

# Differenze fonetiche generali tra voci femminili e voci maschili\*

Antonio Romano

Laboratorio di Fonetica Sperimentale "Arturo Genre", Università di Torino

## 1. Introduzione

Sappiamo dai manuali di fonetica e da testi specialistici, anche recenti, che variabili come l'altezza o il timbro di una voce dipendono da caratteristiche anatomiche come la lunghezza delle pliche vocali o le dimensioni delle cavità sopraglottidali (contributi vari in Schindler O. 2009; Henrich Bernardoni 2021).

Non ci sorprende quindi che a laringe con cartilagini tiroidee ben pronunciate (a cui corrispondono spazi laringei di maggiore estensione) corrispondano voci generalmente più gravi, cioè con valori di frequenza fondamentale ( $f_0$ ) più bassi.

È evidente quindi che, al di là dei condizionamenti culturali e delle abitudini personali, voci di individui che dipendono da fattori biologici / ormonali di tipo maschile tendono ad avere estensioni vocali in intervalli di frequenza più bassi di individui biologicamente femminili<sup>1</sup>.

Per comprendere questo punto, occorre approfondire alcune semplici caratteristiche della produzione dei suoni linguistici alla luce di: 1) la teoria sorgente-filtro (interazione tra caratteristiche laringee della voce ed effetti di risonanza dovuti all'articolazione dei suoni nel condotto vocale, Fant 1960), 2) l'osservazione delle modalità di sviluppo della curva di  $f_0$  nel corso degli enunciati (intonazione di frase), 3) lo studio contrastivo tra modelli diversi di dispersione statistica dei valori individuali di  $f_0$  (ispezio-

---

<sup>1</sup> Sappiamo, sin da Bralley *et alii* (1978) (ma v. anche interventi di G. Motta e D. Cossu), che la somministrazione di testosterone determina una maggiore crescita della laringe (con allungamento della glottide) provocando effetti di abbassamento generale di  $f_0$  nei transessuali gino-androidi (ai quali è d'uso riferirsi come parlanti *AFAB = assigned female at birth*). Da questi stessi studi si apprende che l'opposto non è facile da conseguire nel caso dei transessuali andro-ginoidi (*AMAB = assigned male at birth*) somministrando estrogeni. Ma, al di là dei trattamenti endocrinologici, una compensazione si può ottenere con un'adeguata riabilitazione vocale, consistente in attività che non devono prevedere **solo** un innalzamento generale di  $f_0$  (cfr. Panzeri & Falciati 2003; per altri spazi linguistici v. Spencer 1988).

---

\* Il presente contributo riassume alcuni contenuti presentati nel corso del mio intervento tenuto al corso di formazione «La voce parlata e cantata della persona transgender» organizzato a Torino da *Voce-InForma* nei giorni 2 e 3 dicembre 2023.

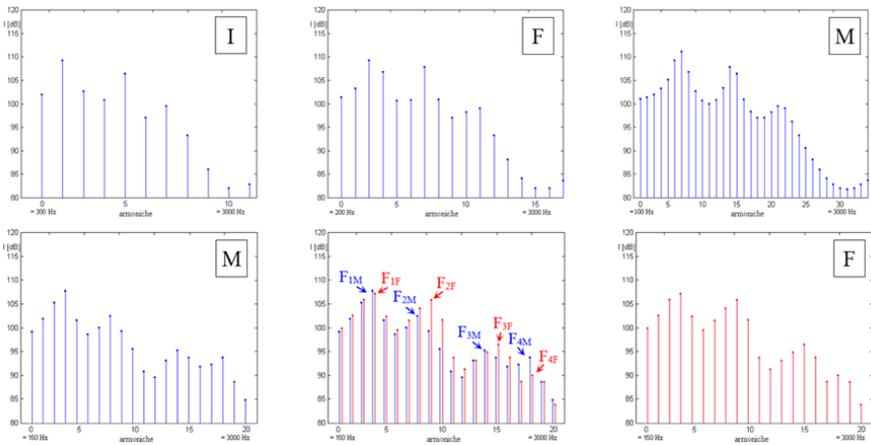


Fig. 1. In alto: strutture armoniche fittizie per profili timbrici simili di voci di tonalità diversa (di tipo infantile, I con  $f_0 = 300$  Hz, femminile, F con  $f_0 = 200$  Hz, maschile, M con  $f_0 = 100$  Hz). In basso: profili timbrici diversi per voci maschili (M, a sinistra) e femminili (F, a destra) a parità di  $f_0$  (= 150 Hz) (al centro: confronto tra i due profili M e F ravvicinati).

ne di curve di *densità di probabilità*, *ddf*).

Riguardo al primo punto (v. §3), ho avuto occasione di chiarire, con gli esempi riportati in Fig. 1 (in alto, cfr. Ferrero *et alii* 1979: 77), la minore definizione dello spettro acustico (decomposizione armonica di Fourier) di suoni prodotti da voci infantili (I, in alto a sinistra), femminili (F, in alto al centro) o, comunque, più acute, rispetto a voci maschili (M) caratterizzate da valori di  $f_0$  più bassi e conseguente *pattern* armonico più fitto (in alto a destra). D'altra parte, come si vede dai grafici della stessa figura (in basso), anche a parità di  $f_0$  e di qualità timbrica percepita, voci di tipo M o F presentano schemi formantici lievemente diversi (con valori di  $F_2$  o  $F_3$ , ma spesso anche delle altre formanti,

generalmente più alti nel caso di queste ultime, v. §3) (Fig. 1).

Per le voci maschili ciò vale prevalentemente nelle posizioni degli enuncianti in cui raggiungono valori più gravi, ma in queste posizioni non si può escludere che anche alcune voci femminili raggiungano limiti inferiori simili. A questo proposito vale la pena ricordare che non sono per nulla normativi campioni (anche di diverse centinaia di parlanti) se basati sulle caratteristiche di pronuncia di vocali isolate (sostenute per pochi secondi, per la realizzazione delle quali si determinano assetti in molti casi innaturali nel corso di produzioni decontestualizzate<sup>2</sup>. Si tratta di dati li-

<sup>2</sup> A questo riguardo si vedano i confronti tra voci di diversa età nel campione di Romano & De Iacovo (2021), v. dopo. Le differenze che

mitati, che non tengono conto degli intervalli esplorati nel corso di compiti linguistici più generali e che, riferendosi esclusivamente a specifiche popolazioni, non sono estendibili ad altri.

Non soddisfano neanche misurazioni di estensioni basate su pochi dati, senza considerare la variazione interna agli enunciati, così come restano generiche e abusive considerazioni del tipo «le voci degli uomini presentano un maggior numero di toni discendenti» (citata da altri lavori in Panzeri & Falciati 2003: 213)<sup>3</sup>.

In un manuale di prosodia tra i più influenti pubblicati negli ultimi anni in Italia si trova una definizione di

---

appaiono in questo campione di voci italiane presentano effetti poco sistematici su media e dispersione dei valori: i dati vocalici dei parlanti considerati meriterebbero inoltre approfondimenti anche in termini di *jitter* e *shimmer*, cioè di leggera variazione di frequenza o di ampiezza in cicli glottidali successivi.

