

PROSODIA DI IERI, PROSODIA DI OGGI. UN ESPERIMENTO DI AUTOTRAPIANTO SUL PARLATO TELEVISIVO

Massimo Pettorino
Università degli Studi di Napoli L'Orientale
mpettorino@unior.it

1. INTRODUZIONE

Gli studi più recenti sulla comunicazione parlata hanno rivolto un sempre maggiore interesse agli aspetti ritmico-prosodici del parlato. I motivi di tale accresciuta attenzione sono vari e di diverso genere.

Certamente la possibilità di analizzare, confrontare e modificare i parametri soprasegmentali di un enunciato ha determinato sempre più la consapevolezza del ruolo fondamentale che essi svolgono nello scambio comunicativo. Da questo punto di vista la stessa suddivisione in due livelli, quello segmentale e quello soprasegmentale, appare ormai inadeguata a rappresentare la complessità dell'interazione faccia a faccia. Per comprendere in che modo e attraverso quali canali l'informazione passa dal parlante all'ascoltatore, infatti, non è più sufficiente analizzare, se pure in maniera sempre più accurata e sofisticata, le caratteristiche articolatorio-acustiche dei modelli consonantici e vocalici e considerare le loro deformazioni come varianti dovute a specifiche situazioni contestuali. Appare invece sempre più evidente il contrario, nel senso che i tratti soprasegmentali sono piuttosto la base sulla quale si va ad adattare il testo. Da questo punto di vista, quindi, le parole non si deformano, ma nascono nella mente del parlante già ben determinate in una certa forma o in un'altra a seconda di vari fattori (intenzione, stato emozionale, attenzione etc). Se l'intenzione fosse diversa, se lo stato emotivo fosse un altro, il parlante non userebbe lo stesso testo in una diversa variante prosodico-intonativa, ma probabilmente userebbe un testo completamente diverso. Le parole, nella comunicazione parlata, o sono così come vengono prodotte oppure, semplicemente, non sono. E anche quando questo accade, la comunicazione rimane attiva e funzionante, affidata a canali diversi. Indagini condotte con metodologie diverse confermano che il testo, le parole, le frasi, costituiscono circa il 10% del messaggio trasmesso, il restante 90% segue evidentemente canali diversi (Mehrabian & Ferris, 1967; Mehrabian, 1972).

I diversi modi in cui uno stesso testo può essere prodotto sembrano quindi assolutamente prevalenti rispetto a un significato "letterale" che si rivela sempre più vago e sfuggente. L'intonazione, le accelerazioni e decelerazioni dell'enunciato, le variazioni di intensità del segnale, la frequenza e la durata delle pause silenti e di quelle che solo apparentemente sono disfluenze ma che in realtà svolgono una importante funzione comunicativa, tutto concorre a mostrare come gli aspetti prosodici non possano essere relegati a varianti diastratiche, diafasiche o diamesiche, se non anche idiosincratiche, di un presunto modello astratto. Al contrario sono essi stessi parte integrante del messaggio, "modelli" di contenuti comunicativi specifici, dipendenti da vari fattori, alcuni dei quali apparentemente scontati e di immediata comprensione (aspetti legati al parlante, al suo stato emozionale, all'intento comunicativo, alla situazione contestuale), altri meno immediati ma altrettanto evidenti e definiti. Tra questi ultimi, il rapporto tra la prosodia di un enunciato e l'epoca in cui esso viene prodotto.

2.1. *Modelli prosodici*

Il rapporto tra le caratteristiche prosodiche dell'enunciato e l'epoca in cui esso si realizza è l'oggetto di una ricerca pluriennale, iniziata a Torino nel 1994 con l'analisi di parlato radiofonico (Pettorino & Giannini, 1994), per poi proseguire verso l'analisi del parlato televisivo in Italia (Giannini & Pettorino, 1998; Pettorino, 2002; Pettorino & Giannini, 2005; Pettorino, 2008a; Pettorino, 2008b) poi in altri paesi (Pettorino & Giannini, 2010).

Gli studi sul parlato dei media hanno poi spinto a riprendere le tematiche connesse all'identificazione del parlante già affrontate in studi precedenti (Pettorino, 1989; Giannini et alii, 1989) e in particolare all'età dello speaker (Giannini & Pettorino, 2011; Pettorino & Giannini, 2011). Ricerche parallele nel campo dell'acquisizione della L2 avevano evidenziato il ruolo centrale giocato dalla prosodia anche in riferimento alla percezione dell' "accento straniero" (De Meo et alii, 2011; Pettorino et alii 2011; Pettorino et alii, 2012; Vitale et alii, 2012). Tali studi hanno quindi spinto a mettere a punto una particolare procedura, quella del trapianto prosodico-intonativo, che consentisse la manipolazione degli aspetti prosodici, con l'obiettivo di per consentire da un lato di verificarne il ruolo sul piano della percezione e dall'altro di superare alcuni ostacoli nel processo di acquisizione di una lingua (Pettorino & Vitale, 2012).

Tali studi hanno quindi spinto a utilizzare la procedura del trapianto prosodico-intonativo, introdotta nel 2007 da Yoon in uno studio sull'inglese L2 di un parlante coreano madrelingua (Yoon, 2007). È stato quindi necessario adattare tale procedura alle specificità linguistiche dell'italiano (Pettorino & Vitale, 2012).

