Università degli studi di Roma
"La Sapienza"

Facoltà di Ingegneria
Corso di laurea : Ingegneria Elettronica
Dipartimento di Scienza e Tecnica dell'Informazione e della Comunicazione

Tesi di Laurea in Comunicazioni Elettriche
Anno Accademico 1998/1999

ANALISI ACUSTICA DELLE CONSONANTI NASALI SINGOLE E GEMINATE IN ITALIANO

Relatore:
Prof. Maria Gabriella Di Benedetto

Laureando:
Marco Mattei
Matr. N. 09090909

INDICE

CAPITOLO 2
IL FENOMENO DELLA GEMINAZIONE E LE CONSONANTI NASALI

CAPITOLO 3
LA BASE DI DATI, IL SOFTWARE E GLI STRUMENTI STATISTICI

Allegato

"ACOUSTIC ANALYSIS OF SINGLETON AND GEMINATE NASALS IN ITALIAN"

INTRODUZIONE

La presente tesi è stata svolta presso il Laboratorio Voce del Dipartimento INFOCOM della Facoltà di Ingegneria dell'Università "La Sapienza" di Roma.

Scopo della ricerca è stato lo studio analitico delle caratteristiche delle consonanti nasali italiane [m, n] e delle vocali [a, i, u] coarticolate con esse. In particolare, nel corso del lavoro, sono stati analizzati e studiati i seguenti punti:

- studio della geminazione;
- studio della nasalizzazione delle vocali;

I campi di applicazione dei risultati ottenuti da lavori sul segnale vocale come il presente sono molteplici: la conoscenza approfondita del segnale vocale permette la realizzazione di algoritmi di compressione sempre migliori, facilitando le possibilità di comunicazione vocale a distanza. Anche il progetto di riconoscitori vocali non può prescindere da studi acustici sul segnale vocale anche in relazione alle regole particolari di ogni lingua. Infine i risultati di queste analisi possono rivelarsi utili anche per l'implamentazione di sintetizzatori vocali, sempre più presenti nelle nuove applicazioni tecnologiche.

I passi che hanno portato dall'inizio al completamento del lavoro possono così riassumersi:

1. Organizzazione della base dati e delle pronunce disponibili.
2. Fase di studio teorico dei segnali vocali corrispondenti alle pronunce della base dati, atto ad individuare le caratteristiche delle consonanti sotto esame e il modo di operare l'analisi futura.
3. Scelta dei parametri caratteristici da estrarre durante il corso dell'analisi.
4. Misurazione di tutti i parametri nel dominio del tempo e nel dominio della frequenza.
5. Sviluppo di software di supporto per l'estrazione automatica di altri parametri utili per l'analisi.
6. Studio teorico dei test statistici necessari all'analisi dei dati acquisiti.
7. Ricerca e studio di software che implementassero i test statistici scelti per l'analisi
8. Analisi statistica dei dati ottenuti dalle misure.
9. Interpretazione dei risultati ottenuti al punto precedente e formulazioni di ipotesi.
10. Classificazione delle consonanti singole/geminate sulla base delle ipotesi fatte.
11. Confronto con altri lavori in letteratura riguardanti la geminazione in italiano ed in altre lingue.

La tesi è stata strutturata come segue:

Nel primo capitolo, vengono descritte la produzione della voce attraverso l'apparato fonatorio e la percezione attraverso l'organo dell'udito. Sono date anche le nozioni fondamentali di acustica, di fonologia ed, infine, di elaborazione numerica del segnale vocale .

Nel secondo capitolo viene trattato a livello teorico il fenomeno della geminazione, uno degli argomenti centrali di tutta la tesi, e quello della nasalizzazione; viene inoltre data una descrizione particolareggiata delle consonanti nasali, oggetto di studio.

Nel terzo capitolo, di preparazione all'analisi, sono descritte la struttura della base dati e gli strumenti software usati nel corso della tesi, primo fra tutti UNICE. Vengono inoltre richiamati i principi teorici delle analisi statistiche utilizzate per l'analisi dei dati.

Il quarto capitolo descrive quindi l'analisi acustica delle consonanti nasali nel tempo, in frequenza e dal punto di vista energetico. In questo capitolo sono inoltre riportate le ipotesi formulate sulla base dell'analisi statistica condotta ed i risultati ottenuti.

In ultimo, il capitolo cinque riguarda il confronto dei risultati ottenuti in questo studio con quelli di altri lavori sulla geminazione (sia sulla lingua italiana sia su altre lingue). Sempre in questo ultimo capitolo vengono forniti alcuni spunti per lavori futuri sulla voce.

Le appendici sono parte integrante e fondamentale di tutta la tesi: esse raccolgono tutti i dati relativi alle misure effettuate con le loro medie e statistiche.

In particolare, nelle appendici A, B e C sono raccolti, rispettivamente, i dati dell'analisi temporale, dell'analisi energetica e dell'analisi in frequenza. Nell'appendice D ci sono i risultati dettagliati delle analisi statistiche condotte.

Nell'appendice E, infine, sono raccolti alcuni listati dei programmi in C utilizzati.

Tutto il materiale descritto: la base dati, i programmi C, i dati relativi a tutte le misure ecc. sono archiviati su cd-rom e sono disponibili presso il laboratorio voce del Dipartimento INFOCOM.

CAPITOLO 1

LA VOCE: FISIOLOGIA, FONETICA, ACUSTICA ED INGEGNERIA

INTRODUZIONE

La voce è indubbiamente la più antica forma di comunicazione possibile tra gli esseri umani ed è ancora quella maggiormente utilizzata. Per questo motivo è facile rendersi conto che vi sono tantissimi aspetti legati alla voce e molte scienze hanno a che fare con essa. In questo primo capitolo saranno quindi esaminati brevemente gli aspetti principali legati alla voce.

Nel primo paragrafo verranno dati dei cenni di fisiologia umana per ciò che concerne gli apparati di percezione e di produzione; nel secondo paragrafo è trattato l'aspetto linguistico, in particolare quello fonetico, della lingua italiana. Nel terzo paragrafo sono dati cenni di fisica acustica. Infine si accennerà ad alcune tecniche ingegneristiche usate per l'analisi del segnale vocale.

1.1 CENNI DI FISIOLOGIA

1.1.1 L'organo dell'udito

Esaminiamo la struttura propriamente anatomica dell'orecchio e i complicati processi di fisiologia neurologica per mezzo dei quali le vibrazioni sonore sono trasmesse, attraverso il nervo uditivo, al cervello, dove vengono interpretate come suoni.

L'orecchio consiste di tre parti:

- ***ORECCHIO ESTERNO,*** che comprende il **PADIGLIONE,** visibile esteriormente, e il **CONDOTTO UDITIVO ESTERNO,** che fa capo alla membrana del timpano; questa parte dell'orecchio raccoglie e dirige i movimenti vibratori dell'aria.

- ***ORECCHIO MEDIO,*** o **CASSA DEL TIMPANO**, che trasforma le vibrazioni dell'aria in vibrazioni liquide; esso consiste di una cassa piena d'aria e comunica con la parte posteriore della cavità delle fosse nasali attraverso la **TROMBA DI EUSTACHIO.** Il timpano ha la forma di un cilindro le cui basi presentano la convessità dell'una rivolta verso l'altra: queste due basi, distanti *3-6* millimetri (alla circonferenza), sono la **MEMBRANA DEL TIMPANO** e il setto dell'orecchio interno.

Queste due pareti e la catena di ossicini che le unisce costituiscono il meccanismo di trasmissione delle vibrazioni sonore all'orecchio interno.

La membrana del timpano ha uno spessore di un decimo di millimetro; quanto alla forma, è approssimativamente quella di un cerchio con un diametro verticale che va da 10 a 11 millimetri. Benché sia tanto sottile, la membrana del timpano è resistentissima grazie allo strato interno di tessuto fibroso posto fra la pelle del condotto uditivo esterno e la mucosa che riveste interamente la cassa del timpano.

- ***ORECCHIO INTERNO,*** la cui parete racchiude gli organi della percezione uditiva. In questa parete sono praticati due fori: la **FINESTRA ROTONDA,** che ha un diametro di 1,5-2 millimetri ed è chiusa da una membrana simile a quella del timpano, e la **FINESTRA OVALE,** cui fa capo la catena di ossicini: il **MARTELLO,** l'**INCUDINE** e la **STAFFA.**

Questa catena trasmette le vibrazioni dell'aria al liquido dell'orecchio interno, che è molto più denso dell'aria. L'equilibrio fra il liquido, l'aria interna e l'aria esterna è mantenuto dai muscoli dell'orecchio medio e da quelli della tromba di Eustachio. E' il gioco della staffa e della membrana della finestra rotonda che determina il movimento del liquido dell'orecchio interno il quale, a sua volta, mette in movimento la membrana basilare in punti dipendenti dalla frequenza dello stimolo sonoro.

E' dunque nell'orecchio interno che si compie quel fenomeno che chiamiamo audizione; ne sono centro le cavità ossee che per la loro forma sono dette **LABIRINTO:** il **VESTIBOLO,** i **CANALI SEMI-CIRCOLARI e** la **CHIOCCIOLA** .

Il **vestibolo**, che è in comunicazione verso l'esterno con la cassa del timpano, verso l'interno con i canali semicircolari e la chiocciola, ha forma ovale ed è lungo 6 millimetri, largo 3 e alto da 4 a 5. Dei **canali,** due sono verticali; uno, quello superiore, di 15 millimetri, è disposto perpendicolarmente, l'altro, quello posteriore, di 18 millimetri parallelamente all'asse della rocca petrosa (l'osso temporale in cui è scavato il labirinto); il terzo canale, quello esterno, di 12 millimetri, è orizzontale.

La chiocciola consiste di tre sezioni: un nucleo, detto **COLUMMELLA** alto circa 3 millimetri, forato da canaletti che accolgono il nervo uditivo **(CANALE AFFERENTE, CANALE SPIRALE e CANALE EFFERENTE),** un tubo cilindrico aperto a una base e chiuso all'altra estremità dopo che s'è avvolto a spirale tre volte attorno al nucleo, terza, infine, una lamella ossea che con il suo bordo interno divide il tubo cilindrico in due rampe di cui una comunica con la cassa del timpano, l'altra col vestibolo.

Il nervo uditivo si dipana nel condotto uditivo interno; il labirinto è in comunicazione con il cervello attraverso l'**ACQUEDOTTO DEL VESTIBOLO**.

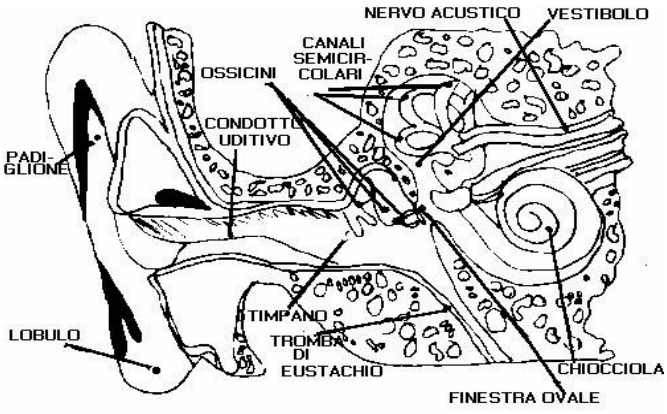


Fig. 1.1 Schematizzazione dell'organo dell'udito.

Le cavità del labirinto contengono un sistema di sacche e di tubi membranosi aderenti a una parte della parete dei canali ossei cui sono ancorati mediante sostegni fibrosi; le sacche sono contenute nei vestiboli, i tubi nelle cavità cilindriche. Questi condotti galleggiano in un liquido, la **PERILINFA**, e sono pieni di un altro liquido, **l'ENDOLINFA**. Le sacche del vestibolo sono in comunicazione fra loro mediante il canale endolinfatico dell'acquedotto vestibolare. Nelle sacche e nei canali sono collocati gli organi sensoriali.

Là dove il nervo uditivo sbocca nelle due sacche vestibolari **(UTRICOLO** e **SACCULO),** la mucosa di rivestimento mostra tre tipi di formazioni cellulari: cellule **BASALI**, cellule **DI SOSTEGNO** e cellule **SENSORIALI**. Nell'utricolo, nel sacculo e nelle ampolle, si trovano dei piccoli cristalli di carbonato di calcio.

Il canale cocleare è appoggiato alla parete del tubo cilindrico, cui è trattenuto dal legamento spirale, e alla lamina spirale, mediante la fasciola striata; esso sta dunque a cavallo delle due rampe della chiocciola da cui è separato mediante la **MEMBRANA DI REISSMER** e la **MEMBRANA BASILARE.**

In perfetto equilibrio sulla membrana basilare si trovano gli organi uditivi. La mucosa del canale cocleare, al livello della parte interna della membrana basilare e in corrispondenza del punto in cui sboccano le ramificazioni terminali del nervo uditivo che spuntano dai **FORAMINA NERVINA** della fasciola striata, si solleva a formare **l'ORGANO DEL CORTI,** il centro del quale è occupato da una serie di arcate. Le fibre nervose passano fra i pilastri che le sostengono. Ai due lati delle arcate si trovano le file delle **CELLULE UDITIVE,** di cui 3.300 sono interne e 18.000 sono esterne, le quali presentano le **CIGLIA UDITIVE** disposte a ferro di cavallo; le sovrasta la **MEMBRANA DEL CORTI.**

Le ampolle su cui si innestano gli archi dei canali semicircolari sono considerate organi del senso dello spazio e dell'equilibrio; la percezione uditiva ha sede nelle vescicole del vestibolo e nella chiocciola. Le prime recepirebbero, pare, le vibrazioni aperiodiche che chiamiamo rumori, mentre le vibrazioni regolari, periodiche, ecciterebbero gli organi della chiocciola e ivi sarebbero percepiti come dei toni o suoni musicali.

Quando un'onda sonora colpisce la membrana del timpano mettendola in vibrazione, il movimento è trasmesso attraverso gli ossicini fino alla finestra ovale. I movimenti della staffa creano una pressione sulla perilinfa del vestibolo e questo scuotimento della perilinfa è a sua volta trasmesso attraverso la membrana di Reissmer all'endolinfa del canale cocleare così da provocare uno spostamento verso il basso sia della membrana basilare che della membrana reticolare e dell'organo del Corti.

Non si conosce ancora in tutti i suoi dettagli la maniera in cui funziona la chiocciola, tuttavia è stato stabilito con sicurezza che si ha uno spostamento massimo della posizione della membrana basilare ad ogni tono puro e che la posizione di questo spostamento varia al variare della frequenza dell'onda sonora che produce lo stimolo. Le onde ad alta frequenza causano uno spostamento massimo della membrana basilare fin vicino la finestra ovale alla base della coclea e le onde a bassa frequenza causano uno spostamento massimo verso la cupola della chiocciola. Quando la coclea è influenzata dalle vibrazioni di un'onda complessa, la membrana basilare viene spostata a dei punti corrispondenti alle frequenze delle componenti dell'onda. A ciascun punto di spostamento le ciglia dell'organo del Corti vengono scosse.

La ricerca dei fatti fisiologici e neurofisiologici che stanno dietro all'audizione, al livello dell'orecchio interno e a quello della corteccia, cioè fin nel centro uditivo del cervello, compete a diverse discipline; quel che interessa la fonetica è soprattutto il modo in cui l'orecchio reagisce ai diversi parametri fisici (frequenza, ampiezza, complessità, periodicità) dell'onda sonora che trasmette il messaggio linguisticamente formato. Il primo problema è pertanto di sapere qual è la gamma di frequenze e di ampiezze all'interno della quale l'orecchio è sensibile alle vibrazioni e alle differenze vibratorie.

1.1.2 Gli apparati di produzione della voce

L'apparato fonatorio dell'essere umano uomo è un insieme composto da un certo numero di organi la funzione primaria dei quali è, per tutti, una funzione eminentemente biologica: la respirazione, la deglutizione, ecc. L'apparato fonatorio umano è un adattamento ai fini comunicativi di organi la cui funzione è stata in origine, e resta tuttora, diversa. Si usa distinguere nell'apparato di fonazione le seguenti parti e funzioni:

-la **REALIZZAZIONE DI UNA CORRENTE D'ARIA** che nell'assoluta maggioranza dei casi è una corrente espiratoria da parte dell'*APPARATO RESPIRATORIO,*

-la **SORGENTE SONORA** responsabile delle vibrazioni periodiche utilizzate per la differenziazione fonetica (il tono glottidale): la *LARINGE,*

-e i **RISUONATORI** o *CAVITÀ SOPRAGLOTTIDALI.*

APPARATO RESPIRATORIO

La **respirazione,** addominale o costale a seconda dei casi, è una condizione essenziale per la formazione dei suoni del linguaggio ma contribuisce ben poco a differenziarli e non c'è bisogno di descriverla.

La ***LARINGE*** è una specie di scatola cartilaginea che forma la parte superiore della trachea; essa è composta di quattro cartilagini: la cricoide che ha forma di anello e ne costituisce la base, il corpo tiroide

che è attaccato alla cricoide per mezzo di due corna, aperte verso l'alto e all'indietro, e le aritenoidi, due piccole piramidi poggiate sul castone della cricoide in modo da poter essere mosse mediante un sistema di muscoli.

La parte posteriore delle aritenoidi (l'apofisi muscolare) è il punto di appoggio dei muscoli che muovono le aritenoidi e comandano così l'apertura e la chiusura della glottide, cioè lo spazio circoscritto dalle due corde vocali e dai loro prolungamenti nelle apofisi vocali.

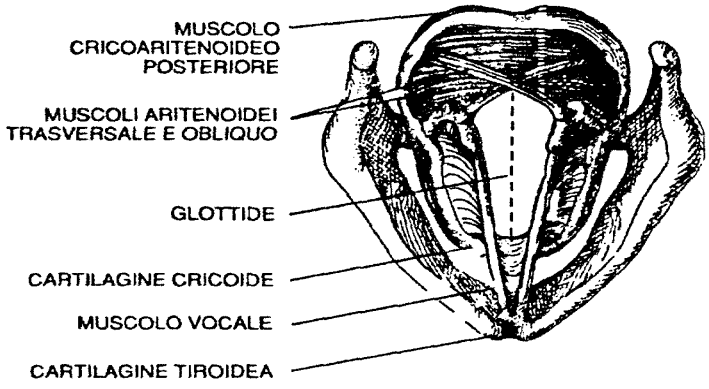


Fig. 1.2 Sezione longitudinale della laringe.

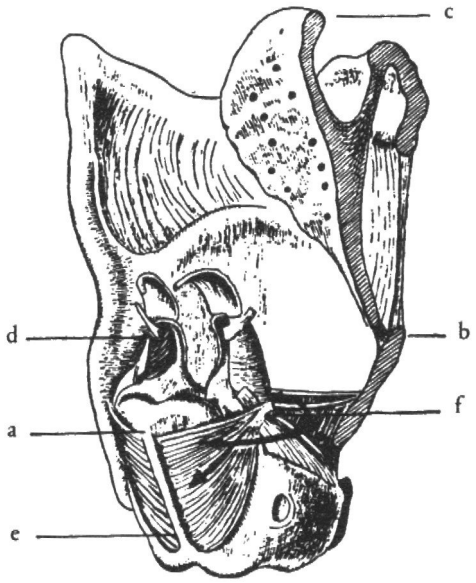


Fig. 1.3 La laringe vista da dietro. a: cartilagine cricoidea; b: cartilagine tiroidea; c: epiglottide; d: aritenoide (sinistra); e, f: muscoli (le frecce indicano le direzioni di movimento).

Tutte le pareti interne della laringe sono rivestite di una mucosa; questo tessuto forma sui lati dell'interno del corpo tiroide due coppie di pieghe che formano due rilievi orizzontali nella laringe. Sono queste pieghe che vengono chiamate **CORDE VOCALI** e **FALSE CORDE VOCALI.**

Le corde vocali sono un muscolo rivestito di mucosa formato da cinque strati di tessuto con proprietà meccaniche differenti, che servono ad assicurarne una vibrazione corretta. Nell'uomo sono lunghe circa 23 mm e nella donna 18 mm, mentre l'apertura media glottale è di circa 5 mm^2 con picchi tipici dell'ordine di 15 mm^2.

Le tasche che si formano entro queste due pieghe si chiamano **VENTRICOLI DI MORGAGNI.** Le corde vocali si riuniscono in avanti nell'angolo della tiroidea; dietro esse sono attaccate alle apofisi vocali delle aritenoidi. Le aritenoidi sono attaccate al castone della cricoidea e sono mobili in più di una direzione: verso l'esterno, in posizione di riposo, verso l'interno, per chiudere la glottide, e verso l'alto e verso il basso. In posizione di riposo esse si trovano a una certa distanza l'una dall'altra in modo che formano un triangolo col vertice nell'angolo della tiroide.

Il meccanismo che muove le aritenoidi è stato studiato e descritto dall'anatomista svedese Bertil Sonesson. E'grazie a questi movimenti delle aritenoidi realizzati mediante un sistema di muscoli che può essere variata la forma della glottide (cfr. fig. 1.4). Si distinguono quattro posizioni principali della glottide (cfr. fig. 1.5):

- la prima, triangolare, è utilizzata durante la normale respirazione;
- la seconda, pentagonale, è quella della respirazione profonda;
- la terza, con i bordi dei labbri incollati uno all'altro, ma con le aritenoidi separate, è quella che si adopera nel bisbiglio (infatti i suoni bisbigliati si formano al passaggio dell'aria attraverso lo stretto canale fra le aritenoidi);
- la quarta posizione della glottide è quella della fonazione: la glottide è chiusa in tutta la sua lunghezza e l'aria in uscita passa con una serie di scosse fra i bordi vibranti delle corde vocali.

Infine è possibile far assumere alle corde vocali una quinta posizione: i bordi possono essere appoggiati uno sull'altro e la conseguenza è una chiusura completa (occlusione) del passaggio dell'aria, questa posizione caratterizza la consonante detta colpo di glottide.

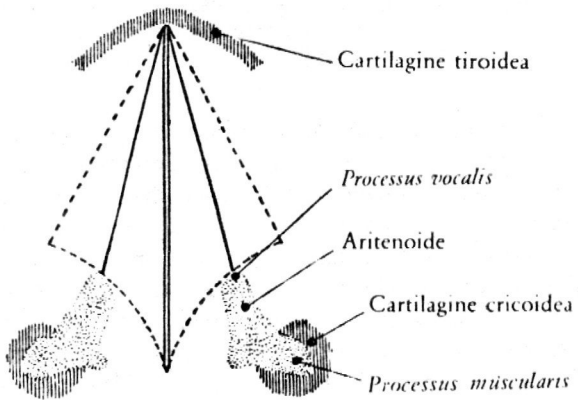


Fig. 1.4. Disegno schematico del meccanismo di apertura e chiusura della glottide. Le due linee più grosse indicano la posizione delle corde vocali durante la respirazione normale, le linee tratteggiate più grosse la posizione durante la respirazione profonda. Le due linee verticali sottili indicano la posizione di fonazione. Le linee tratteggiate sottili indicano la direzione del movimento delle aritenoidi quando la glottide cambia forma. (Da I.Tarneaud).

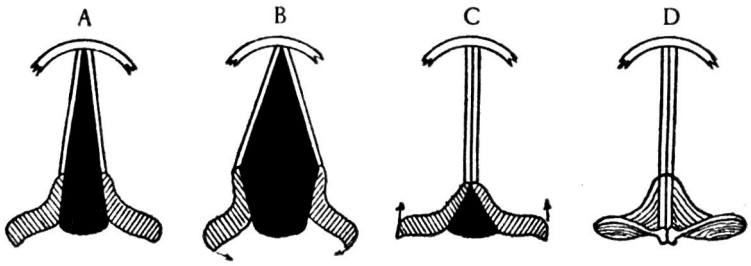


Fig. 1.5. Disegno schematico delle differenti posizioni della glottide: A respirazione normale, B respirazione profonda, C bisbiglio (le corde vocali sono chiuse ma il passaggio fra le aritenoidi resta libero), D fonazione. (Da J. Forchhammcr).

E' dunque grazie alle cartilagini aritenoidi e ai muscoli che ne comandano i movimenti che è possibile far variare la forma, la posizione e la tensione delle corde vocali interessate che possono vibrare o no al passaggio dell'aria attraverso la glottide. Il muscolo cricotiroideo, ad esempio, contribuisce al controllo dell'altezza dei suoni emessi quando le corde vibrano, variandone la tensione longitudinale e provocando così una loro deformazione. La variazione di tensione comporta una modifica delle frequenze di vibrazione delle corde vocali. E' noto, infatti, che le frequenze proprie di risonanza di una corda di lunghezza l soggetta ad una tensione T e fissata agli estremi, sono date dalla:

$$\nu = \frac{n}{2l}\sqrt{\frac{T}{\mu}} \qquad \text{n = 1, 2, 3, ...} \qquad (1.1)$$

ove μ rappresenta la densità lineare della corda. La laringe ha una tendenza naturale ad alzarsi e abbassarsi proporzionalmente all'ampiezza del suono prodotto, compromettendo così la sua emissione con qualità vocali costanti. Ciò può essere evitato impiegando i muscoli estrinseci per cercare di mantenere stazionaria la posizione dello scheletro cartilagineo.

Le ***CAVITÀ SOPRAGLOTTIDALI*** sono la **faringe, la cavità orale e le fosse nasali**.

La **cavità faringea** si estende fino alla glottide e può essere compressa ritraendo la radice della lingua verso la parete della faringe. Mediamente la lunghezza dell'intero condotto vocale è di 17 cm negli uomini.

La **cavità nasale** è principalmente ossea e quindi la sua forma è fissa. Essa può essere isolata dal resto del condotto vocale, se si solleva il **velo palatino** o **palato molle**. Così facendo, si solleva il diaframma rinovelare che mette in comunicazione la cavità nasale con quelle orale e faringale. Quando il condotto vocale è in posizione di riposo, il velo pende, estendendosi verso il basso, e il diaframma rinovelare è dunque aperto. Durante la produzione della maggior parte dei suoni linguistici, il velo è sollevato ed il diaframma è chiuso ma, nel caso di suoni nasali o nasalizzati, esso rimane aperto in modo che l'aria possa passare attraverso la cavità nasale per uscire dalle narici. Nell'uomo la cavità nasale ha una lunghezza e un volume medi rispettivamente di circa 12 cm e 60 cm^3.

La **cavità orale** si trova essenzialmente tra la lingua ed il palato e termina alle labbra. Essa può assumere un grandissimo numero di conformazioni diverse a causa del movimento della mandibola, delle labbra, della lingua e del velo palatino (organi fonatori mobili). Gli organi fonatori fissi sono i denti, gli alveoli ed il palato.

La cavità formata dalla protrusione e dall'arrotondamento delle labbra la si può considerare come quarto risuonatore. E' essenzialmente grazie ai movimenti della lingua che è possibile cambiare la forma e il volume, e di conseguenza l'effetto risuonatore, della faringe e della cavità boccale. Dal punto di vista delle possibilità articolatorie, bisogna distinguere fra il dorso è l'apice della lingua (articolazioni dorsali e apicali). La volta della cavità orale presenta le seguenti regioni (fra parentesi le denominazioni rispettive delle articolazioni che vi si formano):

- i denti (dentali),
- gli alveoli (alveolari),
- il palato duro (palatali, distinte in prepalatali, mediopalatali e postpalatali)
- il palato molle, o velo palatino (velari), con l'ugola o *uvula* (uvulari).

1. labbra
2. denti
3. gengive (alveoli)
4. palato duro
5. palato molle (velo)
6. uvula
7. punta della lingua (apice)
8. parte anteriore della lingua
9. parte posteriore della lingua
10. laringe
11. epiglottide
12. corde vocali

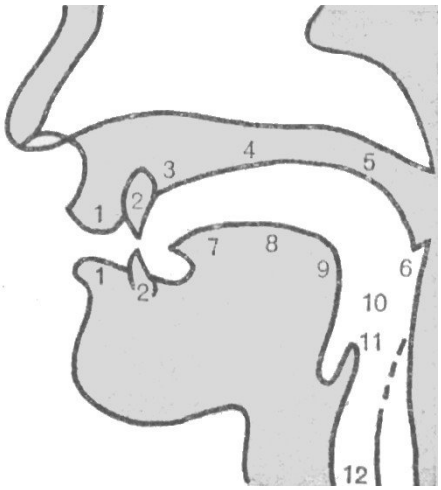


Fig. 1.6 .

Dietro si ha infine la parete posteriore della faringe (faringali). Un'articolazione con la partecipazione delle fosse nasali è detta nasale, o nasalizzata. Le articolazioni realizzate mediante le labbra sono dette labiali e più particolarmente, bilabiali se sono in gioco tutt'e due le labbra, labiodentali se il labbro inferiore va a toccate gli incisivi superiori, o il contrario, come accade talvolta. E' servendosi di combinazioni di questi termini che si arriva a definire abbastanza esattamente la maggior parte dei tipi articolatori che sono impiegati nel linguaggio: apico-dentali, dorso-palatali, dorso-velari, ecc., composti nei quali il primo termine indica l'organo articolante, il secondo il punto di articolazione come vedremo più dettagliatamente nel prossimo paragrafo.

1.2 LA SCIENZA DELLA FONETICA

1.2.1 Generalità

La **fonetica** è la scienza che si occupa dello studio della lingua parlata. Esistono diversi approcci allo studio di questa scienza: quello *articolatorio*, che studia la produzione dei suoni in funzione degli organi fonatori, quello *uditivo* o *percettivo*, che studia le modalità di acquisizione ed elaborazione delle informazioni fonetiche da parte del cervello umano, quello *funzionale (fonologia)* che analizza la struttura di un sistema fonologico dato o i principi generali della determinazione e della descrizione dei fonemi, interessandosi anche al valore e alla funzione che i suoni hanno in relazione con il loro significato. Altro approccio di interesse per noi è quello *acustico*, che studia strumentalmente le caratteristiche fisiche dei suoni.

I linguaggi in uso nel mondo sono composti ad alto livello da *morfemi*, che sono unità portatrici di significato (ad esempio la parola tavolino è articolata nei **morfemi** *tavol*, *in*, e *o*, con /tavol/ che ci dà l'informazione denotativa sull'oggetto, /in/ sul fatto che ci si sta riferendo ad esso con un diminutivo e /o/ sul suo genere, maschile, e numero, singolare), e dai cosiddetti **fonemi** a basso livello. I fonemi sono le unità minime distintive non dotate di senso che, combinandosi fra loro, permettono di formare le unità portatrici di significato o morfemi.

L'insieme dei fonemi di una lingua costituisce il complesso dei suoni elementari previsti dalle sue regole di pronuncia. Le realizzazioni foniche di un fonema sono dette *allòfoni*; ve ne sono teoricamente infiniti, in funzione delle caratteristiche dei diversi parlatori: loro età, sesso, stato d'animo, provenienza, etc. Una delle principali cause della diversità di realizzazione di un fonema da parte di uno stesso parlatore, anche a pochi secondi di distanza, è rappresentata dall'influenza dei fonemi confinanti nella sequenza pronunciata: si parla in tal caso del fenomeno della **coarticolazione**.

Anche nella lingua italiana si trovano numerosissimi allofoni o realizzazioni concrete di un solo suono (basti pensare alla /s/ pronunziata da un settentrionale, da un toscano o da un meridionale); tuttavia i fonemi nell'italiano sono soltanto 28.

Per individuare i fonemi bisogna ricorrere alla **prova linguistica di commutazione:** se esistono almeno due parole in italiano il cui significato varia esclusivamente per la sostituzione di un suono, allora diremo che quel suono è un fonema del sistema fonologico della nostra lingua. Così, nella sequenza *...atto,* potremo avere le coppie *gatto-matto,* o *fatto* e *ratto,* cioè dei significanti diversi, differenziati dai fonemi /g/, /m/, /f/, /r/.

Dato che le lettere del nostro alfabeto sono soltanto ventuno, vuol dire che i segni di trascrizione o **grafemi** non corrispondono esattamente ai suoni e le lettere non coincidono con i fonemi: una lettera può servire per più di un fonema o, viceversa, uno stesso fonema è trascritto con più grafemi; vi sono inoltre dei fonemi trascritti con due o tre lettere (i **digrammi** e i **trigrammi).**

Fonemi	Lettere
/a/	a
/b/	b
/č/ (cero) /k/ (casa)	c (digramma *ci* e *ch*)
/d/	d
/é/ (néro) /è/ (bène)	e
/f/	f
/ǧ/ (gita) /ɣ/ (gara)	g (digramma *gi* e *gh*)
—	h
/i/	i
/l/	l
/ʎ/ (fo*gli*o)	— (digramma *gl* e trigramma *gli*)
/m/	m
/n/	n
/ɲ/ (*gn*omo)	— (digramma *gn*)
/ó/ (póllo) /ò/ (pòco)	o
/p/	p
—	q
/r/	r
/s/ (suono) /ṣ/ (caso)	s
/ʃ/ (*sc*emo)	— (digramma *sc* e trigramma *sci*)
/t/	t
/u/	u
/v/	v
/z/ (pazzo) /ẓ/ (zona)	z

Tab. 1.1 Lettere e fonemi italiani

Le opposizioni fra /s/ sorda (*suono, casa* nella pron. toscana) e /s/ sonora (*smania, rosa* e *casa* nella pron. settentrionale) e fra /z/ sorda (*pazzo, zio* nella pron. toscana) e /z/ sonora (*zero, zio* nella pron. settentrionale) non sono sicuramente avvertite nei vari tipi di italiano regionale: così pure le opposizioni fra vocali aperte e chiuse: /é/ chiusa ed /è/ aperta non sono sempre distinte (si veda la pronuncia settentrionale di bene, vento, pesca con la *e* chiusa); ancora meno sentita la differenza fra /ó/ chiusa e /ò/ aperta, anche negli omografi come *bótte* (recipiente) e *bòtte* (percosse). Pertanto il numero dei fonemi con funzione realmente distintiva nell'italiano contemporaneo è di 24.

Per un uso puramente legato alla fonetica è stato creato, ed è oramai standardizzato, il metodo della trascrizione fonetica, che prevede l'uso di un set di caratteri diverso da quello dell'alfabeto, contenente un carattere per ciascuno dei fonemi (non degli allòfoni) previsti dalle lingue in uso. Una descrizione grafica standard dei suoni delle varie lingue è rappresentata dal sistema International Phonetic Alphabet (I.P.A.).

1.2.2 La fonetica articolatoria

Ogni suono linguistico è compreso in una delle due classi principali chiamate tradizionalmente vocali e consonanti. Riservando l'uso di questi termini al senso più scientifico della fonetica funzionale, in questo contesto si useranno i termini vocoidi e contoidi. Per lo studio dell'articolazione di tutti i fonemi ci si serve di diagrammi che mostrano la posizione dei vari organi coinvolti. In particolare, per i vocoidi si usa il *trapezio fonetico*, e per i contoidi lo *spaccato sagitta*le (sezione di profilo) dell'apparato fonatorio[1].

Articolazione dei vocoidi

Si possono definire **vocoidi** (in termini articolatori) quei suoni sonori, che sono prodotti dall'aria (proveniente dalla glottide) che non incontra alcuna ostruzione (nemmeno parziale) tra gli organi fonatori, né restringimenti tali da produrne la frizione. Il suono caratteristico di ciascun vocoide dipende soprattutto dalle posizioni assunte da due organi fonatori: lingua e labbra. In particolare, dipende dal sollevamento/abbassamento e avanzamento/arretramento della lingua (che può quindi muoversi in uno spazio schematizzato come bidimensionale) e dall'arrotondamento o meno delle labbra (spazio unidimensionale). Le possibili posizioni verticali della lingua rispetto al palato sono cinque: *alto*, *medioalto*, *medio*, *mediobasso* e *basso*; quelle orizzontali sono tre: *palatale*, *prevelare e velare* (o anteriore, cenrale, posteriore). La figura 1.7 mostra, invece, i particolari della posizione delle labbra durante l'articolazione delle tre vocali estreme italiane [i, a, u].

Il **trapezio fonetico** può ben rappresentare, schematicamente, uno spazio tridimensionale dove far "muovere" i vocoidi: sull'asse orizzontale e su quello verticale si rappresenta la rispettiva posizione della lingua[2], mentre un punto disegnato arrotondato o no rappresenta la posizione delle labbra. Nella figura 1.8 è disegnato il trapezio fonetico con i sette vocoidi propri dell'italiano.

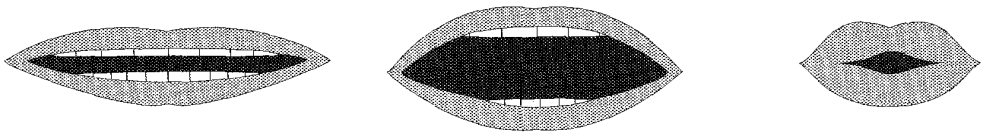


Fig. 1.7 Posizione delle labbra nelle tre articolazioni vocaliche estreme dell'italiano neutro: Labbra non arrotondate per la vocale alta anteriore [i] Labbra non arrotondate per la vocale bassa centrale [a] Labbra arrotondate per la vocale alta posteriore [u] (Canepari, 1992).

[1] Per descrivere adeguatamente le articolazioni di certe consonanti, il metodo fonetico accosta utilmente agli spaccati "sagittali", anche spaccati "ortogonali" (sezioni orizzontali normali al profilo) e spaccati "trasversali" (sezioni verticali di prospetto).

[2] Poiché i movimenti orizzontali della lingua in posizione bassa sono meno ampi, il campo dei possibili punti di articolazione viene racchiuso in un trapezio.

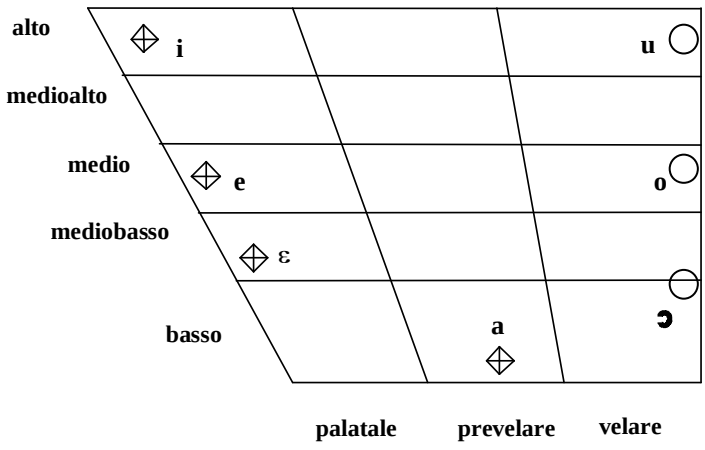


Fig. 1.8 Trapezio fonetico dell'Italiano (Canepari, 1979).

Articolazione dei contoidi

Si possono definire **contoidi** tutti quei suoni che non hanno le caratteristiche dei vocoidi. Infatti, nella produzione della maggior parte dei fenomeni consonantici si ha la formazione di costrizioni al passaggio dell'aria a causa dell'accostamento degli organi mobili contro le altre parti del condotto vocale.