<sup>3</sup> In queste condizioni, nel caso dell'inglese americano (dell'area di Toronto) ad es. si registra che «Older females had a lower  $f_0$  [...] than younger females, but younger and older males did not differ. Shimmer increased with age for males, but neither jitter nor shimmer increased with age for females, whereas noise measures were similar for both ages» (Goy *et alii* 2013). Su simili affermazioni, insieme ad altre piuttosto malferme, si basano lavori come quello di Gucciardo (2007) e Gucciardo *et alii* (2010), i quali – al di là di ottime intuizioni e buoni riferimenti culturali – propongono argomenti spesso esposti con formulazioni impressionistiche e senza riferimenti autorevoli.

estensione tonale come *pitch range*, riferito ai valori minimi e massimi «di un contorno intonativo» (Soriano 2006: 42), laddove per definire la distribuzione dei valori di  $f_0$  in produzioni con diversi livelli di attivazione si notano differenze non trascurabili (Romano 2000) ed è quindi indispensabile riferirsi a un *pitch range* ottenuto mediante valutazioni che si riferiscano a più enunciati e a diverse condizioni di produzione<sup>4</sup>.

D'altra parte, è molto limitante riportare solo indicativamente valori medi, minimi e massimi di  $f_0$  per distinguere voci maschili, femminili o infantili (riferendosi, come nell'esempio ideale proposto sopra, e come fanno alcuni manuali, a valori medi fittizi di tipo 100, 200 e 300 Hz oppure a valori di media e deviazione standard,  $dev.st = \sigma$ ). Come mostrerò, infatti, sulla scia di ricerche passate in rassegna in lavori precedenti, i valori di  $f_0$  si disperdono concentrandosi in alcuni intervalli di maggiore «comodità» della voce, ovvero in intervalli di altezza relativa nei quali occorre soffermarsi nel rispetto di vincoli accentuali, intonativi ed espressivi (cfr. Patterson & Ladd 1999). E ancora, come

---

<sup>4</sup> L'intervallo di variazione di  $f_0$  nelle voci è studiato in numerosi lavori in cui, oltre che come *pitch range* e *pitch span*, è indicato anche come *pitch compass* (v. Jassem 1971). Sulla distinzione tra «estensione» e «tessitura» v. Romano *et alii* (2012).

mostrano esempi di voci maschili di estensione molto acuta e voci femminili particolarmente gravi (si pensi ad es. a figure televisive note, come il giornalista Mario Giordano e la conduttrice Maria De Filippi), non è neanche scontato che individui biologicamente di un determinato sesso presentino voci attribuibili inequivocabilmente a quel sesso.

## 2. Maggiore variabilità nell'altezza delle voci femminili (ed effetto del sesso dell'ascoltatore)

Un parametro che contribuisce a influenzare l'identificazione del sesso di un parlante è anche il sesso dell'ascoltatore (Weston *et alii* 2015)<sup>5</sup>. Ad es. è più facile discriminare tra voci maschili in termini di identificazione (tanto per le donne quanto per gli uomini), mentre nel caso di voci ambigue le

ascoltatrici tendono più spesso a identificarle con voci maschili<sup>6</sup>.

Tuttavia, conoscendo le non linearità della percezione uditiva (che si riflettono sulle scale musicali di tipo logaritmico diffuse in molte culture), è intuitivo attendersi che sensazioni di altezza di scarti melodici simili in una scala più grave corrispondano a variazioni di maggiore estensione in un *range* più alto (si vedano in Fig. 2 le diverse modulazioni di due voci che producono enunciati corrispondenti a uno stesso contenuto frasale con grado di attivazione simile).

Nell'esempio di figura, al di là di due diverse modalità di accentazione che si manifestano con picchi di  $f_0$  più «accentuati» nel caso della voce femminile, le due voci percorrono lo spazio della loro tessitura con variazioni di estensione diversa. Fatte salve le mancate misure (errori di misurazione dell'algoritmo), la voce maschile in questo caso risulta compresa tra circa 103 e 183 Hz, mentre quella femminile (ignorando le false rilevazioni in coda all'ultimo vocoide) si estende da 176 a 307 Hz.

Partendo da queste considerazioni (e dalle limitazioni che prefigurano) proviamo a vedere come sia possibile classificare rigorosamente queste voci anche solo in riferimento alla loro variazione

---

<sup>5</sup> A questo riguardo, un altro aspetto rilevato in Joassin *et alii* (2011), riguarda l'associazione tra voci e volti che costituisce un elemento dell'identificazione: «Human social interactions are shaped by our ability to recognise people. Faces and voices are known to be some of the key features that enable us to identify individual people, and they are rich in information such as gender, age, and body size, that lead to a unique identity for a person. A large body of neuropsychological and neuroimaging research has already determined the various brain regions responsible for face recognition and voice recognition separately, but exactly how our brain goes about combining the two different types of information (visual and auditory) is still unknown».

---

<sup>6</sup> Un esperimento simile, limitato al parlato letto, ma basato su indici acustici segmentali, è quello di Munson *et alii* (2006).

di  $f_0$  in questi due enunciati (v. Fig. 3).

Ora, è vero che la  $f_0$  media del primo è a 146 Hz (con  $\sigma = 19$  Hz) e che l'estensione del suo intervallo di variazione è di 80 Hz (da 103 a 183) ed è pur vero che la  $f_0$  media della seconda è a 241 Hz (con  $\sigma = 41$  Hz) e che il suo *pitch range* è di 131 Hz (da 176 a 307). Tuttavia rileviamo come le due voci si siano soffermate più tempo intorno a certi valori che nel primo caso sono stati a 132, 154 (soprattutto) e 178 Hz mentre nel secondo caso si sono dispersi mag-

giormente, concentrandosi attorno a 180, 208, 246, 272 (soprattutto) e 306 Hz. Certo, si tratta di pochi valori, ma ciò che emerge dall'analisi quantitativa di un gran numero di produzioni di uno stesso locutore (come ho potuto mostrare sin da Romano 2000) si caratterizza comunque per una dispersione quasi mai riconducibile a una *ddp* gaussiana e quindi non rappresentabile con valori quali la media, la deviazione standard ( $\sigma$ ) o, ancor meno, il minimo e il massimo (o l'intervallo tra questi).

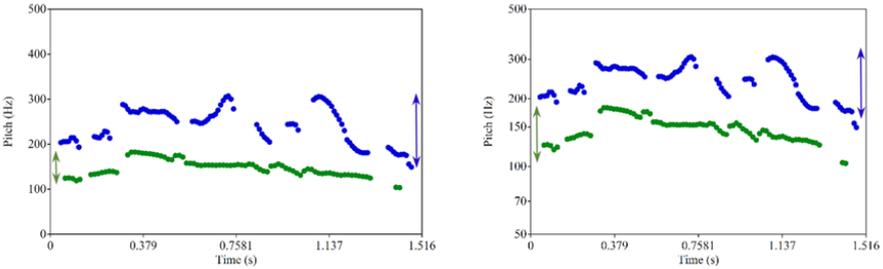


Fig. 2. Profili melodici simili per voci di estensione diversa rappresentati su scale diverse. A sinistra, scala lineare: la voce di tipo femminile (blu) si presenta con un'estensione apparentemente maggiore di quella di tipo maschile (verde). A destra, scala logaritmica: il range di variazione di  $f_0$  nei due casi appare congruente (le variazioni melodiche della voce di tipo maschile risaltano in modo più evidente e in accordo con le sensazioni di altezza che producono).