Dalla procedura del trapianto oggi si è giunti a una sorta di "autotrapianto", tecnica che sarà l'oggetto d'analisi del presente lavoro.

2.2. *Il parlato dei mass media*

La ricerca sull'italiano parlato dell'informazione televisiva ha preso le mosse dall'analisi di un corpus di radiogiornali e telegiornali "antichi" e "moderni", i primi trasmessi negli anni '50 e '60, i secondi alla fine degli anni '90, evidenziando significative differenze tra i due tipi di parlato. Rispetto all'estrema eterogeneità delle tipologie di parlato trasmesso, il materiale utilizzato per le analisi presentava una serie di caratteristiche comuni. Si trattava in entrambi i casi di parlato letto, prodotto da speaker professionisti in uno stesso contesto situazionale, vale a dire in uno studio di registrazione della Rai. Inoltre era un parlato orientato verso l'ascoltatore, altamente intelligibile, prodotto nella varietà standard dell'italiano.

Le analisi spettroacustiche hanno mostrato l'emergere di profondi cambiamenti sul piano ritmico-prosodico e intonativo sia nella comunicazione radiofonica sia in quella televisiva. Il modello di parlato degli anni '90 risulta molto diverso da quello degli anni '50 e '60. Infatti le catene foniche (porzioni di enunciato comprese tra due pause silenti consecutive) risultano mediamente più lunghe, i silenzi si riducono sia di numero sia di durata, il range tonale è più ampio e variato. Tutti questi elementi concorrono a delineare modelli prosodici molto diversi. La tendenza è verso un parlato che potrebbe essere definito come più produttivo: a parità di tempi di fonazione, viene prodotto un numero maggiore di sillabe. Tuttavia, l'accelerazione non viene ottenuta a scapito della qualità articolatoria: la velocità di articolazione (VdA), indice che misura la qualità del gesto articolatorio, rimane invariata (circa 6,5 sill/s). Gli speaker "moderni" continuano quindi a produrre un parlato accurato, asse-

gnando a ciascuna sillaba tutto il tempo necessario al raggiungimento dei diversi target articolatori. L'accelerazione viene ottenuta esclusivamente attraverso la riduzione dei silenzi. Le analisi mostrano infatti l'uso di una diversa strategia nell'uso delle pause da parte dello speaker. Negli anni '50 e '60 le pause silenti sono prevalentemente di natura sintattica, costituendo quindi il principale segnale demarcativo di un confine di frase, sia esso forte o debole. Nel parlato degli anni '90, al contrario, le pause silenti si distribuiscono all'interno della catena parlata quasi esclusivamente in funzione delle esigenze respiratorie e, quindi, con diversa localizzazione rispetto all'organizzazione sintattica del testo. Le analisi, inoltre, evidenziano una notevole differenza anche per quanto riguarda la loro durata: se negli anni '50 e '60 sono presenti due tipologie di pause vuote, una pausa lunga (circa 0,9 s) e una pausa breve (circa 0,4 s), prodotte rispettivamente in corrispondenza dei confini sintattici forti e deboli, negli anni '90 vi è un'unica tipologia di pausa, chiaramente respiratoria, di durata estremamente breve (circa 0,3 s).

Il mancato apporto dei silenzi alla demarcazione dei confini sintattici viene compensato negli anni '90 da un diverso uso dell'andamento intonativo. Infatti, immediatamente prima del silenzio "soppresso", compare un picco intonativo molto marcato, il cui scopo è quello di far sapere all'ascoltatore che in quel punto c'è un confine. In tal modo l'intelligibilità del messaggio è salvaguardata e, per giunta, lo speaker può disporre di un tempo aggiuntivo, corrispondente a quello della pausa soppressa.

Tuttavia i dati ottenuti nella prima fase della ricerca non hanno consentito di valutare quale fosse l'incidenza del testo sulle variazioni ritmico-prosodiche osservate in quanto, come è naturale, gli speaker avevano letto notizie differenti. Il testo poteva aver influito sulla velocità di articolazione e sulla fluidità, l'argomento sul ritmo, la struttura sintattica sulla prosodia. Per superare questa difficoltà, nel 2005 si è fatto ricorso a un artificio.

2.3. 2005: l'artificio

Da alcuni telegiornali degli anni '60 sono state selezionate notizie di vario genere (politica interna ed estera, cronaca, sport) lette dai conduttori in studio. Le notizie sono state trascritte ortograficamente. Successivamente il testo è stato dato a due speaker professionisti del TG2 per realizzare un TG virtuale. I due conduttori hanno letto le notizie "antiche" come in un vero telegiornale. Per preservare la medesima situazione comunicativa, la registrazione ha avuto luogo presso lo studio del TG2 della sede Rai di Saxa Rubra: subito dopo la sigla di chiusura del vero TG2 (edizione delle ore 13), è ripartita la sigla di apertura e i due conduttori hanno letto le notizie degli anni '60 come se fossero notizie di assoluta attualità.

Grazie a tale artificio, è stato possibile disporre di un corpus di parlato costituito da due letture di uno stesso testo, effettuate a distanza di circa 40 anni, nella stessa situazione contestuale e prodotte da una stessa tipologia di parlanti, vale a dire giornalisti professionisti. L'identità del testo ha consentito di annullare quei parametri (argomento, numero e lunghezza delle parole, struttura sintattica etc.) la cui variazione avrebbe potuto influire sull'analisi diacronica degli andamenti ritmico-prosodici dei due modelli di parlato. Il corpus è stato letto da tre speaker per gli anni '60 (Andrea Barbato, Piergiorgio Branzi e Marco Raviart) e da due speaker per il 2005 (Fabio Cappelli e Silvia Vaccarezza).

