenoci è assegnista di ricerca in Linguistica preso la Scuola Normale Superiore di Pisa. Si interessa di analisi acustica e articolatoria dei meccanismi sottostanti il controllo motorio del parlato normale e patologico, con particolare attenzione alla balbuzie evolutiva e alla disprassia verbale.

Chiara Meluzzi è assegnista di ricerca in Linguistica presso l'Università di Pavia. Si occupa principalmente di variazione sociofonetica e fonetica sperimentale nella pronuncia dell'italiano da parte di parlanti nativi e non nativi.

Irene Ricci è responsabile tecnico del Laboratorio ELiTe della Scuola Normale Superiore di Pisa; collabora alle attività di ricerca in linguistica sperimentale occupandosi della realizzazione di protocolli sperimentali e della gestione della strumentazione scientifica.

AISV - Associazione Italiana Scienze della Voce
sito: www.aisv.it
email: aisv@aisv.it| redazione@aisv.it
ISBN: 978-88-97657-19-4

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