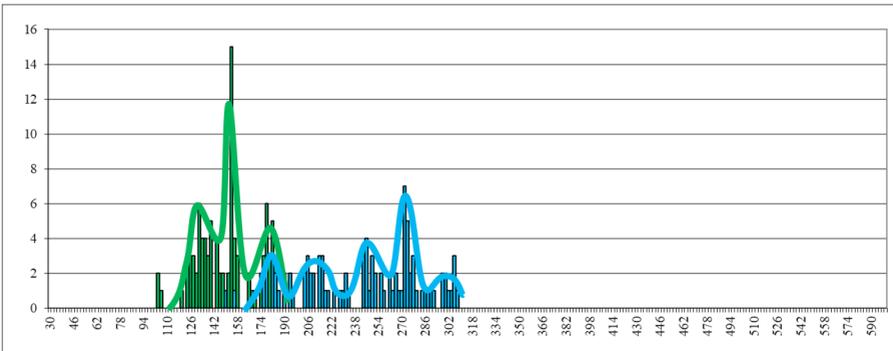


Fig. 3. Istogrammi di densità di probabilità con passo 2 Hz per le due voci di Fig. 2.

Nell'articolo Hudson *et alii* (2007) le  $f_0$  medie di 100 parlanti maschi britannici tra i 18 e 25 anni sono state analizzate statisticamente, mostrando una distribuzione tra i 90 e i 150 Hz. Si tratta di una distribuzione di medie (100 parlanti -> **100** valori) e non di una *ddp* cumulativa.

Nell'articolo di Romano & De Iacovo (2021), invece, di due gruppi di 58 parlanti maschi (29 di un gruppo più giovane, 18-25, e 29 di uno più anziano, 25-50) si sono valutate le dispersioni di valori di  $f_0$  di ciascuno di essi collassate in *ddp* cumulative di più di **150000** valori. La prima ha presentato una  $f_0$  media di 141 Hz ( $\sigma = 44$  Hz) e la seconda una  $f_0$  media di 126 Hz ( $\sigma = 44$  Hz). Si tratta di *ddp* simili a gaussiane ma piuttosto eccentriche (la moda è di 15-20 Hz più bassa). Si riporta in Fig. 4 la somma delle due *ddp* (con l'aggiunta di valori accuratamente rilevati tra 76 e 80 Hz). Si tratta di circa di **312000** valori con

media a 134 Hz e moda a 114 Hz.

Le voci femminili di quest'esperimento, rispettivamente quelle di 78 e 35 parlanti (più di **400000** misurazioni nel primo caso), presentano invece  $f_0$  media 209 Hz ( $\sigma = 49$  Hz) e  $f_0$  media 202 Hz ( $\sigma = 52$  Hz). Tuttavia la forma delle *ddp* non somiglia a una gaussiana, quanto piuttosto alla sovrapposizione di due «campane» pseudo-gaussiane: una più bassa, con moda a 80 Hz, e una più alta, con moda a 198 Hz, nel primo caso e una più bassa e un po' meno distinguibile, con moda sempre attorno a 80 Hz, e un'altra con moda a circa 180 Hz. I minimi delle due campane più alte si situano tra 100 e 130 Hz mentre i massimi, raggiunti con code molto smorzate, si trovano attorno a 380 Hz. Non si tratta quindi di estensioni valutabili rilevando i minimi e i massimi di singole produzioni che non avrebbero fatto apparire la composizionalità delle due *ddp*. Sebbene la campana più grave sia caratterizzata da

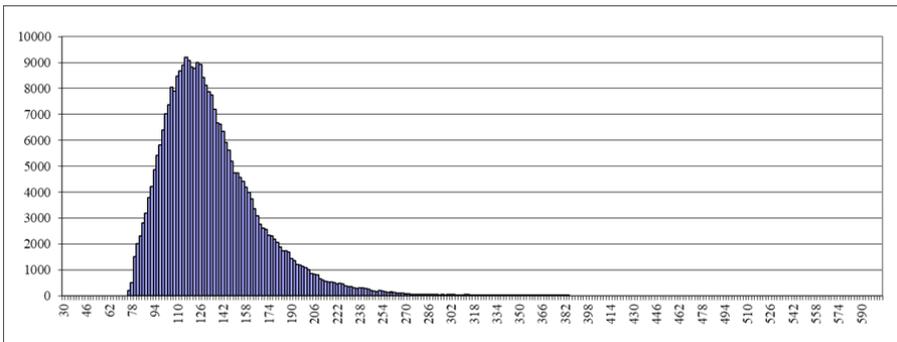


Fig. 4. Istogramma di densità di probabilità con passo 2 Hz per i due gruppi di voci maschili di Romano & De Iacovo (2021) [dati unificati e verificati manualmente tra 76 e 80 Hz].

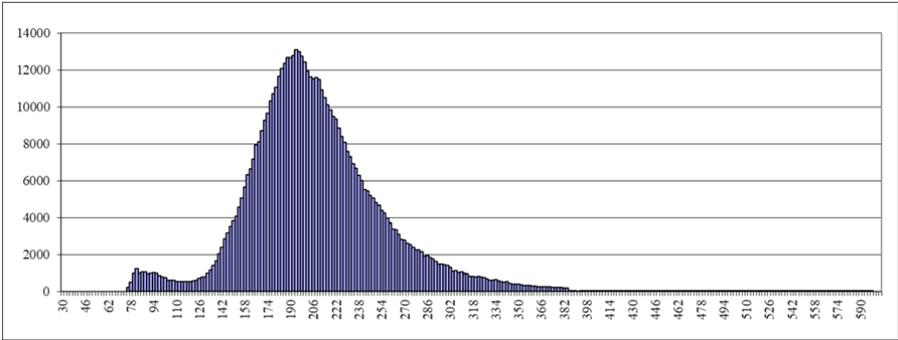


Fig. 5. Istogramma di densità di probabilità con passo 2 Hz per le voci femminili di Romano & De Iacovo (2021) [dati unificati e verificati manualmente tra 76 e 80 Hz].

un numero inferiore di valori, le due campane sembrano rispondere al ricorso da parte di questi gruppi di parlanti a due distinti meccanismi di produzione della voce: uno ( $f_0$ M0 di Roubeau *et alii* 2009 e Henrich Bernardoni 2021), usato meno frequentemente, in un registro di voce cricchiata e un altro (M1), ampiamente esplorato – ma, appunto, con minimi tra 100 e 130 – nel registro di voce modale. Anche in questo caso le due *ddp* sono state unificate e i valori tra 76 e 80 Hz verificati manualmente. Si riporta quindi in Fig. 5 la *ddp* risultante, basata su circa di **584000** valori,

con media a 207 Hz e moda a 192 Hz.

Dal confronto diretto tra le due *ddp* risultanti si può osservare come le voci ricoprano in buona misura lo stesso spazio di variazione di  $f_0$  nelle basse frequenze (v. Fig. 6). Per questi parlanti di italiano, non esiste un modo netto di separare le due *ddp*: tra i 75 Hz e i 265 entrambe le voci coesistono, anche se con diversa probabilità. Fissando una soglia arbitraria di 1000 valori, l'intervallo di maggior ricoprimento si potrebbe circoscrivere tra i 130 e i 200 Hz, con equiprobabilità a circa 155 Hz: intorno

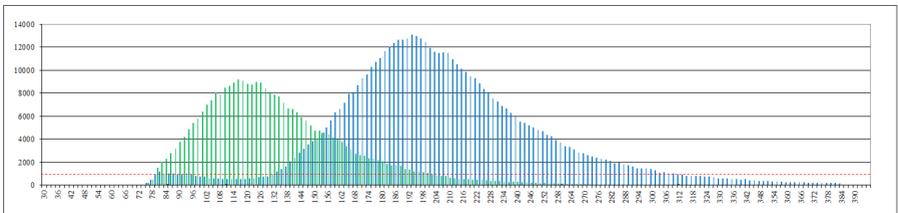


Fig. 6. Istogrammi di densità di probabilità con passo 2 Hz per le voci maschili e femminili di Romano & De Iacovo (2021).

a questi valori (per i quali sono state misurate migliaia di valori, > 4000!) una voce ha le stesse probabilità di essere considerata maschile o femminile (fatti salvi gli altri fattori anticipati)<sup>7</sup>.