Le analisi condotte su questo particolare corpus hanno messo in evidenza vari fenomeni. Per quanto riguarda le pause silenti, le osservazioni fatte in precedenza hanno trovato

una conferma sia per quanto riguarda la loro distribuzione che per quanto riguarda la loro durata. Un esempio significativo è dato dal seguente brano:

Andrea Barbato, TG 1968: "... *mentre si trovavano in Nigeria sulla frontiera con la provincia secessionista del Biafra # insieme a tre tedeschi e due libanesi # Le uniche informazioni di cui si dispone...*".

In questo breve passaggio si può notare come le due pause presenti corrispondano a due confini sintattici, uno debole e l'altro forte. Confrontiamo lo stesso passaggio pronunciato da Fabio Cappelli nel 2005: "... *mentre si trovavano in Nigeria sulla frontiera con la provincia # secessionista del Biafra assieme a tre tedeschi e due libanesi # Le uniche informazioni di cui si dispone...*". Anche in questo caso troviamo due pause silenziose ma, mentre una continua a segnare il confine di frase (*libanesi #*), l'altra si sposta e insorge tra il sostantivo (*provincia*) e l'aggettivo (*secessionista*) a conclusione di una lunga catena fonica, evidentemente per esaurimento della riserva d'aria fonatoria, come testimoniato anche dalla presenza di una profonda inspirazione.

Ancora più marcata è la differenza rispetto al modello del '68 se si esamina lo stesso passo prodotto nel 2005 da Silvia Vaccarezza: "... *mentre si trovavano in Nigeria sulla frontiera con la provincia secessionista del Biafra insieme # a tre tedeschi e due libanesi Le uniche # informazioni di cui si dispone...*".

In questo caso, infatti, nessuna delle due pause silenziose coincide con un confine sintattico (*insieme # a* e *uniche # informazioni*). Da notare come, rispetto al modello del 1968, i due silenzi si siano spostati di una parola a destra, evidentemente con uno scopo comunicativo preciso: la prima parola della frase non può attendere, deve essere prodotta subito in coda alla frase appena terminata in modo da legare il discorso e tenere desta l'attenzione dello spettatore. Il confine verrà comunque segnalato mediante un picco intonativo marcato sia nell'uno che nell'altro caso. Il risultato di questa diversa strategia comunicativa determina un dimezzamento del numero di pause silenziose e una diversa tipologia di durata.

2.4. Prosodia e globalizzazione

Per verificare se i cambiamenti ritmico-prosodici emersi dalle analisi dell'italiano radiofonico e televisivo fossero da considerare come specifici dell'italiano o se fossero presenti anche in altre lingue, nel 2009 è stato condotto uno studio comparativo sul parlato dei mass media in sloveno, inglese, francese del Canada e giapponese (Pettorino & Giannini, 2010). Il corpus è stato tratto da telegiornali degli anni '70 e del 2009 trasmessi da diverse emittenti (Radio Televizija Slovenija, BBC, Canadian Broadcasting Corp., NHK Ni-hon Hōsō Kyōkai). Le voci analizzate sono voci maschili tranne quelle estratte dal telegiornale sloveno.

I risultati delle analisi spettroacustiche sono poi stati confrontati con quelli precedentemente ricavati per l'italiano in cui i testi erano diversi. Per ciascun telegiornale è stato computato il numero delle sillabe e sono state misurate le durate delle catene foniche, le durate sillabiche, le durate delle pause silenziose, i valori massimo e minimo di f_0 per ciascuna catena fonica. Sulla base di tali misure sono poi stati ricavati i valori medi della velocità di articolazione, della velocità di eloquio, della fluenza, della percentuale di silenzio e il range tonale. I risultati dello studio hanno evidenziato come i moderni telegiornali, in tutte le lingue esaminate, sono caratterizzati da catene foniche più lunghe, da pause silenziose meno frequenti, da una velocità di articolazione piuttosto stabile, da pause più brevi, da una minore percen-

tuale di silenzio, da un range tonale più ampio. L'ipotesi di un modello ritmico condiviso dal sistema mediatico globalizzato trovava quindi una conferma dalle analisi sperimentali.