La posizione in cui si forma la costrizione è detta **punto di articolazione** e se ne possono individuare diversi come mostrato in figura 1.9. Per quanto riguarda i punti di articolazione, in italiano, ce ne sono otto fondamentali individuabili:

- **Labiovelari**, che interessano labbra, dorso della lingua e velo palatino (p. es. il fonema /w/ di **u**omo);
- **Bilabiale**, in cui, per realizzare il modo di articolazione, vengono usate entrambe le labbra (p. es. i fonemi /p/ di **p**a**p**a, /b/ di **b**iro, /m/ di **m**ano);
- **Labiodentale**, che prevede l'uso del labbro inferiore e dei denti superiori (p. es. i fonemi /f/ e /v/ **f**a**v**o);
- **Dentale**, in cui sono interessati la punta della lingua e i denti superiori (p. es. i fonemi /s/, /ts/, /d/ e /t/ di **s**en**z**a **d**i **t**e,/dz/ di **z**ero, /z/ di o**s**are);
- **Alveolare**, realizzato con la punta della lingua e gli alveoli che prendono parte all'articolazione (p. es. i fonemi /r/ di **r**ane, /l/ di **l**ana, /n/ di **n**a**n**a);
- **Alveopalatale**, con la lingua alta e con la punta in zona intermedia tra alveoli e palato (p. es. i fonemi /tʃ/ di **c**inta, /dʒ/ di **g**iro e /ʃ/ di **sc**immia);
- **Palatale**, con il dorso della lingua ed il palato (p. es. i fonemi /j/ di **i**eri, /ʎ/ di **gl**i, /ɲ/ di le**gn**o);
- **Velare**, con il dorso della lingua ed il velo (p. es. i fonemi /k/ e /g/ di **c**an**g**uro).

Altri punti di articolazione vengono usati nelle realizzazioni allofoniche, tra i quali è di interesse il punto **prevelare** (p. es. i fonemi /k/ e /g/ seguiti dal fonema /i/ vengono realizzati, a causa dell'effetto

della coarticolazione, sul punto di articolazione prevelare, come in **ch**ina e **gh**iro). Rispetto al punto d'articolazione velare, in tal caso, la parte interessata risulta più spostata verso il palato.

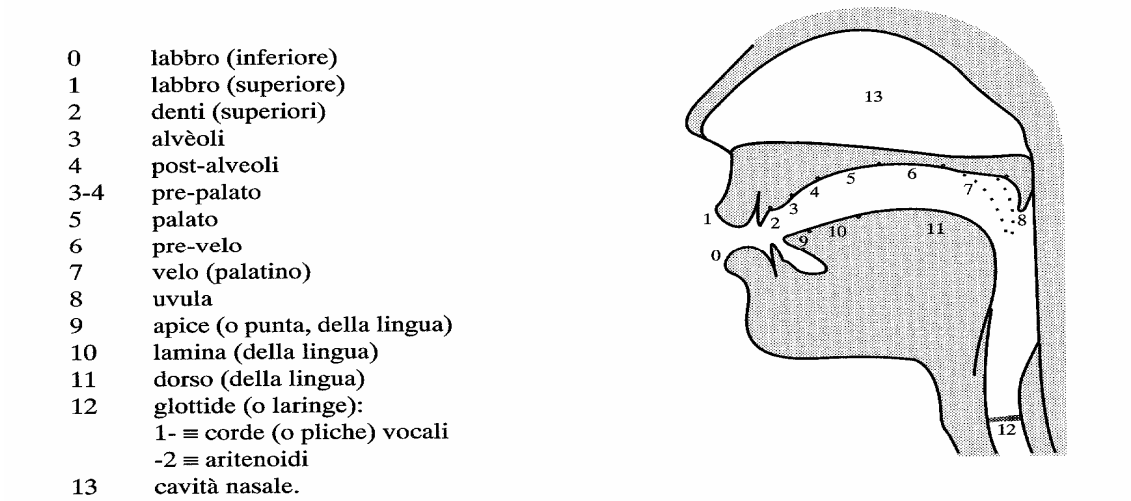


Fig. 1.9 Punti di articolazione (Canepari, 1992).

Il modo in cui la costrizione si realizza è detto **modo di articolazione**. Si distinguono, secondo questo aspetto, i seguenti gruppi di contoidi:

- **Occlusivi**, realizzati bloccando completamente il flusso d'aria, portando a contatto due organi fonatori e rilasciando in seguito velocemente tale costrizione (p. es. i fonemi /t/ e /p/ di **ti**p**o**, /k/ e /d/ di **di**c**o**);
- **Fricativi**, prodotti operando un'occlusione non completa, che causa una particolare frizione dell'aria uscente (p. es. i fonemi /f/ e /v/ di **f**a**v**o, /s/ di **s**ano, /z/ di o**s**are, /ʃ/ di **sc**ena);
- **Affricati**, realizzati da un'occlusione seguita immediatamente da una frizione: si noti che non si tratta di un fonema occlusivo seguito da uno fricativo, il passaggio è rapidissimo e dà luogo ad un suono originale (p. es. i fonemi /ts/ di **z**ucchero, /dz/ di **z**aino, /tʃ/ di **c**ima, /dʒ/ di **g**iugno);
- **Nasali**, prodotti occludendo il tratto vocale orale ma senza tenere il velo schiacciato sulla parete faringale retrostante come per gli altri, in modo che l'aria fluisca dal naso (p. es. i fonemi /m/ e /n/ di **m**a**n**o);
- **Laterali**, realizzati bloccando il flusso d'aria al centro della bocca ma lasciandola fluire lateralmente (p. es. i fonemi /l/ e /λ/ di **l**u**gl**io);
- **Vibranti**, prodotti mediante la vibrazione di un organo mobile (possono essere mono o poli vibranti) (p. es. il fonema /r/ di **r**osa).
- **Approssimanti**, in cui la frizione è molto lieve, al punto che talvolta vengono indicati con il termine di semivocali o di semiconsonanti (p. es. il fonema /j/ di **i**eri e il fonema /w/ di **u**omo);

Suddivisioni dei fonemi di questo tipo permettono di costruire tabelle dove i fonemi sono raggruppati per tratti distintivi misti, come quella per l'italiano, riportata in tabella 1.2.

MODO DI ARTICOLAZIONE	PUNTO DI ARTICOLAZIONE							
	Labio-velari	Bilabiali	Labio-dentali	Dentali	Alveolari	Alveo-palatali	Palatali	Velari
Approssimanti	w						j	
Fricativi			f, v	s, z		ʃ		
Affricati				ts, dz		tʃ, dʒ		
Occlusivi		p, b		t, d				k, g
Vibranti					r			
Laterali					l		ʎ	
Nasali		m			n		ɲ	

Tab. 1.2 Tabella dei contoidi italiani (Muljacic, 1972)

Quindi, dal punto di vista della fonetica articolatoria, le consonanti si distinguono sulla base delle loro tre componenti indispensabili: il tipo di fonazione (sorda o sonora) su cui torneremo nel paragrafo 1.3.2, il modo di articolazione e il punto di articolazione. Per evidenziare il fatto che questo tipo di classificazione non è l'unico possibile, vedremo nel prossimo paragrafo il confronto con la classificazione operata tramite la fonetica binarista.

1.2.3 Fonetica Binarista

Secondo la *fonetica binarista*, dovuta al fonetista Jakobson, esistono una dozzina di tratti distintivi di natura binaria (o opposizioni); cioè, per ogni fonema (qualsiasi lingua esso appartenga), si può univocamente dire se presenta o meno tale tratto distintivo. Tali tratti possono essere scelti in vari modi, ma comunque sempre secondo canoni della fonetica acustica più che della fonetica articolatoria, cioè basandosi sull'analisi strumentale (spettrogrammi, ecc.) dei suoni di una lingua[3]. Una volta individuato l'insieme di tratti giudicato sufficiente a rappresentare l'intero sistema linguistico che si vuole descrivere, una sua rappresentazione alquanto compatta ed esplicativa è data dalla matrice binaria associata a tale sistema. Si tratta di una matrice con una riga per ciascun tratto distintivo e una colonna per ciascun fonema, e con il segno "+" o "-" agli incroci. Alcuni tratti distintivi sono detti *pertinenti* per un fonema, e sono quelli che bastano ad individuarlo univocamente all'interno del sistema linguistico cui appartiene;

[3] Anche se il presupposto su cui si basa la scuola binarista, cioè che ogni realtà linguistica si identifichi tramite una successione di scelte binarie, può apparire più una costruzione ideale che una reale rappresentazione dei processi cognitivi del cervello umano, essa opera una sistematizzazione della fonetica molto utile metodologicamente.

altri sono detti *ridondanti*, e servono a facilitare la "decodifica" del suono da parte dell'ascoltatore, qualora l'informazione connessa con i tratti pertinenti sia degradata. In tale secondo caso, nell'incrocio corrispondente, spesso si lascia la casella vuota o il simbolo viene indicato tra parentesi.

Il binarismo maturo cerca di evitare ad ogni costo i casi di mancata opposizione binaria. In ogni caso, i trenta fonemi italiani (includendo in questo contesto anche le semivocali [*j, w*]) possono specificarsi con undici coppie di tratti distintivi intrinseci (o TDI) [4], come mostrato in tabella 1.3. La media dei TDI per fonema è di 5,8.

FONEMI:	p	b	f	v	t	d	ts	dz	s	z	k	g	c	ɟ	ʃ	m	n	ɲ	l	ʎ	r	i	e	ɛ	a	ɔ	o	u	j	w
1 *Vocalico*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-
2 *Consonantico*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-
3 *Nasale*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+												
4 *Compatto*	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+				-	+	-	-	-	-	+	-	-	-		
5 *Diffuso*																						+	-	-		-	-	+		
6 *Grave*	+	+	+	+	-	-	-	-	-	-	+	+	-	-	-	+	-	-				-	-	-		+	+	+	-	+
7 *Acuto*																	-	+												
8 *Teso*																							+	-		-	+			
9 *Sonoro*	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-															
10 *Continuo*	-	-	+	+	-	-	-	-	+	+			-		+				+		-									
11 *Stridulo*					-	-	+	+																						

Tab. 1.3 I fonemi italiani e i loro TDI secondo (Muljacic,1972).

1.2.4 Gli elementi prosodici

I fonemi da soli non descrivono completamente i "*suoni*" di una lingua, pertanto vanno considerate anche altre caratteristiche che agiscono su tutta la frase, trasmettendo informazione e completando la descrizione del processo di produzione propriamente detto.

Questi altri elementi prendono nome di **caratteristiche soprasegmentali** e sono molto difficili da definire e formalizzare da un punto di vista linguistico. Alcuni esempi sono il tono, l'accento e l'intonazione. Il **tono** non è presente in tutte le lingue, ma solo in quelle, come il cinese mandarino, in cui modifica il significato lessicale e grammaticale delle parole. Esso interessa l'altezza relativa delle parole e delle sillabe all'interno di una frase. L'**accento** si manifesta nel porre in risalto alcune sillabe rispetto alle altre all'interno di una stessa parola, combinando vari fattori quali l'intensità dell'emissione, la lunghezza (durata nel tempo) e l'altezza dei suoni. L'**intonazione** è una combinazione di alcuni fenomeni di carattere locale, come l'accentazione, la durata e l'intensità dei foni pronunciati, e di alcuni fenomeni di carattere globale, che coinvolgono tutta la frase. Tra questi, la differente modulazione della frequenza fondamentale usata per cambiare significato ad una stessa frase, come avviene ad esempio per differenziare una frase affermativa da una interrogativa o per comunicare le nostre emozioni

[4] Secondo la teoria binarista ci sono due tipi di *tratti distintivi*: *prosodici* e *intrinseci*. I primi si possono avere solo sul nucleo sillabico fonologico e si raggruppano in tre classi: altezza, intensità e durata. I secondi, invece, si possono classificare in dodici opposizioni la cui suddivisione e classificazione viene modificata quasi in ogni nuova opera che presenti questa teoria, anche da parte dello stesso Jakobson. In questa sede, ci occuperemo esclusivamente dei tratti distintivi intrinseci e gli altri non saranno più menzionati nel seguito.

all'ascoltatore. Questi contorni "melodici", chiamati anche contorni prosodici, sono caratteristici di ogni lingua, alla stregua dei suoni e delle regole grammaticali. Essi danno un gran contributo alla comprensione delle frasi e sono un aspetto fondamentale della naturalezza della voce umana.

1.3 IL SUONO E L'ACUSTICA DEL SEGNALE VOCALE

Quel che abbiamo l'abitudine di chiamare SUONO non è altro, in realtà, che una variazione della pressione atmosferica registrata dal nostro apparato uditivo mediante il timpano. I movimenti di questa membrana sono trasmessi dagli ossicini dell'orecchio medio all'orecchio interno dove, a condizione che si trovino all'interno del campo di sensibilità dell'orecchio[5], essi diventano segnali che vengono ricevuti dal cervello. Queste variazioni della pressione atmosferica hanno la forma di onde che si propagano nell'aria o, in certi casi, attraverso mezzi diversi, liquidi o corpi solidi; l'osso, per esempio, è un buon conduttore delle onde sonore. Le onde si propagano, nell'aria e alla temperatura di 0°, con una velocità di circa 330 metri al secondo, velocità che varia leggermente in rapporto alla pressione e alla temperatura: a 20°, per esempio, la velocità è di 344 metri al secondo. Queste variazioni di pressione sono dovute all'impulso esercitato sulle particelle dell'aria, che vengono smosse dal loro stato di quiete; il fenomeno inizia sempre con uno stimolo meccanico che mette in vibrazione una massa qualunque, un corpo solido, una certa porzione di un corpo gassoso.

L'energia sonora si propaga nello spazio per onde sferiche e quindi decresce con il quadrato della distanza; in ogni caso, quello che si intende con **segnale vocale acustico** è l'andamento temporale della variazione di pressione acustica nella zona limitrofa ad una persona che parla e perciò, con ottima approssimazione, si può considerare trascurabile la perdita di energia e unidimensionale il segnale generato.

Secondo la teoria acustica **della produzione del segnale vocale**, proposta la prima volta da (Fant, 1960) ed ancora oggi generalmente accettata, il segnale acustico viene generato facendo fluire l'aria nella laringe e/o in altre ostruzioni create nel condotto vocale. Le turbolenze che ne scaturiscono danno origine ad un segnale caratterizzato da un ampio contenuto armonico. Questo viene infine modificato tramite l'azione di filtraggio operata dal condotto vocale.

[5] Come si sa, l'uomo non percepisce tutte le vibrazioni come suoni. Nella musica il limite inferiore è di circa 25 Hz (anche se la frequenza più bassa che sia stata percepita è di 11Hz); mentre il limite superiore varia a seconda dell'età e da individuo a individuo. Un bambino può sentire frequenze fino a 20.000 Hz; in età avanzata non si sentono più le frequenze al di sopra di 12.000, 13.000 Hz. Tutte le frequenze utilizzate dal linguaggio umano si trovano al disotto di 10.000 Hz.

1.3.1 Lo spettro acustico

E' noto da tempo che l'udito avverte principalmente le differenze di frequenza e quelle di ampiezza di oscillazione, ma non quelle di fase. Pertanto, nella maggioranza dei casi, i fenomeni sonori che differiscono fra loro soltanto per le relazioni di fase tra le loro componenti armoniche, vanno considerati come un solo fenomeno sonoro agli effetti dell'ascolto (Franchina, Marietti, 1994)[6]. Si rivela perciò assai utile una rappresentazione grafica del tipo di quella di fig. 1.10, nella quale compaiono soltanto le frequenze delle varie componenti sinusoidali e le corrispondenti ampiezze. L'insieme delle righe dei grafici come quello di fig. 1.10 prende il nome di **spettro acustico**. La prima riga a sinistra rappresenta l'armonica fondamentale (frequenza f1); le altre righe corrispondono alle frequenze $f_2 = 2f_1$ (seconda armonica), $f_3 = 3f_1$ (terza armonica) ecc.

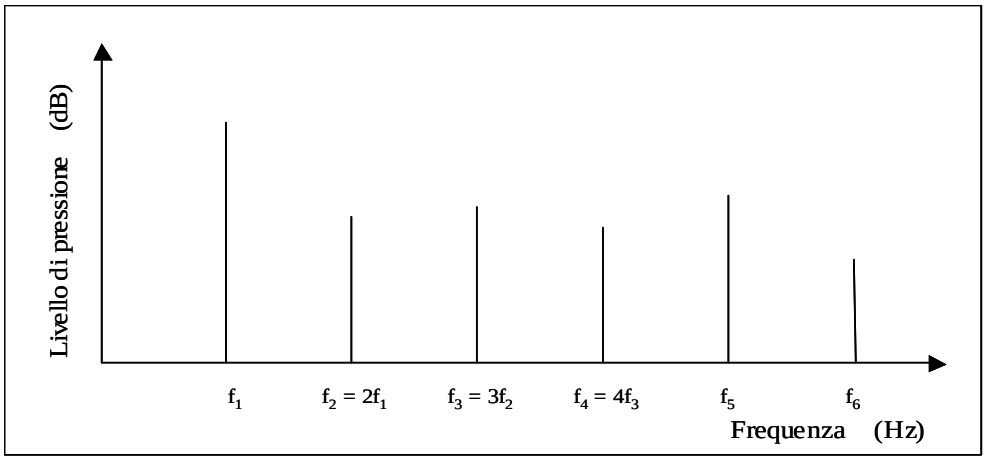


Fig. 1.10 Spettro acustico di un suono complesso.

Queste considerazioni si applicano integralmente soltanto ai fenomeni oscillatori periodici in regime stazionario, condizione quasi mai realizzata nella realtà. Il linguaggio parlato, infatti, è proprio un caso di fenomeno acustico costituito da un gran numero di suoni diversi di breve durata, che si susseguono in rapida successione. Mentre un suono isolato inizia, di regola, con un breve periodo transitorio di attacco ed ha termine con un periodo transitorio di estinzione, nel linguaggio parlato i diversi suoni si succedono senza soluzione di continuità, cosicché il transitorio di estinzione di ciascuno di essi si connette con quello di attacco del suono successivo in modo da costituire quasi un unico transitorio.[7]

---

[6] Questa affermazione va fatta comunque con cautela; infatti, alle relazioni di fase sono legati, in modo più o meno evidente, alcuni importanti aspetti della sensazione uditiva, come la identificazione della direzione di provenienza del suono, come il timbro e la stessa intensità soggettiva, che un tempo si pensava ne fossero indipendenti.

[7] Nel linguaggio parlato, i suoni elementari (foni) aventi carattere relativamente stazionario (vocali, semivocali e alcune consonanti quali [*n, m*]) si alternano con altri suoni consonantici aventi il carattere di brevi transitori (esplosive [*p, b, t, d* ecc.])

Comunque, anche per i fenomeni sonori del tipo ora detto, la rappresentazione mediante lo spettro acustico può riuscire utile, purché si tenga conto in qualche modo dell'evoluzione delle caratteristiche spettrali nel corso del tempo (si ritornerà su quest'argomento nell'ultimo paragrafo).

1.3.2 Suoni sordi e suoni sonori.

Durante la respirazione, il flusso d'aria non incontra ostacoli nel passaggio dalle corde vocali che si trovano in posizione allargata al condotto vocale che è privo di costrizioni. Acusticamente non si percepisce alcun suono. Saranno ora presi in esame i due principali modi di funzionamento dell'apparato di produzione della voce e, a partire da questi, si descriveranno le caratteristiche distintive dei diversi tipi suoni che siamo in grado di produrre e le conseguenti caratteristiche del relativo segnale acustico generato.

Suoni sordi

Le corde vocali possono essere tenute separate tra di loro cosicché l'aria può passare liberamente attraverso la glottide senza far vibrare le corde vocali. Se c'è però la presenza di una costrizione o di un'improvvisa apertura lungo il tratto vocale, si genera l'emissione di suoni chiamati sordi o non vocalizzati, provocati dal moto turbolento del flusso d'aria a valle dell'ostacolo. Acusticamente si percepisce un suono con caratteristiche "rumorose" ad ampio spettro. A seconda della posizione assunta dagli organi mobili del tratto vocale, sono soggetti ad ulteriori classificazioni (per es., sibilanti o plosive, con ulteriore suddivisione a seconda della posizione della costrizione o dell'improvvisa apertura del condotto).

Come esempio di suoni sordi riportiamo le consonanti [p t k f s Σ] in *pane, tondo ,corre, ferro ,sale, scena.*

Suoni sonori

Per la produzione dei suoni sonori, inizialmente le corde vocali sono a contatto l'una con l'altra a causa delle forze presenti e quindi la glottide è chiusa. Quando i polmoni espellono aria, la pressione[8] sotto la glottide aumenta fino a valori che consentono l'allontanamento progressivo delle corde vocali a partire dal basso. Un ulteriore aumento di pressione causa l'apertura della glottide con conseguente passaggio di aria. Le forze elastiche e di altro tipo resistono alla separazione del margine superiore delle corde, ma il flusso d'aria le sovrasta (fig. 1.11).

La legge di Bernoulli asserisce che quando un fluido passa attraverso una strozzatura la pressione ivi presente è minore che nelle sezioni a monte e a valle. Tale riduzione di pressione, accompagnata dalle proprietà elastiche dei tessuti, tende a richiudere le corde vocali. Nel frattempo la pressione sotto la glottide diminuisce anch'essa, dato che la glottide si è aperta per far uscire l'aria. A causa di questi fenomeni, i margini inferiori delle corde vocali cominciano a chiudersi quasi immediatamente, anche se quelli superiori si stanno ancora aprendo.

[8]Generalmente il valore della pressione dell'aria proveniente dai polmoni al livello glottale è di 7 cm H_2O per il parlato normale, 2 cm H_2O per un parlato appena percettibile, e di 20 cm H_2O per un parlato a voce molto alta.

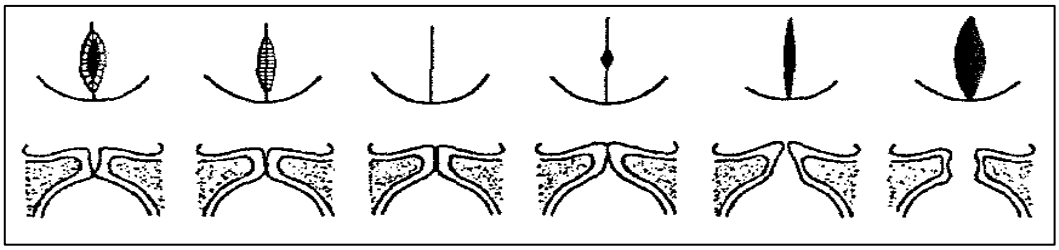


Fig. 1.11 Rappresentazione schematica dello stato di affrontamento delle corde vocali. Parte superiore: sezione longitudinale delle corde vocali (la mancanza del contatto è evidenziata in colore nero); parte inferiore: sezione trasversale.

Questo fatto riduce ulteriormente la forza esercitata dal flusso d'aria e i margini superiori delle corde vocali ritornano allora nella posizione iniziale e chiudono la glottide[9]. A questo punto l'aria torna ad accumularsi al di sotto della glottide e il ciclo così si ripete, alternando le fasi di apertura e di chiusura delle corde vocali[10].

I **suoni sonori** sono dunque quelli prodotti da questo funzionamento delle corde vocali; naturalmente il suono così prodotto può subire modifiche passando attraverso il resto del condotto vocale. Esempio di suoni sonori sono le consonanti [b, d, g, v, z, Z, dZ] di *bene, due, gara ,vetta ,usi, agile* (pronunciato alla toscana); inoltre in italiano sono sempre sonore [m, M, n, ɲ , N, r, l, ʎ] come in *mese, anfora, notte, bagno, àncora, rosa, lupo, figlio*. Le vocali sono tutte suoni sonori.

1.3.3 La frequenza fondamentale o pitch

Il singolo ciclo descritto per i suoni sonori si indica con il nome di **ciclo di fonazione** o **ciclo glottale**, mentre la frequenza con cui vibrano le corde vocali è chiamata **frequenza fondamentale** (F_o) o **pitch**, e la durata del singolo ciclo è detta **periodo di pitch**.

La frequenza fondamentale dell'emissione vocale di un parlatore, il cosiddetto "tono naturale", dipende dalle caratteristiche fisiche delle corde vocali. Varia quindi da parlatore a parlatore e può essere modificata con azioni fisiche, da parte del parlatore, variando il livello di tensione delle corde.

Mediamente il volume d'aria che attraversa il condotto vocale è pari a 1 cm^3/ciclo glottale. Il rapporto tra la durata della fase di apertura delle corde vocali e la durata dell'intero ciclo è variabile tra 0,3 e 0,7. Il valore del rapporto dipende dall'intensità, dalla frequenza con cui vibrano le corde vocali e da quanto è addestrato il soggetto. Infatti, i cantanti professionisti riescono ad ottenere i valori della velocità del

[9]Generalmente tra le corde vocali si realizza un contatto, quando si verifica la chiusura della glottide, per uno spessore di circa 2 - 5 mm.

[10]Il ciclo può anche avere luogo con le corde vocali inizialmente non in contatto tra loro. La pressione dovuta all'effetto di Bernoulli in questo caso fa dapprima avvicinare le corde; la fine della fonazione può avvenire in due modi, a seconda che le corde vocali si rilassino o che vengano forzate a rimanere unite: nel primo caso la vibrazione si esaurisce gradualmente e le corde vocali non si toccano per gli ultimi cicli; nel secondo la vibrazione cessa immediatamente e si ha chiusura glottale anche nell'ultimo ciclo.

volume d'aria minori, ad intensità costante, e a realizzare in questo modo un maggior rendimento nella conversione pressione - suono.

Le corde vocali non imprimono quindi energia all'aria vibrando come le corde di un violino, ma aprendo e chiudendo la glottide, creando "sbuffi" d'aria nell'apparato vocale. L'improvvisa cessazione del flusso d'aria a causa del rapido accostarsi delle corde vocali produce una vibrazione acustica che risuona nel condotto vocale. Tale meccanismo è simile a quello che dà origine al suono prodotto sbattendo le mani. L'istante in cui avviene la completa chiusura della glottide è chiamato **istante di epoch**. Anche se è all'istante di *epoch* che viene prodotto il maggior contributo all'energia sonora responsabile dell'emissione della voce, un altro contributo di minor entità viene dall'aprirsi delle corde vocali che si verifica più lentamente della loro chiusura (Strube, 1974).

L'intensità vocale, o volume, dipende da quanta energia viene impartita dalle vibrazioni delle corde vocali all'aria nell'apparato vocale. Quando la pressione dell'aria aumenta, l'ampiezza delle vibrazioni cresce perché le corde vocali si allargano maggiormente e si richiudono più bruscamente; di conseguenza, durante ciascun ciclo di fonazione, il flusso d'aria attraverso la laringe si interrompe più nettamente e l'intensità del suono prodotto cresce.

L'andamento nel tempo della velocità del volume d'aria, per una voce di intensità normale, è un segnale quasi periodico di forma approssimativamente triangolare caratterizzata da due istanti di discontinuità, uno iniziale ed uno finale, che rappresentano rispettivamente gli istanti di apertura glottale e di *epoch*[11]. Data la natura periodica, il suo spettro è a righe, con componenti periodiche sono multipli interi della frequenza fondamentale. L'inviluppo dello spettro presenta un'attenuazione nelle alte frequenze di circa 12dB/ottava, anche se vi possono essere grandi differenze nelle altezze delle armoniche da soggetto a soggetto e, per lo stesso soggetto, passando da un periodo di *pitch* all'altro. Mediamente, per i soggetti che leggono un testo, l'intervallo di variazione della frequenza fondamentale di rado supera un'ottava nel corso della lettura. Poiché gli uomini hanno corde vocali più lunghe (tra i 20 e 25mm) delle donne e dei bambini (tra i 15 e 20 mm), il loro *pitch* è generalmente più basso. In tabella 1.4 sono illustrate le frequenze fondamentali che la voce può avere nel corso del parlato normale (nel caso del canto la frequenza fondamentale può variare approssimativamente tra i 40Hz e i 1800Hz).

Soggetto	$F_{o\ minima}$ (Hz)	$F_{o\ media}$ (Hz)	$F_{o\ massima}$ (Hz)
Uomini	50	125	200
Donne	150	225	350
Bambini	200	300	500

Tab. 1.4 Valori della frequenza fondamentale minima, media e massima per soggetti adulti maschili, femminili e per bambini (M.I.T., 1986)

[11]Le forze aerodinamiche responsabili delle oscillazioni delle corde vocali sono influenzate dal tratto sopra - glottale. Ciò causa un leggero ritardo dell'andamento nel tempo della velocità del volume d'aria rispetto all'andamento dell'aria nella glottide.

Comunque la frequenza fondamentale normalmente può variare al massimo dell'1%/ms, il che corrisponde, ad esempio, ad un cambiamento del 2% per periodi di pitch adiacenti per F_o=500 Hz e del 20% per F_o=50 Hz. Chiaramente la frequenza di pitch può essere modificata dal parlatore agendo sul livello di tensione delle corde vocali.

1.3.4 Frequenze Formanti

I suoni sonori sono caratterizzati, oltre che dalla F_o anche dalle frequenze formanti. Vediamo, come abbiamo fatto nel precedente paragrafo per la F_o , qual è l'origine fisica delle formanti.

Un risonatore acustico è un sistema fisico che presenta la capacità di alterare la natura di un suono che lo attraversa. Più precisamente nel passaggio di un segnale acustico nel risonatore, alcune frequenze componenti sono attenuate, altre, nelle regioni di risonanza, vengono invece amplificate e irradiate quindi con maggior ampiezza. Per quanto riguarda la voce, le frequenze di risonanza sono dette **frequenze formanti**, e sono determinate dalla forma del condotto vocale che dipende dalla posizione degli organi mobili, dall'età e dal sesso dell'individuo. Donne e bambini hanno un apparato vocale più breve degli uomini e di conseguenza i valori delle frequenze formanti saranno più elevati[12]. Ad esempio, se si schematizza il condotto vocale in posizione "neutrale", come per la vocale /*u*/ nella parola inglese "*but*", assimilandolo ad un tubo uniforme senza perdite chiuso ad un'estremità (la glottide) e aperto all'altra (le labbra), le frequenze di risonanza *v* delle onde stazionarie che vi si generano assumono i valori dati dall'espressione:

$$v = \frac{c}{4l}(2n+1) \qquad \text{n = 1, 2, 3, ...} \qquad (1.2)$$

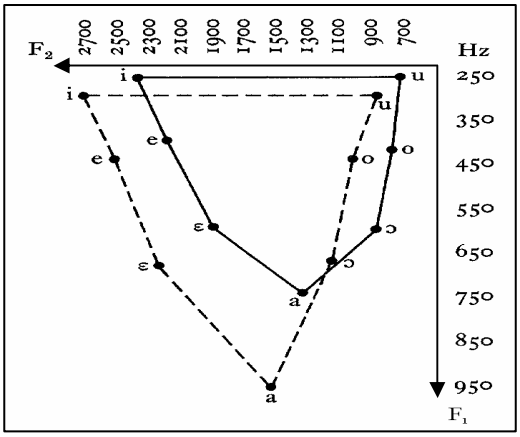


Fig. 1.12 Medie delle prime due formanti dei sette vocoidi tonici italiani: voci maschili (linea continua) e femminili (linea tratteggiata) sovrapposte. (Canepari, 1979).

[12]Un'altra causa da cui dipende la lunghezza e la forma del condotto vocale, e quindi le caratteristiche delle frequenze formanti, è la frequenza fondamentale usata durante l'eloquio. Infatti, i suoi cambiamenti causano un abbassamento od un sollevamento dello scheletro cartilagineo della laringe, provocando perciò una modifica delle lunghezza del condotto vocale.

dove *l* è la lunghezza del condotto vocale (mediamente 17 cm) e *c* la velocità delle onde elastiche nell'aria (circa 340 m/s). Per questi valori si hanno i seguenti valori di *v*: 500 Hz, 1500 Hz, 2500 Hz, ecc.

Tali valori corrispondono ai valori delle frequenze formanti. Per suoni diversi il condotto vocale assume configurazioni differenti, quindi si hanno valori differenti delle frequenze formanti, ciascuno caratteristico di ogni suono.

Vediamo infine più nel dettaglio come il timbro dei vocoidi dipende dalle singole formanti. Per i vocoidi sono fondamentali le prime due formanti (F1 e F2) contando dal basso dopo la fondamentale. Le formanti superiori servono soprattutto per le caratteristiche individuali della voce. Per i vocoidi F1 è bassa (250 Hz circa, per una voce maschile) se sono alti come [i] e [u], alta (intorno ai 750/800 Hz) se sono bassi come [a]. La F1 si sposta gradualmente tra questi due estremi, inversamente all'elevazione della lingua. Invece F2 è determinata dalla lunghezza della cavità orale: più essa è lunga, più F2 è bassa; se poi s'arrotondano le labbra ,come per la [u], la cavità si allunga ulteriormente facendo abbassare F2 ancora di più.

Nella figura 1.12, sono mostrate le medie delle prime due formanti delle vocali italiane così come riportato dal Canepari.

1.3.5 Caratteristiche acustiche generali della voce emessa

La conoscenza delle principali caratteristiche acustiche del linguaggio parlato è un dato preliminare indispensabile nella tecnica delle telecomunicazioni. Menzioniamo brevemente alcuni risultati medi sperimentali.

- La potenza vocale media a lungo termine[13]di un parlatore è dell'ordine di 20 μW con un livello di voce moderato (68 dB è il corrispondente livello di pressione acustica alla distanza di un metro). La massima escursione è compresa fra pochi μW (voce bassa) e oltre 1mW (voce urlata), corrispondente ad un intervallo di circa 24dB;
- Lo spettro acustico medio a lungo termine mostra che i livelli di voce più elevati si hanno nella banda 200÷400 Hz, mentre per frequenze più elevate il livello spettrale di voce decresce di circa 10 dB per ottava.
- La dinamica della voce è di circa 40 dB nel caso di un discorso tenuto a un livello normale.
- Il ritmo di fonazione medio, ossia la rapidità con la quale si succedono gli elementi fonetici nel discorso, si aggira intorno agli 8÷10 fonemi per secondo.

[13] Per media a lungo termine si intende quella che si riferisce a un intervallo di tempo comprendente parecchi fonemi, senza pause di silenzio tra frasi diverse.

1.3.6 Caratteristiche acustiche della sensazione uditiva

Si espongono ora alcune caratteristiche dell'apparato percettivo umano. Tali caratteristiche devono essere tenute sempre presenti nel formulare conclusioni, per non incorrere nell'errore di dare importanza ad aspetti colti visivamente sullo spettrogramma, che però l'orecchio percepisce diversamente (o per nulla!) e che quindi non hanno rilevanza percettiva.

All'interno dell'orecchio vi sono una molteplicità di fibre nervose sensibili alla pressione dell'aria, e in grado di trasformare le onde sonore del segnale acustico in segnale elettrico inviato al cervello. Tali fibre sono in genere sensibili ad una frequenza ben precisa, detta **frequenza caratteristica**, con una banda passante di 100÷150 Hz; fibre vicine hanno frequenze caratteristiche vicine. Ma la caratteristica più importante da rilevare è che il loro funzionamento non è perfettamente lineare, nel senso che componenti a frequenza vicina vengono percepite dando luogo a componenti spurie con frequenza di intermodulazione tra le due originali. Ciò dà luogo al cosiddetto *effetto centro di gravità spettrale*, cioè due formanti a distanza inferiore di 300 Hz vengono percepite come una sola, avente frequenza intermedia tra le due (e spostata verso quella a maggior contenuto energetico). Per compensare il fenomeno della non linearità è stata proposta una scala alternativa a quella delle frequenze per descrivere il segnale vocale, la cui unità di misura è il *Bark*, e la formula di conversione è la seguente:

$$Bark = 13 \cdot \text{arctg}(0.76 \cdot f_{khz}) + 3.5 \cdot \text{arctg}\left(\frac{f_{khz}}{7.5}\right)^2 \qquad (1.3)$$

L'effetto della trasformazione è una compressione dei valori in frequenza (5kHz = 18.54B), con una maggiore conformità alle caratteristiche percettive non lineari dell'orecchio umano come si vede in figura 1.13a.

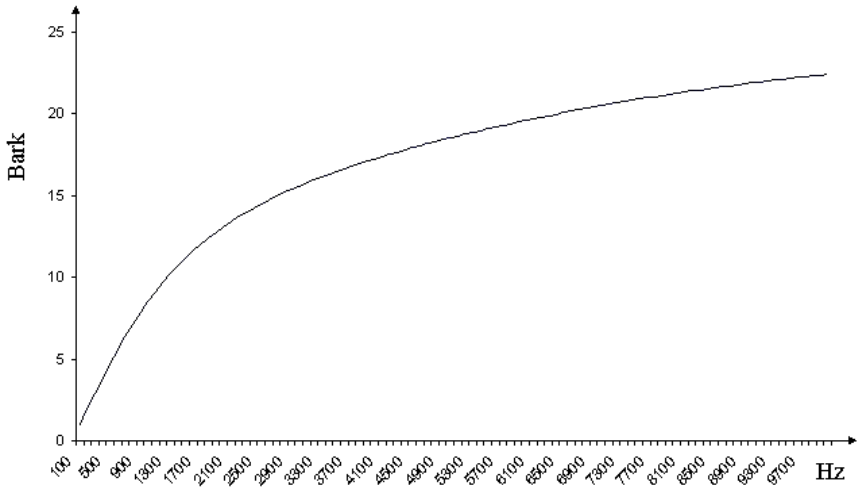
Un altro fenomeno da tenere presente è l'*adattamento*, per cui la risposta ad un suono stazionario è stazionaria per un po', per poi decadere con una costante di decadimento τ di circa 30 ms. Tale caratteristica suggerisce l'idea che il cervello preferisce individuare l'informazione nelle variazioni del segnale in arrivo. Conseguenza dell'adattamento è un altro fenomeno simile, detto del *mascheramento posteriore*, per cui l'orecchio sottoposto ad un suono di test prolungato, poi ad una pausa e poi ad una breve riproposizione del suono, fornisce stavolta una risposta alquanto debole.

Facendo riferimento al caso più semplice, e cioè a quello dei toni puri in regime stazionario, si possono inoltre individuare le seguenti caratteristiche:

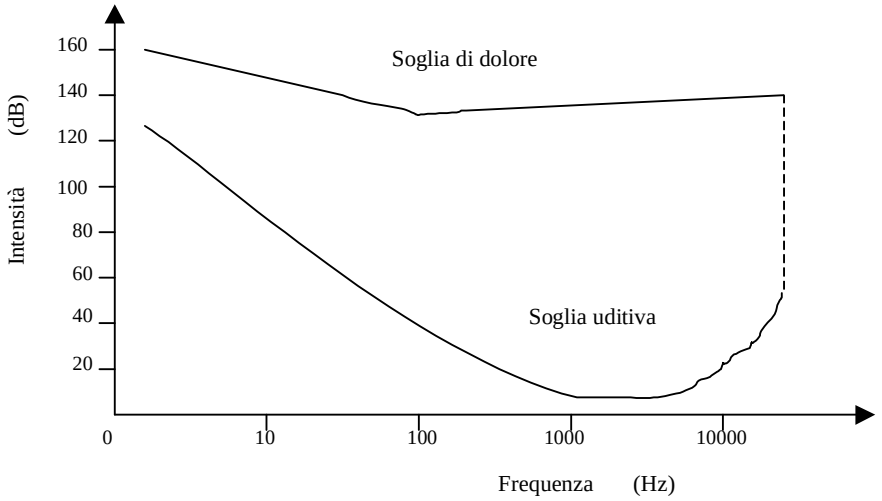
- **Altezza tonale**, caratteristica per la quale i suoni si distinguono in più o meno gravi o acuti. E' legata essenzialmente alla frequenza dell'oscillazione;
- **Intensità soggettiva**. E' legata in modo essenziale sia al livello di pressione dell'onda sinusoidale, sia alla sua frequenza. Il conseguente comportamento dell'udito umano per i suoni puri è illustrato dall'audiogramma normale ottenuto costruendo sperimentalmente, per diversi valori di intensità, le cosiddette curve isofoniche (ovvero di isointensità soggettiva). L'andamento di queste curve (Raccomandazione Internazionale ISO/R226) mostra che, perché una vibrazione sia percepita come suono, bisogna che raggiunga un certo valore minimo di intensità (soglia inferiore di udibilità); al contrario esiste un valore massimo di tollerabilità dell'orecchio, sorpassato il quale si ha una sensazione di sofferenza (soglia del dolore). Inoltre, la sensibilità dell'udito è maggiore per le frequenze acustiche medie (fra

qualche centinaio e qualche migliaio di Hz) che ai due estremi della banda acustica, e che nel campo dei toni gravi molto intensi la sensibilità dell'udito cresce con la pressione acustica più rapidamente che nella restante parte dell'area di udibilità. Un'idea dell'andamento di tali curve è dato in fig. 1.13b.