### 3. Caratteristiche timbriche di voci di tipo maschile e femminile

Nonostante le numerose omissioni teoriche e i difetti procedurali che affliggono il loro studio, Panzeri & Falciati (2003) giungono a conclusioni

---

<sup>7</sup> Osservo tuttavia che anche tra gli 80 e i 100 Hz la soglia di 1000 valori è raggiunta: a 85 Hz le voci di tipo femminile hanno la metà delle probabilità di muoversi in questo *range* rispetto a quelle di tipo maschile (riconquistando però l'equiprobabilità nei rari casi in cui la voce si localizza sotto gli 80 Hz, cioè nel registro di cricchiato, meccanismo M0). Alla luce di tutto ciò, non deve quindi sorprendere se un valore medio di 155 Hz, inferiore a 160 Hz, faccia «già identificare come femminile la voce di un transessuale *male-to-female*»: uno scarto tra un valore medio di  $f_0$  e una soglia già di per sé non è significativo senza indicazione della varianza o, appunto, dell'estensione. In questo caso, inoltre, il valore assunto come indice presenta uno scarto risibile rispetto a una soglia arbitraria (per di più attribuita a Wolfe *et alii* 1990 che, in riferimento all'inglese, cioè per una comunità di parlanti ben precisa, stabiliscono un limite a 155 Hz!). Sorprende invece che un test di attribuzione di genere, condotto su 8 voci italiane e isolando solo produzioni di /a/ (Panzeri & Falciati 2003: 215), abbia potuto confermare la validità di una simile soglia. Approfitto per segnalare che la terminologia riguardo a questi temi è quanto mai soggetta a rapida obsolescenza (basti pensare al cambiamento di titolo che si è reso necessario per la terza edizione, 2019, di Adler *et alii* 2006).

verificabili. Partendo da una premessa sostanzialmente corretta, conforme a una visione semplicistica della teoria-sorgente-filtro, «è stato dimostrato che quando viene eliminata la differenza della frequenza fondamentale, le altre formanti [sic] aiuterebbero l'ascoltatore a percepire la mascolinità o la femminilità della voce» (p. 214). Si tratta evidentemente di un campo esplorato da numerosi autori e con varie metodologie. Si vedano, ad es., Halberstam & Raphael (2004), e riferimenti citati, che ribadiscono l'importanza dell'interazione tra  $f_0$  e l'identificazione dei *pattern* vocalici (e viceversa), e Camargo *et alii* (2012), che indicano la pendenza spettrale come altro fattore dirimente.

D'altra parte l'interazione è nota anche da studi di diversi ricercatori italiani (Barillari *et alii* 1995, Ferrero *et alii* 1995, Maisano 1996, Calamai 2003) che hanno fornito i materiali di riferimento, partendo da Fant (1978) e fonti citate (v. in particolare §7, p. 81 «Male-female difference», v. Fig. 7a), oltre che da Ferrari-Disner (1980), tra i primi a proporre metodi di valutazione oggettivi anche in termini sensoriali.

Oltre a differenze acustiche specifiche, gli effetti percettivi di  $F_2'$  (cioè dei valori efficaci di  $F_2$  pesati da altre componenti energetiche presenti nello spettro acustico) sono stati studiati in Ferrero *et alii* (1996) e, nell'ottica di lavori di normalizzazione, in Pätzold &

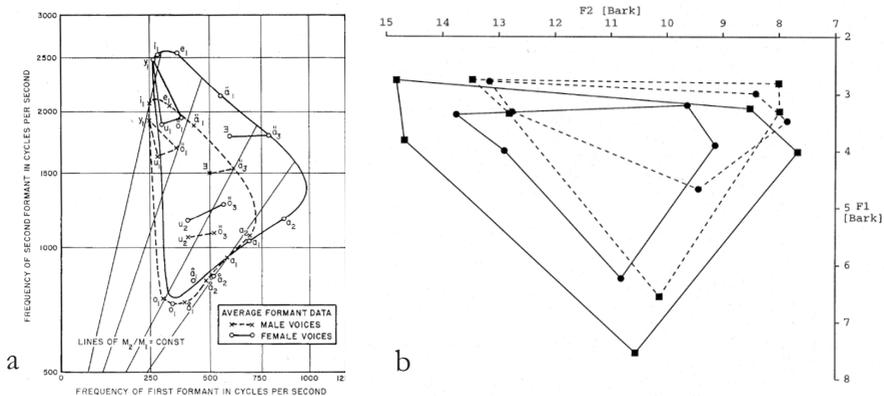


Fig. 7. Differenze tra dispersioni di valori formantici ( $F_1$  e  $F_2$ ) per uomini e donne di lingua svedese (6a) e tedesca (6b) [tratterispettivamente da Fant (1978) e Pätzold & Simpson (1997)].

Simpson (1997) (v. Fig 7b)<sup>8</sup>. Il metodo di confronto tra voci sulla base della superficie occupata dai poligoni definiti dalla dispersione dei valori formantici è stato poi studiato per l'italiano (fiorentino e pisano) da Calamai (2004) e Calderone & Calamai (2007)<sup>9</sup>. La valutazione dello spazio articolatorio si fa

in questi casi considerando la superficie dello spazio acustico-articolatorio dei diversi parlanti e valutando un coefficiente che stabilisce la relazione di proporzione: i poligoni di tipo maschi-

<sup>8</sup> Citando Peterson & Barney (1952), Ferrero *et alii* (1979: 78) sottolinea come si trovino differenze per i valori formantici di una «stessa vocale» pronunciata da un uomo, da una donna o da un bambino «Le differenze uomo-donna per  $F_1$  variano tra il 5% e il 30%, per  $F_2$  tra il 10% e 25% e per  $F_3$  tra il 10% e il 20%». Riguardo a  $F_2$ , già in Ferrero *et alii* (1979: 124) si definisce una «seconda formante percettiva calcolata [...] a partire dai valori delle prime quattro formanti acustiche». In riferimento a Bladon & Fant (1978) si aggiunge poi che questi «hanno fornito una definizione più perfezionata di  $F_2$  nell'ambito di un modello a due formanti delle vocali cardinali» (pp. 124-125).

<sup>9</sup> L'importanza di queste conoscenze sfugge a molti ricercatori di ambito medico-logopedico che non accedono alla lettura fondamentale del campo fonetico e/o non si riferiscono alle ricerche che sono state condotte in Italia per decenni facendo luce su tutti questi aspetti e contribuendo a definire una disciplina. In mancanza di questi riferimenti, lavori come quelli menzionati sopra potrebbero presentare risultati fuorvianti o comunque non rappresentativi dello stato di conoscenze raggiunto in un determinato periodo all'interno di una comunità scientifica (alla quale si approssimano acriticamente, ripartendo da conoscenze relative ad altre epoche e ad altri spazi linguistici, mentre invece, nell'estrema semplicità del loro esperimento, passando in rassegna diversi lavori, Sulpizio *et alii* 2015 hanno messo in evidenza l'essenzialità della dipendenza dalla lingua).

le, essendo associati a condotti vocali stimati mediamente di lunghezza 17,5 cm (Giannini & Pettorino 1992), hanno minore estensione (generalmente in entrambe le dimensioni definite da  $F_1$  e  $F_2$ ) rispetto a quelli di tipo femminile. In questi casi infatti le ridotte dimensioni antropometriche definiscono una minore lunghezza dello spazio risonanziale (15,5 cm in stime medie piuttosto grossolane) a cui corrisponde (essendo la frequenza inversamente proporzionale alla lunghezza d'onda) un «poligono formantico» più esteso.