2.5. Verifiche percettive

Un ulteriore passo avanti nell'indagine è stato fatto nel 2011. Questa volta si è cercato, mediante la manipolazione ritmico-intonativa del segnale, di stabilire se i due modelli evidenziati dalle analisi spettro acustiche trovassero dei riscontri sul piano percettivo. A questo scopo si è proceduto alla raccolta di un corpus di parlato tratto da notiziari radiofonici e televisivi degli anni '50 e del 2011. Sono state selezionate tre notizie di "ieri" e tre notizie di "oggi" di vario argomento (cultura, cronaca, sport). L'analisi spettroacustica del corpus ha confermato la presenza dei due modelli ritmico-intonativi precedentemente individuati. Si è quindi proceduto alla manipolazione, mediante la quale il modello di oggi è stato trapiantato sugli enunciati di ieri e viceversa. Il corpus risultante, costituito da 6 notizie originali e 6 modificate, è stato organizzato in 10 coppie, poste in ordine casuale su un nastro d'ascolto che ha rappresentato l'oggetto di un test percettivo, somministrato a 50 ascoltatori italiani, di varia età. Il test ha mostrato con evidenza che le variazioni diacroniche individuate dalle analisi acustiche sono, sul piano percettivo, ben presenti nella coscienza dell'ascoltatore italiano, il quale è in grado di discriminare tra i due modelli prosodici e di ricondurli al parlato di epoche temporali diverse. Tuttavia è evidente che il tessuto soprasegmentale non è l'unico parametro che permette all'ascoltatore di collocare un enunciato in una determinata epoca. Un altro fattore è infatti il contenuto testuale. Per eliminare questa variabile, è stata utilizzata la tecnica del *low-pass filtering*, con frequenza di taglio a 400 Hz. In tal modo è stata preservata la curva melodica dell'enunciato, in quanto il range tonale si manteneva ampiamente al di sotto di tale soglia, mentre venivano eliminate le frequenze di risonanza che permettevano l'identificazione segmentale e, di conseguenza, il contenuto testuale. I test condotti sul segnale filtrato non modificavano in maniera significativa i risultati del primo test, confermando che il modello ritmico-intonativo era sufficiente a permettere all'ascoltatore di collocare temporalmente un enunciato.

3. LA VOCE E L'ETÀ

Gli studi sul parlato dei mass media prodotto in epoche diverse evidenziavano modelli di parlato molto diversi. A parità di situazione comunicativa, quindi, lo speaker di oggi adotta strategie diverse rispetto al passato. Oltre alla diversa epoca, un ulteriore fattore che sembra incidere sulla variazione del modo di parlare è l'età cronologica dello speaker. Lo step successivo della ricerca si è infatti concentrato sull'analisi del rapporto tra voce e età del parlante.

3.1. I correlati acustici dell'età del parlante

Esiste una vasta letteratura sul rapporto tra voce ed età del parlante. Il processo di invecchiamento è generalmente associato al sopraggiungere di modifiche del tratto vocale (Eppley & Mueller, 2001). I cambiamenti funzionali che si verificano nel sistema respiratorio, fonatorio e sopralaringeo, dalla prima infanzia alla vecchiaia, alterano il modo in cui gli esseri umani parlano al punto che gli ascoltatori sono in grado di giudicare l'età del parlante con notevole precisione (Pear, 1931; Ptacek & Sander, 1966; Shipp & Hollien, 1969; Pettorino & Giannini, 2011; Krauss et alii, 2002).

L'irrigidimento del torace e indebolimento dei muscoli respiratori (Linville, 2004) e la diminuzione della capacità polmonare (Awan, 2006) sono i più significativi cambiamenti nel sistema respiratorio con l'avanzare dell'età. Un altro organo che subisce variazioni anatomiche legate all'invecchiamento è la laringe; quest'ultima è infatti soggetta all'ossificazione e calcificazione delle cartilagini e all'atrofia delle corde vocali. Il segnale vocale cambia altresì in funzione dei cambiamenti che avvengono nel tratto sopralaringale (Linville, 2004; Tompkins et alii, 2006; Wohlert & Smith, 1998).

Tutti questi cambiamenti anatomo-fisiologici determinano variazioni complesse del segnale vocale. Tuttavia, sebbene il rapporto tra voce ed età sia molto complesso da circoscrivere, vi è un generale accordo tra i ricercatori su alcuni aspetti.

Per quanto riguarda il registro tonale, ad esempio, molti studi hanno mostrato differenze tra invecchiamento maschile e femminile. In età avanzata la f_0 aumenta sostanzialmente negli uomini (circa 35 Hz), mentre diminuisce nelle donne (circa 10-15 Hz). Indipendentemente dal sesso, invece, la voce con gli anni diventa più tremula e roca a causa di una f_0 meno stabile (Linville, 1987; Jacques & Rastatter, 1990; Traunmüller & van Bezooijen, 1994).

Altri correlati acustici dell'età sono un aumento di jitter e shimmer (Linville, 1987; Ramig & Ringel, 1983; Dehqan et alii, 2012), un abbassamento delle formanti (Linville, 2001; Linville, 1987), l'allungamento delle vocali e della fase di tenuta delle occlusive (Ptacek & Sander 1966), l'aumento della deviazione standard di f_0 (Linville, 1987; Jacques & Rastatter 1990; Traunmüller & van Bezooijen, 1994).

3.2. Il metodo

Uno dei punti deboli di queste ricerche è nella metodologia seguita per la raccolta del corpus. I materiali più largamente usati per le indagini sono essenzialmente di due tipi: o uno stesso testo (brani, frasi, parole o vocali isolate) letto da diversi parlanti appartenenti a fasce d'età diverse, o un corpus di parlato spontaneo prodotto da parlanti di diverse fasce di età. Tuttavia tali metodologie non assicurano che i risultati delle analisi siano legati esclusivamente alla diversa età del parlante. Va considerato, infatti, che il tipo di parlato dipende da numerose variabili quali, ad esempio, la situazione contestuale, il rapporto con l'interlocutore, lo stile formale o informale, l'argomento, il parlato letto o spontaneo e altri ancora. Per poter confrontare parlati diversi e poter attribuire le cause delle eventuali differenze a uno specifico parametro, è necessario che tutte le altre variabili siano state precedentemente azzerate: solo in questo modo, infatti, è lecito concludere che le variazioni riscontrate sono state determinate proprio da quell'unico parametro lasciato in libera variazione. Per quanto riguarda l'età del parlante, quindi, sarebbe necessario poter disporre di uno stesso parlante che, a distanza di molti anni, pronunzi le stesse frasi nella medesima situazione comunicativa, rivolto a uno stesso interlocutore con il quale abbia lo stesso grado di familiarità. È quanto si è provato a fare in un esperimento condotto nel 2007, i cui risultati sono stati presentati al convegno nazionale AISV del 2008 (Giannini & Pettorino, 2009).