- **Timbro**, caratteristica per la quale suoni di stessa altezza e stessa intensità possono essere assai spesso facilmente distinti (ad esempio una stessa nota musicale emessa con uguale intensità da due diversi strumenti musicali). E' legata principalmente alla struttura spettrale del suono complesso ma anche ad altri parametri fra cui l'intensità globale.



a)



b)

Fig. 1.13 a) Conversione di scala Hz/Bark b)Il campo di sensibilità dell'orecchio umano alle vibrazioni.

1.4 L'INGEGNERIA: IL SEGNALE VOCALE ELETTRICO E LA SUA ELABORAZIONE

L'elaborazione analogica, e ancor più quella digitale del segnale vocale elettrico hanno portato grandi cambiamenti nella nostra vita quotidiana: si pensi a tutti i sistemi di telefonia e di comunicazione vocale, ai riconoscitori vocali che ormai sono a corredo di molti apparecchi hi-tech e dei computer (soprattutto negli USA), ai sintetizzatori vocali. Per questo motivo, ma anche per rendere più chiara la descrizione del lavoro svolto per la presente tesi, trattiamo in questo paragrafo i fondamenti dell'approccio ingegneristico al segnale vocale.

1.4.1 I sistemi numerici di elaborazione del segnale

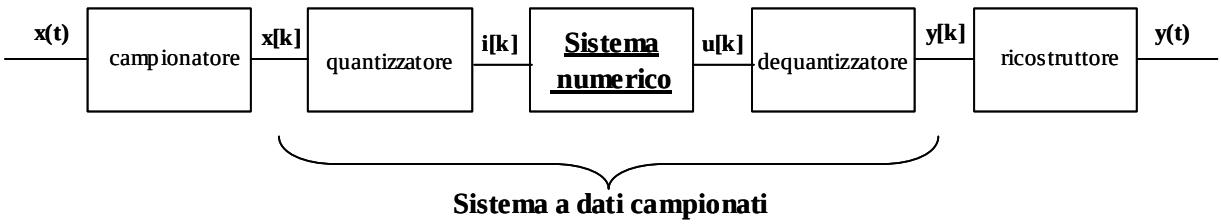


Fig. 1.14 Struttura di elaborazione per segnali unidimensionali.

Nel terzo paragrafo, si è definito il segnale vocale acustico come l'andamento temporale della variazione di pressione acustica nella zona limitrofa al parlatore. Questo segnale, per essere elaborato, viene trasdotto da un microfono, che lo trasforma in un segnale elettrico chiamato anch'esso vocale. La qualità del segnale riprodotto dipende quindi, in primo luogo, dalle caratteristiche del microfono. In pratica, il trasduttore di sorgente si limita a generare un segnale elettrico in qualche modo proporzionale a quello prodotto dalla sorgente. In questa sede, si farà conto che non ci sia perdita di segnale né degradazione di esso nel passaggio dalla forma d'onda acustica a quella elettrica (trasduttore ideale); nel seguito, ci si riferirà indifferentemente all'una o all'altra forma con il nome generico di segnale vocale.

La figura 1.14 rappresenta un generico sistema di comunicazione numerico. Nel caso più generale (e certamente per la voce), il segnale completamente numerico (sia in tempo che in ampiezza) i[k], che entra nel sistema di elaborazione numerica (ad esempio un calcolatore o un DSP), deriva dal rispettivo segnale analogico sul quale si è operato un campionamento e una quantizzazione. Nel passaggio dal sistema analogico a quello a dati campionati, sotto l'ipotesi che x(t) sia limitato in banda e che si siano rispettate

le condizioni del teorema del campionamento[14], non c'è alcuna degradazione (almeno teorica) del segnale. I segnali limitati nel tempo hanno banda infinita e quindi si ha comunque una perdita di qualità nel segnale campionato. In pratica si sceglie la frequenza di campionamento a seconda della banda di frequenze che contiene informazioni importanti per la specifica applicazione per cui è progettato il sistema di elaborazione. La qualità del segnale riprodotto dipende quindi anche dalla frequenza di campionamento scelta[15]. Il segnale vocale è sempre di banda base e può ritenersi membro di un processo aleatorio spesso assumibile come stazionario ed ergodico (caratterizzato quindi da proprietà comuni a tutti i membri del processo quali larghezza di banda, spettro di densità di potenza, ecc.). Inoltre, a questa categoria di segnali è applicabile il teorema del campionamento (P. Mandarini, 1990), e dunque ciascuno di essi è rappresentabile completamente attraverso la sequenza dei suoi campioni, presi a distanza temporale opportuna (si ricorda che, al massimo, la voce umana copre solo i primi 10kHz della banda acustica).

Per quanto riguarda il passaggio dal sistema a dati campionati a quello completamente numerico, è inevitabile una degradazione del segnale già in linea teorica (il cosiddetto rumore di quantizzazione). Questo è causato dal dover necessariamente usare un numero finito di registri di memorizzazione o una lunghezza di parola finita, rispettivamente per un'elaborazione via hardware o via software. Ciò nonostante, l'elaborazione numerica presenta dei vantaggi notevolissimi e, addirittura, a parità di costi, spesso superiore anche come qualità a quella puramente analogica.

1.4.2 Un modello per la generazione del segnale vocale

Alla base dell'approccio ingegneristico allo studio di fenomeni fisici c'è spesso la creazione di un modello del sistema. Riportiamo in figura 1.15 lo schema a blocchi dell'apparato fonatorio umano generalmente accettato. Il filtro digitale H(z) tiene conto dell'influenza esercitata dall'atteggiamento assunto dagli organi fonatori ed è in genere una funzione con soli poli (anche se tale ipotesi non è verificata ad esempio nella produzione di suoni nasali). In pratica, tale influenza corrisponde a modificare le frequenze di risonanza delle cavità del tratto vocale, che hanno l'effetto di far assumere allo spettro del segnale uscente una forma particolare, esaltandone energeticamente alcune bande di frequenza rispetto ad altre. Per quanto riguarda lo studio dei suoni nasali è usuale considerare il filtro H(z) come il parallelo di due filtri corrispondenti agli effetti del passaggio del segnale nella bocca e nel naso. L'amplificatore pilotato dal parametro G_0 tiene conto del livello energetico del segnale.

---

[14] Per rappresentare un segnale limitato in banda con banda pari a W, è sufficiente estrarre i campioni del segnale alla Frequenza di Nyquist pari a $F_N = 2W$ (quindi con un periodo $T = 1/2W$). Questa è la minima frequenza richiesta per ricostruire correttamente il segnale, valida, ovviamente, solo per un campionamento ideale.

[15] Ad esempio, nelle comunicazioni, spesso, il segnale vocale deve essere trasdotto, trasmesso e riprodotto, al solo scopo di rendere completamente riconoscibile il significato della locuzione e l'identità del parlatore, e ciò definisce una particolare esigenza di qualità che qualifica il segnale vocale come telefonico (nella pratica, quello con banda compresa tra 300 e 3400 Hz).

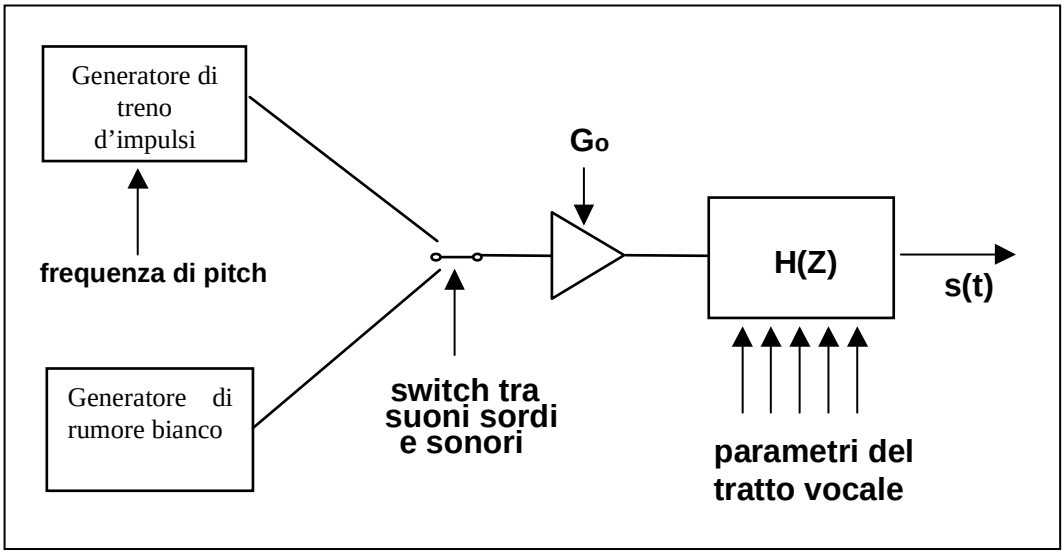


Fig. 1.15 Modello tempo-discreto dell'apparato fonatorio per la generazione di parlato.

1.4.3 Sottocampionamento e Sovracampionamento

Si è ritenuto utile riportare le due tecniche di base utilizzate per modificare la frequenza di campionamento. Esse, infatti, si sono rivelate essenziali (vedi paragrafo 3.2.4) per digitalizzare la base dati usata per l'analisi delle consonanti nasali.

Si definisce **sottocampionamento** l'operazione atta a ridurre il tasso di campionamento di un segnale. La tecnica prevede, ovviamente, una *decimazione* dei campioni nel tempo. Questo comporta, in frequenza, che la banda del segnale originale dopo sottocampionamento, aumenti proporzionalmente col fattore di decimazione M.

Le formule (1.4) esprimono il legame tra una sequenza e il rispettivo segnale analogico dal quale è derivata con campionamento di periodo T. Come si vede, lo spettro della sequenza è periodico di periodo 2π e l'asse delle "frequenze analogiche" Ω si trasforma nel nuovo asse delle "frequenze numeriche" secondo la relazione $\omega = \Omega T$, quindi, come mostrato in figura 1.16, perché la sequenza non sia affetta da *aliasing* occorre che sia $\Omega_0 \leq \pi/T$ (teorema del campionamento).

$$x[n] = x_a(nT) \qquad X(e^{j\omega}) = \frac{1}{T}\sum_{k=-\infty}^{+\infty} X_a\left(j\frac{\omega}{T} - j2\pi\frac{k}{T}\right) \tag{1.4}$$

$$x_d[n] = x[M \cdot n] = x_a(M \cdot nT) \qquad X_d(e^{j\omega}) = \frac{1}{M}\sum_{i=0}^{M-1} X\left(e^{j\left(\frac{\omega}{M} - i\frac{2\pi}{M}\right)}\right) \tag{1.5}$$

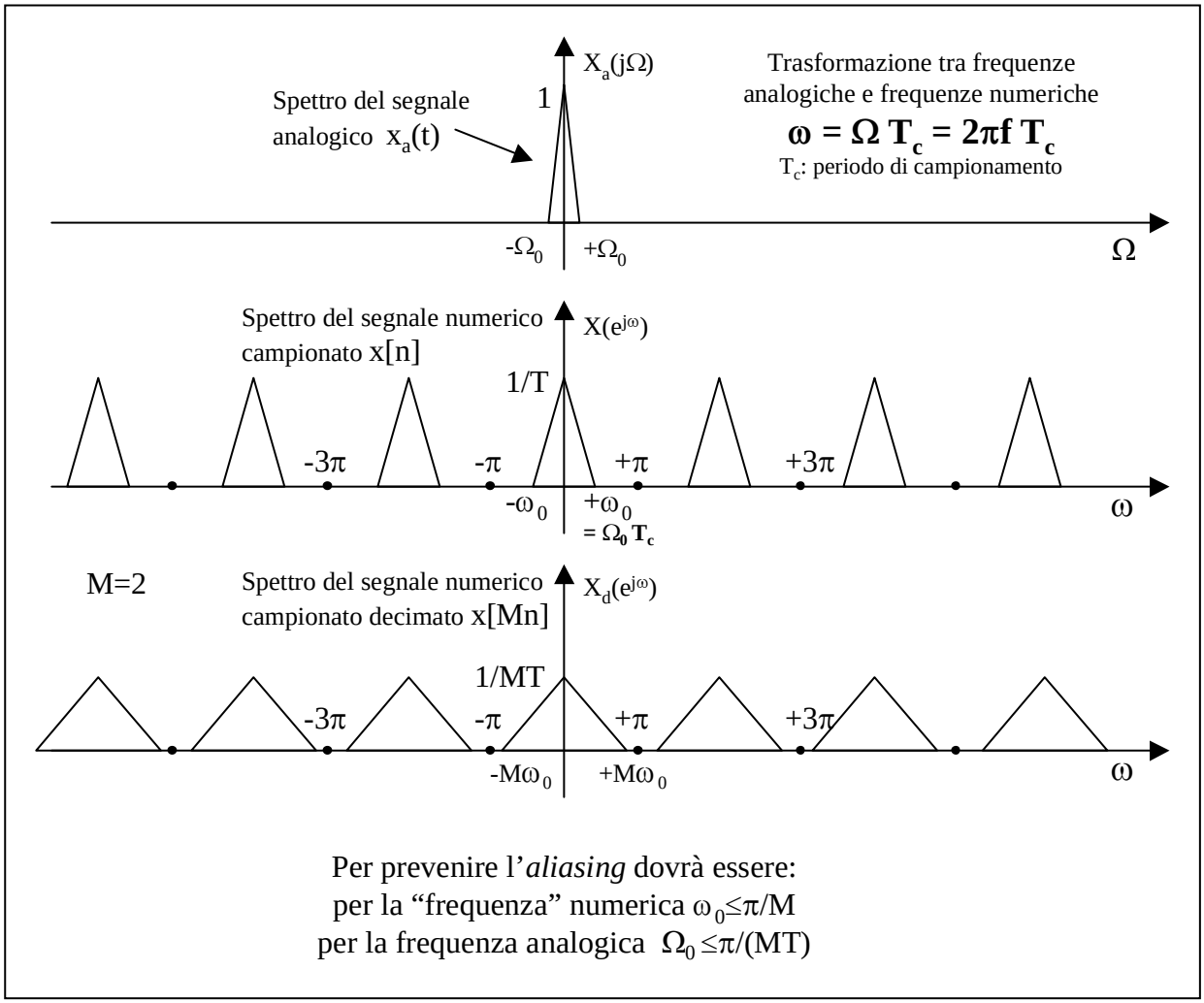


Fig. 1.16 Tecnica di sottocampionamento: legame (in frequenza) tra una sequenza, il segnale analogico dal quale è stata campionata e tra la stessa sequenza e decimata di un fattore M.

Le formule (1.5) esprimono, invece, il legame tra la stessa sequenza di prima e quella decimata di un fattore M (costruita cioè dalla prima prendendo un solo campione ogni M). Anche in questo caso il legame è chiaro: lo spettro originale viene espanso sull'asse ω di un fattore M. Perché non sussista aliasing la condizione è stavolta $\omega_0 \leq \pi/M$. Si veda la figura 1.16 a tal proposito.

L'operazione inversa della decimazione è l'*interpolazione*, detto anche **sovracampionamento**, che prevede l'inserzione di (L-1) campioni fittizi pari a 0 tra ciascuna coppia di campioni consecutivi della sequenza. Nel dominio della frequenza l'effetto di questa operazione consiste nel distanziare le repliche dello spettro a distanza L x (distanza originaria).

1.4.4 Lo studio nel dominio della frequenza: l'analisi spettrale

Il segnale vocale può essere utilmente studiato con vari approcci, per dedurne le caratteristiche e associarle ai vari fonemi e addirittura ai vari modi di articolazione. Le tecniche più usate sono quelle che prevedono lo studio nel dominio del tempo o nel dominio della frequenza, effettuando eventualmente elaborazioni ulteriori tese a evidenziare alcune proprietà particolari del segnale.

Come noto la trasformata di Fourier di un segnale s(t) è detta *spettro* del segnale, per cui per quanto riguarda lo studio in frequenza si parla in genere di **analisi spettrale** del segnale. La tecnica seguita in questa sede prevede il campionamento del segnale s(t) e il suo studio tramite elaborazioni di tipo numerico (Trasformata discreta di Fourier, DFT). Ricordiamo brevemente l'espressione matematica della DFT, che prevede, usando una finestra (detta normalmente **frame**) di N campioni del segnale s(t), il calcolo di N campioni in frequenza della trasformata di Fourier F[s(t)], nella banda propria del segnale:

$$DFT(s(nT)) = S(k) = \sum_{n=0}^{N-1} s(nT) e^{-\frac{j2\pi knT}{N}}$$

$$s(nT) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) e^{\frac{j2\pi knT}{N}} \qquad \text{per k (n), da 0 a (N-1)} \qquad (1.6)$$

dove T è l'inverso della frequenza di campionamento. Se la banda del segnale s(t) è B e si è scelta una frequenza di campionamento T=1/2B, gli S(k) sono i campioni della sua trasformata continua di Fourier a distanza B/N.

Per l'analisi del segnale vocale, nella scelta della lunghezza del frame, occorre tener presente che bisogna eseguire un'analisi in intervalli di tempo sufficientemente brevi, da poter associare le caratteristiche del segnale a quelle del condotto vocale, ma sufficientemente lunghi perché le caratteristiche del segnale possano essere considerate stazionarie in tale intervallo, con sufficiente approssimazione. Si deve inoltre tenere presente il principio generale in base al quale la risoluzione in frequenza è tanto migliore quanto più grande è il frame scelto. Se, infatti, come detto, N è il numero di campioni di un frame analizzato e se indichiamo con f_c la frequenza di campionamento utilizzata per il segnale, la risoluzione in frequenza che si ha quando vengono calcolati gli spettri è data dalla formula:

$$Risoluzione\_frequenziale = \frac{f_c}{N} \qquad (1.7)$$

Per trovare un compromesso tra le due esigenze opposte di località dello spettro e di risoluzione in frequenza, si usa un ***fattore di sovrapposizione S*** *tra frame adiacenti* non nullo, ma compreso tra 0 e 1. In

pratica, ogni N·(1-S) campioni è analizzata una finestra di segnale lunga N campioni[16]. I parametri su cui si può agire sono quindi la *dimensione dei frame* N che determina la *risoluzione in frequenza* dello spettro (tanti campioni vi sono in un frame e tanti campioni vi sono nella DFT di quel frame) ed il *fattore di sovrapposizione* S dei frame che influisce sulla *risoluzione temporale* dello spettrogramma (più i frame sono sovrapposti, più frame vi saranno in un segnale lungo T).

Nell'analisi del segnale vocale risulta particolarmente utile l'osservazione dell'evoluzione temporale delle caratteristiche spettrali di un segnale. Ciò è possibile tramite lo **spettrogramma** ottenuto affiancando gli spettri locali di finestre contigue. Nello spettrogramma, la grandezza riportata in ordinata è la *frequenza*, sulle ascisse è riportato il *tempo* (tutti i vari frame analizzati) mentre l'*ampiezza* dello spettro è data dall'annerimento, maggiore o minore, sul disegno. Se N, ampiezza della finestra, è "grande" (128 o 256 campioni) lo *spettrogramma* viene detto **narrow band**, in quanto il passo di approssimazione della trasformata di Fourier è piccolo (circa 40 o circa 20Hz rispettivamente, per B=5kHz), se "breve" (16 o 32 campioni) viene invece detto **wide band**, avendo passo di approssimazione grande (circa 300 o circa 150Hz per B=5kHz). Chiaramente per i motivi precedentemente illustrati, lo spettrogramma wide band riesce a mostrare caratteristiche di breve durata del segnale al prezzo di una minore accuratezza nel campionamento in frequenza. Lo spettro narrow band è quindi particolarmente adatto per l'analisi dei segmenti fonici; lo spettro wide band, invece, è molto utile nello studio dei contoidi e nell'analisi delle caratteristiche individuali della voce, come il tono e l'intonazione.

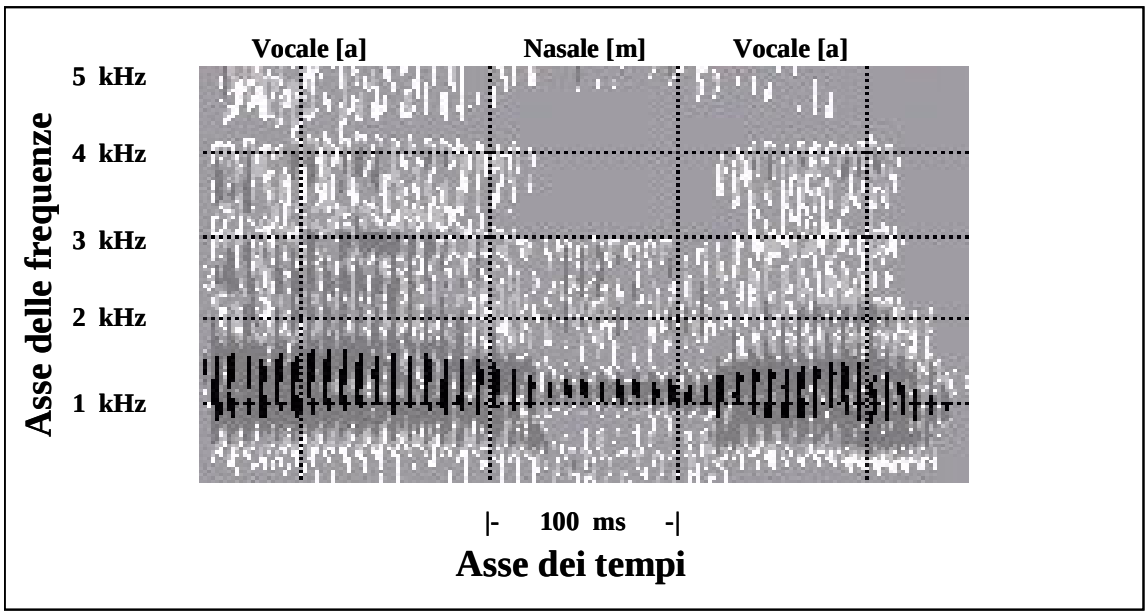


Fig. 1.17 Esempio di spettrogramma.

Esistono vari approcci nell'analisi del segnale vocale. Oltre l'analisi spettrale attuata mediante DFT, altre tipiche elaborazioni possibili sono la costruzione della funzione di *zero crossing*, per evidenziare i

[16] Se, ad esempio, il numero di campioni per finestra è pari a 128 e il fattore di sovrapposizione è 3/4, il risultato è che ogni 32 campioni (128·1/4) viene analizzato un tratto di segnale lungo 128.

momenti di silenzio fonetico, gli algoritmi di *pitch tracking*, il calcolo dell'energia locale del segnale, l'analisi LPC (*Linear Predictive Coding*), che aiuta molto nell'individuazione delle formanti, l'autocorrelazione, l'estrazione dei parametri statistici classici (covarianza, valore medio, ...). Sulle analisi tramite FFT e LPC in particolare conviene soffermarsi, visto l'uso estensivo che se ne farà nel seguito.

Analisi con la FFT

L'analisi in frequenza del segnale vocale può essere condotta eseguendo direttamente la FFT delle sequenze di campioni contenuti in ogni frame. Poiché la FFT di una sequenza di lunghezza "m" è la trasformata di Fourier del segnale periodico di periodo "m", ottenuto replicando la sequenza di durata finita (figura 1.18), essa conterrà delle componenti in frequenza spurie, non legate al segnale originario, ma semplicemente introdotte dalle brusche variazioni di ampiezza dovute alle repliche della sequenza di durata "m".

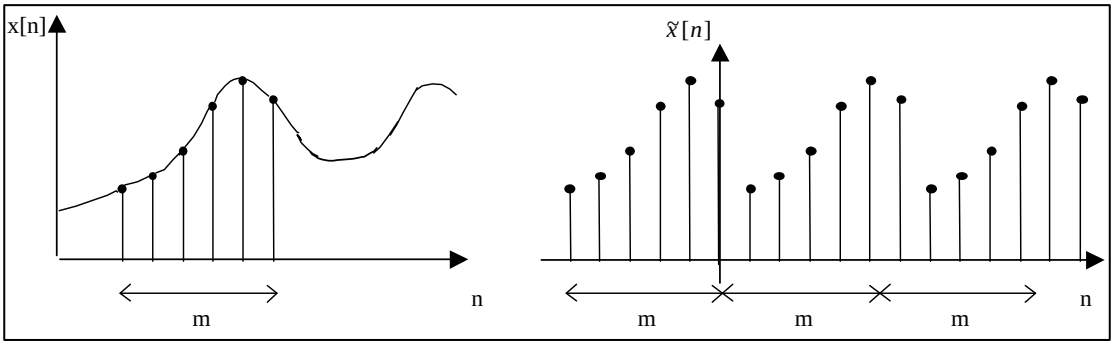


Fig. 1.18 Sequenza *x[n]* di durata finita "m" e corrispondente sequenza periodica di periodo "m".

Per prevenire la formazione di queste frequenze spurie, il tratto di segnale contenuto nel frame di cui si vuole calcolare la DFT, viene modulato con un'opportuna finestra, che attenua il segnale agli estremi dell'intervallo. La funzione di modulazione impiegata è la **finestra di Hamming** (o del coseno rialzato), la cui espressione è:

$$w(n) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \qquad \text{con } 0 \le n \le N\text{-}1 \qquad (1.8)$$

che moltiplicata per il tratto di segnale contenuto nel frame, ne preserva la parte centrale. Con l'impiego della finestratura, si rende ancora più necessario lo slittamento di ciascun frame di almeno N/2 campioni (se N è il numero di campioni per frame), per non perdere le informazioni del segnale agli estremi del frame stesso. Infatti, con questa accortezza, i campioni che si trovano attenuati agli estremi di un frame, risulteranno praticamente inalterati all'interno di quelli immediatamente precedente e successivo.

Un'ulteriore operazione da compiere, prima di visualizzare la DFT del segnale, è quella di **preenfasi**, ottenuta con un filtro la cui risposta impulsiva è: $h(n) = \delta(n) - \alpha\,\delta(n-1)$. L'effetto che si desidera ottenere è quello di una *enfatizzazione* dello spettro tramite una trasformazione tesa ad esaltare l'importanza del contenuto energetico in alta frequenza, altrimenti poco visibile graficamente (ma non per questo meno importante dal punto di vista percettivo). Con un valore di α pari a 0.95, come comunemente si usa, le

basse frequenze vengono attenuate notevolmente (fino a oltre 20 dB), mentre al limite della banda di lavoro si ha un'amplificazione di circa 6 dB.

L'andamento del modulo della funzione di trasferimento del filtro di preenfasi è riportato in figura 1.19.

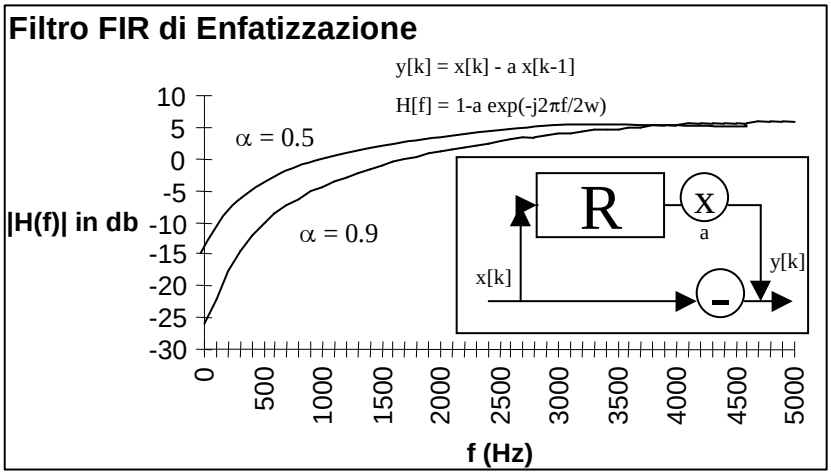


Fig. 1.19 Funzione di trasferimento del filtro di preenfasi.

Nella figura 1.20 è riportata, come esempio, la rappresentazione di un frame di segnale sinusoidale modulato con la finestra di Hamming. La grandezza nella parte inferiore della figura, è ovviamente il modulo quadrato, previa preenfasi, della DFT del segnale in questione; mentre lo spettro di un segnale perfettamente sinusoidale è formato da un'unica riga, lo spettro della sinusoide "finestrata" ha una banda più larga.

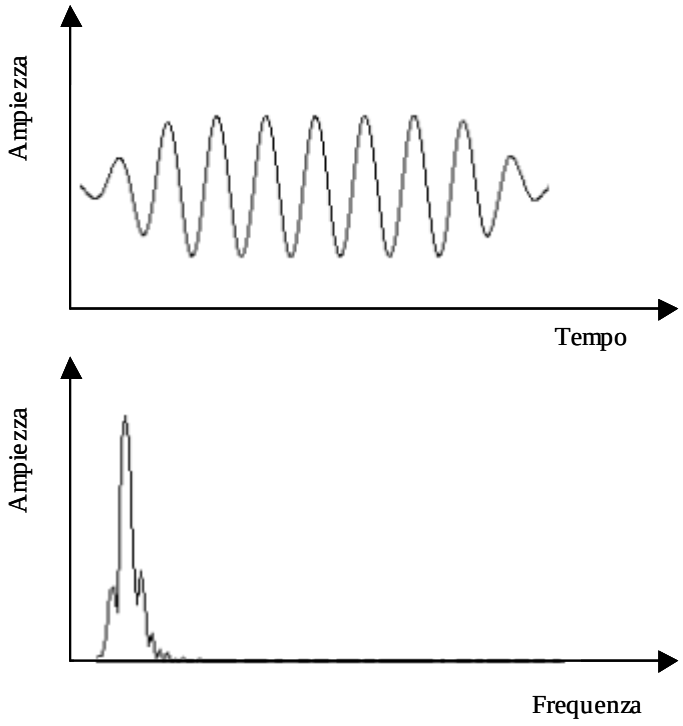


Fig. 1.20 Sinusoide finestrata secondo Hamming e sua DFT preenfatizzata.

Analisi LPC

Una delle più efficaci tecniche di analisi del segnale vocale è quella della predizione lineare. L'importanza di tale metodo risiede nella capacità di fornire una stima accurata dei parametri del tratto vocale e delle frequenze formanti, e nella sua velocità di calcolo.

Il problema fondamentale della predizione lineare è quello di esprimere il generico campione del segnale come una combinazione lineare dei "p" campioni immediatamente precedenti:

$$\tilde{s}(n) = \sum_{j=1}^{p} a_j s(n-j) \tag{1.9}$$

I coefficienti incogniti a_j della combinazione lineare prendono nome di **coefficienti di predizione**. Il problema della determinazione dei coefficienti incogniti è affrontato con il criterio di minimizzazione dell'errore quadratico medio di predizione.

Tale errore è definito come

$$E_{n,m} = \sum_{i=0}^{m} e_n^2(i) = \sum_{i=0}^{m} \left(s(n+i) - \tilde{s}(n+i)\right)^2 \tag{1.10}$$

dove n è il primo campione della finestra di ampiezza m. Sostituendo nella precedente relazione, l'espressione del campione predetto, si ottiene

$$E_{n,m} = \sum_{i=0}^{m} \left(s(n+i) - \sum_{j=1}^{p} a_j s(n+i-j) \right)^2 \tag{1.11}$$

La minimizzazione dell'errore quadratico medio, si ottiene imponendo uguali a zero le sue derivate parziali rispetto alle p incognite a_j per j = 1, 2, …, p. Così facendo si ottiene un sistema lineare di p equazioni in p incognite, che, risolto, dà proprio i coefficienti cercati.

$$\sum_{i=0}^{m} s(n+i-k) \cdot s(n+m) = \sum_{j=0}^{p} a_j \sum_{i=0}^{m} s(n+i-k) \cdot s(n+i-j) \tag{1.12}$$

Ricordando il modello di figura 1.15, chiamando u(n) il segnale in ingresso all'amplificatore, quello in uscita sarà $G_0u(n)$, ricordando poi che $s(z)=G_0u(z)H(z)$, si può stabilire l'uguaglianza $s(z)/H(z)=G_ou(z)$. Come già specificato la H(z) è una funzione di soli poli e perciò può essere scritta

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p'} \alpha_k \cdot z^{-k}} \tag{1.13}$$

per cui antitrasformando l'uguaglianza impostata sopra si ottiene:

$$s(n) = \sum_{k=1}^{p'} \alpha_k \cdot s(n-k) + G_o \cdot u(n) \tag{1.14}$$

Il risultato fondamentale di questa analisi è il seguente: si può dimostrare che se un segnale vocale è generato con una sintesi come quella descritta dall'equazione (1.14), allora i p coefficienti che minimizzano l'errore quadratico medio in una finestra di larghezza m coincidono con gli α_k coefficienti del filtro che modella il tratto vocale. Questo importante risultato comporta anche che l'errore commesso usando l'approssimazione lineare di cui sopra è pari a $G_ou(n)$, cioè un treno di impulsi, di piccola ampiezza per la maggior parte del tempo. Inoltre ciò comporta che il calcolo dei coefficienti di predizione

fornisce automaticamente i parametri del filtro H(z), che con G_0 e il bit di selezione per i suoni sordi o sonori costituisce una rappresentazione completa del segnale vocale, frame per frame. Come sempre accade in questi algoritmi, esiste un trade-off tra la complessità di calcolo e l'accuratezza della rappresentazione: più sono i poli della H(z) (cioè i coefficienti dell'approssimazione tramite combinazione lineare), maggiore è la complessità di calcolo per la soluzione del sistema lineare (1.12).

La funzione $H(z)G_0$ fornisce lo spettro del segnale approssimante, ovvero uno spettro approssimato del segnale s(t). Tale spettro viene detto LPC e presenta la caratteristica utile di essere molto fedele nell'individuare i massimi dello spettro reale (ma poco per quanto riguarda i suoi minimi). Inoltre il parametro p permette di controllare la precisione dell'approssimazione, nel senso che un p elevato permette di evidenziare nello spettro LPC anche massimi vicini dello spettro, che altrimenti sarebbero stati fusi in un unico picco situato in una regione intermedia. Tenendo presente queste considerazioni, risulta evidente come le osservazioni fatte durante l'analisi LPC vadano sempre interpretate tenendo presente i limiti e le approssimazioni descritte. Tuttavia tale analisi riesce particolarmente utile nel processo di individuazione delle formanti (si tratta infatti di picchi ben distanziati), e costituisce un utile strumento anche per l'individuazione del pitch.

CAPITOLO 2

IL FENOMENO DELLA GEMINAZIONE E LE CONSONANTI NASALI

INTRODUZIONE

Il fenomeno della geminazione è una caratteristica molto rara nelle lingue. Tra le lingue che presentano questo fenomeno, l'Italiano[1] è quella col maggior numero di parlatori e probabilmente quella che ne fa l'uso più estensivo. Per questo la lingua italiana presenta un grande interesse per chi si occupa di questo argomento in modo scientifico.

In questo capitolo, dopo aver definito in maniera chiara il problema della geminazione, si traccerà un quadro della situazione attuale sullo studio di questo fenomeno, sia dal punto di vista fonetico sia dal punto di vista ingegneristico. Infine, saranno descritte nel dettaglio le consonanti nasali oggetto di studio, prima di affrontarne l'analisi acustica.

[1] Pochissime lingue hanno, come l'Italiano, molte geminate. L'Hindi e il Finnico sono tra queste. Il Francese conosce consonanti geminate solo in qualche raro caso, mentre né l'Inglese, né il Tedesco, né lo Spagnolo e il Portoghese possiedono questa caratteristica. L'assenza di geminazione nella pronuncia dell'italiano da parte di parlatori non nativi è forse l'errore più tipico nella casistica dei cosiddetti *cross-language errors* (errori fonetici commessi da parlatori di una lingua che non sia la propria lingua madre) relativamente all'Italiano.

2.1 LA GEMINAZIONE

Introdurre il fenomeno della geminazione non è cosa semplice poiché esso è ancora in fase di studio e quindi non vi sono definizioni universalmente accettate.

In Italiano, vi sono diverse **coppie minime**, ossia coppie di parole che possono essere distinte solo per la presenza o l'assenza della geminazione in una delle consonanti. Un esempio di ciò è dato dalla coppia minima *pane, panne.* Il Malmberg dà la seguente definizione: "Se una consonante è scissa in due parti da una frontiera sillabica, la chiamiamo *geminata*" (Malmberg, 1974).

La geminazione si esprime graficamente scrivendo due volte la lettera relativa alla consonante geminata, ma non c'è sempre un'esatta corrispondenza tra la pronuncia e la grafia.

Conviene sin da ora precisare la terminologia che si userà nel seguito della tesi: si chiamerà **rafforzamento sintattico** o **geminazione** il fenomeno fonetico, mentre con la parola **raddoppiamento** si preferirà indicare l'espediente grafico che serve a trascriverlo, inoltre sarà detta *singola* la consonante che non subisce il fenomeno del rafforzamento e *geminata* quella che lo subisce. Nel seguito, si useranno molto spesso anche i termini "pronuncia singola" e "pronuncia geminata". Con essi si indicheranno tutti gli effetti che comporta la geminazione sull'intera parola (in particolare sulla consonante e sui fonemi adiacenti).

2.1.1 La geminazione dal punto di vista grammaticale

Non esiste, come appena accennato, una "corrispondenza biunivoca" tra pronuncia geminata di una consonante e corrispondente trascrizione grafica. Se consideriamo forme come *accorrere*, *eccellere*, *accanto* ecc., si ricava che si tratta di processi assimilativi già attuatisi nel latino classico (rispettivamente, negli esempi citati, da ad+currere, ex+cellere, ad+canto). Queste parole sono pronunciate [ak 'kor re re, et·'tʃel le re, ak·'kan:to] e anche la grafia ne tiene conto. Invece, nei casi di *a capire*, *va via*, *tu sai*, ecc., si vede agire lo stesso principio a livello di pronuncia per cui sarà [a kka'pi:re, va v·'vi·a, tu s·'sa·i], ma questa volta la grafia non ne tiene conto. In ogni modo, questo fenomeno è giustificatissimo perché non si parla pronunciando singole parole staccate, come potrebbe far supporre la scrittura, bensì emettendo intere fonie che formano la cosiddetta "catena parlata" (Canepari, 1979).

Si esporranno ora, in maniera schematica, alcune regole pratiche per scrivere e/o pronunciare le geminate in italiano.