Esempi di confronto tra sistemi vocalici di parlanti ben differenziati (per l'area barese) sono proposti anche in Romano & Manco (2004). I valori della frequenza delle prime due formanti discussi in questo studio (v. Fig. 8) sono compatibili con quelli presentati da Ferrero *et alii* (1979)<sup>10</sup>.

<sup>10</sup> I grafici di riferimento per voci maschili e femminili si trovano alle pp. 122-123 (v. Appendice D), insieme a molte considerazioni che – ignorando gli studi di Fant e quello di Pätzold & Simpson (1997) – devono essere sfuggite a Panzeri & Falciani (2003). Pur dedicando a F. Ferrero il loro articolo (e avendo inserito tra i riferimenti proprio questo manuale) non sembrano infatti aver preso in considerazione i passaggi che li avrebbero scoraggiati a basare le loro valutazioni su  $F_2$  e  $F_3$  (e non su  $F_1$  o  $F_2$ ). In particolare, in Ferrero *et alii* (1979) si dice che il diagramma di esistenza  $F_3$ – $F_2$  «dimostra l'inadeguatezza di  $F_3$  a differenziare tra loro, altrettanto nettamente di  $F_1$ – $F_2$ , le sette vocali italiane» (pp. 123-124). Vi si legge anche: «è stato dimostrato che le formanti superiori alle prime due aggiungono poco» (nella definizione del timbro).

L'area dei poligoni definiti dai centroidi delle dispersioni di valori presentate dai 4 speaker considerati è di circa 1089000 Hz<sup>2</sup> e 1024000 Hz<sup>2</sup> per i due locutori (M) e di circa 1474000 Hz<sup>2</sup> e 1317000 Hz<sup>2</sup> per le due locutrici (F). Si hanno quindi coefficienti compresi tra 1,3 e 1,4 per passare da una dispersione tipica di queste voci maschili a quella delle voci di tipo femminile che riproducono contrasti simili (si vedano le aree di esistenza dei fonemi)<sup>11</sup>.

Come si vede dai grafici di Fig. 9, tuttavia, l'espansione deve interessare diversamente i vari timbri, dato che ad es. per le realizzazioni di /u/ i valori delle formanti misurati per i locutori e le locutrici risultano molto simili, mentre è soprattutto per /i/ e per /a/ che si creano delle divergenze acusticamente molto rilevanti.

<sup>11</sup> Per offrire un altro riferimento, rispetto ai valori qui presentati, riassumo in Appendice I quelli prototipici desumibili dai grafici di Ferrero *et alii* (1979: 122-123). Con quei valori, le superfici dei poligoni valgono rispettivamente circa 790000 Hz<sup>2</sup> (M) vs. 1076000 Hz<sup>2</sup> e il valore del coefficiente di conversione spaziale M → F risulta 1,36 (compatibile con quanto visto sopra). L'applicazione di questo coefficiente richiede però di lavorare a partire da diagrammi di esistenza simili, mentre naturalmente questo non è facile da conseguire in un Paese come l'Italia caratterizzato da forte variazione dialettale (in termini geografici). I valori rilevati per  $F_2$  di /a/ in Panzeri & Falciani (2003) sembrano quindi corrispondere ai valori medi che si riscontrano in voci tipizzate come M o F e sono in accordo con i valori riportati anche negli studi citati sopra (apparsi per la verità in anni seguenti). Alla

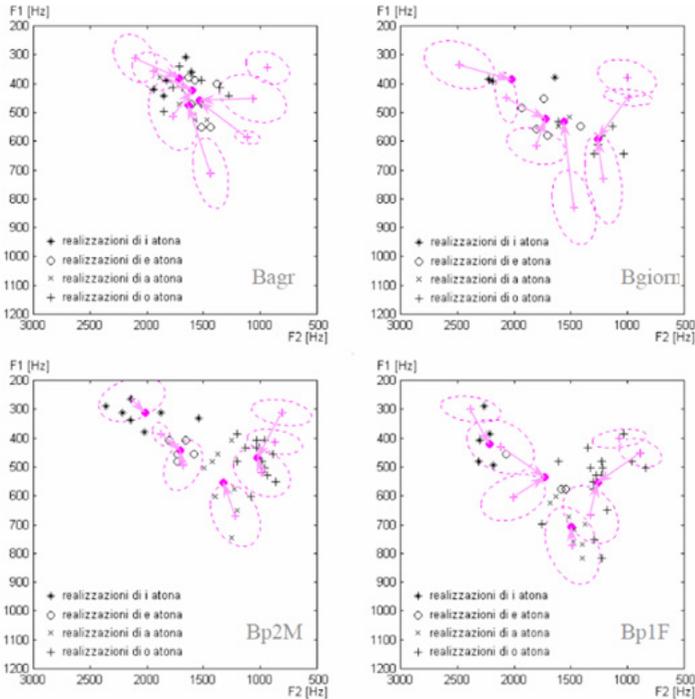


Fig. 8. Dispersioni di valori formantici ( $F_1$  e  $F_2$ ) per uomini e donne baresi (un agronomo e una giornalista, in alto, e due studenti, M e F, in basso) [tratte da Romano & Manco 2004].

Il lavoro di mappatura che ristabilisce, a un livello solo teorico, le somiglianze che emergono in termini

differenza di  $F_2$  per /a/ tra M (ca. 1250 Hz) e F (ca. 1450) rispondono mediamente bene i valori medi ricavabili dai valori di  $F_2$  tabulati per i gruppi di controllo ( $F_{2M} = 1241$  Hz e  $F_{2F} = 1503$ ). Tuttavia la dispersione di valori per le voci maschili ( $\sigma$  per  $F_{2M} = 226$  Hz vs.  $\sigma$  per  $F_{2F} = 145$  Hz) differisce notevolmente da quella osservata negli studi citati sopra (v. anche le voci maschili di Fig. 7, con  $F_{2M}$  media = 1332 Hz e  $\sigma = 138$  vs.  $F_{2F}$  media = 1478 con  $\sigma = 149$ ) e testimonia di una variazione che interessa potenzialmente anche rilevanti differenze nelle rese fonetiche (alcuni speaker di controllo hanno valori di  $F_2$

perceptivi tra sistema di contrasti realizzato da voci maschili e femminili deve quindi prevedere una pondera-

per /a/ che fanno pensare a rese di tipo [a]. L'uniformità nella variazione diatonica perseguita invece negli studi riassunti nella rassegna di Calamai (2004) (o nei dati riportati nei diagrammi di Fig. 7), oltre a limitare le differenze tra M e F per una maggiore convergenza intra-comunità si legano anche a una variazione più contenuta nei valori medi nelle realizzazioni di questa vocale rappresentata da allofoni simili, laddove una pratica riabilitativa dovrebbe poter considerare anche le specificità geografica del campione di riferimento, assicurando anche il passaggio da un allofono all'altro (v. dopo).