Seguendo la metodologia già sperimentata nei lavori sul parlato televisivo, il noto giornalista Piero Angela ha letto, a distanza di 40 anni, lo stesso testo da lui letto nel corso di un telegiornale del 1968. Per mantenere una stessa situazione contestuale, tutto si è svolto in uno studio di registrazione della sede RAI di Saxa Rubra. Questo artificio ci ha permesso di azzerare le numerose variabili dovute sia al testo che al parlante e di ottenere due registra-

zioni, in un certo senso sovrapponibili, che consentono un confronto puntuale, anche se limitato a un solo parlante. Le analisi spettroacustiche del corpus hanno evidenziato numerose modifiche occorse con l'avanzare dell'età, tra le quali:

- un rallentamento articolatorio di circa una sillaba/s;
- pause silenti più frequenti e più lunghe;
- una f_0 più alta e più variata;
- un aumento della porzione vocalica, soprattutto nelle sillabe toniche chiuse;
- sul piano segmentale, un ampliamento dell'area vocalica, causato da un arretramento e abbassamento delle vocali posteriori.

3.3. Prosodia e identità

Va sottolineato anche che, accanto alle differenze riscontrate tra la voce del 1968 e quella del 2007, nell'analisi di questo corpus particolare sono emerse alcune sorprendenti somiglianze: “alcune frasi sono state pronunciate, a distanza di 40 anni, con lo stesso andamento tonale, la stessa velocità, le stesse pause, le stesse esitazioni. Se le differenze trasmettono informazione sull'età del parlante, le similarità ci dicono evidentemente che quel parlante è Piero Angela in entrambi i casi. Se le prime sono dovute alle trasformazioni fisiologiche dovute al passare degli anni, le seconde sono dovute a caratteristiche che si mantengono costanti nel tempo e che sono frutto delle scelte fonatorie che ciascun parlante opera molto presto, presumibilmente già durante la fase dell'apprendimento del linguaggio, e che restano poi, in buona misura, immutate nel tempo” (Giannini & Pettorino, 2009: 177). Quel lavoro si concludeva con un'ipotesi di sviluppo della ricerca, vale a dire verificare se, mediante un esperimento di manipolazione ritmico-intonativa, fosse possibile “invecchiare” la voce di Piero Angela del 1968 e, al contrario, “ringiovanire” quella del 2007. Per procedere in questa direzione, è stato necessario perfezionare la tecnica del trapianto prosodico.

3.4. Il trapianto

Mediante la tecnica del trapianto prosodico-intonativo le caratteristiche soprasegmentali (intonazione, intensità, durate, dislocazione e durata delle pause silenti) di una voce A (donatore) sono clonate e trasferite alla voce B (ricevente) senza compromettere la percezione dell'identità del parlante. La tecnica del trapianto è basata sull'algoritmo PSOLA, Pitch-Synchronous Overlap and Add (Charpentier & Moulines, 1989), implementati nel software Praat (Boersma, 2001).

Per verificare il ruolo svolto dalla prosodia e dall'intonazione nella capacità, da parte dell'ascoltatore, di valutare l'età dello speaker, si è proceduto a una sorta di autotrapianto, in quanto le caratteristiche ritmico-prosodiche della voce di Piero Angela quarantenne sono state trapiantate sulla voce dello stesso Angela, ma ottantenne. Il trapianto ha avuto come oggetto tre notizie tratte dal telegiornale del 1968 e le stesse notizie rilette dopo 40 anni. Gli enunciati sono stati segmentati ed etichettati in 4 livelli: foni, sillabe, porzioni vocaliche e consonantiche, intervalli tra due consecutivi punti di inizio vocalici (VtoV). Per poter effettuare il trapianto, i livelli “foni” sono stati duplicati e modificati, in modo che ogni segmento del 1968 avesse un elemento corrispondente nell'enunciato del 2007. Poiché la tecnica del trapianto richiede che il numero di segmenti sia uguale nelle due voci, quella del donatore e quella del ricevente, in caso di non corrispondenza si è proceduto all'inserimento di micro-segmenti là dove era necessario. Grazie al trapianto gli enunciati prodotti nel 1968 sono stati “invecchiati”, trasferendo su di essi il contorno intonativo e le durate di foni e

pause silenti dei corrispondenti enunciati del 2007. Al contrario, questi ultimi sono stati “ringiovaniti” perché hanno assunto le caratteristiche prosodiche degli enunciati del 1968.

Il corpus risultante è quindi costituito da:

- 3 enunciati originali prodotti nel 1968 (denominati “Y”);
- 3 enunciati originali prodotti nel 2007 (denominati “O”);
- 3 enunciati del 1968 con prosodia del 2007 (denominati “tO”);
- 3 enunciati del 2007 con prosodia del 1968 (denominati “tY”);
- 1 voce di controllo, età 27 anni (cY);
- 1 voce di controllo, età 85 anni (cO).