Per quanto riguarda il raddoppiamento nella grafia, si possono ricordare le seguenti norme:

a) Non si raddoppiano mai le consonanti iniziali e finali.
b) Dinanzi a -ione, g e z non si raddoppiano mai (p. es. *ragione, azione,* …).
c) Non si raddoppiano sc, gn, gl, mentre, per rafforzare ch e gh si raddoppiano solo la c e la g (p. es. *ricche, agghiacciante,* …).
d) Il raddoppiamento di q è cq (tranne soqquadro).
e) Si raddoppiano i prefissi a, e, o, da, se, su, so, ra, fra, sopra, sovra, contra (ma non contro!) ecc. (p. es. *sebbene, supporre, frattanto,* …).

Per quanto riguarda, invece, la pronuncia all'interno delle frasi, il rafforzamento sintattico è prodotto da alcune forme uscenti in vocale e legate, semanticamente e foneticamente, alla parola seguente, che comincia con una delle consonanti che possono ricorrere geminate anche all'interno delle parole. Si riassumono le principali forme che si pronunciano rafforzate:

a) La vocale a, e i monosillabi "forti" da, su, tra, fra (p. es. *tra noi, fra mesi,* …).
b) I monosillabi che hanno accento grafico, come dà, dì, là, già, giù, sé, ciò, più ecc. (p. es. *dà tutto, già lo vedo, ciò fu fatto,* …).
c) I verbi ho, ha, do, fa, fu, va (p. es. *do tutto, fa male,* …).
d) Le parole che, chi, qui, qua, se, ma, o, e, tu ecc. (p. es. *chi sa!, qui sotto,* …).
e) I polisillabi tronchi, con l'accento sull'ultima sillaba, come perché, poiché, però, andò, caffè, farà ecc.
f) I quattro bisillabi piani come, dove, sopra, qualche (p. es. *sopra tutto*).

E' importante, infine, vedere quali forme non producono il rafforzamento sintattico. Esse sono, i monosillabi "deboli" la, le, lo, i, li; i monosillabi apostrofati nella scrittura come di', va' ecc. o le esclamazioni; inoltre di, ne, me, mi, te, ti, se, si, ce, ci, ve, vi, glie, gli.

2.1.2 La geminazione dal punto di vista fonetico

I suoni del linguaggio si distinguono gli uni dagli altri non solo per i loro tratti puramente qualitativi, ma anche per quel che concerne la "*quantità*". Già all'inizio del secolo, fonetisti come E.A.Mayer avevano intuito l'importanza linguistica degli aspetti quantitativi come la lunghezza o durata di un fonema, o anche l'intensità (energia) articolatoria. Una vocale, per esempio, è generalmente più lunga davanti ad una spirante che davanti ad un'occlusiva o davanti ad una sonora che davanti ad una sorda, più lunga anche davanti a [r] che davanti alle nasali e a [l] (Malmberg, 1974). Ancora, una vocale anteriore è spesso un po' più breve di una vocale posteriore. Per le consonanti valgono regole simili. Una sorda è normalmente più lunga di una sonora e così via. Tutti questi esempi fanno pensare che la misura di quantità relative, basate cioè sul confronto dei risultati ottenuti per differenti suoni nella stessa posizione o per lo stesso suono in posizioni diverse, sia forse molto più interessante delle misure assolute; inoltre, non tutte le variazioni di quantità "misurabili" hanno un valore linguistico propriamente detto, nel senso che non tutte portano differenze di significato. Perciò, l'osservazione condotta sulla reazione percettiva dell'uomo può dare l'auspicabile e definitiva oggettività rispetto al valore linguistico di quantità misurabili come energia e lunghezza. Queste considerazioni sono alla base delle teorie sulla geminazione.

Abbiamo introdotto il fenomeno del rafforzamento sintattico parlando di coppie minime (p. es. *fato* vs. *fatto, casa* vs. *cassa, eco* vs. *ecco*, ecc.) ma in Italiano, anche per le così dette coppie sub-minime (p. es. *l'ho dato* vs. *lodato, tra monti* vs. *tramonti, né gare* vs. *negare*)[2], solo una corretta pronuncia del fenomeno del rafforzamento permette di eliminare i conflitti omofonici.

[2] Siccome la grafia non segna questo fenomeno, i settentrionali, che pure si sforzano di pronunciare "lunghe" le consonanti segnate doppie nella grafia, non producono geminata in questi casi, non accorgendosi (o al più

Secondo Muljacic, i fonemi consonantici che possono ricorrere singoli o geminati sono quindici. Essi sono: [f, v, s, p, t, k, b, d, g, m, n, l, r, tΣ, dj]. Nelle descrizioni dell'italiano, i fonetisti si combattono su due punti di vista diametralmente opposti riguardo alla geminazione già dalla fine degli anni trenta, e la polemica non pare ancora esaurita (Muljacic, 1972). Un primo gruppo di studiosi (detti anche ***monofonematisti***) si rende fautore di una *classe speciale* composta di quindici fonemi chiamati lunghi, rafforzati o intensi o, recentemente, anche tesi. Un secondo gruppo, tra i quali il Muljacic stesso, (chiamati, per contrapposizione, ***bifonematisti***), considerano invece che quello che distingue una singola da una geminata, non è l'opposizione tra un fonema semplice e uno rafforzato, ma la presenza di un *fonema in più*. In pratica, una geminata sarebbe una consonante singola ripetuta due volte. Secondo i bifonematisti, quindi, l'ortografia denoterebbe (sebbene in modo imperfetto) lo stato di fatto fonologico.

In più di uno studio è stato messo in evidenza che le consonanti geminate influenzano con la loro presenza anche gli altri fonemi delle sillabe cui appartengono, anche se non è ancora chiarissimo quando e come questo avvenga.

Da quanto detto finora si capisce che il problema della geminazione è decisamente complesso e comprende aspetti diversi e interdipendenti.

2.1.3 La geminazione dal punto di vista acustico-ingegneristico

L'approccio ingegneristico allo studio di problemi linguistici non gode sicuramente della secolare tradizione della fonetica. Per questo motivo non esistono molti lavori "tecnici" sul fenomeno della geminazione in italiano. Bisogna inoltre dire che gli studi ingegneristici sulla geminazione hanno finalità solitamente diverse da quelli fonetici. Infatti gli studi acustici sono solitamente orientati alla ricerca di risultati sperimentali statistici utili per *riconoscere in maniera automatica* la presenza o l'assenza di geminazione, e per la caratterizzazione di certi parametri perché si possa *produrre con voce sintetica* una consonante geminata. Tuttavia questi studi possono risultare utili anche per dirimere controversie più squisitamente teoriche e per avere un quadro più chiaro del problema della geminazione.

Nel presente lavoro, che ovviamente affronta la geminazione da un punto di vista acustico ingegneristico, ci si è serviti di programmi di elaborazione del segnale vocale (implementati su calcolatore) e si è analizzata una base dati costruita appositamente. Sia i programmi sia la base dati saranno descritti nel prossimo capitolo.

Non sono noti, al momento, lavori estensivi sulla geminazione delle consonanti nasali. Lavori precedenti a questo, con simile impostazione e metodologia, sono riportati in bibliografia (A.Vannucci 1993; R.Rossetti, 1993; F.Argiolas, 1995; F.Macrì 1995; M.Giovanardi, 1998; A.Esposito e M.G. Di Benedetto, 1999). In particolare, si ricordano le tesi svolte negli anni 1993, 1995 e 1998 presso lo stesso laboratorio voce del dipartimento INFOCOM dell'Università di Roma "La Sapienza" presso il quale è stata svolta anche questa. Esse si sono occupate delle consonanti occlusive [p, b, t, d, c, g] ([A.Vannucci 1993; R.Rossetti, 1993), delle consonati liquide [l, r] (F.Argiolas, 1995; F.Macrì, 1995) e delle consonanti fricative [f, v, s, z, ʃ](M.Giovanardi, 1998), e sono state un ottimo punto di riferimento. Questa tesi,

---

giudicandolo uno spiacevole abuso linguistico) che i centrali e i meridionali lo fanno, felici come sarebbero di poter eliminare tutte le doppie anche all'interno delle parole (Canepari, 1979).

quindi, si colloca nell'ambito di un progetto più ampio (progetto GEMMA) per lo studio della geminazione di tutte le consonanti italiane e per la loro caratterizzazione tramite parametri acustici. Dopo aver descritto l'analisi delle consonanti nasali nel quarto capitolo, nel quinto sarà operato un confronto tra tutte le consonanti studiate nei lavori appena citati; verrà infine operato un confronto anche con i risultati ottenuti in studi riguardanti la geminazione in altre lingue.

2.2 LE CONSONANTI NASALI IN ITALIANO

I contoidi *nasali* italiani (*ŋ, n, m)* sono contenuti in *gnomo, annunciare, manca*, e si chiamano nasali in quanto l'aria esce solo dal naso perché nella cavità orale, o buccale, s'è formata un'occlusione completa tra gli organi fonatori mentre il velo palatino è nella posizione abbassata, come nel respiro (vedi fig.2.1). Per i nasali italiani visti sopra, l'occlusione avviene tra le labbra per *m* o tra una parte della lingua e un punto della volta palatina per gli altri, come vedremo più avanti. Generalmente i nasali sono sonori.

Le consonanti nasali che esistono anche nella forma geminata sono *n* e *m.* In mancanza di una regola ben precisa e riconosciuta, le versioni geminate saranno indicate sempre aggiungendo i due punti alla lettera che indica la sua versione singola (p. es. /m:/, /n:/); benché questa notazione di norma indichi un allungamento del fonema corrispondente, in questo caso vuole solo essere una caratterizzazione grafica di una consonante geminata. Vediamo in dettaglio i contoidi nasali a seconda del punto di articolazione.

2.2.1 Nasale bibiale

La figura 2.1 rappresenta il contoide bilabiale nasale [m] che differisce dall'occlusivo sonoro [b] per il fatto che il velo palatino è nella posizione abbassata, come per tutte le nasali.

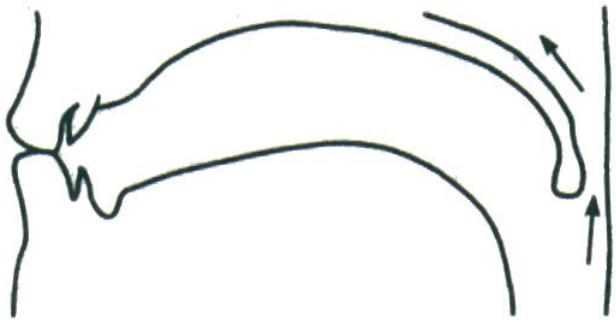


Fig. 2.1 Nasale bibiale [m]

2.2.2 Nasale Labiodentale

Nella figura 2.2 si vede il nasale labiodentale che in italiano viene articolato automaticamente per assimilazione anticipatoria, o regressiva, quando n sia seguito da f o v: (es. anfora, invidia).

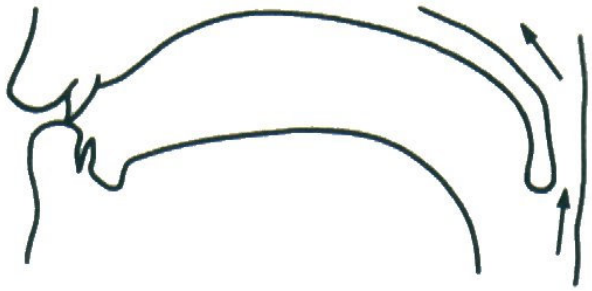


Fig. 2.2 Nasale labiodentale [n]

2.2.3 Nasale Dentale

Il nasale sonoro [n] è articolato come dentale per assimilazione ai dentali seguenti come in: *pranzo, intenso*. La parte della lingua che interviene nell'articolazione è la punta e/o la corona. La figura 2.3 dà l'articolazione del fono nasale.

Fig. 2.3 Nasale dentale [n]

2.2.4 Nasale Alveolare

L'articolazione normale di [n], quando non è influenzata dai contoidi seguenti, è alveolare: *nonno* (fig. 2.4), mentre in *tanto* essa è dentale perché s'assimila al punto d'articolazione dell'occlusivo dentale seguente. Comunque si usa lo stesso simbolo dato che l'impressione uditiva è praticamente la stessa.

Fig. 2.4 Nasale alveolare [n]

2.2.5 Nasale Prepalatale

Chiameremo *pre palatali* i suoni che s'articolano con la punta e la corona della lingua (o corona e predorso) contro la zona del palato duro che è stata definita prepalato**.** In italiano *n* seguita dagli alveopalatali si articola in questo modo (come in *conscio, mangio)* per assimilazione preparatoria.

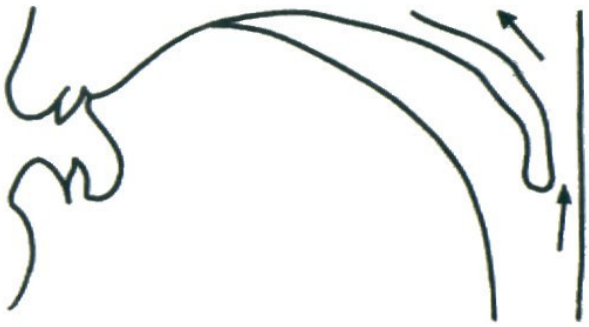


Fig. 2.5 Nasale prepalatale [n]

2.2.6 Nasale Palatale

Al punto d'articolazione *palatale* appartiene il contoide nasale di *pegno.*

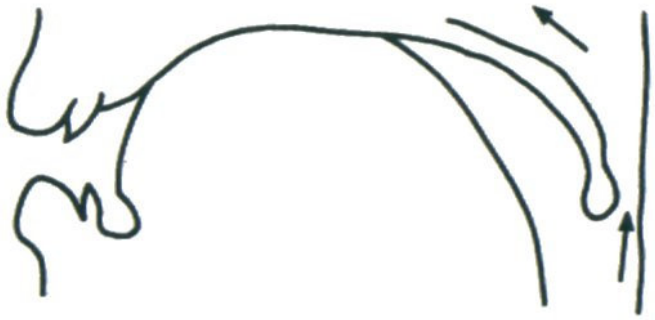


Fig. 2.6 Nasale palatale [n]

2.2.7 Nasale Velare

Consideriamo la parola *fango*: in questo esempio vediamo che la *n* che precede gli occlusivi velari si assimila a questi, divenendo a sua volta velare. Anche in italiano quindi esiste questo suono anche se senza valore distintivo, mentre in inglese e tedesco, per esempio, esso può avere la capacità di distinguere segni linguistici con significati diversi, per cui in queste lingue [ŋ] è spesso considerato *fonema,* mentre in italiano è solo un *allòfono combinatorio,* o *tassòfono,* che dipende dal contesto dei foni vicini.

Fig. 2.7 Nasale velare [n]

2.3 LA NASALIZZAZIONE DELLE VOCALI.

Nel parlato vi sono diverse situazioni in cui si può avere l'accoppiamento della cavità nasale al tratto vocale. Naturalmente una di queste situazioni si verifica quando si pronunciano le consonanti nasali. Può però accadere che per problemi anatomici o funzionali l'accoppiamento sia permanente o che esso, anche in situazioni normali, permanga abbastanza, dopo la pronuncia di una nasale, da influenzare la vocale che segue. Si può avere anche un effetto anticipatorio in cui l'accoppiamento comincia prima della pronuncia della nasale ed è la vocale precedente ad essere, per così dire, nasalizzata. Questo fenomeno della nasalizzazione delle vocali è presente in molte lingue (ad esempio l'Inglese) ma presenta notevoli differenze tra una lingua e l'altra (Stevens et al, 1987). Alcune lingue come il Francese ed il Portoghese distinguono dal punto di vista fonetico vocali nasali e non nasali. Dal punto di vista acustico, una teoria che renda possibile la predizione del suono prodotto in una vocale nasalizzata, deve necessariamente riguardare il calcolo della risposta di un sistema che modellizzi l'accoppiamento del tratto vocale al tratto nasale al livello del velo faringeo.

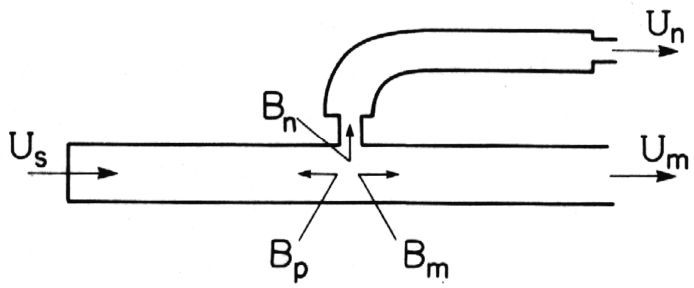


Fig. 2.8 Il modello acustico dell'accoppiamento naso-bocca proposto da Stevens.

Dalla figura 2.8 si può avere un'idea di un modello di questo accoppiamento. Il modello in esame è quello proposto da Stevens (1998). U_s,U_n, e U_m rappresentano rispettivamente le velocità dei volumi d'aria nella glottide, nelle narici e nella bocca mentre B_p, B_n, e B_m sono le suscettanze acustiche nel punto di accoppiamento guardando rispettivamente nella faringe, nella cavità nasale e nella cavità orale.

Secondo Fant (1960) e Fujimura e Lindqvist (1971) la principale differenza tra la funzione di trasferimento per il sistema che modellizza una vocale nasalizzata e quella per il sistema relativo ad una non nasalizzata, è la presenza di formanti aggiuntive come conseguenza, appunto, dell'accoppiamento acustico del tratto nasale a quello orale. Stevens (Stevens et al, 1987) sostiene che le vocali che sono percepite come nasali abbiano in comune la proprietà che la prima formante è sostituita da due picchi a distanza di 200-400Hz dovuti ad una combinazione polo-zero nella funzione di trasferimento. Per le lingue in cui la nasalizzazione delle vocali è molto forte e addirittura distintiva sono stati condotti molti studi. Per l'Italiano, in cui la nasalizzazione delle consonanti non è sicuramente così evidente come in

altre lingue, non si conoscono studi né acustici né percettivi su questo argomento. Le nostre considerazioni quindi partono semplicemente dalla consapevolezza che vi sia la possibilità teorica di un accoppiamento del tratto nasale che influenzi anche le vocali e dall'osservazione, durante la misura dei parametri frequenziali, di un picco aggiuntivo attorno ai 500Hz ricorrente nelle vocali [i] e [u]. Con ogni probabilità questo picco, che è stato misurato ogni volta che è stato rilevato, è dovuto all'effetto di nasalizzazione della vocale. Dei picchi correlati alla nasalizzazione sono presenti a frequenze analoghe nella funzione di trasferimento per le [i] e le [u] calcolata dal modello teorico di Maeda (1993). In questo modello un parametro fondamentale per il calcolo delle funzioni di trasferimento è il grado di accoppiamento nasale, denominato NC. Questo parametro rappresenta fisicamente l'area del passaggio velofaringeo ed è misurato in cm^2. Nelle figure 2.9 e 2.10 sono riportate le funzioni di trasferimento per le vocali [i] e [u] in versione nasalizzata e non. Le NF rappresentano le formanti di nasalizzazione mentre le Z sono gli zeri della funzione di trasferimento. Risultati analoghi, per quanto riguarda le risonanze delle funzioni di trasferimento delle vocali nasalizzate si ottengono dal modello di Stevens.

I risultati dell'analisi della nasalizzazione verranno illustrati nel capitolo quattro.

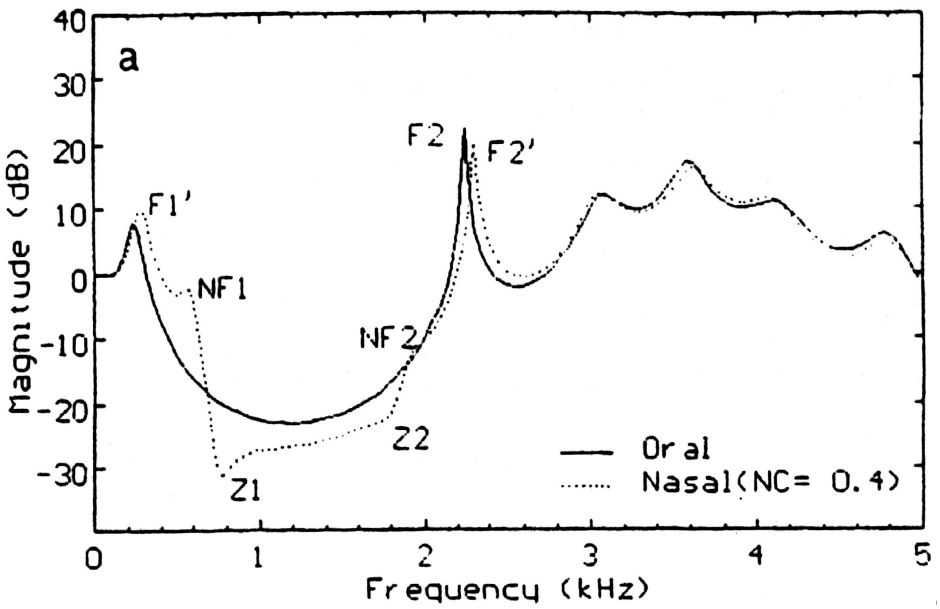


Fig. 2.9 Funzioni di trasferimento per [i] "orale" (linea continua) e nasalizzata (linea tratteggiata). NC rappresenta l'accoppiamento nasale ed è misurato in cm^2.

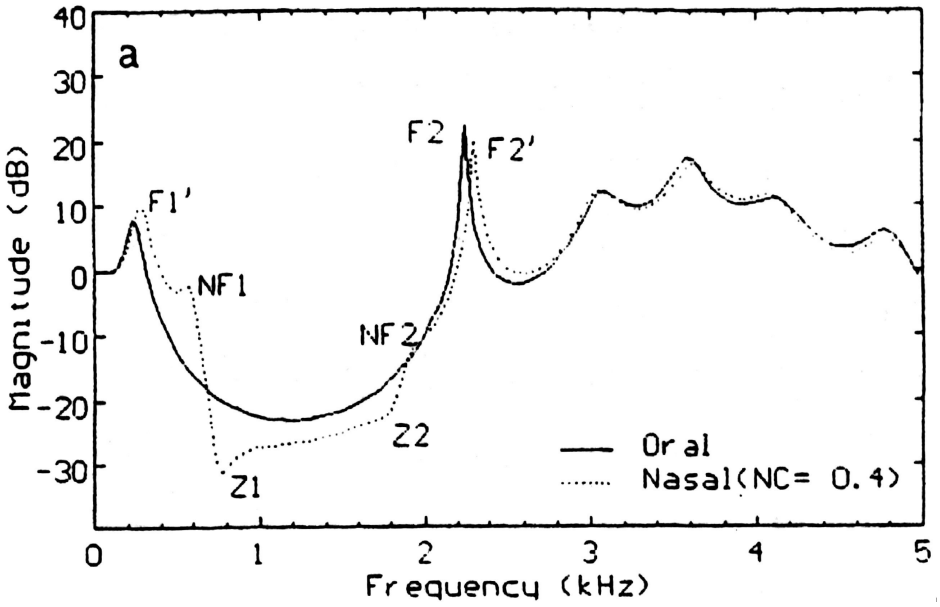


Fig. 2.10 Funzioni di trasferimento per [u] "orale" (linea continua) e nasalizzata (linea tratteggiata). NC rappresenta l'accoppiamento nasale ed è misurato in cm^2

CAPITOLO 3

LA BASE DATI, IL SOFTWARE E GLI STRUMENTI STATISTICI

INTRODUZIONE

In questo capitolo sarà fornita innanzi tutto un'accurata descrizione della base dati utilizzata. Inoltre, dopo aver parlato nel primo capitolo delle proprietà acustiche ed elettriche del segnale vocale, descrivendo i metodi d'analisi da un punto di vista strettamente matematico, si riprenderà il discorso sull'analisi della voce, spiegando quali sono le potenzialità di UNICE (il software utilizzato a questo scopo) e al tempo stesso come esse siano state sfruttate nella pratica per il nostro lavoro. Verranno, poi, brevemente descritti anche gli altri software usati ed infine sarà dedicato un paragrafo alla descrizione dei test statistici utilizzati in sede di analisi.

3.1 LA BASE DATI

Per caratterizzare i fonemi /m/ e /n/ nelle loro versioni sia singole sia geminate, si è resa necessaria la raccolta di un certo numero di pronunce, in modo da poter realizzare un'opportuna base di dati in italiano.

3.1.1 Criteri di scelta dei parlatori

Per caratterizzare la pronuncia corretta dei fonemi consonantici nasali si è ritenuto sufficiente un campione di sei parlatori; infatti, la presente ricerca non è orientata in maniera diretta al riconoscimento vocale e quindi non si è creduto di dover focalizzare troppo l'attenzione sulle sfumature che differenziano diversi parlatori. La scelta dei parlatori si è basata su diversi criteri. Un primo criterio è quello dell'equilibrio tra i sessi: sono stati dunque scelti tre uomini e tre donne. Un secondo aspetto che si è preso in considerazione per operare la scelta dei parlatori è stato quello riguardante la loro età. Si sono scelti parlatori di varie età, in modo da coprire l'arco tra i 20 e i 50 anni, senza, quindi, scendere sotto i

vent'anni, età alla quale la voce dell'essere umano è completamente formata. Ovviamente si sono cercati parlatori che non presentassero particolari difetti di pronuncia o inflessioni di qualsiasi tipo e quindi la scelta è caduta su persone con un livello di istruzione medio-alto. Nella tabella 3.1 riportiamo i dati relativi ai 6 parlatori (in ordine alfabetico) che si sono prestati per l'esperimento.

Parlatore (Sesso)	**Luogo/data di nascita**	**Formaz. Fonetica della famiglia**	**Frequentazione delle scuole**	**Professione**
AI (m)	Salerno 3/9/1967	Salerno	Roma	Studente universitario
AV (f)	Roma 13/6/1968	Potenza/Roma	Roma	Studente universitario
EZ (f)	Roma 21/4/1967	Lombardia	Roma	Studente universitario
GD (f)	Napoli 16/3/1958	Napoli	Parigi	Professore universitario
FM (m)	Roma 18/4/1967	Roma	Roma	Studente universitario
PM (m)	Roma 13/2/40	Napoli	Napoli	Professore universitario

Tab. 3.1 Dati relativi ai 6 parlatori che hanno contribuito alla formazione della base di dati. Nelle cinque colonne ci sono, rispettivamente, il nome e il cognome, il luogo e la data di nascita, il luogo di formazione fonetica della famiglia, il luogo in cui si sono frequentate le scuole primarie e secondarie (dove quindi si è appresa la lingua), la professione.

3.1.2 Altri particolari sulla base di dati

A ciascuno dei parlatori appena elencati è stato chiesto di emettere un certo numero di pronunce di parole contenenti i fonemi /m/ e /n/. Queste parole, in forma di segnale elettrico trasdotto da microfono, sono state memorizzate su nastri magnetici. Le parole scelte sono delle sequenze fonetiche prive di contenuto semantico, per ottenere una pronuncia il più possibile *neutra*. In particolare la scelta delle sequenze è stata influenzata soprattutto dagli aspetti in cui si articola lo studio: la coarticolazione e la geminazione. Per quanto riguarda la geminazione, si è scelto di far pronunciare ai soggetti tutte le parole contenenti le consonanti nasali sia in versione singola che geminata. Per quanto riguarda invece la coarticolazione, si è limitato lo studio al solo caso vocalico, scegliendo le tre vocali che costituiscono gli estremi del trapezio fonetico: /a/, /i/ e /u/.[1]

[1] Ad essere precisi, va osservato (vedi figura 1.8) che il trapezio fonetico ha, ovviamente, quattro vertici, e che mentre la /i/ e la /u/ italiane si trovano proprio in corrispondenza dei due superiori, altrettanto non può dirsi per la

Per ogni parola della base dati, per ciascun parlatore, sono state registrate tre versioni, al fine di ottenere in sede di elaborazione dei dati, dei valori medi non alterati da eventuali fenomeni aleatori. Sono state eliminate e quindi fatte ripetere alcune pronunce palesemente scorrette.

Considerando che la base dati doveva contenere pronunce relative alle due consonanti m e n, sono state registrate: *(2 fonemi) x (2 geminazioni) x (6 parlatori) x (3 ripetizioni) x (3 vocali) = (216 parole)*. Le parole sono state costruite secondo la struttura VCV per le versioni singole e VCCV per le corrispondenti geminate, cioè vocale-consonante-vocale, tipica dell'italiano (che solo raramente prevede parole terminanti con consonante). L'accento delle parole è stato posto sulla prima vocale, visto che la stragrande maggioranza delle parole italiane è piana. L'elenco completo delle parole componenti la base dati italiana è mostrato nella tabella 3.2.

		Consonante			
		m		**n**	
Vocale	**a**	*ama*	*amma*	*ana*	*anna*
	i	*imi*	*immi*	*ini*	*inni*
	u	*umu*	*ummu*	*unu*	*unnu*

Tab. 3.2 Elenco completo delle pronunce relative alla base dati sulle nasali italiane

La base dati è stata costruita pensando a sviluppi molteplici non tutti esauriti dal presente lavoro. La base dati, infatti, comprende anche pronunce relative ai fonemi consonantici occlusivi, liquidi, fricativi e affricati. In particolare, sulle prime due classi di fonemi è stato già eseguito un lavoro di analisi e di percezione. Per quanto riguarda i fonemi fricativi oltre ad un lavoro di analisi è stato eseguito un lavoro di sintesi. La disponibilità di dati e risultati di lavori precedenti permetterà nel capitolo cinque, di operare un confronto tra diverse serie fonematiche.

3.1.3 La registrazione della base dati

Le registrazioni del materiale acustico sono tutte avvenute nel Laboratorio Voce del Dipartimento INFOCOM della Facoltà di Ingegneria dell'Università di Roma La Sapienza, basandosi su suggerimenti di esperti e su preziosi consigli pratici contenuti nel testo "Microphones" (Clifford, 1986). I supporti tecnologici usati sono costituiti da una camera silente, un microfono omnidirezionale, un impianto stereo e varie cassette magnetiche. Vediamo più in particolare le specifiche tecniche del materiale utilizzato.

- Camera silente: *Mini Cabina Amplisilence* della Amplifon, con pareti interne fonoassorbenti per eliminare il riverbero della voce e una capacità di abbattimento dei rumori esterni di circa 30 dB alle frequenze di interesse.
- Microfono: SONY ECM 144, omnidirezionale (per catturare il suono proveniente da qualsiasi direzione), con risposta in frequenza piatta fino a 15kHz, mono, della tipologia a condensatore, con

/a/. Essa, infatti, si trova al centro dei due vertici inferiori del trapezio fonetico, i quali rappresentano due vocali leggermente diverse da quella tipica italiana, una è palatale e l'altra è velare.

-55.3 dBm/mbar di sensibilità (potenza del segnale generato, in mW, in presenza di un suono di 1 mbar di pressione). La scelta di questo particolare strumento è stata guidata dalla consultazione del testo “Microphones” (Clifford, 1986), testo assolutamente esauriente in materia.

- Impianto stereo: KENWOOD KT-48L con possibilità di regolazione del volume di registrazione (caratteristica che assicura l’assenza del dispositivo di regolazione automatica del volume d’ingresso, di cui sono dotati molti moderni apparecchi stereo, e che opera un filtraggio imprevedibile del segnale d’ingresso, al fine di evitare la saturazione della dinamica del dispositivo e del nastro).
- Cassette magnetiche: TDK SA 50 o 60, di buona qualità e con una risposta in frequenza praticamente piatta fino ad oltre 10kHz.

Il pannello posto sulla parte frontale della camera silente permette di collegare il microfono allo stereo tenendo la porta sigillata, consentendo l’assorbimento degli echi da parte delle pareti della camera e isolando parzialmente il microfono dai rumori esterni. Il vetro della parete frontale permette, poi, di instaurare una comunicazione visiva con il parlatore seduto sullo sgabello all’interno della camera silente. Grazie a tale caratteristica si sono realizzate le registrazioni mostrando ai soggetti dell’esperimento le parole da pronunciare mediante cartelli. Si noti che, essendo mono il segnale prodotto dal microfono durante la registrazione, è stato sfruttato un solo canale dello stereo e una sola pista delle cassette magnetiche, senza che ciò abbia avuto particolari conseguenze (nelle fasi di riascolto in cuffia per comodità è stato fatto in modo di distribuire il segnale su entrambi i canali d’uscita in modo da avere un effetto acustico più verosimile). Le registrazioni effettuate, come detto, su nastro magnetico sono state digitalizzate in un secondo tempo usando UNICE, un software usato anche per il resto dell’analisi. Si desidera fare una puntualizzazione riguardo quest’ultimo punto e spiegare per quale motivo si sono effettuate le registrazioni su nastro magnetico. Evidentemente se le registrazioni fossero state effettuate direttamente sul computer sarebbero state sicuramente più “pulite” e silenziose; bisogna tuttavia tenere presente anche che le sedute di registrazione sono risultate molto stancanti per i parlatori e quindi, per evitare che le pronunce fossero affette dal fattore “stanchezza” si è cercato di velocizzare il più possibile le operazioni. Scartando per questo motivo la possibilità di registrare, controllare e catalogare contestualmente ciascuna pronuncia, una buona procedura sarebbe stata sicuramente quella di registrare di continuo l’intera seduta di registrazione sull’Hard Disk del computer per poi andare a scegliere e catalogare in un secondo tempo le pronunce corrette. Quando è stato registrato il database (1992) non si disponeva di HD così capienti e nemmeno di registratori digitali a costo contenuto e quindi, considerato anche che il fruscio introdotto dal nastro non risultava così fastidioso per gli scopi preposti, si è scelto di registrare l’intera seduta e di digitalizzare le parole in un secondo tempo.

Si rimanda alle tesi citate in bibliografia per altri particolari riguardanti le modalità di esecuzione delle registrazioni (R.Rossetti, 1993; A.Vannucci, 1993).

3.2 UNICE: IL SOFTWARE PER L’ANALISI DEL SEGNALE VOCALE

UNICE è un software per l’ambiente MS-DOS progettato e realizzato dalla società francese Vecsys. Il programma sfrutta le routine del sistema di gestione della scheda per PC-IBM *AU21* prodotta dalla

OROS. Questa è dotata di un chip di campionamento e tenuta a 16 bit, capace di lavorare fino ad una frequenza massima di 128kHz, di un filtro analogico con banda pari a 20kHz e del chip per il *DSP TMS320C25* della Texas Instruments. Per quanto riguarda l'interfacciamento con l'ambiente esterno, la scheda dispone di: un ingresso microfonico (MIC), un ingresso e un'uscita per il collegamento diretto con un sistema di riproduzione, registrazione e amplificazione del segnale audio (LINE IN / LINE OUT) e, infine, un'uscita per la cuffia (PHONES) (si rimanda per ulteriori specifiche tecniche al manuale di riferimento OROS citato in bibliografia). Grazie a questo dispositivo hardware, è possibile ottenere una velocità di elaborazione che consente di visualizzare gli spettrogrammi in tempo reale.

Le principali funzioni di UNICE sono:

- Registrazione (da microfono o da ingresso esterno) e digitalizzazione di un segnale analogico.
- Visualizzazione dell'andamento del segnale nel tempo
- Ascolto in cuffia o su supporto esterno
- Visualizzazione e calcolo degli spettri e degli spettrogrammi in tempo reale con differenti tecniche (*FFT* a banda stretta e larga, *LPC*).
- Visualizzazione e calcolo della frequenza di *pitch.*
- Visualizzazione dell'energia a breve termine del segnale.

Il programma UNICE è descritto sommariamente nel relativo manuale di utilizzo (Vecsys, 1989). Pertanto, si cercherà qui di seguito di metterne in luce le caratteristiche più rilevanti e le potenzialità maggiormente utili per il presente lavoro.

3.2.1 L'analisi temporale con UNICE

UNICE memorizza il segnale digitale in due file separati, con lo stesso nome ma con estensioni diverse. Il primo, con estensione ***.sig*** (da signal), contiene i dati veri e propri, ossia i campioni, mentre il secondo, con estensione **.key**, contiene le informazioni necessarie all'interpretazione dei dati. Il formato adottato per i file .sig, consiste in una semplice sequenza di campioni (tanti quant'è la frequenza di campionamento adottata moltiplicata per la durata del segnale in secondi[2]), ognuno dei quali è rappresentato con *16 bit in complemento a 2, senza alcuna intestazione.* Il file .key che viene automaticamente creato, contiene la frequenza di campionamento, il numero di campioni ed eventuali segmentazioni ed etichettature. In pratica, l'insieme dei due file equivale al più conosciuto e utilizzato formato *wave [.wav]*, che ha, però, intestazione e campioni del segnale in un unico file. La struttura di un file *.key* è mostrata in figura 3.1. Come detto in esso sono memorizzate anche le segmentazioni e le relative etichettature: UNICE permette, infatti, semplicemente con il mouse, di segmentare il segnale esattamente in corrispondenza di un ben preciso campione, evidenziandolo nella forma d'onda temporale con una barra rossa verticale. Questa possibilità si è rilevata utilissima per l'analisi temporale e si è rivelato molto comodo anche il fatto che la segmentazione sia memorizzata nel file *.key*.

[2] L'unico vincolo per il numero di campioni del segnale è che deve essere un multiplo intero di 128, dato che, come si spiegherà fra poco, Unice divide il segnale in frame di 128 campioni l'uno.

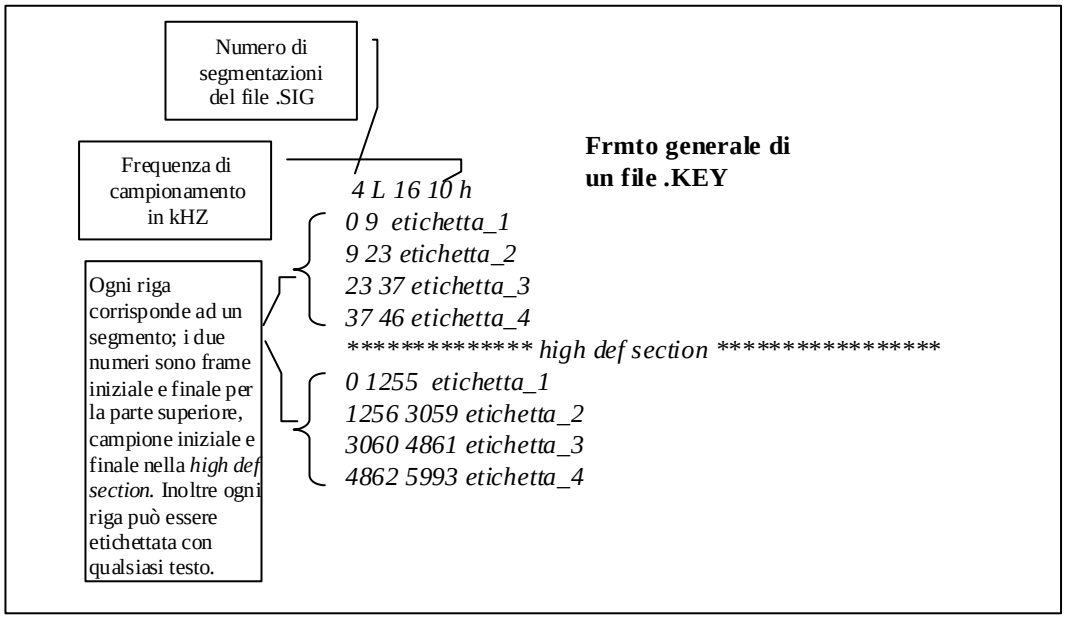


Fig. 3.1 Formato generale di un file .key usato da UNICE allo scopo di memorizzare le informazioni relative alle caratteristiche e alle segmentazioni di un file di voce.