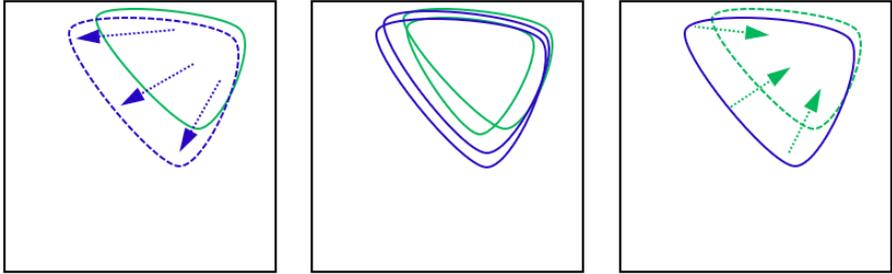


Fig. 9. Spazi vocalici idealizzati di voci maschili (M, verde) e femminili (F, blu) (in riferimento a quelli definiti da locutori e locutrici di Fig. 7). Il passaggio M>F dovrebbe prevedere un'estensione dello spazio vocalico (come nel grafico a sinistra), mentre il passaggio F>M dovrebbe prevedere una contrazione (come nel grafico a destra).

zione di questo coefficiente (della frazione che vogliamo restituire alle due dimensioni).

Ovviamente questo può ottenersi, diversamente, attraverso una mappatura su un piano formantico «deformato» in modo da far coincidere meglio le aree per mezzo di misure indirette (sottraendo costanti e introducendo coefficienti).

Diverse soluzioni erano già state proposte sin dagli anni '80, al punto che per rispecchiare «meglio il modo in cui suonano all'orecchio», Uguzzoni (1992) suggeriva di lavorare con scale di rappresentazione (in bark) in cui si riportano i valori di *spread* (cioè  $F_2 - F_1$ ) vs. *flat* ( $F_2 + F_1$ ), valutando anche, in un'ottica di valutazione congiunta tra valori risonanziali e componenti armoniche, « $F_1 - F_0$ » (Uguzzoni 1992: 32-33).

Tuttavia, come hanno mostrato Ferrero *et alii* (1995: 263, sulla base dei valori della tabella di p. 261 riprodotta

in Appendice II), «la normalizzazione “ $F_2 - F_1$  in Bark” è utile per le vocali posteriori /u, o, ɔ/ e la centrale /a/ per le quali la differenza tra VM e VF non è significativa ( $p < 005$ ). Non è così per le vocali anteriori /i, e, ε/ per le quali la distanza tra  $F_2$  e  $F_1$  nelle vocali femminili è significativamente maggiore rispetto a quelle maschili»<sup>12</sup>.

Ecco quindi che sono soprattutto le vocali anteriori che, nel passaggio andro-ginoide, dovrebbero essere articolate in modo da aumentare le dimensioni dello spazio di articolazione

<sup>12</sup> Sempre al livello teorico, per una rappresentazione maggiormente compatibile con valutazioni sensoriali, Ferrero *et alii* (1995: 263, v. ultimo diagramma in Appendice II) propongono di basarsi su diagrammi che abbiano in ascissa la differenza  $F_2 - F_1$ , in bark, e in ordinata  $F_1 - F_0$ , in bark. A questo riguardo, si noti che una delle formule usate per il passaggio da Hz a bark è oggi quella di Traunmüller:  $f_{\text{bark}} = [(26,81 * f_{\text{Hz}}) / (1960 + f_{\text{Hz}})] - 0,53$ .

(v. Fig. 9 a sinistra), e nel passaggio gino-androide dovrebbe invece ricentrarsi in uno spazio acustico contratto con valori di formanti più bassi (v. Fig. 9 a destra)

Come sia possibile perseguire questi due opposti scopi resta una materia complessa. L'eventuale obiettivo potrebbe essere raggiunto adottando metodi di compensazione articolatoria che risultano idealmente più facilmente risolvibili nel caso  $M > F$ , per via del fatto che un condotto vocale anatomicamente più lungo può essere sfruttato in una sua dimensione ridotta. Risulta invece oggettivamente più difficile nel caso  $F > M$  abbassare le formanti aumentando la lunghezza del condotto, anche se resta vero che una disposizione più avanzata nell'articolazione dei suoni anteriori, così come una maggiore protrusione labiale o una maggiore apertura ottenuta con l'abbassamento della mandibola (nonché della radice della lingua e della laringe), potrebbero offrire un *setting* articolatorio maggiormente orientato verso questi risultati. In questa condizione però abbiamo già visto che i trattamenti farmacologici sortiscono effetti migliori sulla laringe. E, infatti, per tornare ai limiti di una teoria sorgente-filtro banalizzata a una separazione totale dei contributi di laringe e cavità sopraglottidali, giocherebbe (e gioca) un ruolo fondamentale un lavoro sull'assetto del-

la laringe, condizionata in modo da produrre e muovere armoniche tali da stimolare le risonanze rispettivamente più acute o più gravi nell'inviluppo risonanziale del condotto vocale (v. Fig. 1) con dinamiche tali da assicurare uno spazzolamento (*sweeping*) di un *range* di  $f_0$  corrispondente a quello del modello di voce desiderato (v. Fig. 6) purché compatibile con le proprie condizioni anatomiche.

#### 4. Conclusioni

Con il breve intervento riassunto in questo contributo ho inteso ristabilire un ordine di priorità in merito alla consultazione di studi scientifici precedenti sulle caratteristiche vocali di parlanti maschi e femmine di varietà di italiano. Lo studio non pretende di essere esaustivo, ma, sulla base di un certo numero di esempi e riflessioni, ha mostrato come non sia prioritaria la ricerca di metodi di studio che – ignorando lo stato dell'arte in materia – ripartano da zero e fondino la loro conoscenza su metodi sviluppati per le comunità linguistiche maggiormente rappresentate nelle riviste internazionali.

I dati riportati e le discussioni scaturite dall'osservazione di questi mostrano come spesso – purtroppo ancora in tempi recenti – questi condizionamenti pregiudichino ricerche condotte in ambito medico-logopedico, nel quale è particolarmente

sconsigliabile che i gruppi di ricerca che lavorano su questi temi sensibili non partano da un'adeguata formazione interdisciplinare. Questa si è già rivelata imprescindibile sin dai primi lavori in questo campo avviati in Italia in seno al *Gruppo di Fonetica Sperimentale* dell'*Associazione Italiana di Acustica* e proseguiti poi anche all'interno dell'*Associazione Italiana di Scienze della Voce* e, in parte, del *Gruppo di Studio sulla Comunicazione Parlata* della *Società di Linguistica Italiana*.

Il quadro qui offerto si è soffermato dapprima su valutazioni di misurazioni di  $f_0$  eseguite su parlanti italiani con voci di tipo maschile e femminile appartenenti a un campione di dimensioni mai raggiunte finora. Sulla base dei dati discussi, ha permesso di concludere considerando con maggiore cautela le reali estensioni dei valori assunti da questa variabile prima di incoraggiare gli interessati a spingere la propria voce su valori molto spesso solo stereotipati sulla base di intuizioni non verificate.

In un secondo tempo, l'articolo si è concentrato sul confronto dei timbri oggettivi e delle sensazioni da questi generate in funzione di altri fattori, vincoli strutturali e limitazioni statistiche. Alla discussione dei valori delle formanti studiate da vari autori per voci di tipo maschile e femminile nei sistemi di varie comunità italiane (Firenze, Padova, Bari e Lecce) si è aggiunta una

valutazione delle possibilità di normalizzazione applicate in Italia alla fine degli anni '90. L'obiettivo, oltre che tener conto le varie scale di conversione, utili ai fini di un confronto tra le distinte dispersioni di valori determinate da caratteristiche anatomiche, è ovviamente soprattutto quello di riportare la qualità acustica dei timbri alle possibilità plastiche del condotto vocale.