3.5. *Il test*

Si è quindi proceduto alla somministrazione di un test percettivo. 70 studenti universitari italiani, di età compresa tra 23 e 26 anni, hanno partecipato al test. In una prima fase, di tipo comparativo, sono state fatte ascoltare 16 coppie di enunciati, composte da due voci che leggono la stessa notizia, con diverse combinazioni di età. Per ogni coppia l’ascoltatore doveva dire se il parlante più giovane corrispondeva alla prima voce, alla seconda o se i due parlanti avevano la stessa età. Il test è stato progettato con varie combinazioni di voci, sia originali che modificate. Nella seconda fase del test agli stessi studenti è stato chiesto di ascoltare 14 enunciati e di indicare per ciascun enunciato l’età del parlante, scegliendo tra sette fasce di età:

1. 26-35 anni;
2. 36-45 anni;
3. 46-55 anni;
4. 56-65 anni;
5. 66-75 anni;
6. 76-85 anni;
7. 86-95 anni.

4. I DATI

I risultati del test comparativo indicano che, quando le coppie di voci erano composte da voci originali, la quasi totalità delle risposte era corretta, sia che una delle due voci era più giovane dell’altra sia che erano entrambe di una stessa età. Nel caso di coppie composte da voci originali e modificate, i risultati confermano il ruolo della prosodia e dell’intonazione sulla percezione dell’età del parlante. Per quanto riguarda l’effetto di invecchiamento, la figura 1 mostra i dati relativi alle coppie del tipo “vecchio-giovane” e “vecchio-invecchiato”: nel primo caso la voce del 1968 viene riconosciuta come più giovane nell’84% dei casi, nel secondo caso le due voci vengono ritenute avere una stessa età nel 60% dei casi.

In figura 2 sono riportati i dati relativi alle coppie del tipo “giovane-giovane” e “giovane-invecchiato”. Se nel primo caso le risposte corrette (stessa età) raggiungono il 96%, nel secondo caso si fermano al 38%, segno evidente della buona riuscita dell’invecchiamento sul piano percettivo.

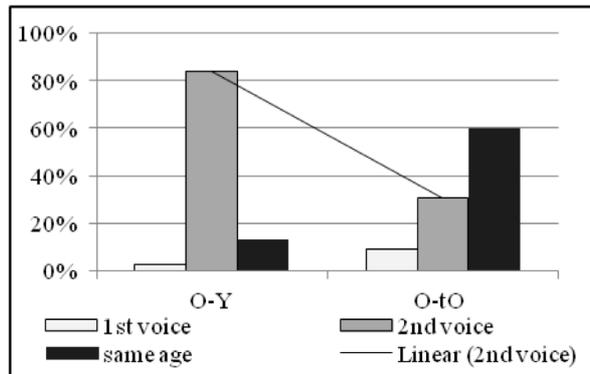


Figura 1: "invecchiamento" rispetto alla voce del 2007. Le risposte date alle coppie O-Y e O-tO sono significativamente diverse ($p < 0.01$).

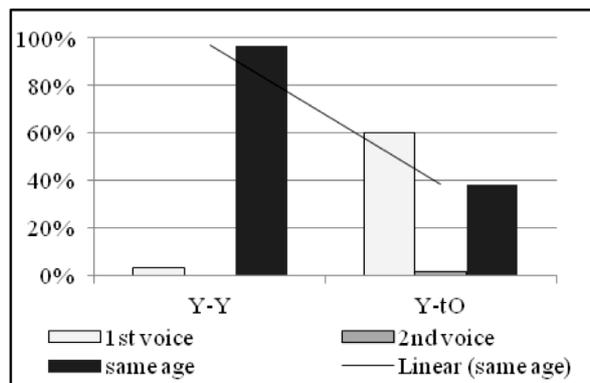


Figura 2: "invecchiamento" rispetto alla voce del 1968. Le risposte date alle coppie Y-Y e Y-tO sono significativamente diverse ($p < 0.01$).

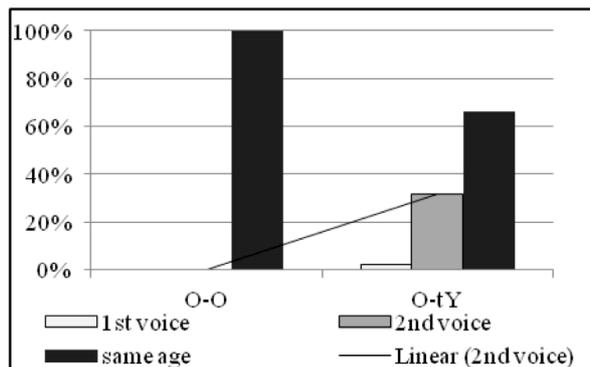


Figura 3: "ringiovanimento" rispetto alla voce del 2007. Le risposte date alle coppie O-O e O-tY sono significativamente diverse ($p < 0.01$).

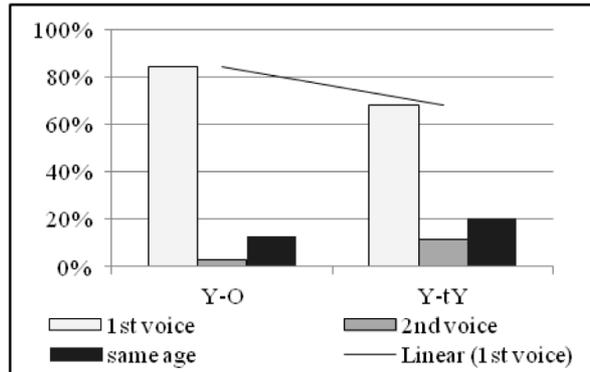


Figure 4: "ringiovanimento" rispetto alla voce del 1968.