3.2.2 Il metodo della *"short-time analysis"*

Si è già avuto modo di rilevare nel primo capitolo che, nello studio del segnale vocale interessano soprattutto le caratteristiche locali, in modo da poter associare le variazioni del tratto vocale alle variazioni del segnale nel tempo e in frequenza. Sarebbe di scarsa utilità conoscere l'energia oppure la FFT di un segnale nella sua totalità. Per questo, quella che si usa in genere è la tecnica chiamata *short-time analysis* (Rabiner e Schafer, 1978) con la quale si prende in considerazione di volta in volta, solo una sequenza di campioni relativi ad una parte del segnale. Matematicamente la sequenza può essere rappresentata come

$$Q_n = \sum_{m=-\infty}^{\infty} T[x(m)] \cdot w_N(n-m) \tag{3.1}$$

dove T[ ] rappresenta una generica trasformazione (lineare o non lineare) operata sul segnale vocale che può dipendere da alcuni parametri e $w_N(n)$ è una finestra rettangolare di ampiezza N (cioè, con soli N campioni pari a 1 e tutti gli altri identicamente nulli) traslata in corrispondenza del campione di indice n. Esempi di analisi di questo tipo sono la FFT narrow-band o wide-band, l'analisi LPC, l'analisi condotta con la short-time energy. In quest'ultimo caso, ad esempio, si ha

$$E_n = \sum_{m=n-N+1}^{n} x^2(m) \tag{3.2}$$

dove l'operazione T[ ] è semplicemente il quadrato. E_n rappresenta l'energia del segnale, considerato per soli N campioni consecutivi alla volta. La "finestra" di analisi viene ogni volta traslata in avanti di un campione ed E_n di nuovo calcolata. Il suo significato è alquanto diverso della semplice energia *totale* del

segnale, ottenuta quadrando e sommando *tutti* i campioni. Un esempio dell'andamento dell'energia a breve termine per un segnale vocale è mostrato in figura 3.2.

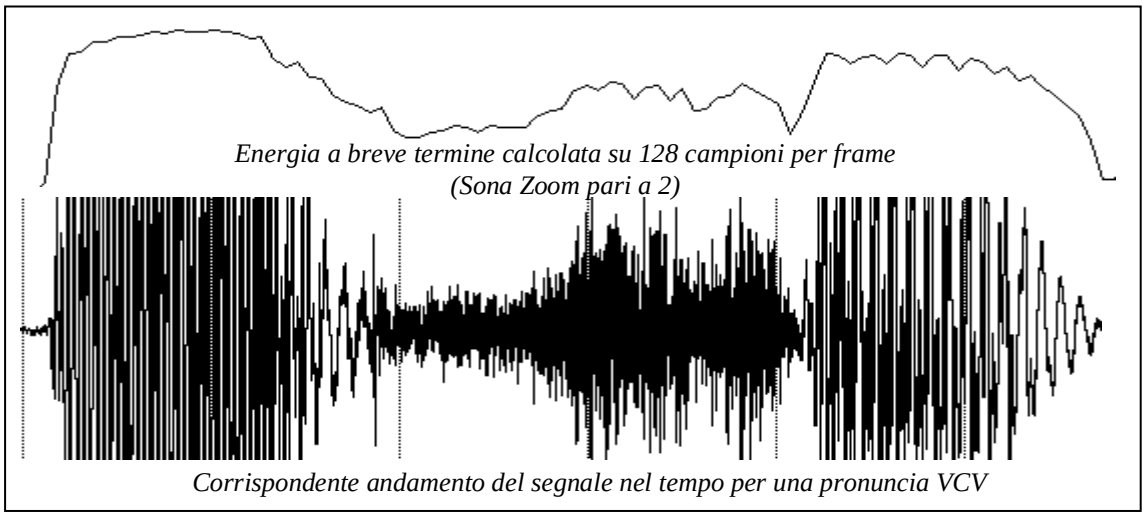


Fig. 3.2 Esempio di visualizzazione di energia a breve termine con il programma UNICE; sono visualizzati sullo stesso asse temporale circa 550 ms di segnale.

Per una corretta analisi è importante chiedersi:

a) quanto dovrebbe essere ***l'ampiezza della finestra*** o ***frame*** N e su cosa influisce questo numero;

b) se sia il caso di calcolare Q_n per tutti gli n, o se si può saltare il calcolo di alcuni elementi della sequenza e ripeterlo solo per dei multipli interi di n: in pratica, quanto dovrebbe essere ***il fattore di sovrapposizione*** tra frame adiacenti.

La risposta a queste domande dipende evidentemente dal tipo di analisi che si vuole effettuare. Per chiarire questo concetto consideriamo l'esempio del calcolo della FFT. Le FFT narrow-band e wide-band, costituiscono un esempio di trasformazione T[·] e si differenziano per il fatto che la prima ha una finestra di analisi di ampiezza maggiore della seconda. In figura 3.3 sono mostrati tre esempi di FFT (a 128, a 256 e a 512 campioni), per uno stesso segnale di voce campionata a 16000Hz (rappresentante una vocale). Alle tre FFT corrispondono, rispettivamente, una finestra temporale di analisi di 8, 16 e 32 ms. Potremmo affermare che: la prima è una wide-band, la terza è una narrow-band mentre la seconda è una via di mezzo tra le altre due. La figura 3.3 è molto esplicativa: risulta evidente che, più la finestra è ampia più la risoluzione in frequenza aumenta, anche se, ovviamente, ne risente la velocità di calcolo (il numero di punti della FFT è più grande). Lo svantaggio maggiore, in ogni modo, è quello di una diminuzione di risoluzione temporale. Il fattore di sovrapposizione può essere usato per raggiungere dei buoni compromessi tra le due esigenze.

Nel caso delle analisi riguardanti il segnale vocale, bisogna considerare che, mediamente, per la voce di un uomo il periodo di pitch è di 8 ms mentre per una donna è di 4.4 ms (vedi tabella 1.4), e che la lunghezza di un fonema è in media di 150 ms. Perciò, occorre prestare molta attenzione nello scegliere l'ampiezza N della finestra di analisi in funzione del caso oggetto di studio: per esempio, se si è

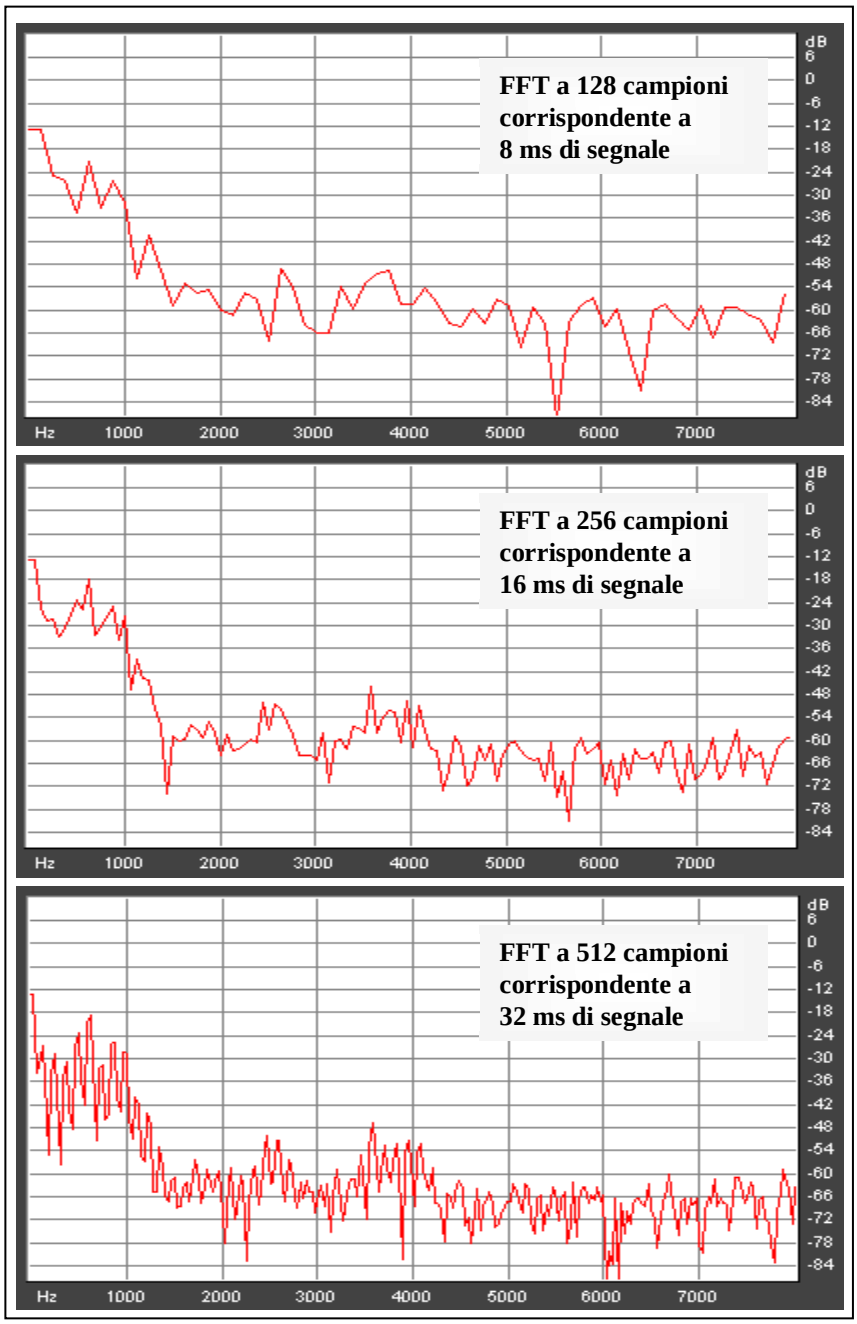


Fig. 3.3 Spettri FFT a 128, 256 e 512 campioni per un segmento di vocale campionato a 16000 Hz.

interessati alle caratteristiche prosodiche delle parole servirà sicuramente una finestra più ampia che per analizzare i singoli fonemi, per i quali servirà, a loro volta, una finestra più ampia che per l'analisi delle zone di transizione o di brusche variazioni nel segnale e così via. Per finire, si affermerà che un giusto compromesso tra tutte le esigenze è, come al solito, la soluzione ottimale; per arrivare a questo, tuttavia, si devono conoscere a fondo tutti i vantaggi e gli svantaggi delle scelte che ci si presentano.

UNICE gestisce la short time analysis suddividendo *il segnale nel tempo in frame* la cui lunghezza N varia in funzione della frequenza di campionamento f_c (in kHz) secondo la semplice relazione:

$$N = \frac{128 \cdot f_c}{10} \tag{3.3}$$

La durata di ciascun frame, uguale a N/f_c, è, invece, fissa e pari a 12.8 ms (comprendendo quindi 128 campioni per una frequenza di campionamento di 10kHz).

3.2.3 L'analisi in frequenza con UNICE

Per l'analisi in frequenza Unice mette a disposizione sia uno spettrogramma a tutto schermo del tipo di quello mostrato in figura 1.17 che un'altra finestra di analisi più piccola, a sua volta suddivisa in due semifinestre, dove sono visualizzati gli spettri e/o il segnale nel tempo, frame per frame, come mostrato in figura 3.4. Ricordiamo che lo spettrogramma è un diagramma tridimensionale: tempo (o meglio frame essendo la dimensione temporale quantizzata in pacchetti) sulle ascisse, frequenza sulle ordinate e ampiezza, visualizzata tramite una tonalità di grigio, più scura se l'ampiezza è più alta (Oppenheim, Schafer, 1975).

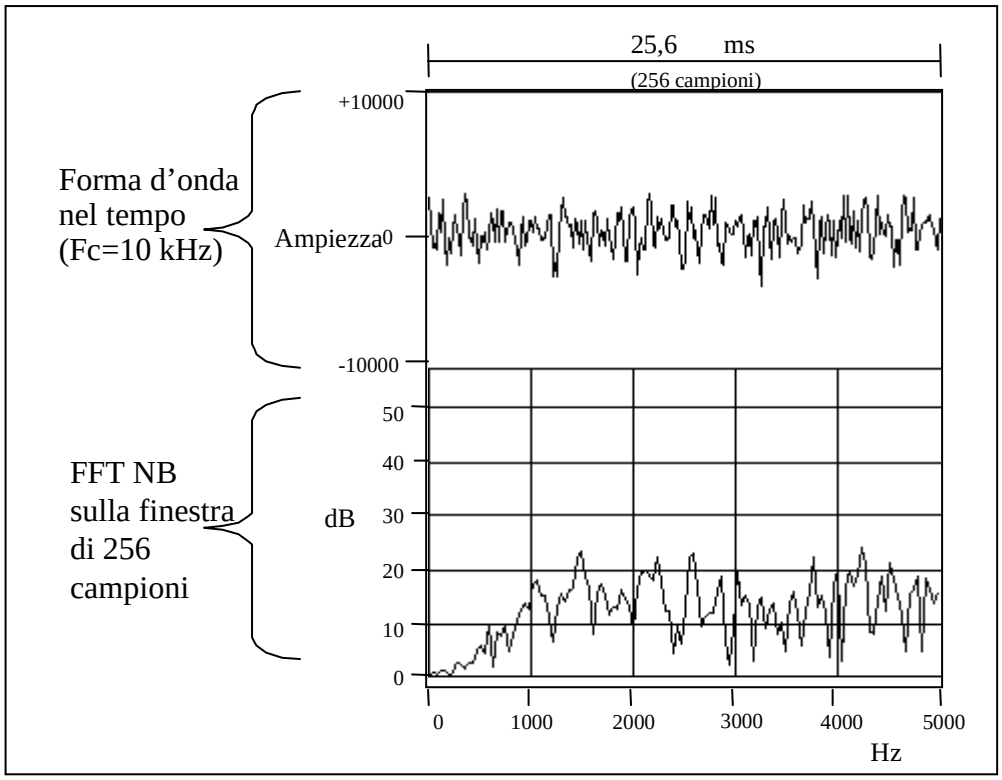


Fig. 3.4 Esempio di segnale + spettro visualizzato da UNICE relativamente ad un solo frame di analisi.

Le modalità di calcolo della FFT usate dal programma per questo tipo di analisi sono le seguenti:

1) FFT a banda stretta, realizzata a partire da 256 campioni nel tempo, precedentemente finestrati con finestra di Hamming, che restituisce 128 campioni in frequenza (e non 256, per effetto della simmetria che presenta lo spettro di un segnale campionato). Nel caso di un segnale campionato a 10Khz si ha una risoluzione in frequenza di 39.0625Hz (vedi formula 1.7).

2) FFT a banda larga, realizzata a partire da 60 campioni nel tempo precedentemente finestrati con finestra di Hamming, che restituisce ancora 128 campioni in frequenza. In effetti, teoricamente dovrebbero essere 30, ma, usando l'artificio di considerare 98 campioni nulli seguiti dai 60 campioni di cui si vuole la FFT a banda larga (FFT WB) e poi ancora da 98 campioni nulli, e calcolando su questi 256 campioni totali una FFT a banda stretta (FFT NB), si ottengono di fatto 128 campioni in frequenza, come mostrato in figura 3.5. Questa tecnica, per cui si aggiungono dei campioni nulli, è chiamata "zero padding" e non consente di aumentare la risoluzione in frequenza ma solo di migliorare la visualizzazione della FFT. La risoluzione effettiva in frequenza sarà di 188Hz (formula 1.7 con N pari a 60) per una frequenza di campionamento di 10kHz. Per maggiori dettagli sullo zero padding rimandiamo a Oppenheim e Schafer (1975).

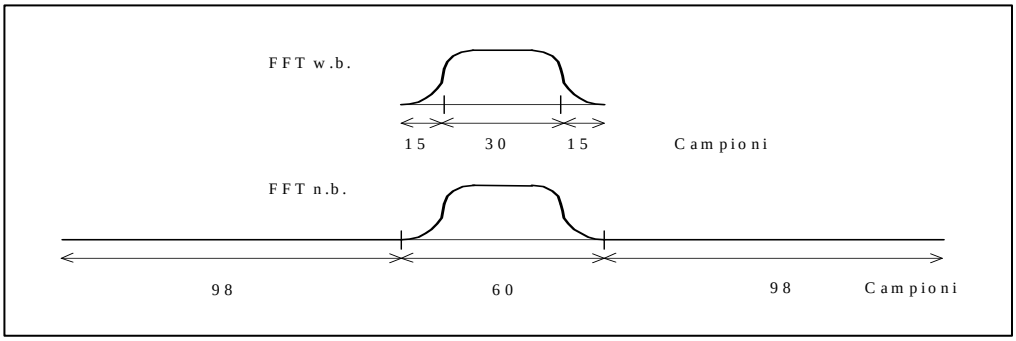


Fig. 3.5 FFT wide band e narrow band con *zero-padding*.

Quindi la FFT NB consente un'analisi più accurata in frequenza (essendo maggiore la risoluzione che offre) rispetto alla FFT WB, tuttavia, quando le caratteristiche del segnale subiscono variazioni repentine (in conseguenza di altrettanto rapide variazioni del tratto vocale), la FFT WB permette di isolare più selettivamente la zona di analisi, in virtù dei soli 60 campioni nel tempo di cui fa uso per il calcolo.

Nell'analisi in frequenza oltre alla FFT è disponibile anche l'LPC, con cui vengono calcolati i coefficienti di predizione su 256 campioni precedentemente finestrati con finestra di Hamming.

In tutti i casi è prevista un'*enfatizzazione* dello spettro con coefficiente pari a 0.95 tesa ad esaltare l'importanza del contenuto energetico in alta frequenza (vedere par. 1.4.4 per i dettagli).

Né la durata, né la posizione dell'inizio o della fine di un frame sono modificabili. Per ovviare a questa limitazione, che non consentirebbe di analizzare sequenze di campioni appartenenti a frame diversi, ma adiacenti, si può far uso delle opzioni offerte dal *Sona Zoom*. Si può impostare, infatti, il parametro Sona Zoom in una scala di valori tra 1 e 8: la dimensione della finestra tramite la quale viene condotta l'analisi visiva è 12.8/SZ, fino perciò ad un minimo di 1.6 ms.

La scala temporale di visualizzazione dipende dal fattore di Sona Zoom. Se esso vale 1, per ogni frame viene visualizzata una sola FFT: l'analisi in frequenza viene quindi ripetuta ogni 12.8 ms. Quando Sona Zoom è impostato ad *i*, per ogni frame vengono visualizzate *i* FFT. Chiaramente se si è impostato il parametro Sona Zoom pari a 1 allora lo spettro narrow mostrato in una semifinestra è calcolato esattamente sulla porzione di segnale visualizzata nell'altra semifinestra. Se ci si sposta di un frame a destra o a sinistra in Sona Zoom pari a SZ (per SZ=1, 2,...8), lo spettro narrow mostrato nella finestra di

visualizzazione dettagliata è calcolato in una finestra di 25.6ms che si sovrappone alla precedente per un fattore pari a

$$S = \frac{(2 \cdot SZ - 1)}{(2 \cdot SZ)} \tag{3.4}$$

In tal modo si può impostare il passo di spostamento della finestra di calcolo degli spettrogrammi tra otto valori diversi.

A conclusione delle caratteristiche di UNICE riguardo all'analisi in frequenza pensiamo sia interessante far vedere analogie e differenze tra i tre tipi di analisi FFT NB, FFT WB e LPC sia sulla base degli spettrogrammi (per l'analisi complessiva di una pronuncia VCV) sia sulla base degli spettri (per l'analisi di un singolo frame posto al centro di una vocale).

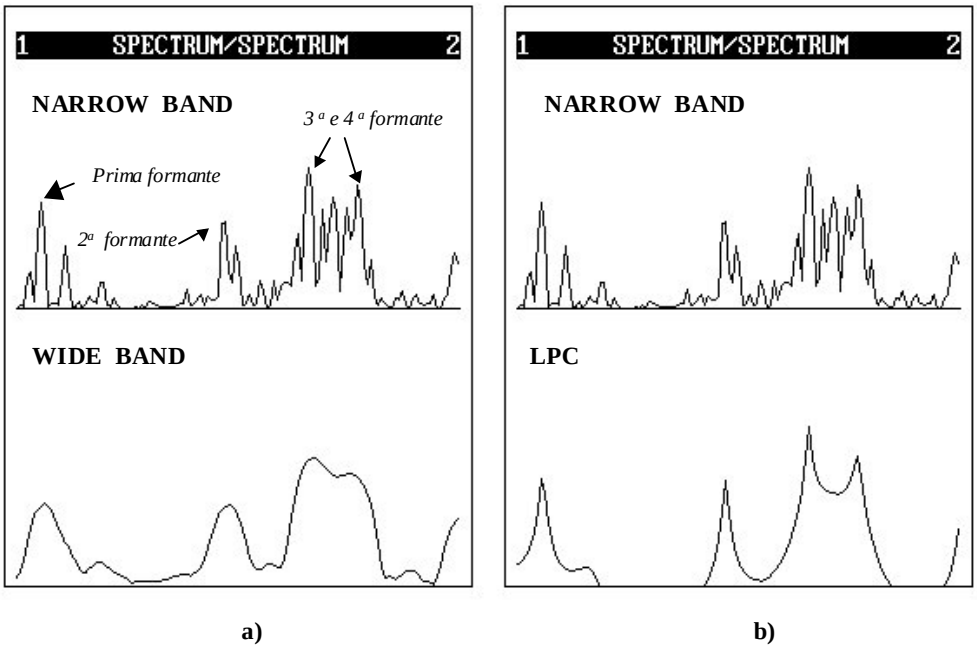


Fig. 3.6 Spettri FFT NB, FFT WB e LPC relativi al frame centrale della vocale /i/ della pronuncia /imi/ (nella figura 3.7 sono mostrati i rispettivi spettrogrammi costruiti affiancando gli spettri di tutti i frame).

A questo scopo, si osservino attentamente i grafici di figura 3.6 dove sono mostrati gli spettri calcolati al centro di una vocale. Si supponga di voler ricercare i valori delle formanti. Ci si accorge facilmente che (figura 3.6a), nello spettro WB la risoluzione in frequenza minore comporta anche una minore accuratezza nella ricerca del valore esatto delle formanti, tanto è che due picchi "vicini" non vengono distinti (correndo il rischio di perdere qualche formante rispetto al NB); di contro, con il WB si vede meglio l'inviluppo o la forma dello spettro e quindi questo spettro risulta più indicato per osservare più in generale picchi e larghezze di banda. La stessa cosa può dirsi per la figura 3.6b dove si può notare come lo spettro LPC "assomigli" al WB. La differenza sostanziale è che il primo individua meglio i picchi delle formanti, ma peggio, senza dubbio, la forma dello spettro e quindi le larghezze di banda.

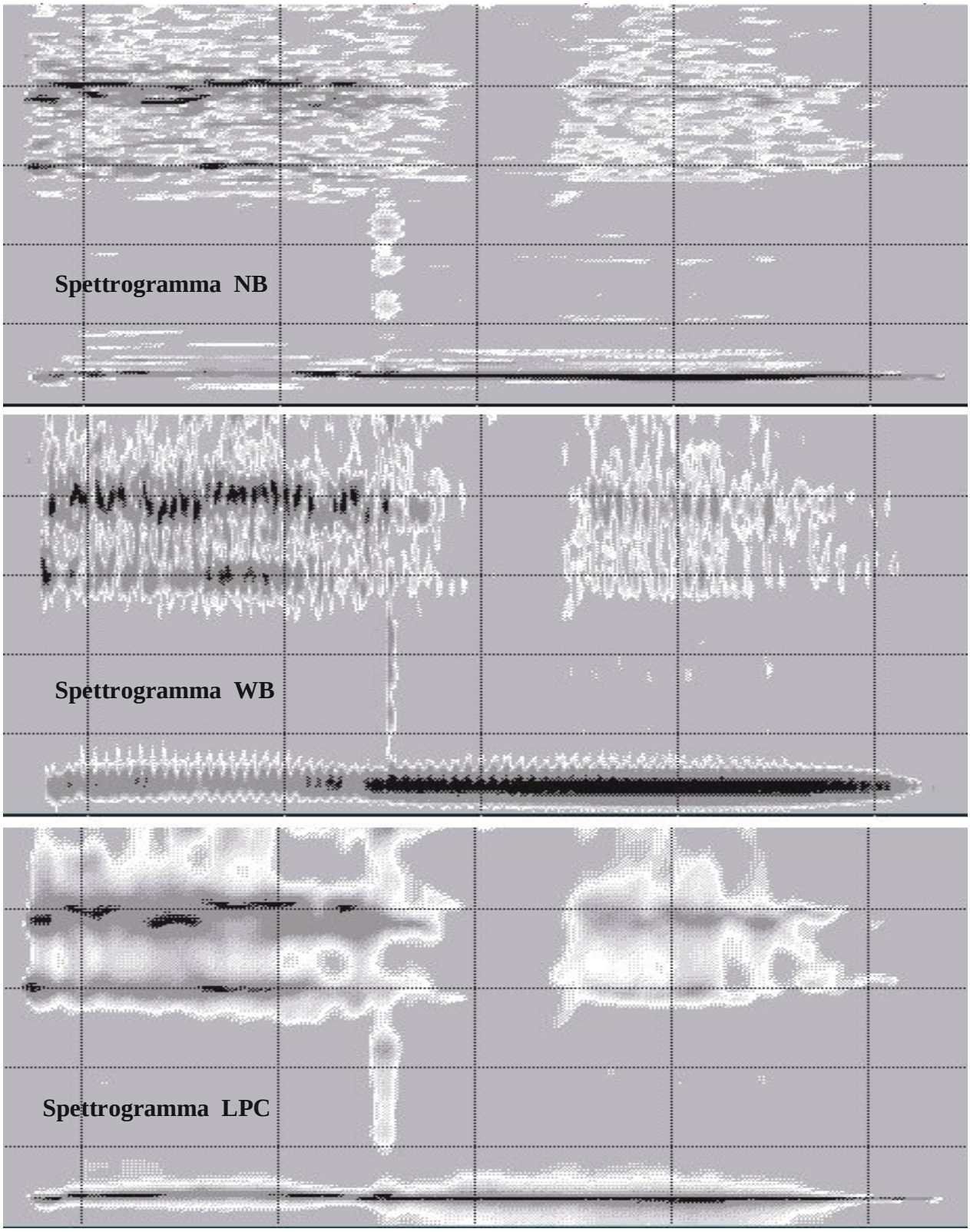


Fig. 3.7 Spettrogrammi FFT NB e WB e LPC calcolati da UNICE sulla pronuncia della parola /ini/. Sona Zoom è stato impostato a 2: l'intervallo temporale visualizzato corrisponde a circa 550 ms (in figura 3.6 sono mostrati gli spettri di un solo frame posto al centro della prima vocale).

Sulla base della figura 3.7, si può fare un ulteriore confronto tra i metodi a disposizione per l'analisi. Questa volta il calcolo è globale: tutti gli spettri del tipo di figura 3.6 (uno o più, dipendendo dal fattore di sovrapposizione, per ogni frame) sono affiancati l'uno all'altro per la durata dell'intera pronuncia. Valgono, quindi, le stesse considerazioni appena fatte riguardo alla differenza tra i vari tipi di spettro. Si può aggiungere che l'effetto di una risoluzione frequenziale più o meno alta e la conseguente risoluzione temporale, rispettivamente, più o meno bassa creano due effetti visivi opposti: uno spettrogramma NB a righe orizzontali contro uno spettrogramma WB a righe verticali. Quello LPC invece risulta più uniforme rispetto ad entrambi gli assi. Nel prossimo capitolo si focalizzerà maggiormente l'attenzione su come le potenzialità di Unice siano state sfruttate.

3.2.4 La digitalizzazione e l'archiviazione della base dati con UNICE

Le registrazioni sono state tutte digitalizzate e archiviate su floppy disk, utilizzando lo stesso stereo usato in fase di registrazione, collegato in uscita con la scheda di acquisizione del PC. Ciascun floppy è stato marcato con un descrittore del suo contenuto e con un numero progressivo ed è stato archiviato.

UNICE dispone di un menù per l'attivazione della procedura di registrazione, che permette di scegliere la fonte del suono (cavo esterno o direttamente microfono), l'amplificazione d'ingresso (in dB), il nome del file prodotto e la frequenza di campionamento. La scheda, all'atto dell'acquisizione del segnale, si occupa di convertire il segnale analogico presente all'ingresso MIC o LINE IN in un segnale digitale, filtrandolo e campionandolo ad una frequenza che può essere selezionata dall'utente in un intervallo di valori tra 10 e 20kHz.

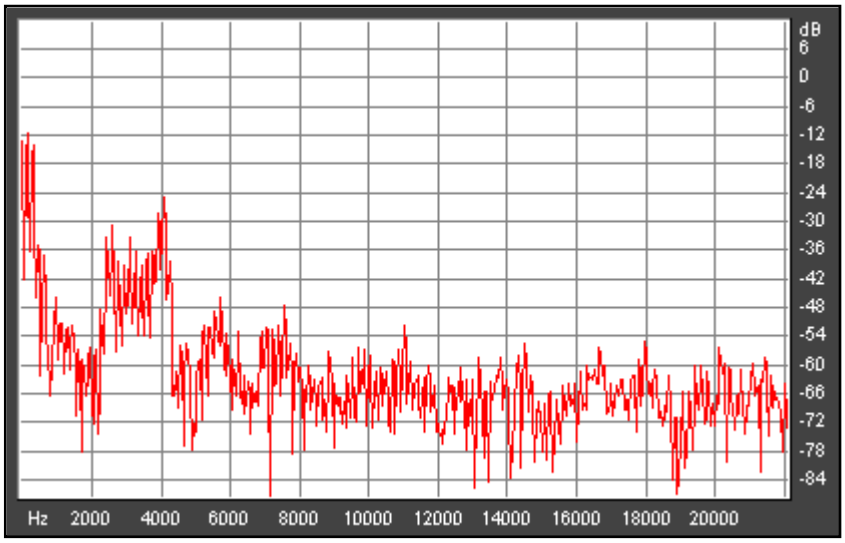


Fig. 3.8 Esempio di spettro FFT a 1024 campioni per una vocale il cui segnale è stato campionato a 44kHz (la finestra temporale è cioè di circa 23 ms). Nella stessa modalità di registrazione il livello di rumore in assenza di segnale è risultato pari a circa -70 dB.

Per la digitalizzazione del database utilizzato nella presente analisi, l'ingresso della scheda è stato collegato all'uscita del canale dello stereo utilizzato in fase di registrazione e la frequenza di

campionamento è stata impostata a *10kHz*, ricordando che l'informazione fonetica captata dall'orecchio umano risiede quasi completamente nei primi 5kHz dello spettro del segnale vocale, soprattutto per quanto riguarda i suoni oggetto dello studio. In figura 3.8 è mostrato lo spettro di un segnale di voce campionato a 44kHz proprio per far notare come le frequenze superiori a 5kHz siano decisamente inferiori di ampiezza a quelle della restante banda.

Per quanto riguarda il nome dei file prodotti, ogni pronuncia è stata archiviata con il nome VCVXNC.SIG (VCCVXNC.SIG nel caso di pronuncia geminata), usando per V e C i grafemi della vocale e della consonante della pronuncia, per X il numero della ripetizione e per N e C la sigla di identificazione del parlatore. Avendo provveduto, già in fase di registrazione, a limitare le differenze di volume tra parlatori, in fase di acquisizione si è usata per tutti la stessa amplificazione d'ingresso, alquanto limitata per non rischiare la saturazione delle capacità elettriche della scheda e la conseguente distorsione del segnale prodotto.

Il software di gestione della scheda consente di ovviare alla mancanza di un filtro hardware con banda regolabile con la tecnica dell'oversampling. Se si campionasse direttamente a 10kHz senza filtrare prima a 5kHz, il segnale che si andrebbe a studiare risentirebbe dell'aliasing causato dalla presenza delle componenti della voce a frequenza compresa tra i 5 e i 20kHz. Fissando invece un fattore di oversampling opportuno, si riesce ad aggirare questo problema via software, tramite la tecnica della decimazione. In pratica, il convertitore A/D campiona il segnale, filtrato a 20kHz dall'unico filtro hardware disponibile sulla scheda (figura 3.9a), ad una frequenza reale calcolata come (fattore di oversampling) x (frequenza di campionamento richiesta). Scegliendo perciò tale fattore pari a 40kHz / (frequenza di campionamento richiesta) si eviterà il fenomeno dell'aliasing (figura 3.9b, frequenza di campionamento = 10kHz, fattore di oversampling = 4).

Se ora si opera un filtraggio numerico a 10kHz, la sequenza di campioni che si ottiene ha uno spettro come quello di figura 3.9c, visto che il filtraggio numerico "taglia" sia lo spettro originale sia le sue repliche in frequenza. La sequenza costruita *decimando* la sequenza originaria, ovvero prendendo un campione ogni N, ha lo stesso spettro di figura 3.9c, ma con le repliche ravvicinate a distanza f_C/N (essendo f_C la frequenza di campionamento reale). Per N=4 lo spettro che si ottiene è quello di figura 3.9d, che è lo spettro della sequenza di campioni del segnale vocale originario campionato alla frequenza desiderata di 10kHz, ma senza aliasing.

3.2.5 Altre funzionalità di UNICE

Si sono ampiamente descritte le caratteristiche di UNICE riguardanti la forma d'onda nel tempo, l'energia a breve termine e gli spettri (e spettrogrammi) FFT e LPC. Ora si descriveranno brevemente le altre caratteristiche degne di nota:

- Con il **programma Spectro.exe** dato a corredo di UNICE è possibile registrare su file il calcolo del modulo degli spettri: ricevuti in ingresso uno o più file con estensione *.sig* e alcuni parametri che specificano il tipo di elaborazione da effettuare, si producono in uscita uno o più file con estensione *.fft* contenenti sequenze di gruppi di 128 byte. Ogni gruppo di 128 byte rappresenta il risultato di una FFT a banda larga o stretta oppure di una predizione lineare. Ogni byte contiene un intero senza segno rappresentante il modulo della DFT espresso in 3/8 di dB.

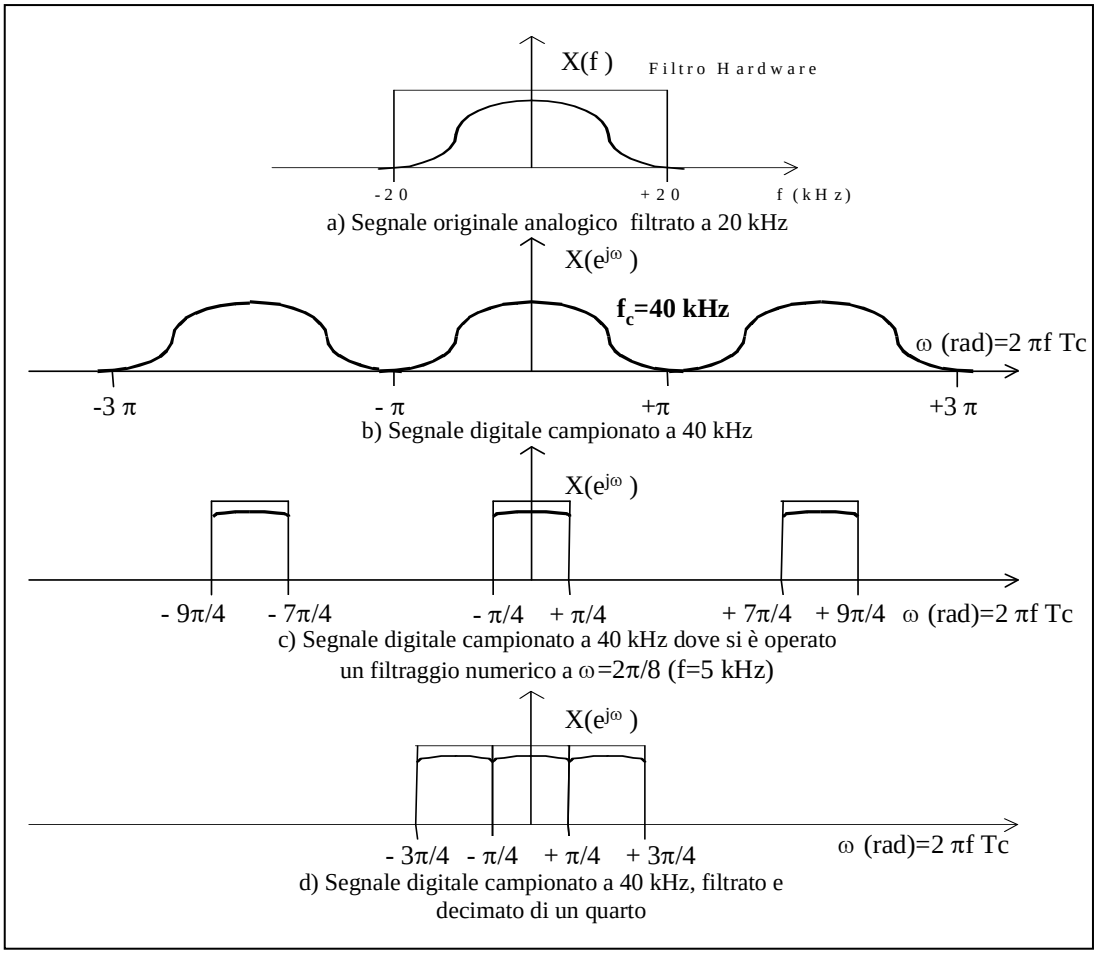


Fig. 3.9 Esemplificazione della tecnica usata per digitalizzare la base di dati con l'uso della decimazione nel tempo.

Se viene specificata l'opzione */MN* o */ML* con */Zn* (*n*=1...8), vengono calcolate 3·*n* FFT narrow band o LPC per ogni frame, mentre con */MW /Zn* vengono calcolate 6·*n* FFT wide band per ogni frame, ciascuna delle quali costituita, come si è già detto, da 128 campioni[3]. Si è quindi reso indispensabile individuare quale delle tre FFT calcolate da Spectro, in Sona Zoom 1, (narrow o LPC) corrispondesse all'unica visualizzata da UNICE a pari condizioni e, analogamente, quale delle sei ottenute da Spectro wide, in Sona Zoom 1, fosse quella visualizzata da UNICE. Si è visto sperimentalmente che, per ogni terna di FFT narrow band o LPC determinata da Spectro, la FFT visualizzata da UNICE è sempre la seconda. Sottolineiamo a beneficio di chi, per ulteriori sviluppi di questo lavoro, si trovasse ad utilizzare questa utility di Unice che, per tutte le analisi effettuate, è sempre stata utilizzata l'opzione */MN* con *n*=1.

- Con il **programma F0.exe** è possibile calcolare e registrare su file con estensione *.f0* la frequenza di pitch per ogni frame, con campioni a 16 bit in complemento a 2. Inoltre UNICE permette di visualizzare (sotto lo spettrogramma) l'andamento della F0 per tutto il segnale. Questa funzione non è stata utilizzata perché il calcolo del pitch tramite questo programma non

[3]UNICE con le stesse opzioni specificate in Spectro calcola *n* FFT per frame in modalità *narrow* o *wide* o LPC.

è sempre così accurato e si è preferito controllarlo di volta in volta come verrà puntualizzato nel capitolo 4.