Rispetto alle ricerche condotte alcuni decenni fa, in riferimento a quelle di studiosi interessati ad altri spazi linguistici, con diversi condizionamenti culturali e socio-fonetici, e grazie alla disponibilità di risultati ottenuti recentemente in lavori sperimentali condotti proficuamente in Italia da specialisti del settore, il ventaglio di soluzioni proponibili è cambiato considerevolmente. Le conoscenze oggi disponibili grazie alla facilità di circolazione delle informazioni e la maggiore diffusione di strumenti che agevolano le misurazioni e le valutazioni statistiche possono infatti rivelarsi utili per comprendere meglio la materia, aiutando a predisporre metodologie efficaci per un *counselling* professionale.

### Appendice bibliografica

Adler R.K., Hirsch S. & Mordaunt M./ Pickering J. (2006). *Voice and Communication Therapy for the Transgender/ Transsexual Client*. San Diego: Plural (3<sup>a</sup> ed. 2019, > *Voice and Communication Therapy for the Transgender/ Gender Diverse Client*).

- Barillari U., Giannini A., Pettorino M. & Siani M. (1995). «Indici acustici e movimenti articolatori nella voce maschile e femminile». *Studi Italiani di Linguistica Teorica e Applicata*, XXIV/2, 265-276.
- Calamai S. (2003). «Vocali d'Italia. Una prima rassegna». In: P. Cosi, E. Magno Caldognetto, A. Zamboni (a cura di), *Voce Canto Parlato. Studi in onore di Franco Ferrero*, Padova: Unipress, 49-57.
- Calamai S. (2004). «Vocali fiorentine e vocali pisane a confronto». In: F. Albano-Leoni, F. Cutugno, M. Pettorino & R. Savy (a cura di), *Il Parlato Italiano* (Atti del Convegno Naz., Napoli, 13-15 febbraio 2003), Napoli: D'Auria (CD-ROM).
- Calderone B. & Calamai S. (2007). «Lo spazio delle vocali». In: V. Giordani, V. Bruseghini & P. Cosi (a cura di), *Scienze Vocali e del Linguaggio - Metodologie di Valutazione e Risorse Linguistiche* (Atti del III Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce, ITC-IRST Povo - Trento, 29 nov. - 1 dic. 2006), Torriana (RN): EDK, 77-89.
- Camargo Z., Madureira S., Pessoa A. & Rusilo L. (2012). «Voice Quality and Gender: Some Insights on Correlations between Perceptual and Acoustic Dimensions». *Proc. Speech Prosody 2012* (Shanghai, 22-25 maggio 2012), 115-118.
- Fant G. (1960). *Acoustic Theory of Speech Production*. L'Aja: Mouton.
- Fant G. (1978). «Vowel perception and specification». *Rivista Italiana di Acustica*, II(2), 69-87.
- Ferrari-Disner S. (1980). «Evaluation of vowel normalization procedures». *Journal of the Acoustical Society of America*, 67, 253-261.
- Ferrero F., Genre A., Boë L.J. & Contini M. (1979). *Nozioni di fonetica acustica*. Torino: Omega.
- Ferrero F., Magno Caldognetto E. & Cosi P. (1995). «Le vocali al femminile». In: G. Marcato (a cura di), *Dialettologia al femminile. Atti del Conv. Intl. di Studi* (Sappada, 26-30 giugno 1995), Padova: Cleup, 413-436.
- Ferrero F., Magno Caldognetto E. & Cosi P. (1996). «Sui piani formantici acustici e uditivi delle vocali di uomo, donna e bambino». In: A. Peretti & P. Simonetti (a cura di), *Atti del XXIV Convegno Naz. AIA* (Trento, 12-14 giugno 1996), 169-178.
- Giannini A. & Pettorino M. (1992). *La fonetica sperimentale*. Napoli: Ed. Scientifiche Italiane.
- Goy H., Fernandes D.N., Pichora-Fuller M.K. & van Lieshout P. (2013). «Normative Voice Data for Younger and Older Adults». *Journal of Voice*, 27 (5), 545-555.
- Gucciardo A.G. (2007). *Voce e sessualità*. Torino: Omega.
- Gucciardo A.G., Merendi E., Paolillo N.P., Fussi F. (2010). «Sulla voce del transgender /Transessuale. Alcune osservazioni di sesso-vocologia generale

e artistica». In: F. Fussi (a cura di), *La voce del cantante*, VI, Torino: Omega, 503-509.

Halberstam B. & Raphael L.J. (2004). «Vowel normalization: the role of fundamental frequency and upper formants». *Journal of Phonetics*, 32, 423-434.

Henrich Bernardoni N. (2021). «La voce umana, dal respiro al canto». *Bollettino del LFSAG*, 7, 43-57 [www.lfsag.unito.it/ricerca/phonews/07/7\_3.pdf]

Hudson T., de Jong G., McDougall K., Harrison Ph. & Nolan F. (2007). «F0 Statistics for 100 Young Male Speakers of Standard Southern British English». *Proc. of ICPHS XVI* (Saarbrücken, 6-10 agosto 2007), 1809-1812.

Jassem W. (1971). «Pitch and compass of the speaking voice». *Journal of the International Phonetic Association*, 1, 59-68.

Joassin F., Pesenti M., Muraige P., Verreckett E., Bruyer R. & Campanella S. (2011). «Cross-modal interactions between human faces and voices involved in person recognition». *Cortex*, 47, 3, 367-376.

Kachel S., Radtke A., Skuk V.G., Zäske R., Simpson A.P. & Stefens M.C. (2018). «Investigating the common set of acoustic parameters in sexual orientation groups: A voice averaging approach». *PLOS ONE*, 13(12), e0208686.

Maisano L. (1996). «Sistemi vocalici maschili e femminili a confronto». In: A. Peretti & P. Simonetti (a cura di), *Atti del XXIV Convegno Naz. ALA* (Trento, 12-14 giugno 1996), 183-186.

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». *Journal of Phonetics*, 34(2), 202-240.

Oates J. & Dacakis G. (1983). «Speech Pathology Considerations in the Management of Transsexualism: A Review». *British Journal of Disorders of Communication*, 18(3), 139-151.

Panzeri M. & Falciati L. (2003) «Percezione della voce di parlanti transessuali: Requisiti acustici e percettivi per l'identificazione del genere». In: P. Cosi, E. Magno Caldognetto, A. Zamboni (a cura di), *Voce Canto Parlato. Studi in onore di Franco Ferrero*, Padova: Unipress, 213-218.

Patterson D. & Ladd R. (1999). «Pitch range modeling». *ICPhS99*, 1169-1172.

Pätzold M. & Simpson A.P. (1997). «Acoustic analysis of German vowels in the Kiel Corpus of Read Speech». *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel*, 32, 215-247.

Ralph C. Bralley, Glen L. Bull, Cheryl Harris Gore, Milton T. Edgerton (1978). «Evaluation of vocal pitch in male transsexuals». *Journal of Communication Disorders*, 11(5), 443-449

Romano A. (2000). «Statistiche di frequenza fondamentale per uno stesso locutore in diverse condizioni di produzione». *Atti del 28° Convegno Nazionale dell'Associazione Italiana di*

*Acustica* (Trani, 10-13 giugno 2000), 249-252.

Romano A. & De Iacovo V. (2021). «Statistiche di f0 per 200 parlanti di italiano». *Bollettino LFSAG*, 8, 21-33 [http://www.lfsag.unito.it/ricerca/phonews/08/8\_2.pdf]

Romano A. & Manco F. (2004). «Incidenza di Fenomeni di Riduzione Vocalica nel Parlato Spontaneo a Bari e a Lecce». In: F. Albano Leoni, F. Cutugno, M. Pettorino & R. Savy (a cura di), *Il Parlato Italiano* (Atti del Convegno Naz., Napoli, 13-15 Febbraio 2003), Napoli: D'Auria (CD-ROM).