Per quanto riguarda l'effetto di "ringiovanimento", i risultati del test mostrano che il trapianto prosodico, pur se valido, non è così efficace come nel caso dell'invecchiamento. Nella coppia "vecchio-ringiovanito", la voce trapiantata è riconosciuta come più giovane dal 32% di ascoltatori, mentre nella coppia "giovane-ringiovanito" le due voci sono ritenute avere la stessa età nel 20% dei casi e, dato interessante, la voce modificata è nel 12% dei casi ritenuta la più giovane (figure 3 e 4).

Per quanto riguarda la seconda fase del test, quella in cui gli ascoltatori dovevano assegnare ciascuna voce a una fascia di età, la figura 5 mostra la distribuzione delle risposte per ciascuna fascia. Anche in questo caso i dati mostrano come gli effetti del trapianto prosodico siano rilevanti sul piano percettivo in relazione al riconoscimento dell'età del parlante, anche se l'effetto "invecchiamento" è più efficace dell'effetto "ringiovanimento".

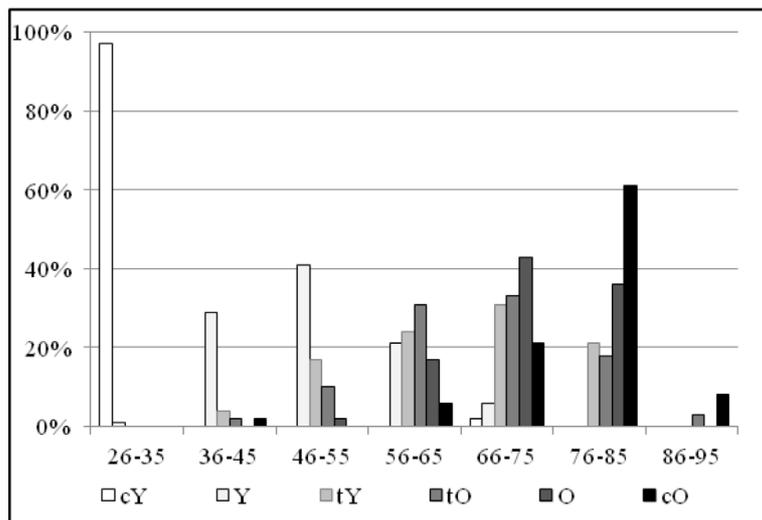


Figure 5: età percepita delle voci di controllo, originali e trapiantate.

CONCLUSIONI

L'analisi spettroacustica effettuata su un corpus di parlato letto, prodotto dallo stesso speaker a distanza di 40 anni ha dimostrato l'esistenza di una relazione tra alcune caratteristiche prosodico-intonative e l'età del parlante. Gli effetti di tali differenze sono stati valutati a livello percettivo utilizzando la tecnica del trapianto prosodico.

I giudizi espressi dagli ascoltatori riguardo alle voci trapiantate suggeriscono diverse considerazioni:

1. sul piano percettivo, gli ascoltatori sono in grado di riconoscere l'età di chi parla basandosi su specifiche caratteristiche prosodiche
2. la tecnica del trapianto è una procedura efficace allo scopo della manipolazione dell'età del parlante
3. mediante il trapianto, l'età percepita varia, anche se l'invecchiamento è più efficace del ringiovanimento.
4. sul piano acustico l'avanzare dell'età comporta un rallentamento della velocità di articolazione e di eloquio, un innalzamento del registro e un ampliamento del range tonale.

Inoltre, con l'età vi è una percentuale maggiore di porzione vocalica e tale incremento rende il ritmo dell'enunciato più simile al modello di lingue isomoraiche. Questi dati suggeriscono la possibilità di estendere la ricerca su altre lingue appartenenti ad altri gruppi ritmici.

BIBLIOGRAFIA

Boersma, P. (2001), Praat, a system for doing phonetics by computer, *Glott International* 5 (9/10), 341-345.

Charpentier, F. & Moulines, E. (1990), Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones, *Speech Communication* 9, 453-467.

Dehqan, A., Scherer, R.C., Dashti, G., Ansari-Moghaddam, A. & Fanaie, S. (2012), The effects of aging on acoustic parameters of voice, *Folia Phoniatria et Logopaedica*, 64, 265-270.

Eppley, B.D. & Mueller, P.B. (2001), Chronological Age Judgments of Elderly Speakers: The Effects of Listeners' Age, *Contemporary Issues in Communication Science and Disorders*, 28, 5-8.

Giannini A. & Pettorino, M. (1998), I cambiamenti dell'italiano radiofonico negli ultimi 50 anni: aspetti ritmico-prosodici e segmentali, in *Atti delle IX Giornate di Studio del Gruppo di Fonetica Sperimentale (GFS)*, Venezia, 65-81.

Giannini, A. & Pettorino, M. (2009), L'età della voce, in *Atti del IV Convegno Nazionale AISV*, Arcavacata di Rende, Cosenza, 165-178.