- Infine, una caratteristica di grande utilità di UNICE, è **l'editor di segmenti di segnale**. Questo menu permette di selezionare un numero qualsiasi di frame (purché siano tutti visualizzati sullo schermo) tramite il mouse, definendo un *segmento* di segnale. Tale segmento può essere riascoltato permettendo di "ascoltare" le caratteristiche acustiche di una parte della pronuncia, al fine di confermare le osservazioni fatte sulla base dell'osservazione visiva del segnale nel tempo o in frequenza. In particolare l'uso di tale funzionalità si è rivelato utile riscontro alle scelte operate nell'individuazione dei punti di transizione consonante/vocale o vocale/consonante. Il segmento può essere poi salvato in un file a se stante e, qualora si scelga il nome di un file già esistente, il salvataggio avverrà in coda a tale file.

3.3 DESCRIZIONE DEGLI ALTRI SOFTWARE UTILIZZATI

Oltre UNICE, per lo svolgimento di questo studio, sono stati utilizzati molti altri software. Riteniamo opportuno dare una sommaria descrizione di questi software e di quali tra le loro funzionalità siano state sfruttate, sia per rendere il più possibile "trasparente" il metodo di lavoro, sia nella speranza che queste annotazioni possano risultare utili e pratici suggerimenti per lavori e sviluppi futuri.

Sono stati utilizzati i seguenti tipi di software:

1. Software di ***statistica***. Dopo una veloce panoramica su diversi software di statistica, per la maggior parte freeware scaricati dalla rete, si è scelto di utilizzare ***Statgraphics plus 2.1.*** Questo software, infatti, risulta decisamente più completo degli altri visionati ed inoltre permette, con qualche piccolo accorgimento, di importare dati in formato Excel. Le analisi effettuate con Statgraphics plus sono: analisi della varianza mono e multivariata, test di Spearman, regressione.
2. Software di ***grafica: Paint Shop Pro 5.0***. Nell'analisi delle pronunce della base dati ci si è trovati a dover gestire le diverse informazioni che UNICE visualizzava di volta in volta sullo schermo, rivelandosi spesso necessaria la possibilità di poter vedere più immagini vicine e confrontarle. Il programma Paint Shop Pro si è rivelato di grande utilità pratica. Infatti, permette (dopo aver lanciato la proprietà capture) di "catturare" in qualsiasi momento immagini di qualsiasi dimensione dallo schermo e salvarle in uno dei formati tra quelli disponibili oggi (.bmp, .gif, .jpg, .pcx ecc.). Dopo aver salvato un'immagine è possibile elaborarla e aggiustarla secondo le esigenze, ad esempio invertire i colori o renderla bianco e nero o ridimensionarla ecc. Tutte le immagini spettrografiche, energetiche o della forma d'onda temporale presenti nella tesi sono state prelevate ed aggiustate in questo modo.
3. I ***compilatori C*** usati per scrivere i programmi di supporto all'analisi sono stati: Turbo C versione 2.0 (1988) e C++ versione 3.0 (1992), entrambi della Borland. Questi compilatori benché ormai superati, si sono rivelati pratici, veloci ed efficaci in virtù del fatto che per le piccole utility scritte con essi, non sono mai servite funzioni particolari non comprese nello standard dell'ANSI C;

inoltre, anche se a volte si sarebbe potuto usare qualche funzione particolare più avanzata, si è sempre e comunque scelto di restare il più possibile nello standard per permettere sia portabilità su qualsiasi compilatore (anche non Borland), sia maggiore facilità di comprensione per chi volesse leggere e utilizzare i programmi.

4. Il pacchetto Microsoft ***Office*** nelle versioni 97 e 2000 ed in particolare il ***Visual Basic***. Questo pacchetto, comune e diffuso oramai in tutto il mondo, rappresenta un vero e proprio ambiente di lavoro integrato sotto il sistema operativo Windows. Si compone di più programmi che permettono varie funzionalità. Per l'utilizzo che se ne è fatto durante lo svolgimento della tesi, tre sono gli applicativi che sono stati fondamentali: Excel, Power Point e Word.

   Il primo, *Excel*, è stato ampiamente utilizzato per la costruzione di tutte le tabelle della tesi, e in particolare per le appendici A, B e C. Oltre alla normale formattazione di una tabella, per altro resa molto veloce dalla possibilità di automatizzare molte procedure, il programma è in grado di definire elaborazioni matematiche tramite formule che collegano tra loro le caselle delle tabelle stesse. Queste potenzialità sono state usate per le elaborazioni statistiche più semplici come medie e deviazioni standard ed altre, direttamente "sul posto" senza cioè utilizzare Statgraphics o il compilatore C.

   Il secondo, *Power Point*, si è rilevato strumento di grande utilità per eseguire disegni e figure. In particolare, quasi tutti i disegni che compaiono in questo lavoro sono stati eseguiti con l'aiuto di questo programma.

   Il terzo, *Word*, è quello che ha permesso la stesura di tutta la tesi, integrando grafici, tabelle e disegni con il testo per formare un'unica unità.

La versione 2000 del pacchetto Office è solo leggermente diversa dalla precedente ma presenta delle piccole novità che si sono rivelate particolarmente comode. Tuttavia si sono utilizzate entrambe le versioni perché al momento della stesura di questa tesi la versione 2000 è disponibile solo in versione inglese e quindi, per tutte le funzionalità strettamente legate alla lingua, come la correzione ortografica, si è dovuta utilizzare la versione 97.

Dedichiamo infine alcune righe di commento al *Visual Basic*[4]. Il Visual Basic è un linguaggio di programmazione nuovo, orientato agli oggetti, che non offre a tutt'oggi le stesse potenzialità di altri suoi simili quali il C++, ma che ha il suo punto forte nella possibilità di gestire facilmente oggetti grafici e testuali compatibili con Office. Infatti, in qualsiasi applicativo di Office si ha la possibilità di "registrare" una macro contenente un certo numero di operazioni che si vogliono automatizzare e di identificarla con un nome specifico. Da questo momento in poi è possibile richiamare la macro, anche con un semplice pulsante, per riprodurre nella stessa sequenza le operazioni registrate. Le macro sono delle vere e proprie routine registrate in codice Visual Basic; conoscendo il linguaggio si può modificarne il codice come con un vero e proprio compilatore. La modifica del codice delle macro, registrate semplicemente eseguendo delle operazioni, si è resa necessaria in tutti quei casi in cui si doveva ripetere l'operazione per un certo

[4] Quello di cui si sta parlando è il linguaggio Visual Basic for Application, leggermente diverso dal Visual Basic "puro" in quanto, pur avendo le stesse funzioni di base, è già predisposto ad essere usato con i programmi di Office, tanto che è compreso nel pacchetto d'installazione di ognuno di essi.

numero di volte: in pratica le modifiche si sono limitate all'inserimento del codice registrato in dei cicli iterativi.

Vogliamo sottolineare che nell'esecuzione delle macro, le istruzioni Visual Basic vengono interpretate e non c'è pertanto il programma oggetto. Nel corso della presente tesi si sono spesso utilizzate queste potenzialità che hanno permesso da una parte notevoli risparmi di tempo e dall'altra hanno minimizzato la probabilità di sbagliare operazioni anche semplici, ma molto ripetitive.

3.4 GLI STRUMENTI STATISTICI PER L'ANALISI DEI DATI

Uno dei maggiori problemi associati alle misure e valutazioni di qualsiasi aspetto del comportamento umano è la sua intrinseca variabilità. **Variabilità**, semplicemente, significa che valori ottenuti dalla misura di un parametro non saranno le stesse in differenti circostanze, rendendo impossibile la decisione di quale sia quello "giusto". Si può in ogni modo pensare che l'uomo, pur producendo una certa variabilità, riporti la sua intenzione a prodursi in atti identici di comportamento. Questa "intenzione", potrebbe considerarsi un'astrazione che non contenga variabilità. Si rendono allora necessari dei metodi automatici che muovano dai dati variabili misurati verso "invarianti" astrazioni. Di questo aspetto così importante per lo studio condotto in questa tesi, si occupa la statistica alla quale si è voluto dedicare questo intero paragrafo.

3.4.1 Media aritmetica e deviazione standard

La più semplice statistica che si che si può estrarre da *n* dati raccolti è la media aritmetica, così definita:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.5}$$

dove *n* è il numero di campioni mentre x_i rappresenta il valore dell' i-esimo campione.

L'attendibilità della media aritmetica quale valore rappresentativo di un insieme di misure di un parametro dipende dal numero di campioni misurati e dal *range* di variabilità di ciascuno. Un'indicazione sul *range* della maggior parte dei valori può essere data dalla deviazione standard calcolata nel modo seguente:

$$StD = \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}} \tag{3.6}$$

A proposito di questa formula bisogna fare alcune precisazioni. Considerando che le misure vengono eseguite su un campione finito della popolazione, e che quest'ultima segue generalmente una distribuzione continua di probabilità per il parametro in esame, si ha che la deviazione standard, come statistica di campione, approssima in maniera più o meno precisa la radice della varianza (incognita) della popolazione (StD $\approx \sigma$). Quando si ha a che fare con un numero piccolo di campioni (in genere <30) la formula (3.6) costituisce una buona approssimazione; mentre, se il numero di campioni è grande (in genere >30), si usa normalmente la stessa formula con al denominatore n al posto di (n-1). Nel presente lavoro si è sempre usata l'approssimazione per piccoli campioni considerando che i dati da mediare raramente hanno superato il numero di qualche decina di unità e che in pratica, al crescere di n, non sussiste alcuna differenza tra le due definizioni (Dillon e Goldstein, 1984; M.Spiegel, 1988).

3.4.2 Il test di analisi della varianza: l'ANOVA

Introduzione

L'analisi della varianza è la metodologia statistica usata per individuare e quantificare l'eventuale influenza delle tipologie prese in considerazione (sesso, pronuncia singola o geminata, vocali, consonanti) sulle misure rilevate dei diversi parametri scelti per l'analisi delle consonanti nasali.

Nel presente sottoparagrafo illustreremo tale metodologia basandoci su "Introduzione alla statistica" di T.H. Wonnacot, R.J. Wonnacot (1972), utilizzando alcune esemplificazioni classiche per questo tipo di trattazione e cercando di utilizzare, quando possibile, dei concetti intuitivi anziché lunghe dimostrazioni.

Analisi della varianza a un fattore

La significatività dei risultati di un'indagine può essere compresa mediante il seguente esempio: vogliamo confrontare tra loro tre macchine (A, B e C), le quali, essendo azionate da uomini e a causa di altre ragioni inesplicabili, danno luogo ad un prodotto orario soggetto a fluttuazioni casuali. Nella speranza di "mediare" e quindi di ridurre gli effetti di tali fluttuazioni, si effettua un campione casuale di 5 ore per ciascuna macchina, i cui risultati sono raccolti nella Tabella 3.3, insieme alle relative medie.

Macchine o numero del campione	*Campione della macchina i*					$\overline{X}_i$
i = 1	48,4	49,7	48,7	48,5	47,7	48,6
= 2	56,1	56,3	56,9	57,6	55,1	56,4
= 3	52,1	51,1	51,6	52,1	51,1	51,6

$$Media delle \overline{X} = \overline{\overline{X}} = 52,2$$

Tab. 3.3 Campioni dei prodotti da 3 macchine

La prima domanda che ci poniamo è: "Le macchine sono realmente differenti?". In altre parole, si vuole stabilire se le medie campionarie $\overline{X}_i$ nella Tabella 3.3 differiscono tra loro a causa della differenza nelle medie μ_i delle popolazioni da cui provengono (μ_i rappresenta la produzione media per tutto il periodo di vita della macchina *i*) oppure se queste differenze tra le $\overline{X}_i$ possono essere ragionevolmente attribuite solamente alle fluttuazioni casuali.

A scopo illustrativo, si supponga che siano stati effettuati tre esperimenti campionari su una macchina, i cui risultati sono raccolti nella Tabella 3.4. Come previsto, le fluttuazioni statistiche campionarie causano piccole differenze nelle medie dei campioni anche se le μ sono identiche.

N. del campione	*Valori campionari*					$\overline{X}_i$
i = 1	51,7	53,0	52,0	51,8	51,0	51,9
= 2	52,1	52,3	52,9	53,6	51,1	52,4
= 3	52,8	51,8	52,3	52,8	51,8	52,3

$\overline{\overline{X}} = 52{,}2$

Tab. 3.4 Tre campioni del prodotto di una macchina

Ne segue che la domanda può essere posta in altri termini: "Le differenze tra le $\overline{X}$ della Tabella 3.3 sono dello stesso ordine di grandezza di quelle della Tabella 3.4 (e così attribuibili alle fluttuazioni casuali), o risultano sufficientemente grandi da indicare una differenza effettiva tra le medie delle corrispondenti popolazioni?". In prima approssimazione, questa seconda spiegazione sembra la più plausibile, ma è chiaro che occorre sviluppare un test formale che fornisca elementi per rispondere con maggior rigore.

L'ipotesi di "nessuna differenza" tra le medie delle popolazioni diviene l'ipotesi nulla:

$$H_0: \mu_1 = \mu_2 = \mu_3 \tag{3.7}$$

L'ipotesi alternativa è che qualcuna delle μ (ma *non* necessariamente tutte) siano realmente differenti.

$$H_1: \mu_i \neq \mu_j \qquad \text{per qualche } i \text{ e } j \tag{3.8}$$

Per sviluppare un test plausibile di questa ipotesi, dobbiamo trovare in primo luogo una misura numerica del grado in cui le medie campionarie differiscono. A tal fine, consideriamo le tre medie campionarie nell'ultima colonna della Tabella 3.3 e ne calcoliamo la varianza; occorre sottolineare, in proposito, che stiamo calcolando la varianza delle medie campionarie e non la varianza di tutti i valori della tabella.

Avremo pertanto:

$$s^2_{\bar{X}} = \frac{1}{(r-1)} \sum_{i=1}^{r} (\bar{X}_i - \bar{\bar{X}})^2$$

$$= ½ [(48,6 - 52,2)^2 + (56,4 - 52,2)^2 + (51,6 - 52,2)^2] = 15,5 \qquad (3.9)$$

in cui r = numero delle righe (cioè numero delle medie campionarie) e

$$\bar{\bar{X}} = \text{media delle } \bar{X} = \frac{1}{r} \sum_{i=1}^{r} \bar{X}_i = 52,2 \qquad (3.10)$$

Tuttavia $s^2_{\bar{X}}$ non esaurisce la questione, poiché, se consideriamo ad esempio i dati della seguente Tabella 3.5, è facile osservare che essi, pur presentando un $s^2_{\bar{X}}$ uguale a quello della Tabella 3.3, si riferiscono a macchine con maggiore variabilità, che producono grandi fluttuazioni casuali nell'ambito di ciascuna riga.

Macchine	*Prodotto campionario della macchina i*					$\bar{X}_i$
i = 1	54,6	45,7	56,7	37,7	48,3	48,6
= 2	53,4	57,5	54,3	52,3	64,5	56,4
= 3	56,7	44,7	50,6	56,5	49,5	51,6
						$\bar{\bar{X}} = 52,2$

Tab. 3.5 Campioni della produzione di 3 macchine diverse

Le implicazioni di tale fatto sono rappresentate nella Figura 3.10. Più in particolare, nella Figura 3.l0-a le macchine presentano una variabilità tale che tutte le produzioni campionarie potrebbero essere state ottenute da macchine della stessa popolazione, cioè le differenze nelle medie campionarie possono essere spiegate dal caso. D'altra parte le (stesse) differenze delle medie campionarie possono difficilmente essere spiegate dal caso nella Figura 3.10-b, poiché in quest'ultimo esempio le macchine *non* presentano una variabilità accentuata.

Abbiamo ora degli elementi per poter operare i confronti. Per quanto riguarda il caso rappresentato nella Figura 3.10-b, concludiamo che i valori delle μ sono diversi e rifiutiamo H_0 poiché la varianza delle medie campionarie $s^2_{\bar{X}}$ è grande *relativamente alla* varianza casuale.

Occorre tuttavia predisporre un indice per misurare la variazione dovuta al caso. Intuitivamente, ci sembra che essa possa interpretarsi come dispersione (o varianza) dei valori osservati *entro* ciascun campione, e quindi calcoliamo senz'altro la varianza entro il primo campione nella Tabella 3.3

$$s_1^2 = \frac{1}{(n-1)} \sum_{j=1}^{n} (X_{1j} - \bar{X}_1)^2 = \frac{(48,4 - 48,6)^2 + ...}{4} = 0,52 \qquad (3.11)$$

in cui X_{1j} è il j-mo valore osservato nel primo campione.

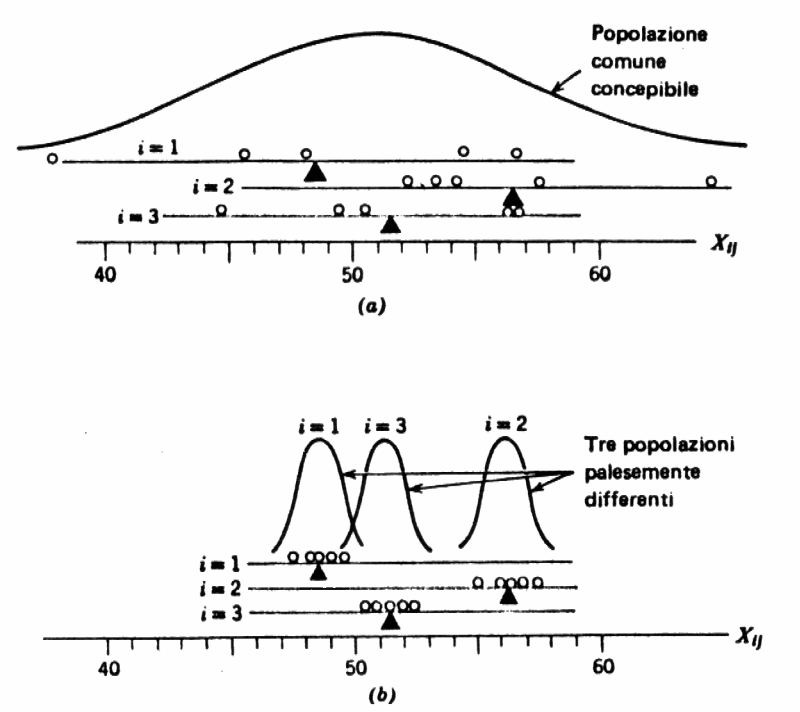


Fig. 3.10 (a) Grafico della Tabella 3.5; (b) Grafico della Tabella 3.3

Allo stesso modo calcoliamo la varianza della fluttuazione casuale entro il secondo (s_2^2) e il terzo campione (s_3^2). La media aritmetica semplice di queste varianze

$$s_p^2 = \frac{1}{r}\sum_{i=1}^{r} s_i^2 = \frac{0{,}52 + 0{,}87 + 0{,}25}{3} = 0{,}547 \qquad (3.12)$$

si assume come misura della fluttuazione casuale, e viene chiamata "varianza comune". Si noti che da ciascuno degli *r* campioni otteniamo una varianza campionaria con (n - 1) gradi di libertà, cosicché la varianza comune s_p^2 ha *r(n* - 1) gradi di libertà.

A questo punto possiamo porci la questione fondamentale consistente nel decidere se $s_{\bar{X}}^2$ è grande relativamente a s_p^2. L'esame del rapporto

$$F = \frac{n s_{\bar{X}}^2}{s_p^2} \qquad (3.13)$$

chiamato rapporto delle varianze, ci aiuta a risolvere la questione. Si noti che si è introdotto *n* nel numeratore in modo che, se H_0 è vera, il rapporto avrà, in media, un valore vicino ad 1; questo dipende dalla relazione che esiste tra la varianza delle medie campionarie e quella della popolazione. Accadrà peraltro che, a causa delle fluttuazioni statistiche, il rapporto stesso risulterà qualche volta superiore e qualche volta inferiore all'unità.

Se H_0 non è vera (e i valori di μ non sono gli stessi), allora $n s_{\bar{X}}^2$ sarà relativamente grande in confronto ad s_p^2 e il valore di *F* nella (3.13) risulterà più grande di 1. Formalmente si rifiuta l'ipotesi H_0 se il valore calcolato di *F* risulta significativamente maggiore di 1.

Il test formale di H_0, come del resto qualsiasi altro test, richiede la conoscenza della distribuzione della

statistica osservata se H_0 è vera. Tale statistica, che si indica in questo caso con il simbolo F, ha una distribuzione che, nel caso particolare sopra esaminato, assume la forma della curva rappresentata nella Figura 3.11, nella quale abbiamo anche indicato il valore critico $F_{.05}$ che lascia a destra il 5% della distribuzione. Pertanto, se H_0 è vera, vi è solamente una probabilità del 5 % che si possa osservare un valore di F superiore a 3,89; se si ottiene un valore superiore a 3,89 si rifiuta di conseguenza H_0. Natura1mente è anche possibile essere molto sfortunati ed osservare un valore di F superiore a 3,89 pur essendo H_0 vera, preferiamo tuttavia assumere H_0 come falsa.

Per illustrare questo procedimento, consideriamo le tre serie di risultati campionari nelle Tabelle 3.3, 3.4 e 3.5 e in ciascun caso ci chiediamo se le differenze che abbiamo rilevato per la produzione delle macchine siano statisticamente significative. In altre parole, in ciascun caso vogliamo provare H_0 : $\mu_1=\mu_2=\mu_3$ contro l'ipotesi alternativa che non siano uguali.

Per i dati della Tabella 3.4 una valutazione della (3.13) è:

$$F = \frac{ns_{\bar{X}}^2}{s_p^2} = \frac{0,35}{0,547} = 0,64 \qquad (3.14)$$

Poiché il risultato è inferiore al valore critico di $F_{.05}$ = 3,89 concludiamo che le differenze osservate tra le medie possono essere spiegate ragionevolmente solo da variazioni casuali. La cosa non sorprende perché i tre campioni della Tabella 3.4 sono stati ottenuti dalla stessa macchina.

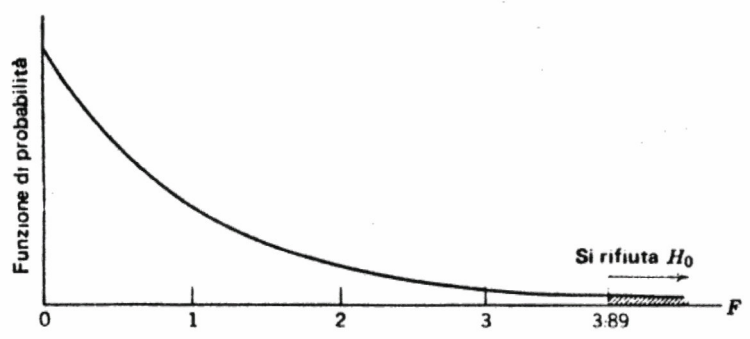


Fig. 3.11 Distribuzione di F quando H0 è vera (con 2,12 gradi di libertà).

Per i dati della Tabella 3.5 il rapporto F è

$$F = \frac{77,4}{35,7} = 2,17 \qquad (3.15)$$

in questo caso, la differenza fra le medie campionarie (cioè il numeratore) è molto più grande; ma la stessa cosa accade anche per la variazione casuale (il che si riflette nel denominatore). Anche questa volta il valore di F risulta inferiore al valore critico 3,89.

Infine per i dati della Tabella 3.3, il rapporto F è pari a

$$F = \frac{77,4}{0,547} = 141 \qquad (3.16)$$

In quest'ultimo caso, la differenza tra le medie campionarie è molto grande se confrontata con la variazione casuale, il che dà luogo ad un rapporto *F* che eccede di gran lunga il valore critico 3,89, e quindi l'ipotesi H_0 viene rifiutata.

Questi tre test confermano le conclusioni intuitive già sviluppate in precedenza. La Tabella 3.3 fornisce l'unico caso nel quale concludiamo che le popolazioni hanno medie diverse.

La distribuzione di F

Poiché questa distribuzione è importante, è bene esaminarla dettagliatamente. La distribuzione di *F* mostrata nella Figura 3.11 non è che una delle tante possibili, dato che ne esistono diverse in dipendenza dei gradi di libertà *(r - 1)* del numeratore e dei gradi di libertà *r(n - 1)* del denominatore. In questa sede possiamo vederne solo intuitivamente il perché. In effetti, maggiori sono i gradi di libertà nel calcolo del numeratore e del denominatore, più queste due stime di varianze risulteranno vicine al loro valore esatto: di conseguenza, il loro rapporto risulterà più vicino all'unità, come può desumersi dalla Figura 3.12.

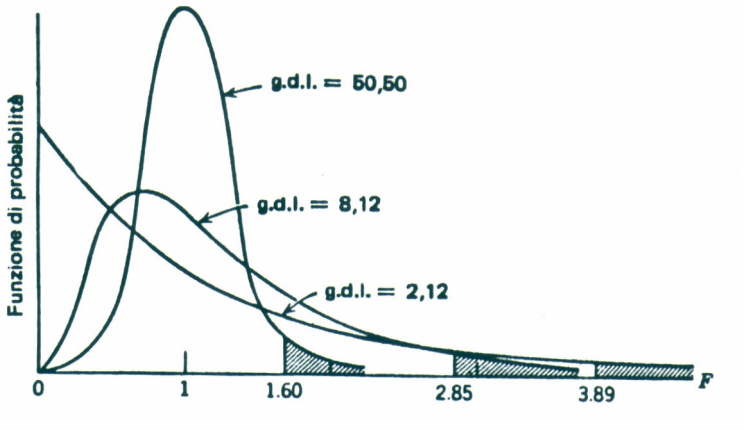


Fig. 3.12 Distribuzione di F con diversi gradi di libertà al numeratore e al denominatore.

Si noti come il punto critico (per il rifiuto di H_0) si sposti verso 1 quando aumentano i gradi di libertà.

Si potrebbe compilare tutto un insieme di tabelle di *F*, ciascuna corrispondente ad una diversa combinazione di gradi di libertà. In pratica, però, ciò non appare necessario dato che in genere si richiedono solamente i valori critici al 5% e all'1%. Come risultato di un test di anova, a volte, al posto di F viene fornito p, la probabilità di osservare un valore di F maggiore di quello effettivamente calcolato.

La tabella ANOVA

In questa sezione è sintetizzato il modo con cui vengono effettuati i calcoli di cui si è appena parlato. Il modello è riassunto nella Tabella 3.6 e nella colonna 2 viene assunta l'ipotesi che tutti i campioni siano estratti da popolazioni normali con la stessa varianza σ^2, ma, ovviamente, con medie che possono essere o no uguali (sono proprio le possibili differenze fra le medie che dobbiamo esaminare).

I calcoli che ne risultano vengono esposti nella Tabella 3.7, chiamata tabella **ANOVA** (abbreviazione per ANalysis Of VAriance). Nella prima riga sono raccolti i calcoli per il numeratore di *F*, nella seconda riga le elaborazioni per il denominatore; nella parte (b) di questa stessa tabella sono riportati i valori per l'esempio specifico delle tre macchine della Tabella 3.3.

(1) *Popolazione*	(2) *Distribuzione ipotizzata*	(3) *Valori campionari osservati*
1	$N(\mu_1,\sigma^2)$	X_{1j} $(j=1 \ldots n)$
2	$N(\mu_2,\sigma^2)$	X_{2j} $(j=1 \ldots n)$
3	$N(\mu_3,\sigma^2)$	X_{3j} $(j=1 \ldots n)$
.		
.		
.		
In generale:		
i	$N(\mu_i,\sigma^2)$	X_{ij} $(j=1 \ldots n)$

Tab. 3.6 Sommario delle ipotesi

a) Tabella ANOVA in generale

(1) *Fonte di variazione*	(2) *Devianza: somma dei quadrati*	(3) *Gradi di libertà*	(4) *Varianza (MSS)*	(5) *F(rapporto)*
Tra le righe "spiegata" dalle differenze tra le $\overline{X}_i$	$n\sum_{i=1}^{r}(\overline{X}_i-\overline{\overline{X}})^2=SS_r$	(r-1)	$MSS_r=SS_r/(r-1)$ $=ns^2_{\overline{X}}$	$\frac{\text{varianza spiegata}}{\text{varianza non spiegata}}=F$
Entro le righe; variazione residua, casuale "non spiegata"	$\sum_{i=1}^{r}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2=SS_u$	r (n-1)	$MSS_u=SS_u/r(n-1)$ $=s_p^2$	
Totale	$\sum_i\sum_j(X_{ij}-\overline{\overline{X}})^2$	(nr - 1)		

b)Tabella ANOVA, per i valori della Tabella 3.3

(1) *Fonte di variazione*	(2) *Devianza*	(3) *Gradi di libertà*	(4) *Varianza*	(5) *F(rapporto)*
Tra le macchine; "spiegata"	154,8	2	77,4	77,4 / 0,547 = 141
Entro le macchine; "non spiegata"	6,56	12	0,547	
Totale	161	14		

Tab. 3.7 a) tabella ANOVA in generale; b) tabella ANOVA per i valori della Tabella 3.3

La tabella ANOVA ci fornisce inoltre due utili controlli intermedi per i nostri calcoli. Il primo riguarda i gradi di libertà della colonna 3. L'altro è relativo alla somma dei quadrati nella colonna 2, poiché la somma dei quadrati *tra* le righe aggiunta alla somma dei quadrati *entro* le righe deve dare come risultato la somma totale dei quadrati. In definitiva:

$$\sum_i \sum_j (X_{ij} - \overline{\overline{X}})^2 = n\sum_i \overline{X}_i - \overline{\overline{X}})^2 + \sum_i \sum_j (X_{ij} - \overline{X}_i)^2 \qquad (3.17)$$

In altre parole la variazione totale è uguale alla somma della variazione spiegata con la variazione non spiegata.

Quando ogni variazione (devianza) viene divisa per i corrispondenti gradi di libertà si ha la varianza. La varianza tra le righe è "spiegata" dal fatto che le righe possono provenire da diverse popolazioni (per esempio, macchine che si comportano in modo diverso). La varianza *entro* le righe è "non spiegata" poiché dipende dalle variazioni casuali che assumono i valori, variazioni che non possono essere spiegate sistematicamente (dalle differenze nelle macchine). Perciò qualche volta ci si riferisce ad *F* come ad un rapporto tra varianze.

$$F = \frac{Varianza \quad spiegata}{Varianza \quad non \quad spiegata} \qquad (3.18)$$

Le considerazioni precedenti ci suggeriscono un mezzo possibile per rafforzare il test *F*. Si supponga che le tre macchine dell'esempio siano sensibili alle differenze di temperatura. Allora si può introdurre esplicitamente la temperatura nella nostra analisi. Se parte delle variazioni non spiegate possono essere ora spiegate dalla temperatura, il denominatore della (3.13) si ridurrà, dando luogo ad un valore di *F* più grande del precedente, il che ci metterà a disposizione un test più potente per le macchine (cioè saremo in una posizione più forte per rifiutare H_0). Ne segue che l'introduzione di altre spiegazioni della varianza ci permetterà di determinare se una specifica causa (quella delle diverse macchine) è importante o meno. Ciò ci conduce all'esame dell'argomento "**ANOVA** a due fattori".

Analisi della varianza a due fattori

Riferendoci sempre all'esempio delle macchine, vediamo come si possa tenere conto del fatto che parte della varianza comune è dovuta al fattore umano.

Si supponga che le produzioni campionarie nella Tabella 3.5 siano state ottenute da cinque diversi operatori e che ogni operatore produca uno dei valori campionari su ciascuna macchina. In tali condizioni, conviene raggruppare i dati precedenti mediante una classificazione a due caratteri (a seconda della macchina *e* dell'operatore) ed ottenere la Tabella 3.8.

Operatore / *Macchine*	j =1	2	3	4	5	*Media della macchina* $\overline{X}_{i.}$
i = 1	56,7	45,7	48,3	54,6	37,7	48,6
2	64,5	53,4	54,3	57,5	52,3	56,4
3	56,7	50,6	49,5	56,5	44,7	51,6
Media dell'operatore $X_{j.}$	59,3	49,9	50,7	56,2	44,9	$\overline{\overline{X}} = 52{,}2$

Tab. 3.8 Campioni della produzione (X_{ij}) di tre diverse macchine (come nella Tabella 3.5 ma ordinate secondo l'operatore)

E' necessario a questo punto complicare la notazione poiché ci interessa sia la media di ciascun operatore ($\overline{X}_{.j}$, media di ciascuna colonna) sia la media di ciascuna macchina ($\overline{X}_{i.}$, media di ciascuna riga) [5].

Ora il quadro è più chiaro: alcuni operatori sono efficienti (il primo e il quarto), mentre altri non lo sono. Le macchine dopo tutto non presentano una notevole variabilità poiché si osserva soltanto una grande differenza nell'efficienza degli operatori. Pertanto, se potessimo tenere conto di quest'ultima circostanza, riusciremmo a ridurre la nostra varianza non spiegata (o casuale) al denominatore della (13.18). E poiché il numeratore rimarrà invariato, il rapporto *F* risulterà di conseguenza così grande da consentirci, forse, di rifiutare l'ipotesi H_0. In tale caso, apparirebbe chiaramente che un'altra influenza (differenza negli operatori) sarebbe responsabile della maggior parte delle difficoltà della nostra analisi della varianza della sezione precedente; superando questa difficoltà speriamo di ottenere un test molto più potente per le nostre macchine.

L'analisi appare come un'estensione dell'analisi della varianza (ANOVA) ad un fattore, ed è sintetizzata nella Tabella 3.9.

Naturalmente in questa tabella, la lettera minuscola *c* rappresenta il numero delle colonne nella Tabella 3.8 e sostituisce *n* nella Tabella 3.5, mentre, come nel caso precedente, le diverse componenti delle variazioni della seconda colonna hanno per somma la variazione totale in fondo a questa colonna, cioè

$$\sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij}-\overline{\overline{X}})^2 = c\sum_{i=1}^{r}(\overline{X}_i-\overline{\overline{X}})^2 + r\sum_{j=1}^{c}(\overline{X}_{.j}-\overline{\overline{X}})^2 + \sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij}-\overline{X}_{i.}-\overline{X}_{.j}+\overline{\overline{X}})^2 \qquad (3.19)$$

Questa formula ci dice che la variazione totale è pari alla variazione delle macchine (righe) sommata alla variazione dell'operatore (colonna) e alla variazione casuale

[5] Il punto indica l'indice rispetto al quale si effettua la sommatoria. Per esempio, il punto sostituisce *j* in $\overline{X}_{i.} = \frac{1}{n}\sum_{j} X_{ij}$.

(1) *Fonte delle variazioni*	(2) *Devianza; Somma dei quadrati (SS)*	(3) *Gradi di libertà*	(4) *Varianza (MSS)*	(5) *F*
Tra le righe: **Spiegata** dalle differenze tra le macchine; cioè differenze tra le $\overline{X}_{i.}$	$SS_r = c\sum_{i=1}^{r}(\overline{X}_{i.} - \overline{\overline{X}})^2$	r - 1	$MSS_r = \frac{SS_r}{r-1} = cs^2_{\overline{X}_{i.}}$	$\frac{MSS_r}{MSS_u}$
Tra le colonne: **Spiegata** dalle differenze tra gli operatori. Cioè differenze nelle $\overline{X}_{.j}$	$SS_c = r\sum_{j=1}^{c}(\overline{X}_{.j} - \overline{\overline{X}})^2$	c - 1	$MSS_c = \frac{SS_c}{c-1} = rs^2_{\overline{X}_{.j}}$	$\frac{MSS_c}{MSS_u}$
Non spiegata: cioè residuo risultante da fluttuazioni casuali.	$SS_u = \sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{\overline{X}})^2$	(r - 1) (c - 1)	$MSS_u = \frac{SS_u}{(r-1)(c-1)} = s_p^2$	
Totale	$SS = \sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij} - \overline{\overline{X}})^2$	rc - 1		

Tab. 3.9 ANOVA a due fattori

Notiamo che la variazione dovuta all'operatore è definita analogamente a quella dovuta alla macchina, con l'unica differenza che, in questo caso, la variazione dovuta all'operatore è data dalla variazione registrata dalle medie per *colonna*. La (3.19) viene stabilita mediante una complessa serie di passaggi, simili a quelli necessari per stabilire la (3.17) nel caso semplice.

Prova delle ipotesi

Avendo scisso nella (3.19) la variazione totale in componenti, possiamo ora verificare se si è prodotta una differenza significativa fra le macchine *o* fra gli operatori, tenendo conto, in ambedue i test dell'influenza estranea dell'altro fattore.

Iniziamo col verificare l'ipotesi della differenza fra le macchine, costruendo il rapporto

$$F = \frac{Mss_r}{Mss_u} = \frac{Varianza\ \ spiegata\ \ delle\ \ macchine}{Varianza\ \ non\ \ spiegata} \qquad (3.20)$$

il quale, se H_0 è vera, ha una distribuzione *F*. Così se il valore di *F* osservato, calcolato nella (3.20), supera il valore critico di *F* possiamo rifiutare l'ipotesi nulla, concludendo che c'è una differenza tra le medie per righe della popolazione. I calcoli sono sviluppati nella Tabella 3.10.

Dalla Tabella 3.10 si ottiene che la (3.20) è pari a:

$$F = \frac{77{,}4}{5{,}9} = 13{,}1 \qquad (3.21)$$

(1) *Fonte di variazione*	(2) *Devianza (SS)*	(3) *Gradi di libertà*	(4) *Varianza (MSS)*	(5) *F*	(6) *F critico*
Tra le macchine	154,8	2	77,4	13,1	4,46
Tra gli operatori	381,6	4	95,4	16,2	3,84
Residuo	47,3	8	5,9		
Totale	583,7	14			

Tab. 3.10 ANOVA a due criteri. (Per i dati si veda Tab.3.8)

Poiché il valore ottenuto supera il valore critico[6] di *F* (4,46), rifiutiamo l'ipotesi nulla che le macchine siano simili.

Se confrontiamo il risultato ora ottenuto con il test *F* nella (3.15), in cui non eravamo in grado di rifiutare l'ipotesi nulla, osserviamo che mentre il numeratore rimane invariato, la variazione casuale nel denominatore è molto più piccola, poiché si è tenuto conto degli effetti delle differenze tra gli operatori. Ciò ci ha dato una grande "potenza"[7] in senso statistico, che ci ha permesso il rifiuto dell'ipotesi nulla.

Allo stesso modo potremmo sottoporre a test l'ipotesi nulla che gli operatori lavorino nella stessa maniera. Ancora una volta *F* e' il rapporto tra una varianza spiegata e una non spiegata, ma questa volta, naturalmente, il numeratore è la varianza stimata attraverso le differenze tra le colonne.

$$F = \frac{Varianza\ \ spiegata\ \ dagli\ \ operatori}{Varianza\ \ non\ \ spiegata} = \frac{Mss_r}{Mss_u} = \frac{95,4}{5,9} = 16,2 \qquad (3.22)$$

In questo caso abbiamo isolato l'azione dovuta alle macchine, perciò abbiamo ottenuto un test più potente per confrontare l'azione degli operatori. Poiché il valore osservato di *F* è pari a 16,2 ed è superiore al valore critico[8] di *F (*3,84), rifiutiamo l'ipotesi nulla concludendo che gli operatori in realtà lavorano in modo diverso.

[6] 2 e 8 gradi di libertà, e livello di significatività del 5 %.

[7] A rigor di termini, abbiamo un test più potente poiché abbiamo ridotto la varianza non spiegata; ciò facendo abbiamo guadagnato più di quello che avevamo perso riducendo i gradi di libertà al denominatore di 4.

[8] Diverso dal test precedente poiché ora i gradi di libertà sono 4 e 8.