Romano A., Cesari U., Mignano M., Schindler O. & Venero I. (2012). «Voice Quality» / «La qualità della voce». In: A. Paoloni & M. Falcone (a cura di), *La voce nelle applicazioni* (Atti dell'VIII Convegno dell'Associazione Italiana Scienze della Voce, Roma, 25-27 gennaio 2012), Roma: Bulzoni, 75 (art. int. CD 35 pp.).

Roubeau B., Henrich N. & Castellenigo M. (2009). «Laryngeal Vibratory Mechanisms: The Notion of Vocal Register Revisited». *Journal of Voice*, 23(4), 425-438.

Schindler O. (a cura di) (2009). *La voce: fisiologia, patologia clinica e terapia*. Padova: Piccin.

Spencer L. (1988). «Speech characteristics of male-to-female transsexuals: A perceptual and acoustic study». *Folia Phoniatrica*, 40(1), 31-42.

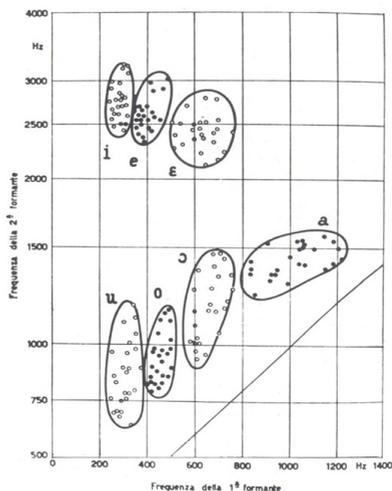
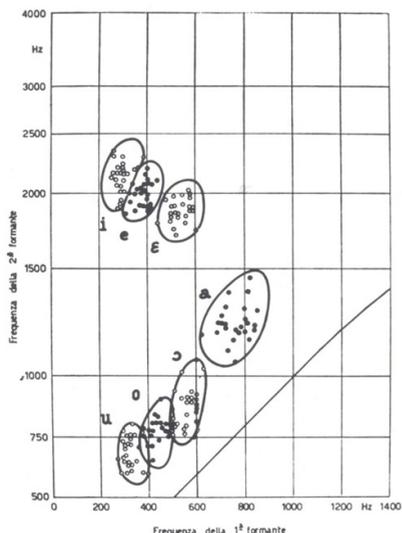
Sulpizio S., Fasoli F., Maass A., Paladino M.P., Vespignani F., Eyssel F. & Benteler D. (2015). «The Sound of Voice: Voice-Based Categorization of Speakers' Sexual Orientation within and across Languages», *PLOS ONE*, 10(7), e0128882.

Uguzzoni A. (1992). «Rappresentazione e classificazione delle vocali in termini uditivi». *Atti delle 3<sup>e</sup> Giornate di Studio del GFS* (Padova, 19-20 nov. 1992), 27-35.

Wolfe V.I., Ratusnik D.L., Smith F.H. & Northrop, G. (1990). «Intonation and fundamental frequency in male-to-female transsexuals». *Journal of Speech & Hearing Disorders*, 55(1), 43-50.

Weston Ph.S.J. et alii (2015). «Discrimination of voice gender in the human auditory cortex». *NeuroImage*, 105, 208-214.

## Appendice I



Diagrammi F1-F2 riprodotti da Ferrero *et alii* (1979: 122-123) (a loro volta ripresi da quelli pubblicati alle pp. 11 e 12 di F. Ferrero «Caratteristiche acustiche di fonemi vocalici italiani», *Parole e Metodi*, 3, 1972, pp. 9-31). A sinistra grafico con le aree di esistenza dei 7 fonemi definita sulla base delle realizzazioni di 25 informatori maschili. A destra le aree di esistenza ottenute partendo dalle realizzazioni di 25 soggetti femminili (v. valori approssimativi nella tabella sotto).

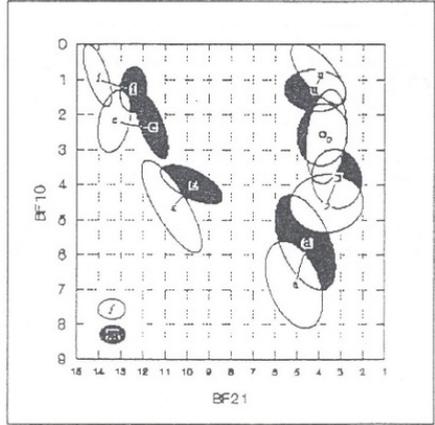
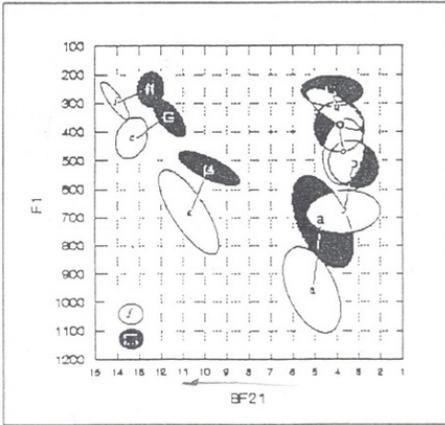
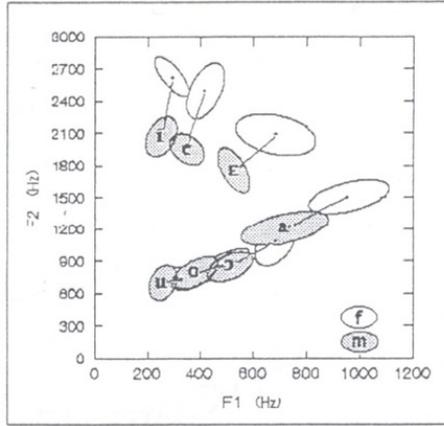
	i	e	ε	<b>a</b>	ɔ	o	u
F1 (M)	300±95	400±70	550±95	<b>770±145</b>	550±65	450±80	320±70
F2 (M)	2200±270	2000±250	1850±250	<b>1250±225</b>	900±180	750±150	650±140
F1 (F)	300±40	410±55	620±100	<b>1050±220</b>	650±120	450±70	300±50
F2 (F)	2880±350	2700±360	2450±400	<b>1450±200</b>	1250±270	950±220	850±270

## Appendice II

Tabella di p. 261 di Ferrero *et alii* (1995) relativa a suoni prodotti da 20 informatori padovani (10 M e 10 F)<sup>13</sup>:

	i	e	ε	<b>a</b>	ɔ	o	u
F1 (M)	251±31	351±34	526±34	<b>716±85</b>	514±42	383±46	258±28
F2 (M)	2071±98	1953±76	1755±110	<b>1208±83</b>	847±71	788±80	747±52
F1 (F)	290±34	420±37	681±77	<b>956±78</b>	681±39	470±59	321±38
F2 (F)	2630±96	2498±134	2086±101	<b>1507±82</b>	1065±108	854±85	747±52

<sup>13</sup> I valori di deviazione indicati risultano più contenuti rispetto a quelli della tabella sopra perché basati su ellissi di equiprobabilità al 75%.



Diagrammi formantici riprodotti da Ferrero *et alii* (1995). In alto su scale lineari in Hz (da p. 249), a confronto con diagrammi con diverse variabili e normalizzazioni (ascissa BF21 = F2-F1, in bark): in basso a sinistra diagramma con F1 in Hz, da p. 262, e a destra diagramma con BF10 = F1-F0, in bark (da p. 263).