Giannini, A. & Pettorino, M. (2011), L'età percepita, in B. Gili Fivela, A. Stella, L. Garraipa, M. Grimaldi (eds.), *Contesto comunicativo e variabilità nella produzione e percezione della lingua*. Roma, Bulzoni Editore.

- Giannini A., Pettorino, M. & Cinque U. (1989), Speaker's Identification by Voice, *Eurospeech 89*, Paris, Vol. 1, 283-286.
- Jacques, R. & Rastatter, M. (1990), Recognition of speaker age from selected acoustic features as perceived by normal young and older listeners, *Folia Phoniatria*, 42, 118-124.
- Krauss, R.M., Freyberg, R. & Morsella, E. (2002), Inferring speakers' physical attributes from their voices", *Journal of Experimental Social Psychology*, 38, 618-625.
- Linville, S.E. (1987), Acoustic-perceptual studies of aging voice in women, *Journal of Voice*, 1, 44-48.
- Linville, S.E. (2001), *Vocal Aging*, Singular.
- Linville, S.E. (2004), The aging voice, *The American Speech-Language-Hearing Association (ASHA), Leader*, 12-21.
- Mehrabian, A. & Ferris, S. (1967), Inference of attitudes from nonverbal communication in two channels, *Journal of consulting psychology* 31 (3), 248-252.
- Mehrabian, A. (1972), *Non verbal communication*, Chicago, Aldine Atherton.
- Pear, T.H. (1931), *Voice and Personality*, London, Chapman & Hall.
- Pettorino M. (1989), Identificazione del peso e dell'altezza del parlante: un'indagine acustico-percettiva, in *Atti del XVII Convegno Nazionale AIA*, Parma, 433-438.
- Pettorino, M. (2002), I cambiamenti della lingua italiana, in *La voce come bene culturale* (A. De Dominicis, editor), Roma, Carocci, 141-157.
- Pettorino, M. (2008), Buonasera, buonasera dal TG2: verso la fonetica storica sperimentale, in *La Comunicazione Parlata*, vol. 2, Napoli, Liguori, 1205-1213.
- Pettorino, M. (2008), Towards an Experimental Historical Phonetics: Italian speech is changing, in *Language Design. Journal of Theoretical and Experimental Linguistics. Special Issue 1: New Trends in Experimental Phonetics: Selected Papers From the IV International Conference on Experimental Phonetics*, 1, 201-209.
- Pettorino, M. & Giannini, A. (1994), Aspetti prosodici del parlato radiofonico, in *Atti delle IV Giornate di Studio del Gruppo di Fonetica Sperimentale (GFS)*, vol. XXI, Torino, 19-28.
- Pettorino, M. & Giannini, A. (2005), Italian TV Speech: a diachronic analysis, in *Proceedings of IX International Symposium on Social Communication*, vol. II, Santiago de Cuba, 849-853.
- Pettorino, M. & Giannini, A. (2010), Il parlato dei mass media, in *La Comunicazione Parlata 3*, vol. 2, Napoli, OPAR, Università degli studi di Napoli l'Orientale, 71-83.
- Pettorino, M. & Giannini, A. (2011). The Speaker's Age: a Perceptual Study, in *Proceedings of the 17th ICPHS*, Hong Kong, 1582-1585.
- Pettorino, M., Pellegrino, E. & Maffia, M. (2014), "Young" and "Old" Voice: the prosodic auto-transplantation technique for speaker's age recognition, in *Proceedings of the 7th In-*

ternational Conference on Speech Prosody (N. Campbell, D. Gibbon & D. Hirst, editors), 135-139.

Pettorino, M., Pellegrino, E., Salvati, L., Vitale, M. & De Meo, A. (2012), La voce dei media. Trapianti ritmico-intonativi per un'analisi diacronica dell'italiano parlato, in *Coesistenze linguistiche nell'Italia pre- e postunitaria*, Atti del XLV Congresso internazionale di studi della SLI (T. Telmon, G. Raimondi & L. Revelli, editors), Roma, Bulzoni Editore, 777-792.

Pettorino, M., & Vitale, M. (2012). Transplanting native prosody into second language speech. In M.G. Busà Maria Grazia, A. Stella (eds.), *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*, Padova, CLEUP, 11-16.

Ptacek, P.H. & Sander, E.K. (1966), Age recognition from voice, *Journal of Speech and Hearing Research*, 9, 273-277.

Ramig, L.A. & Ringel, R.L. (1983), Effects of physiological aging on selected acoustic characteristics of voice, *Journal of Speech and Hearing Research*, 26, 22-30.

Shipp, T. & Hollien, H. (1969), Perception of the aging male voice, *Journal of Speech and Hearing Research*, 12, 703-710.

Tompkins, C.A., Scharp, V.L. & Meigh, K.M. (2006), Communication Disorders in The Encyclopedia of Aging (R. Schultz editor), Fourth Edition, Springer Publishing, Vol. I A-K, 234-242.

Traunmüller, H. & van Bezooijen, R. (1994), The auditory perception of children's age and sex, *Proceedings of ICSLP 1994*, Yokohama, Japan, 3, 1171-1174.

Wohlert, A. & Smith, A. (1998), Spatiotemporal stability of lip movement in older adult speaker, *Journal of Speech, Language and Hearing Research*, 41, 41-50.

Yoon, K. (2007). Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody. *Journal of the Modern British & American Language & Literature* 25 (4), 197-215.