C'è un argomento che può essere ulteriormente chiarito. Nel test a un fattore, abbiamo calcolato la varianza non spiegata ricercando la variabilità degli *n* valori osservati entro un campione, cioè entro l'intera riga nella Tabella 3.5. In un test a due criteri di classificazione (Tabella 3.8), però, avendo scisso le osservazioni per colonna e per riga, siamo rimasti con una sola osservazione per ciascuna casella: ad esempio, c'è una sola osservazione (57,5) del prodotto ottenuto dall'operatore 4 sulla macchina 2. Non possiamo allora calcolare la variazione entro tale casella. Cosa faremo? Ci chiediamo: "Se non ci sono errori casuali, come potremmo prevedere la produzione dell'operatore 4 sulla macchina 2?" Notiamo incidentalmente che questa è una macchina migliore della media ($\overline{X}_{2.}$ = 56,4) e con un operatore relativamente efficiente ($\overline{X}_{.4}$ = 56,2) e quindi, in ogni caso, dovremmo prevedere un prodotto superiore alla media. Questa osservazione può essere facilmente usata per prevedere $\hat{X}_{2,4}$. In effetti, se stimiamo in ciascuna casella l'elemento casuale come differenza tra il nostro valore osservato (X_{ij}) e il corrispondente valore stimato $\hat{X}_{ij}$, otterremo un insieme d'elementi casuali la cui somma dei quadrati sarà esattamente la variazione non spiegata SS_u (l'ultimo termine nell'equazione (3.19) che appare anche nella colonna 2 della Tabella 3.9); dividendo per i gradi di libertà si otterrà la varianza non spiegata usata nel denominatore di ambedue i test condotti sull'ultimo esempio considerato.

In dettaglio, il valore previsto $\hat{X}_{ij}$ è così definito:

$$\hat{X}_{ij} = \overline{\overline{X}} + \text{correzione dovuta al comportamento della macchina +}$$

$$+ \text{correzione dovuta al comportamento dell'operatore} =$$

$$= \overline{\overline{X}} + (\overline{X}_{i\cdot} - \overline{\overline{X}}) + (\overline{X}_{.j} - \overline{\overline{X}}) \qquad (3.23)$$

Nel nostro esempio

$$\hat{X}_{2,4} = 52,2 + (56,4 - 52,2) + (56,2 - 52,2) = 52,2 + 4,2 + 4,0 = 60,4$$

Così, la previsione del comportamento dell'operatore 4 sulla macchina 2 si calcola correggendo il comportamento medio (52,2) con il grado in cui la macchina è superiore alla media (4,2) e il grado in cui lo è l'operatore (4,0). Semplificando i valori $\overline{\overline{X}}$ nella (3.23):

$$\hat{X}_{ij} = \overline{X}_{i.} + \overline{X}_{.j} - \overline{\overline{X}} \qquad (3.24)$$

e l'elemento casuale, che è la differenza tra il valore teorico e quello osservato, diviene:

$$X_{ij} - \hat{X}_{ij} = X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{\overline{X}} \qquad (3.25)$$

Notiamo che questo elemento casuale è il prodotto non spiegato dopo aver introdotto le correzioni per la macchina *i* e l'operatore *j*.

Nel nostro esempio:

$$X_{2,4} - \hat{X}_{2,4} = 57,5 - 60,4 = -2,9 \qquad (3.26)$$

Così questo prodotto osservato è di 2,9 al di sotto del previsto, e deve rimanere non spiegato (risultato delle influenze casuali). La variazione non spiegata (SS_u) viene ad essere uguale alla somma dei prodotti di tutti gli elementi casuali come nella (3.25).

Cenni sull'analisi della varianza a più fattori e sul problema dell'interazione

Abbiamo dunque visto quali sono le differenze tra l'analisi della varianza ad un fattore e quella a due fattori. E' facile, a questo punto, immaginare che, in presenza di più fattori di variabilità dei dati, potranno complicarsi in maniera notevole le formule ma il principio dell'analisi della varianza rimarra lo stesso.

Essendo queste pagine semplicemente a supporto di un lavoro di analisi di dati sperimentali, non ci addentreremo nell'analisi della varianza multifattoriale, che pure abbiamo usato in maniera sistematica nelle analisi che verranno a breve descritte, rimandando ai testi citati in bibliografia per maggiori dettagli. Riteniamo comunque di aver fornito gli elementi essenziali alla comprensione di quanto verrà detto di seguito. Questo stesso discorso è valido anche per il concetto di interazione di cui daremo solo un breve cenno.Sottolineiamo a proposito la difficoltà nel trovare una trattazione esauriente ed approfondita riguardo all'interazione, anche in testi considerati capisaldi della letteratura sull'analisi statistica.

Nel calcolare la produzione prevista $\hat{X}_{ij}$, nell'ultimo esempio fatto, abbiamo supposto che non ci sia interazione tra i due fattori cosa che invece avverrebbe, ad esempio se alcuni operatori lavorassero bene con alcune macchine e non con altre.

La presenza dell'interazione richiede un modello più complesso ed osservazioni in numero maggiore. Se per ogni combinazione dei due fattori sono disponibili *n* osservazioni, queste ultime possono essere considerate come un campione casuale estratto da una popolazione caratterizzata dai livelli *i* e *j* dei fattori ed aventi media μ_{ij}. Anche in questo caso, il valore della singola osservazione X_{ijk} può essere scomposta come la parte dovuta al primo fattore (la macchina), la parte dovuta al secondo fattore (l'operatore) e la parte dovuta alle fluttuazioni casuali.

Gli effetti che concernono i livelli del singolo fattore sono chiamati effetti principali. Se l'effetto del livello *i* del primo fattore sul valore atteso di X_{ijk} è costante al variare del livello *j* del secondo fattore, gli effetti dei due fattori sono additivi. Altrimenti tra i due fattori c'è interazione: l'effetto dei fattori non è la somma dell'effetto del primo fattore e del secondo fattore ed esiste un ulteriore fattore correttivo. Per maggiori dettagli sull'interazione rimandiamo al testo di Cicchitelli (1984).

3.4.3 Misura della correlazione: il test di Spearman

Un problema che spesso si presenta quando si affronta l'analisi di dati sperimentali, è quello di capire se tra due serie di dati relativi a due parametri di un certo evento vi sia correlazione; si vuole capire, cioè, se esiste una relazione diretta o inversa tra i parametri. Ha interesse, inoltre, quantificare il grado di correlazione.

Vi sono diversi test statistici che permettono di dare una risposta alle domande appena formulate; uno di questi è il **test di correlazione di Spearman.**

In questo test il grado di correlazione è indicato dal coefficiente $\mathbf{r_s}$ (***Spearman Rank Correlation Coefficient***). Il valore di questo coefficiente è sempre compreso tra –1 e +1 ed il grado di correlazione massimo corrisponde a 1 (in modulo) mentre il grado di correlazione minimo corrisponde al valore 0. Un valore positivo del coefficiente indica, inoltre, che, in media, all'aumentare di una grandezza aumenta anche l'altra, mentre un coefficiente negativo è indice di un comportamento esattamente opposto.

I passi da seguire per il calcolo del coefficiente $\mathbf{r_s}$ sono i seguenti:

1. Mantenendo in due vettori distinti (di *n* dati ciascuno) le due serie di dati, per ognuna si calcola un vettore di ranghi secondo le seguenti regole (si veda l'esempio di tabella 3.11):
   a) ha rango ***1*** l'elemento del vettore con il valore più basso, …, rango ***n*** l'elemento del vettore con valore più alto;
   b) se ***k*** valori sono coincidenti essi hanno stesso rango pari alla media aritmetica dei ranghi che avrebbero avuto se fossero stati diversi ma comunque adiacenti rispetto all'ordinamento.
2. Per ogni riga i dei vettori, si calcola la quantità d_i sottraendo al rango relativo al dato della i-esima riga del primo vettore quello relativo al dato della i-esima riga del secondo vettore (tab. 3.11).
3. Si calcola il coefficiente di correlazione secondo la formula

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} \qquad (3.27)$$

Si noti che questo test, diversamente da altri test di correlazione, non si basa direttamente sui valori dei dati ma ne sfrutta l'ordine.

Infine è importante ricordare che, accanto al valore di $\mathbf{r_s}$, il test fornisce anche il valore p (come per l'ANOVA) che indica se il valore del coefficiente di correlazione trovato è statisticamente significativo.

parametro X		parametro Y			
score	rank	score	rank	d_i	
31	3	79	7	-4	
40	9.5	92	10	-0.5	
26	1	74	3	-2	
33	4.5	78	5.5	-1	
39	8	82	8	0	
40	9.5	86	9	0.5	
37	7	77	4	3	***rs = +0.66***
33	4.5	78	5.5	-1	
35	6	72	1	5	
30	2	73	2	0	

Tab. 3.11 Esempio di calcolo dei vettori dei ranghi e differenze i-esime tra i ranghi della riga i, allo scopo finale del calcolo di r_s.

3.4.4 Criteri di classificazione

Nell'analisi statistica di dati sperimentali, dopo aver investigato su quali tra i parametri presi in considerazione influenzino significativamente il fenomeno che si sta studiando, ci si può porre la domanda se sia possibile classificare i dati rispetto al fenomeno stesso dal valore di qualcuno dei parametri e con quale precisione. In altre parole può avere interesse la misura della separabilità dei dati in due o più gruppi rispetto al fenomeno. Considereremo il solo caso di classificazione in due gruppi. Facciamo un semplice esempio per chiarire quanto appena detto. Supponiamo di misurare l'altezza di un certo numero di persone, uomini e donne. Supponiamo quindi di trovare, ad esempio applicando ai dati un test di ANOVA, che l'altezza di un individuo è significativamente dipendente dal sesso. A questo punto ci si può chiedere se sia possibile e con quale precisione, dedurre il sesso di una persona conoscendo la sua altezza.

Il criterio di classificazione preso in considerazione nel presente lavoro è il *criterio MLC* (Maximum Likelihood Criterion o Criterio di massima probabilità).

Si suppone che le misure dei parametri di un certo insieme omogeneo, siano statisticamente descrivibili tramite una gaussiana, con un valore medio *m* e una varianza σ^2. Riportiamo di seguito l'espressione d.d.p della gaussiana :

$$p(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{3.28}$$

Questa ipotesi è molto ragionevole ogniqualvolta si tenti di descrivere un qualsiasi fenomeno naturale e non rappresenta perciò in alcun modo una limitazione. Il *criterio MLC* prevede, come misura della separabilità, il calcolo della percentuale di errori commessi operando una classificazione a posteriori di ciascuno dei dati nei due gruppi, secondo un criterio di massima verosimiglianza.

Si procede come segue:

1. Si dividono i dati in due gruppi a seconda dell'aspetto su cui si vuole basare la classificazione (nel nostro esempio, il sesso).
2. Si calcolano *m* e σ delle gaussiane relative ai due gruppi (nel nostro esempio uomini e donne).
3. Si classifica ciascuno dei dati come appartenente ad un gruppo o all'altro a seconda di quale delle funzioni gaussiane relative ai due gruppi sia maggiore, quando la si valuti in quel punto.
4. Si calcola il numero di errori commessi sfruttando il fatto che si conosce già il gruppo di appartenenza di ogni dato in esame.

Il procedimento è a posteriori proprio per questa ultima ragione: si conosce già il gruppo di appartenenza di ogni dato in esame; inoltre i parametri delle due gaussiane sono calcolati proprio tramite i dati che si vanno a classificare.

La figura 4.9 descrive graficamente il procedimento sopra esposto. Come si vede tale tecnica porta all'individuazione di una frontiera tra i due gruppi. Tutti i dati per cui la misurazione del parametro X dia un valore superiore alla frontiera segnata in figura saranno classificate come appartenenti al gruppo 2, le altre al gruppo 1.

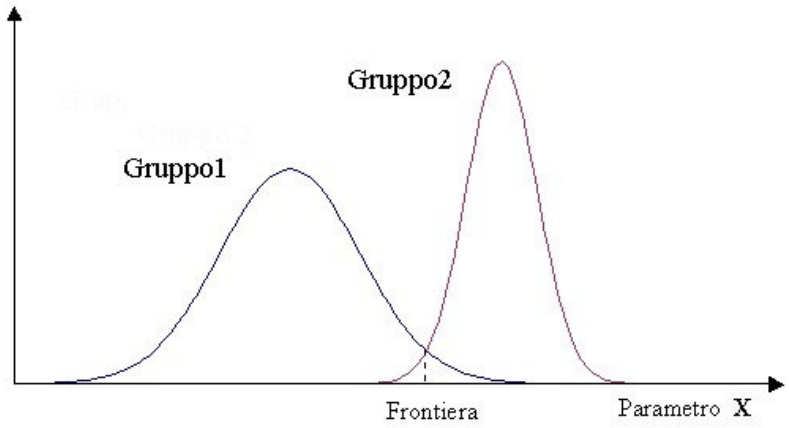


Fig. 3.13 Criterio MLC.

Il criterio MLC è già stato utilizzato per la classificazione in diversi lavori tra quelli del progetto GEMMA. Il test è stato implementato con un programma scritto in Pascal. Si rimanda per maggiori dettagli a A. Vannucci, 1993 e R. Rossetti, 1993.

CAPITOLO 4

L'ANALISI ACUSTICA DELLE CONSONANTI NASALI: METODOLOGIA E RISULTATI

INTRODUZIONE

Nei primi tre capitoli sono state ampiamente descritte le tecniche di analisi del segnale vocale, sia da un punto di vista teorico sia da un punto di vista più applicativo.

Ovviamente, sarebbe riduttivo pensare di poter racchiudere in poche decine di pagine anche solo i fondamenti dell'elaborazione numerica del segnale vocale ed è chiaro che lo scopo dei primi tre capitoli è solo quello di fornire gli elementi minimi indispensabili alla comprensione del lavoro svolto.

In questo capitolo sarà descritta la metodologia dell'analisi condotta sulle consonanti nasali italiane sulla base delle misure effettuate su parametri nel dominio del tempo, della frequenza ed energetici. Saranno quindi illustrati i risultati ottenuti con particolare riguardo a quelli relativi al fenomeno della geminazione. Per non appesantire troppo il testo, tutti i dati relativi alle misure eseguite e alle elaborazioni statistiche sono stati raccolti nelle prime quattro appendici (A, B, C e D). Infine, nell'ultima appendice (E), si trovano i listati in linguaggio C dei programmi scritti a supporto dell'analisi.

4.1 I PARAMETRI SCELTI PER L'ANALISI ED I CRITERI DI MISURA

Nella scelta dei parametri da misurare e dei criteri da seguire per le misurazioni, si è dovuto tenere presente che, come già detto, il presente lavoro si colloca nel contesto di uno studio più ampio sulla

geminazione, condotto su tutte le consonanti italiane (progetto GEMMA). Per questo motivo è stato preso un set di misure standard, comuni a tutti i lavori finora effettuati sulla base di dati del progetto GEMMA. Si è poi ritenuto di dover prendere in considerazione anche la misurazione di altri parametri specifici per la classe di consonanti che si stava analizzando nel presente studio. Nei prossimi paragrafi saranno descritti tutti i parametri misurati e verrà esplicitamente indicato se i parametri misurati sono non standard.

4.1.1 Le misure nel dominio del tempo ed i criteri di segmentazione

Le misure nel dominio del tempo, condotte sul database del progetto GEMMA sono:

- La durata della prima vocale della pronuncia che indicheremo nel seguito come V1d
- La durata della consonante che indicheremo nel seguito come Cd
- La durata della seconda vocale della pronuncia che indicheremo nel seguito come V2d
- La durata totale della pronuncia che indicheremo nel seguito come Utd (utterance duration)

Ricordiamo che il database utilizzato è composto da pronunce del tipo VCV (vocale – consonante - vocale) oppure VCCV. Le consonanti nasali, quando sono pronunciate in questo contesto (VCV o VCCV), non presentano brusche variazioni (come accade invece per le occlusive): si passa "dolcemente" con continuità dalla prima vocale alla consonante e poi ancora dalla consonante alla seconda vocale.

Non si è notata alcuna particolare differenza tra le durate delle due zone di transizione VC e CV, anche se non si sono eseguite misure oggettive su questa proprietà. In ogni caso le due zone di transizione sono sempre comprese nell'intervallo 15÷30 ms corrispondente a 1÷3 frame di analisi[1]. Per misurare le durate dei singoli fonemi, si è dovuto scegliere come comportarsi rispetto alle zone di transizione. Considerato che ciò che interessa in questa sede è il confronto tra le durate dei fonemi, si è deciso di non considerare le zone di transizione e inglobare le loro durate in parte sulla vocale ed in parte sulla consonante. In effetti, dato che la media tra le durate di tutti i fonemi della base dati è di 144 ms [2], le zone di transizione rappresentano appena il 5÷10% di un fonema. L'importante quindi è che il criterio adottato per operare la separazione tra vocale e consonante, sia uniforme, così che i risultati finali non risentiranno di quest'approssimazione.

In accordo alle scelte appena menzionate, si è quindi segmentata ogni pronuncia in tre sezioni corrispondenti ad altrettanti fonemi: il primo e l'ultimo segmento sono le due vocali e quello centrale la consonante (sia essa singola o geminata).

La determinazione delle durate dei fonemi si è dunque tradotta nell'individuazione dei seguenti campioni del segnale nel tempo:

---

[1] Ricordiamo che la frequenza di campionamento utilizzata per digitalizzare tutte le pronunce della base dati è di 10kHz, e perciò il *frame* standard utilizzato da UNICE è di 12.8 ms (corrispondente a 128 campioni).

[2] Questo risultato è in perfetto accordo con i dati riportati in letteratura per un parlato a velocità normale (6÷8 fonemi per secondo). Infatti, per le pronunce analizzate, risulta un "ritmo fonetico" medio di 6.94 fonemi per secondo.

1 Campione di attacco della prima vocale (***V1 onset***);

2 Campione di attacco della consonante (***C onset***) o di fine della prima vocale (***V1 offset***);

3 Campione finale della consonante (***C offset***) o di inizio della seconda vocale (***V2 onset***);

4 Campione di fine della seconda vocale (***V2 offset***).

Si ricorda che UNICE memorizza i campioni della forma d'onda temporale in un file (estensione *.sig*), mentre riserva ad un altro file (con estensione *.key*) il compito di contenere le altre informazioni tra cui la segmentazione. Questa particolare caratteristica è stata sfruttata per il calcolo delle lunghezze dei fonemi.

Infatti, operativamente, si è dovuto semplicemente, con un click di mouse, posizionare quattro marker in corrispondenza dei quattro punti sopra indicati. In un secondo momento, poi, tramite un piccolo programma scritto in C (*Durate.C)*, utilizzando i dati del file *.key* e sapendo la frequenza di campionamento del segnale, è stato possibile calcolare la durate di ogni fonema.

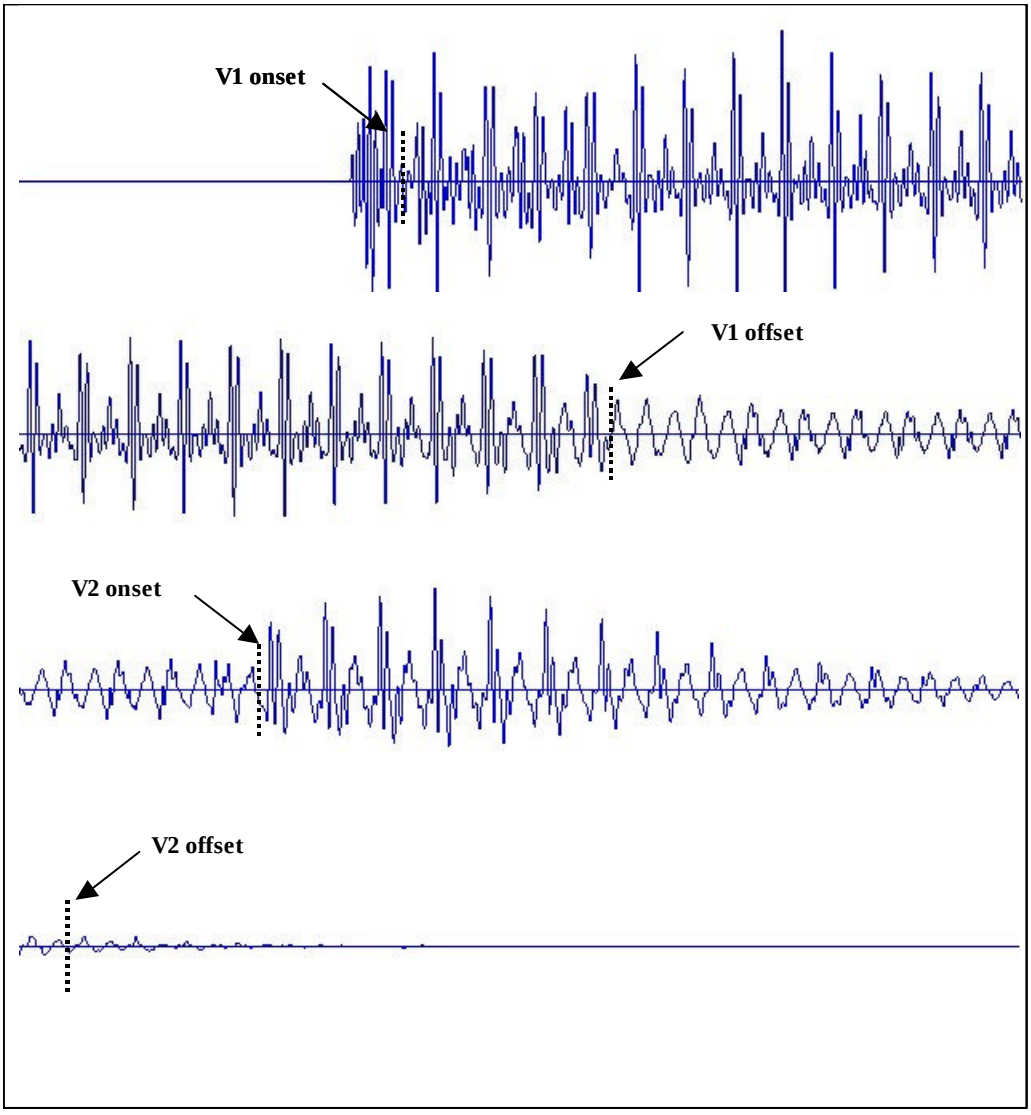


Fig. 4.1 Esempio di segmentazione per una pronuncia singola, di un parlatore femminile: AMA1EZ. Una riga intera corrisponde a circa 125 ms di segnale.

Vediamo ora nel particolare quali problemi si sono incontrati nella segmentazione delle pronunce e quali criteri si sono seguiti per individuare i campioni sopra indicati.

Per identificare in maniera più oggettiva possibile i campioni di separazione tra i fonemi, ci si è serviti dello spettrogramma (sia NB sia WB), del grafico dell'energia a breve termine e, soprattutto, dell'andamento del segnale nel tempo. Spesso si è anche eseguita la verifica della segmentazione con una prova d'ascolto, esaminando le differenze derivanti dall'aggiunta o dall'eliminazione di un frame. L'insieme di queste informazioni ha reso quasi sempre univoco il punto in cui interrompere un fonema per farne iniziare un altro.

Per quanto riguarda la scelta dei due punti di campionamento all'interno della pronuncia, ci si è basati soprattutto sulla forma d'onda nel tempo e sulle transizioni delle formanti. Si sono trovate difficoltà sicuramente maggiori per la separazione V-C rispetto a quella C-V. In particolare, le difficoltà maggiori si sono avute con le pronunce che presentavano la vocale [u]. Infatti, la forma d'onda nel tempo della [u], non si distanzia molto da quella delle consonanti nasali. A volte è stato di aiuto il grafico dell'energia. Nei casi peggiori (per la verità molto pochi rispetto alla totalità delle pronunce con la vocale [u]) la prova d'ascolto ha assunto una rilevanza maggiore e ci si è basati anche sul cambiamento di pendenza della forma d'onda nel tempo.

Nella scelta del punto di inizio e di fine pronuncia (V1 onset e V2 offset, rispettivamente) si sono incontrati problemi diversi. Per quanto riguarda l'attacco, soprattutto nelle pronunce con la vocale [a], si è trovato che spesso i primi millisecondi del segnale erano solo "colpi di glottide" anziché vocale (in pratica dei suoni sonori ma molto sporchi) e per questo sono stati considerati *non* facenti parte della pronuncia. Bisogna dire, d'altra parte che, in alcuni casi in cui questo effetto era particolarmente evidente, anche come durata, e la transizione tra la forma d'onda spuria e quella vocalica era graduale, togliendo completamente i colpi di glottide ed ascoltando la pronuncia si aveva l'impressione di un attacco innaturale. In questi casi si è deciso di mantenere una parte dell'attacco. In generale, comunque, l'inizio della quasi stazionarietà della prima vocale è stato facilmente individuabile sia nella forma d'onda temporale (notando la comparsa di forti picchi crescenti rapidamente d'ampiezza fino alla stazionarietà con periodo immediatamente individuabile), sia nello spettrogramma con comparsa improvvisa di formanti o bande di energia molto intense.

Anche la scelta di V2 offset ha spesso richiesto un'attenzione particolare, questa volta a causa del lento decadimento della vocale conclusiva dovuto all'intonazione discendente di fine parola. L'istante di fine pronuncia si è posto generalmente dove il periodo non aveva più la forma tipica della vocale stazionaria e anche le formanti dalla seconda in poi scomparivano. Accadeva però, relativamente di frequente che per diversi periodi l'ampiezza del segnale tendeva lentamente a zero, senza tuttavia mostrare, da un certo punto in poi, le caratteristiche tipiche di una vocale né sullo spettrogramma né all'ascolto. Si è deciso in questi casi di collocare il campione V2offset nel punto in cui l'ampiezza si abbassava di una certa percentuale (85÷90%) sotto il picco massimo.

Nelle figure 4.1 e 4.2 sono mostrati due esempi di segmentazioni operate: la prima è una pronuncia femminile di una consonante singola; la seconda è una pronuncia maschile di una consonante geminata. In particolare nella prima si può notare il "colpo di glottide" in inizio di pronuncia mentre nella seconda si vede abbastanza bene il decadimento lento verso lo zero alla fine della pronuncia stessa.

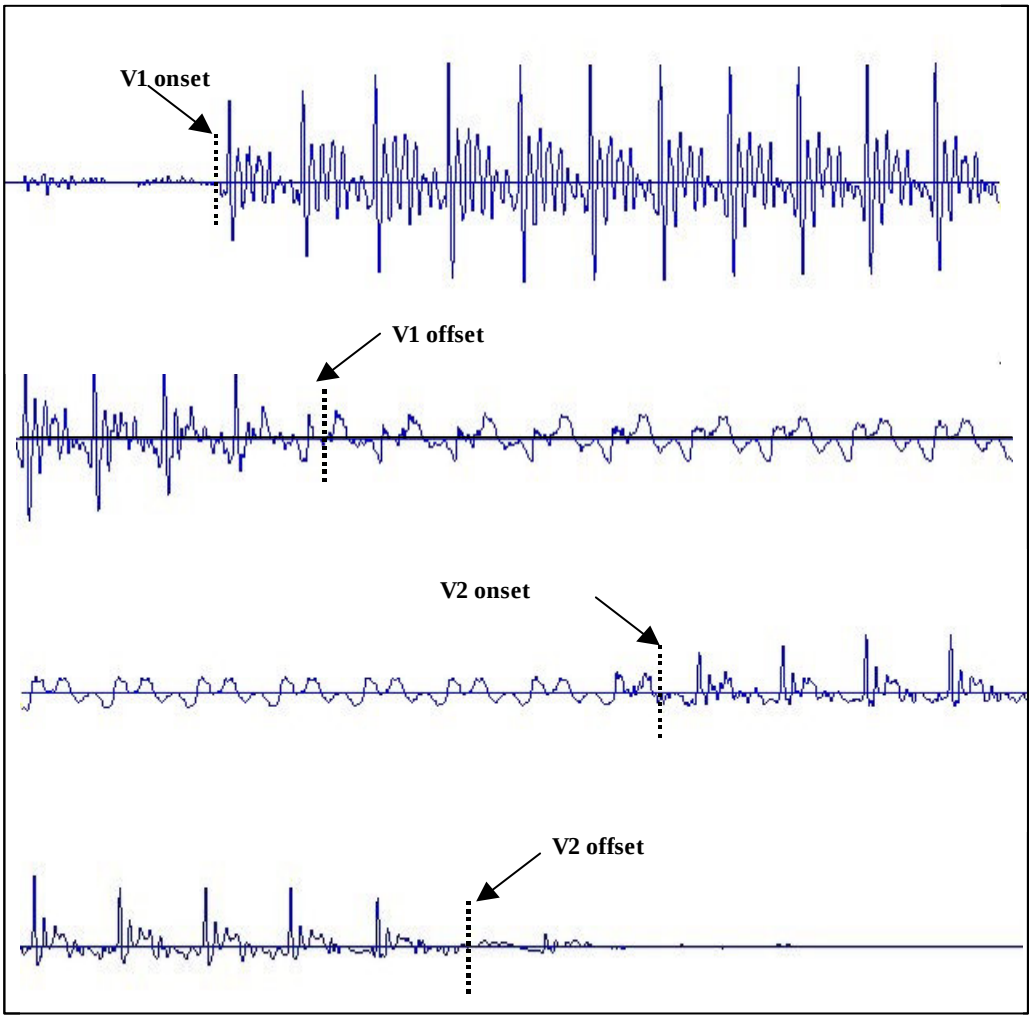


Fig. 4.2 Esempio di segmentazione per una pronuncia geminata, di un parlatore maschile: ANNA1PM. Una riga intera corrisponde a circa 125 ms di segnale.

Concludiamo questo paragrafo dicendo che, per eliminare le incertezze e rendere le misure il più possibile coerenti tra loro, a distanza di tempo, si sono effettuate nuovamente alcune segmentazioni, per poi confrontarle con quelle precedenti: il risultato è stato che nel 90% dei casi gli istanti presi erano praticamente coincidenti, mentre nel restante 10% le differenze restavano modeste (al massimo uno o due periodi di pitch di variazione) a conferma della bontà delle misure effettuate.

Il listato completo del programma *Durate.C* si trova nell'appendice E, mentre il relativo file è archiviato su cdrom insieme al resto del lavoro e conservato al dipartimento INFOCOM.

4.1.2 Le misure nel dominio della frequenza ed i criteri di misurazione del pitch e delle formanti

Le misure standard GEMMA eseguite nel dominio della frequenza sono quelle relative alla frequenza di pitch e alle prime tre formanti tutte con le relative ampiezze, in particolari frame della pronuncia, e sono di seguito elencate:

1. F0, A0, F1, A1, F2, A2, F3, A3, nel frame centrale della V1 (V1 center)
2. F0, A0, F1, A1, F2, A2, F3, A3, nel frame finale della V1 (V1 offset)
3. F0, A0, F1, A1, F2, A2, F3, A3, nel frame di transizione tra V1 e C (V1-C transition)
4. F0 e A0, nel frame iniziale delle consonanti sonore (C onset)
5. F0 e A0, nel frame centrale delle consonanti sonore (C center)
6. F0 e A0, nel frame finale delle consonanti sonore (C offset)
7. F0, A0, F1, A1, F2, A2, F3, A3, nel frame iniziale della V2 (V2 onset)
8. F0, A0, F1, A1, F2, A2, F3, A3, nel frame centrale della V2 (V2 center)

Sono state inoltre effettuate altre misure, non standard, specifiche per la classe di consonanti in esame in questo studio. Queste misure riguardano la formante di nasalizzazione Fn e la sua ampiezza in alcuni frame, come indicato di seguito:

9. Fn e An, nel frame centrale della V1 (V1 center)
10. Fn e An , nel frame finale della V1 (V1 offset)
11. Fn e An, , nel frame di transizione tra V1 e C (V1-C transition)
12. Fn e An nel frame iniziale della V2 (V2 onset)
13. Fn e An, nel frame centrale della V2 (V2 center)

Questi parametri sono stati misurati solo per le vocali [i] e [u] poiché lo spettro delle [a] è caratterizzato alle basse frequenze da molti picchi e ciò rendeva troppo incerta l'individuazione della formante di nasalizzazione.

In figura 4.3 sono descritte schematicamente le posizioni dei frame scelti per la misurazione dei parametri sopra indicati. Sono stati scelti otto frame che sono sembrati di particolare interesse sia per lo studio della stazionarietà dei fonemi sia per quello delle transizioni.

Si noti in particolare come i frame V1 offset, V1-C transition e C onset siano tra loro sovrapposti per metà, per coprire in totale 51.2 ms di segnale (ovvero 512 campioni consecutivi); per l'altra zona di transizione sono pure coperti 51.2 ms ma con un frame in meno in quanto non c'è sovrapposizione tra i due frame di analisi. La scelta di considerare un frame in più nella zona di transizione da V1 a C è dipesa dal fatto che, dopo aver portato a termine la segmentazione e l'analisi temporale, si era notato che quella poteva essere una zona di particolare interesse per lo studio della geminazione.

Vediamo ora i criteri secondo i quali sono stati individuati gli otto frame sopra indicati. Si tenga presente che questi frame sono stati scelti dopo che le pronunce erano già state segmentate.

1. Per quanto riguarda i frame centrali (V1center, C center e V2 center) bisogna dire che non sono esattamente al centro dei fonemi ma piuttosto in punti in cui il fonema è in condizioni di stazionarietà.
2. Per quanto riguarda la prima transizione si è scelto il frame V1-C transition in modo che almeno metà del frame fosse una vocale. Questo perché in questo frame sono stati misurati i parametri tipici delle vocali. Gli altri due frame (V1 offset e C onset) sono stati individuati di conseguenza.
3. Per quanto riguarda invece l'altra transizione, si è scelto il frame C offset come il frame più a destra che presentasse almeno tre quarti di consonante. Il frame V2 onset è stato individuato di conseguenza.

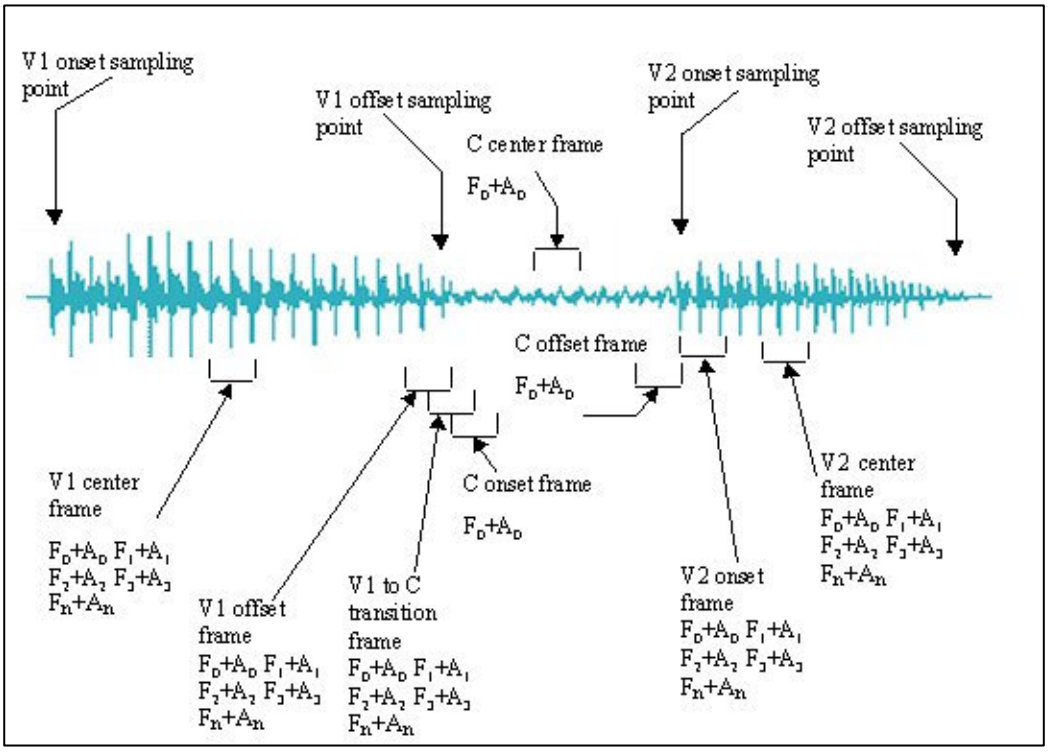


Fig. 4.3 Descrizione grafica dei punti (frame) della pronuncia dove si sono eseguite le misure in frequenza e dei corrispondenti parametri calcolati. Si noti la sovrapposizione tra frame adiacenti nella zona di transizione tra V1 e C.

Per la misurazione dei parametri in frequenza si sono invece adottati i seguenti criteri:

- Per il *calcolo del pitch* si è generalmente fatto uso dell'algoritmo automatico di UNICE. Quando UNICE mostrava discontinuità nell'andamento del pitch nell'arco dell'intera pronuncia, con salti dell'ordine delle decine di Hertz, si è ricorso al metodo più semplice ma sicuro: il calcolo dell'inverso del periodo. L'individuazione di quest'ultimo in pratica è stata sempre

possibile, potendosi calcolare immediatamente come differenza tra gli istanti temporali di due picchi susseguenti in due periodi consecutivi[3].

- *L'ampiezza del periodo* è stata misurata sulla prima armonica degli spettri NB (si ricorda che la risoluzione in frequenza del NB di UNICE è di circa 40 Hz, troppo alta per individuare la F0 ma sufficiente per valutarne l'ampiezza).
- Per il calcolo delle *prime due frequenze formanti* e delle loro ampiezze ci si è serviti contemporaneamente delle informazioni derivanti dallo spettrogramma (in modalità WB) e dallo spettro NB[4]. Il primo era utile per visualizzare con un solo colpo d'occhio l'andamento delle formanti durante tutta la pronuncia (migliore risoluzione temporale), mentre il secondo era indispensabile per calcolare i picchi delle formanti con precisione (migliore risoluzione in frequenza). La F1 e la F2 sono state spessissimo individuate con una probabilità di sbagliare minima. Nei casi di /a/ e /u/ le due formanti sono vicine e questo ha richiesto maggiore attenzione, soprattutto nel caso della /a/. Infatti per certi parlatori (in particolare femminili) nelle pronunce con la vocale [a] e soprattutto nei frame al centro della vocale (dove l'energia è più alta che non nelle transizioni), le due formanti si univano a formare un unico picco molto alto. In questi casi ci si è aiutati con l'andamento delle due formanti nei frame adiacenti dove l'energia minore non ne permetteva la fusione.
- Il *calcolo della F3* è stato più difficoltoso sia perché a quelle frequenze la variabilità è più alta, sia perché, altrettanto spesso, diversi erano i picchi di intensità confrontabile. A parte questo, valgono le stesse considerazioni fatte al capoverso precedente.

Concludiamo questo sottoparagrafo dicendo che le misure complessivamente eseguite in frequenza sono state circa 9500 (44 parametri x 216 pronunce), tutte eseguite manualmente. Per far ciò sono state necessarie diverse settimane di lavoro. Per quanto onerosa, la scelta di misurare manualmente tutte le formanti, senza ricorrere ad algoritmi automatici è stata necessaria. Infatti già in lavori precedenti a questo era stata riscontrata la bassa affidabilità di algoritmi automatici, dovuta al fatto che sono veramente molti i parametri che influenzano la scelta di un picco anziché di un altro come formante, non ultimo l'andamento delle formanti in tutta la pronuncia.

[3] A volte, non è stato possibile trovare in due periodi adiacenti due picchi puliti che lo individuassero in maniera esatta, pur essendo evidente che il periodo terminava. In questi casi, si è calcolata la distanza temporale tra picchi distanti tra loro più di un periodo e poi si è diviso per il numero dei periodi presi in considerazione (una sorta di periodo medio a breve termine tra tre o quattro adiacenti).

[4] Lo spettro LPC si è rivelato invece di grande aiuto all'inizio dell'analisi in frequenza per imparare a discernere tra tutti i picchi del NB (vedere fig. 3.6b) quali erano le formanti vere e proprie. In fase di misura vera e propria però, di solito lo spettro NB dava informazioni più precise (non bisogna dimenticare che la tecnica LPC studia un'approssimazione, nel senso dei minimi quadrati, del segnale).

4.1.3 Scelta e misura dei parametri energetici

I parametri standard scelti per l'analisi energetica sono elencati di seguito:

1) *Energia totale della prima vocale*, EtotV1, data dalla semplice formula

$$E_{totV1} = \sum_{i=t1}^{t2} x^2(i) \tag{4.1}$$

dove x(i) è l'iesimo campione del segnale e t1 e t2 sono gli istanti di V1 onset e di V1 offset.

2) *Potenza media della prima vocale*, PmV1, data da

$$P_{mV1} = \frac{E_{totV1}}{t_2 - t_1} \tag{4.2}$$

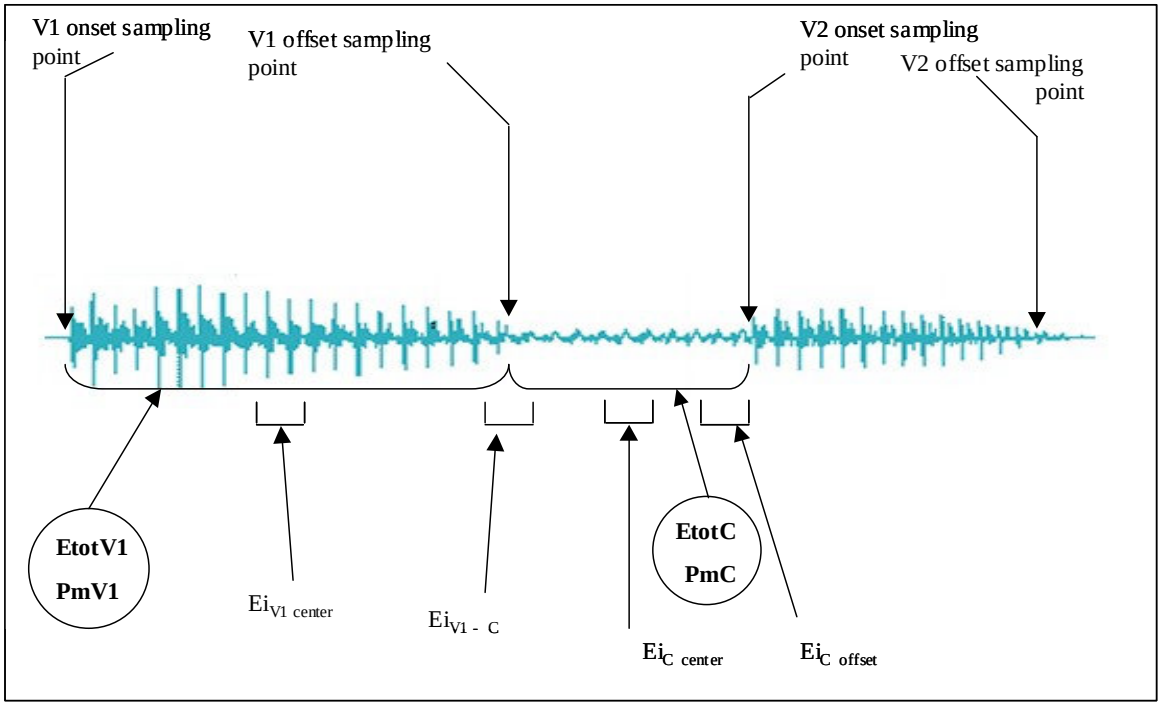


Fig. 4.4 Descrizione grafica dei punti (frame) della pronuncia dove si sono eseguite le misure energetiche e dei corrispondenti parametri calcolati.

3) *Energia totale della consonante*, EtotC, data ancora dalla (4.1) in cui, però, t1 e t2 corrispondono rispettivamente agli istanti C onset e C offset.
4) *Potenza media della consonante*, PmC, calcolata tramite una formula analoga alla (4.2) dove a numeratore figura, ovviamente, l'energia della consonante.
5) Energia istantanea al *centro di V1*, EiV1cent, data dalla (4.1), ma calcolata in una finestra temporale di 256 campioni posizionata al centro[5] di V1.

[5] Diversamente da quanto detto relativamente alla scelta dei frame centrali nell'analisi in frequenza, in questo caso con "centro" si intende proprio che i 256 campioni sono presi a metà del fonema.

6) Energia istantanea alla *transizione V1-C*, EiV1-C; la finestra temporale di 256 campioni è centrata questa volta sul campione corrispondente a V1 offset.
7) Energia istantanea al *centro di C*, EiCcent; la finestra temporale di 256 campioni è centrata nel mezzo di C.
8) Energia istantanea alla *fine di C*, EiCoffset; la finestra di 256 campioni è posizionata in modo che l'ultimo campione della finestra temporale sia quello corrispondente a V2 onset.

Tutti i parametri sono espressi in dB. La figura 4.4 riassume graficamente i punti della pronuncia dove sono stati valutati i parametri energetici.

I parametri energetici, diversamente da quelli frequenza, sono stati calcolati in maniera automatica con il programma *Energie.C*, scritto appositamente in C. La prima versione del programma è stata scritta da Giovanardi, 1998, per l'analisi delle consonanti fricative. Partendo da questa base sono state effettuate diverse modifiche per adattare il programma all'analisi delle consonanti nasali. Il programma *Energie.C* calcola i vari parametri energetici, sia dal domino temporale sia da quello frequenziale, sfruttando nel primo caso le informazioni contenute nei file *.sig* e *.key* (in particolare, la segmentazione tramite campioni) e nel secondo caso quelle contenute nei file *.fft* e *.key* (la segmentazione tramite frame). Il listato completo del codice è riportato in appendice E.

Concludiamo con una nota di terminologia: tutti i parametri sono stati calcolati sfruttando il concetto di energia a breve termine, esposto nel paragrafo 3.2.2; tuttavia, alcuni parametri misurati all'interno di uno o due frame al massimo sono stati chiamati "istantanei", per distinguerli da quelli "totali" riferiti a tutto un fonema[6].

[6] Per chiarire la terminologia adottata, si fa notare che in genere per istantanei (nell'analisi della voce) si intendono parametri misurati in intervalli di tempo di circa 1 ms (cioè molto più piccoli della durata di un fonema), mentre si dà agli altri il nome di parametri a breve termine. Per l'analisi energetica in esame sarebbero tutti parametri a breve termine, ma l'uso della parola istantanea servirà a distinguere meglio i due gruppi.

4.2 L'ANALISI STATISTICA ED I RISULTATI

Questo paragrafo è, sicuramente, il punto nodale di tutta la tesi. Verranno trattati i test statistici utilizzati per l'analisi e verranno esposti i risultati ottenuti. Il test di ANOVA sarà di qui in avanti frequentemente menzionato. A tale proposito premettiamo che quando si parlerà di analisi multivariata si intenderà implicitamente, salvo diversa indicazione, che è stata condotta un'ANOVA considerando come fattori di variabilità il **tipo** (pronuncia singola e geminata), il **sesso** (uomini e donne) la **consonante** ([m] e [n]) e la **vocale** ([a], [i] e [u]).

4.2.1 Elaborazioni statistiche e risultati dell'analisi nel tempo

Scopo dell'analisi temporale è:

- Indagare su quali siano i fattori che influenzano le durate dei fonemi e dell'intera pronuncia.
- Cercare se vi siano relazioni tra le durate dei vari fonemi.
- In particolare, in rapporto alla geminazione, vedere se sia possibile (ed in che misura), dal valore di qualche parametro temporale, distinguere una pronuncia singola da una geminata.

Medie aritmetiche e deviazioni standard

Dopo aver raccolto tutti i dati nel dominio del tempo, per avere le prime indicazioni si sono calcolate le medie aritmetiche e le deviazioni standard dei parametri misurati.

Si sono effettuate medie e deviazioni standard prima solo rispetto alle ripetizioni, poi rispetto a parlatori dello stesso sesso e alle ripetizioni, poi ancora rispetto a tutti i parlatori e alle ripetizioni; infine, statistiche globali rispetto ad una sola vocale o rispetto ad una sola consonante o globali senza alcuna distinzione. In tutte le statistiche appena elencate, si sono sempre elaborate separatamente le due classi delle pronunce singole e geminate. I dati originali con le loro medie e deviazioni standard sono tutti raccolti nell'appendice A, divisi in ventidue tabelle. Anche se non sono riportate tra il testo, esse sono parte integrante e fondamentale della tesi, la quale, come lavoro di analisi, si prefiggeva, tra gli altri, anche lo scopo di prelevare e raccogliere metodicamente dati che possano essere utili in futuro a chi si volesse occupare di argomenti correlati.

Osservando la tabella 4.1 riportata di seguito possiamo fare le prime considerazioni.

La ***media*** della durata di un fonema tra tutte le pronunce è di circa 144 ms, come già anticipato all'inizio del paragrafo 4.1.1, e corrisponde ad un ritmo del parlato di 6.94 fonemi per secondo.

La ***durata di V2*** non sembra essere molto differente tra i due gruppi delle singole e delle geminate; al contrario, le ***durate dei primi due fonemi*** della parola sono alquanto diverse nei due tipi di pronuncia. Inoltre, mentre V1d ha una durata in media maggiore nelle singole, un comportamento opposto si nota per

Cd. Per questo motivo si è ritenuto che il rapporto Cd/V1d potesse avere una certa importanza e si è deciso di calcolarne le medie[7] e di prenderlo in considerazione per ulteriori indagini.

	V1d	Cd	V2d	UTd	Cd/V1d
Singole	183.52	90.64	130.05	404.20	0.51
(StD)	*27.45*	*14.14*	*25.43*	*45.07*	*0.12*
Geminate	124.56	211.75	124.26	460.57	1.78
(StD)	*20.95*	*33.33*	*25.43*	*43.02*	*0.56*

Tab. 4.1 Medie e deviazioni standard (StD) rispetto a tutti i parlatori [6], le ripetizioni [3], le vocali [a, i, u] e le consonanti [m,n] per il gruppo delle singole (108 pronunce) e per quello delle geminate (108 pronunce). Cd/V1 è il rapporto tra le durate di C e di V1; tutte le misure di durata sono in ms.

Analisi della varianza

Naturalmente bisogna ora investigare sulla significatività delle differenze tra le medie appena osservate. Come illustrato nel paragrafo 3.4, un mezzo molto efficace per questo tipo di indagine è l'ANOVA. Riportiamo di seguito i risultati principali dell'indagine compiuta attraverso i test di ANOVA, rimandando per maggiori dettagli all'appendice D.

Dall'analisi della varianza multivariata condotta su **V1d** emerge che la durata della prima vocale dipende in modo statisticamente significativo dal tipo, dal sesso, dalla vocale e dalla consonante. In particolare V1d è maggiore per gli uomini rispetto alle donne, maggiore nelle pronunce con la [n] piuttosto che nelle pronunce con la [m], maggiore nelle pronunce con la [u] piuttosto che in quelle con la [i] e la [a]. Soprattutto, dal nostro punto di vista, è interessante che V1d sia maggiore nelle singole che non nelle geminate, con un alto livello di significatività (p=0,0000).

L'analisi multivariata della varianza per **Cd** mostra che la lunghezza della consonante non dipende significativamente dalla consonante stessa ([m] o [n]). Dipende invece significativamente dal sesso (p=0,0234), dalla vocale (p=0,0003) e soprattutto dal tipo (p=0,0000). In particolare la consonante ha durata maggiore per gli uomini rispetto alle donne, nelle pronunce con la vocale [i] rispetto a quelle con le altre vocali e per le geminate rispetto alle singole. Il divario maggiore tra le medie si nota nel confronto tra le singole e le geminate.

Si è condotta una ANOVA anche su **V2d** ed è risultato che la durata della seconda vocale non dipende in maniera statisticamente significativa né dal sesso né dal tipo (p=0,0666). E' stata trovata invece dipendenza dalla consonante e dalla vocale.

L'ANOVA su **Utd** ha dimostrato che la durata dell'intera pronuncia è influenzata in maniera statisticamente significativa da tutti i fattori presi in considerazione (consonante, sesso, tipo, vocale). Tuttavia, mentre per il sesso e per la consonante la maggiore lunghezza della pronuncia è abbastanza

[7] Si sono calcolati cioè prima tutti i rapporti Cd/V1d e poi su di essi è stata eseguita la media. Invertendo le operazioni si sarebbe trovato il rapporto tra le medie, più semplice, ma che in questo caso sarebbe stato meno significativo.

uniformemente distribuita tra i fonemi questo non accade per il tipo. Infatti, mentre, come già osservato, V1d è maggiore per le singole, Cd è maggiore per le geminate e V2d non dipende dalla geminazione.

C'è quindi un effetto di compensazione che tuttavia non è completo.

L'ANOVA su **Cd/V1d** fornisce risultati che ormai possiamo aspettarci: questo rapporto risulta essere significativamente influenzato solo dalla vocale e dal tipo. Nella tabella 4.2 sono riportate le medie rispetto a tutti i parlatori e alle ripetizioni che possono essere utili per avere, nella maggior parte dei casi, un immediato riscontro dei risultati dei test di anova appena descritti.

	V1d	**(StD)**	**Cd**	**(StD)**	**V2d**	**(StD)**	**Utd**	**(StD)**	**Cd/V1d**	**(StD)**
ama	157.9	*10.2*	86.7	*8.3*	105.9	*16.7*	350.5	*20.4*	0.55	*0.1*
amma	117.8	*15.4*	210.1	*27.9*	106.3	*21.1*	434.1	*41.2*	1.82	*0.4*
ana	201.0	*20.8*	79.6	*12.4*	140.5	*21.5*	421.0	*37.7*	0.40	*0.1*
anna	133.8	*16.6*	201.5	*23.5*	126.8	*30.4*	462.0	*53.3*	1.54	*0.3*
imi	171.7	*23.0*	96.9	*10.7*	120.7	*25.3*	389.4	*31.1*	0.58	*0.1*
immi	118.2	*22.7*	227.1	*32.0*	120.2	*23.6*	465.5	*40.6*	2.01	*0.6*
ini	187.3	*32.6*	88.0	*12.9*	141.8	*21.5*	417.1	*44.0*	0.48	*0.1*
inni	117.8	*21.8*	229.1	*37.6*	132.3	*24.4*	479.2	*30.9*	2.08	*0.9*
umu	182.0	*23.9*	102.7	*16.2*	131.5	*22.1*	416.2	*42.5*	0.58	*0.1*
ummu	131.8	*24.1*	200.6	*36.0*	130.9	*20.4*	463.4	*41.4*	1.59	*0.4*
unu	201.2	*23.1*	89.9	*12.0*	139.9	*25.7*	431.0	*39.4*	0.45	*0.1*
unnu	128.0	*19.5*	202.1	*32.0*	129.1	*25.0*	459.2	*40.7*	1.62	*0.4*

Tab. 4.2 Riepilogo delle misure di durata (in ms): medie (e deviazioni standard) rispetto a tutti i parlatori e a tutte le ripetizioni, eseguite per gruppi appartenenti alla stessa vocale e alla stessa consonante, tenendo separate le singole dalle geminate.

Si è voluto vedere, essendo presenti delle interazioni tra i fattori, se i risultati ottenuti dai test di anova per quanto riguarda la geminazione fossero validi solo in generale o anche prendendo dei subset delle misure. A tal proposito è stata condotta un'analisi monovariata i cui risultati sono riportati nella tabella 4.3.

Dalla tabella si vede come la dipendenza di V1d, Cd e Utd non dipenda dal contesto vocalico-consonantico.

		A				I				U			
		V1d	Cd	V2d	Utd	V1d	Cd	V2d	Utd	V1d	Cd	V2d	Utd
m	F ratio	**84.98**	**324.43**	0.00	**59.67**	**49.34**	**268.26**	0.00	**39.92**	**39.40**	**111.01**	0.01	**11.40**
	p value	**0.00**	**0.00**	0.95	**0.00**	**0.00**	**0.00**	0.95	**0.00**	**0.00**	**0.00**	0.93	**0.00**
		V1d	Cd	V2d	Utd	V1d	Cd	V2d	Utd	V1d	Cd	V2d	Utd
n	F ratio	**14.91**	**377.91**	2.43	**7.10**	**56.54**	**227.17**	1.54	**23.95**	**105.92**	**194.06**	1.64	**4.44**
	p value	**0.00**	**0.00**	0.13	**0.01**	**0.00**	**0.00**	0.22	**0.00**	**0.00**	**0.00**	0.21	**0.04**

Tab. 4.3 Risultati del test di anova condotto su sottogruppi. Il sottogruppi sono dati da tutte le pronunce con la vocale indicata in ascissa e la consonante in ordinata. In grassetto sono indicati i valori statisticamente significativi.

Avendo appurato la significatività di questi parametri per la geminazione vale la pena quantificarne le differenze.

$$\Delta V1d = V1d_{gem} - V1d_{sin} = -58.96ms$$
$$\Delta V1d\% = \frac{\Delta V1d}{V1d_{sin}} = -32.13\% \quad (4.3)$$

$$\Delta Cd = Cd_{gem} - Cd_{sin} = +133.63ms$$
$$\Delta Cd\% = \frac{\Delta Cd}{Cd_{sin}} = +121.12\% \quad (4.4)$$

$$\Delta Utd = Utd_{gem} - Utd_{sin} = +56.36ms$$
$$\Delta Utd\% = \frac{\Delta Utd}{Utd_{sin}} = +13.94\% \quad (4.5)$$

Le correlazioni tra le durate dei fonemi all'interno di una parola

Nell'analisi delle durate dei fonemi condotta nel sottoparagrafo precedente si è trovato che la V1 e la C cambiano in modo inverso le lunghezze medie nel passare da una pronuncia singola ad una geminata, e si sono quantificati i valori dei cambiamenti. E' naturale, quindi, aspettarsi che ci sia un certo grado di correlazione tra queste durate. Sarebbe utile poter misurare questo grado di correlazione e quantificarlo se esiste. A questo scopo, si è condotto il calcolo dello Spearman Rank Correlation Coefficient $\mathbf{r_s}$ tra tutti i parametri di durata. La spiegazione di come venga calcolato e di come vada interpretato questo coefficiente è già stata fornita nel paragrafo 3.4.3.

Per comodità di rappresentazione nelle tabelle riportate nel seguito, si utilizzeranno le matrici di coefficienti $\mathbf{r_s}$ che, ovviamente, avranno le proprietà di essere simmetriche e di avere la diagonale principale unitaria.

Per prima cosa si è eseguito il test di Spearman prendendo in considerazione solo le pronunce singole e poi una seconda volta con solo le pronunce geminate. Il risultato è mostrato nella tabella 4.4.

	V1d sin	Cd sin	V2d sin	V1d gem	Cd gem	V2d gem
V1d sin	*1.00*	*-0.14*	**0.45**			
Cd sin	*-0.14*	*1.00*	*-0.09*			
V2d sin	**0.45**	*-0.09*	*1.00*			
V1d gem				*1.00*	**-0.28**	**0.39**
Cd gem				**-0.28**	*1.00*	*-0.15*
V2d gem				**0.39**	*-0.15*	*1.00*

Tab. 4.4 Matrice di correlazione dei coefficienti r_s. Ogni elemento della matrice rappresenta il coefficiente di correlazione r_s tra la variabile riga e la variabile colonna. Sono presi in considerazione i valori di durata dei fonemi di tutte le pronunce considerando i gruppi singole e geminate separatamente. I valori in grassetto sono quelli statisticamente significativi ($p<0.05$).

Commentiamo brevemente i risultati del test di Spearman.

- Per le geminate esiste, seppure debole, una *correlazione negativa tra durata di C e durata di V1*. Per le singole, invece questa correlazione non è significativa.
- Non è presente alcuna correlazione tra Cd e V2d.
- Esiste invece, sia per le singole sia per le geminate, una discreta *correlazione positiva tra le durate delle due vocali.*

Quest'ultima è probabilmente imputabile alla struttura ritmica del parlato che produce delle compensazioni tra le lunghezza dei fonemi.

Vediamo ora cosa cambia nei coefficienti di correlazione considerando tutte le pronunce (singole e geminate) assieme. Il risultato del calcolo è mostrato in tabella 4.5.

	V1d	Cd	V2d
V1d	*1.00*	**-0.77**	**0.35**
Cd	**-0.77**	*1.00*	*-0.17*
V2d	**0.35**	*-0.17*	*1.00*

Tab. 4.5 Matrice di correlazione (secondo il coefficiente r_s) tra i valori di durata dei fonemi di tutte le pronunce (singole e geminate assieme).

Dai risultati del test di Spearman si vede che :

- Emerge una forte *correlazione negativa tra Cd e V1d*;
- Non c'è alcuna correlazione tra *Cd e V2d*, come nei casi precedenti;
- la *correlazione tra V1d e V2d* resta pressoché invariata rispetto ai due casi precedenti.

Il fatto che una correlazione negativa sia presente solo considerando insieme le pronunce singole e quelle geminate ci permette di affermare che l'allungamento della consonante e l'accorciamento della

vocale sono proprio un effetto della geminazione. Sulla interpretazione di questo risultato torneremo nell'ultimo capitolo. Possiamo però affermare fin d'ora che i risultati delle analisi condotte indicano come primo correlato acustico della geminazione *il rapporto tra le durate della prima vocale e della consonante*.

Classificazione delle pronunce

Come ultimo passo dell'analisi vogliamo vedere se sia possibile classificare efficacemente il tipo delle pronunce sulla base dei parametri di durata che sono risultati significativi per la geminazione.

Abbiamo utilizzato a tal proposito il Maximum Likelihood Criterion, già introdotto nel paragrafo 3.4.4.

I risultati sono illustrati nelle seguenti tabelle

MLC

	Cd			**Cd/V1d**		
Contesto	E.P.P.	Errori	Err. %	E.P.P.	Errori	Err. %
Totale	130	1/216	0.46	0.798	1/216	0.46
Uomini	129	0/108	0.00	0.798	0/108	0.00
Donne	129	0/108	0.92	0.824	2/108	1.85
[a]	122	0/72	0.00	0.733	0/72	0.00
[i]	132	0/72	0.00	0.831	1/72	1.39
[u]	133	1/72	1.39	0.821	1/72	1.39
[m]	133	0/108	0.00	0.842	1/108	0.92
[n]	124	0/108	0.00	0.690	1/108	0.92

Tab. 4.6 Criteri MLC per la classificazione del tipo, condotti sulla base dei parametri di durata Cd e Cd/V1d. E.P.P. rappresenta il punto di equiprobabilità o di separazione delle due gaussiane. Per Cd è espresso in ms mentre è un valore assoluto per Cd/V1d .

Come si vede dalla tabella 4.6, la classificazione sulla base di Cd e di Cd/V1d fornisce ottimi risultati, considerando anche che la capacità di discernimento tra singole e geminate non raggiunge lo zero percento di errori neanche negli ascoltatori madrelingua.

Si nota per Cd/V1d un punto di equiprobabilità di valore pari a 0,8 circa. Si tornerà su questo risultato nell'ultimo capitolo.

MLC						
	V1d			Utd		
Contesto	E.P.P.	Errori	Err.%	E.P.P.	Errori	Err.%
Totale	153	21/216	9.72	432	57/216	26.38
Uomini	155	8/108	7.41	439	36/108	33.33
Donne	149	18/108	16.67	423	21/108	19.44
[a]	151	6/72	8.33	418	19/72	26.39
[i]	147	7/72	9.72	438	12772	16.67
[u]	160	8/72	11.11	442	20/72	27.78
[m]	147	13/108	12.04	420	25/108	23.16
[n]	159	4/108	3.70	447	34/108	31.48

Tab. 4.7 Criteri MLC per la classificazione del tipo, condotti sulla base dei parametri di durata V1d e Utd. E.P.P. rappresenta il punto di equiprobabilità o di separazione delle due gaussiane. E.P.P. è espresso in ms .

Dalla tabella 4.7 si vede chiaramente che, sebbene siano significativamente correlati alla geminazione, i parametri V1d e Utd non consentono una soddisfacente classificazione del tipo. In particolare Utd, risentendo dell'effetto di compensazione tra V1d e Cd, è il parametro meno adatto per la classificazione.

4.2.2 Elaborazioni statistiche e risultati dell'analisi in frequenza

Scopo dell'analisi in frequenza è:

- Indagare su quali siano i fattori che influenzano i parametri frequenziali con particolare attenzione al fenomeno della geminazione.
- Appurare se c'è una dipendenza delle formanti dal punto della pronuncia in cui vengono misurate.
- Indagare sul fenomeno della nasalizzazione.

Medie aritmetiche e deviazioni standard

Le medie e le deviazioni standard dei dati relativi alle misure in frequenza sono state condotte rispetto alle ripetizioni, ai parlatori, al sesso ed infine sulla totalità dei dati. La raccolta di tutti i dati elaborati si trova nell'appendice C. Questa volta la lettura dei dati si presenta più complessa e meno immediata rispetto a quanto accadeva nell'analisi temporale. Passiamo quindi direttamente all'analisi della varianza per vedere se qualche media è significativamente diversa da qualche altra o, in altre parole, se i parametri frequenziali sono influenzati da qualche fattore.

Analisi della varianza

A causa della grande quantità di dati raccolti, l'analisi della varianza per i parametri frequenziali è risultata molto onerosa. In questa sede ricorderemo solo come è stata impostata l'analisi dei dati in frequenza e i risultati principali, rimandando all'appendice D per tutti i dettagli.

Per prima cosa si è condotto un test di ANOVA multivariato per ciascuno dei dieci parametri frequenziali (F0, A0, F1 ,A1, F2, A2, F3, A3, Fn, An) in ciascuno dei frame in cui era stato misurato, per un totale di (2x8+8x5 =) 56 test. Ci occuperemo ora solo dei risultati riguardanti le formanti e le loro ampiezza ripromettendoci di tornare sulla nasalizzazione più avanti.

Il **sesso** influenza in modo statisticamente significativo il pitch e tutte le formanti in tutti i frame analizzati. In particolare il pitch e le formanti sono maggiori per le donne che non per gli uomini come era prevedibile da considerazioni di tipo acustico-fisiologhche.

Per A0 si trova che è significativamente maggiore per le donne (rispetto agli uomini) nella prima vocale mentre lo è per gli uomini nella seconda. A2 è invece significativamente maggiore per le donne in tutti i frame, sia della prima sia della seconda vocale.

Per quanto riguarda la **vocale,** si è trovato che, naturalmente, le formanti dipendono significativamente da essa. Un risultato meno ovvio ma anch'esso teoricamente prevedibile riguarda F0 che è risultata significativamente minore per le [a]. Questo fenomeno, per il quale il pitch dipende dall'altezza delle vocali, è conosciuto col nome di F0 intrinseca.. Per maggiori dettagli si consulti Flammia (1988).

Per le ampiezze si nota che A2 è significativamente maggiore per le [a]. Questo risultato può essere spiegato in relazione all'affiliazione delle formanti alla cavità anteriore o posteriore alla costrizione del tratto vocale.

I parametri che sono influenzati in modo statisticamente significativo dalla **consonante** sono F2 e A3 (in quattro dei cinque frame in cui è stata misurata). Per quanto riguarda F2, essa è sempre maggiore nelle pronunce con la [n]. Anche A3 risulta maggiore nelle pronunce con la [n]. Quest'ultimo effetto è dovuto alla labializzazione delle pronunce con la [m].

Infine vediamo se e come la **geminazione** influenza i parametri frequenziali. Non è stato trovato alcun effetto statisticamente significativo della geminazione sui parametri frequenziali eccetto lievi variazioni del pitch e della prima formante in alcuni frame. In particolare F0 è circa 12 Hz più alta nelle geminate (+8%) nei frame tra la prima vocale e la consonante (V1offset,V1-C transition e Conset). F1 è risultata più alta di 15 Hz nelle geminate (+3%) ma solo nel frame V2 onset. Si noti inoltre che una variazione di 15 Hz di F1 potrebbe essere percettivamente non rilevante, essendo questo valore molto prossimo alla minima variazione percettibile indicata da Kewley-Port e Watson (1994).

Per quanto riguarda la geminazione possiamo quindi concludere, sulla base dei test condotti, che sui parametri frequenzaili non appaiono effetti di particolare importanza per la descrizione di questo fenomeno. Non si ritengono pertanto utili ulteriori indagini tramite test di correlazione e criteri di classificazione come nel caso dell'analisi nel tempo.

Altri test di ANOVA sono stati condotti per determinare se le formanti subiscono variazioni significative al variare del punto della pronuncia in cui sono misurate. Per questo tipo di analisi si è preferito studiare il problema separando uomini e donne. Per ciascuno dei due gruppi si è operata un'analisi della varianza multifattoriale considerando come fattori la **vocale** e la **posizione** ossia il frame in cui era misurato il parametro. Questo tipo di analisi è stato effettuato su tutte le formanti e le loro ampiezze. Un primo risultato di queste analisi è che il pitch è significativamente dipendente dalla posizione, sia per gli uomini che per le donne. In particolare, come si può vedere dai grafici di figura 4.5, si ha un andamento decrescente. Questo è sicuramente dovuto all'intonazione della pronuncia, calante verso la fine. In altre parole, questo effetto non si sarebbe avuto se la pronuncia fosse stata tronca, ossia accentata sull'ultima sillaba. Questo effetto si riflette anche sulle ampiezze A0, A1, A2 e A3 che risultano significativamente dipendenti dalla posizione sia per gli uomini che per le donne. Nella quasi totalità dei casi l'andamento è decrescente anche per le ampiezze. Possiamo quindi dire che si ha un'intonazione calante durante la pronuncia e con energia decrescente. Sottolineiamo che quest'ultima affermazione è valida solo per le due vocali perché si basa sull'analisi delle formanti vocaliche.

Un altro risultato evidenziato da queste analisi è che, per le donne si è trovata dipendenza dalla posizione anche per F1, mentre per gli uomini si è trovata dipendenza dalla posizione per F2. Nei grafici delle figure 4.6 e 4.7 sono illustrati gli andamenti di queste formanti che sono le uniche ad avere differenze significative al variare dei frame.

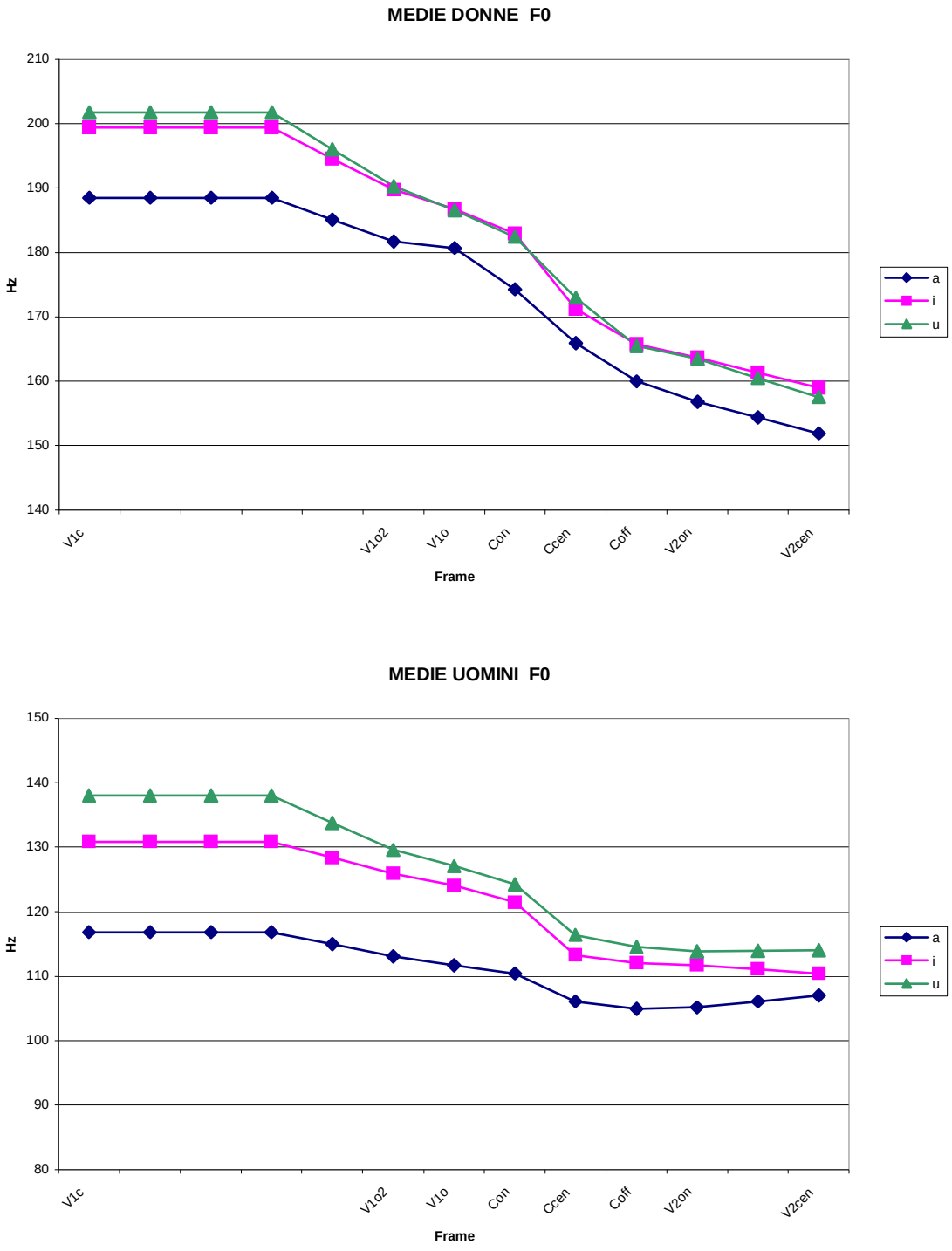
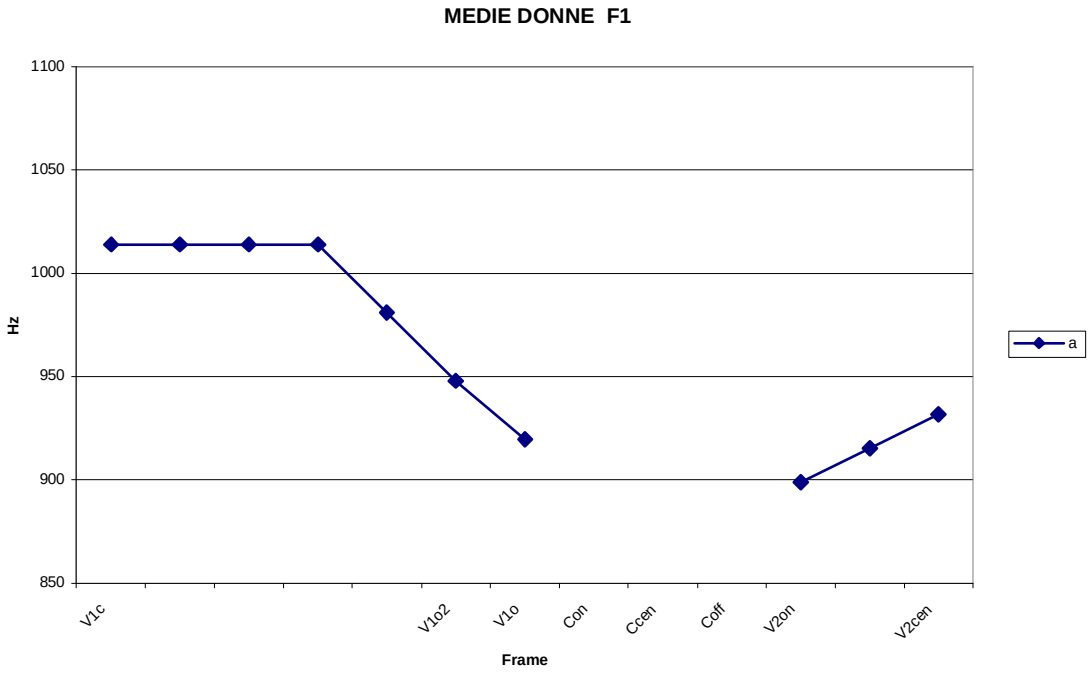


Fig. 4.5 Grafici degli andamenti medi del pitch F0 nelle pronunce dei parlatori maschili e femminili. In ascissa è riportato il frame mentre in ordinata la frequenza in Hz. Notare che il range di frequenze rappresentato in ordinata è diverso nei due grafici ma la sua ampiezza è sempre di 70Hz.



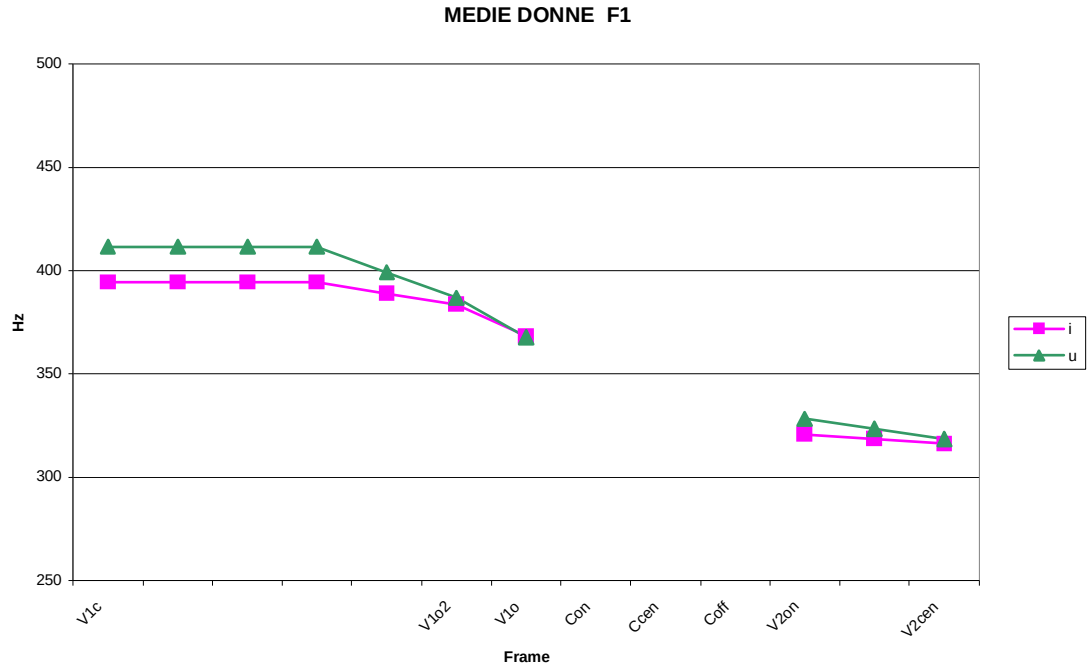


Fig. 4.6 Grafici degli andamenti medi della formante F1 nelle pronunce dei parlatori femminili. In ascissa è riportato il frame mentre in ordinata la frequenza in Hz. Notare che il range di frequenze rappresentato in ordinata è diverso nei due grafici.

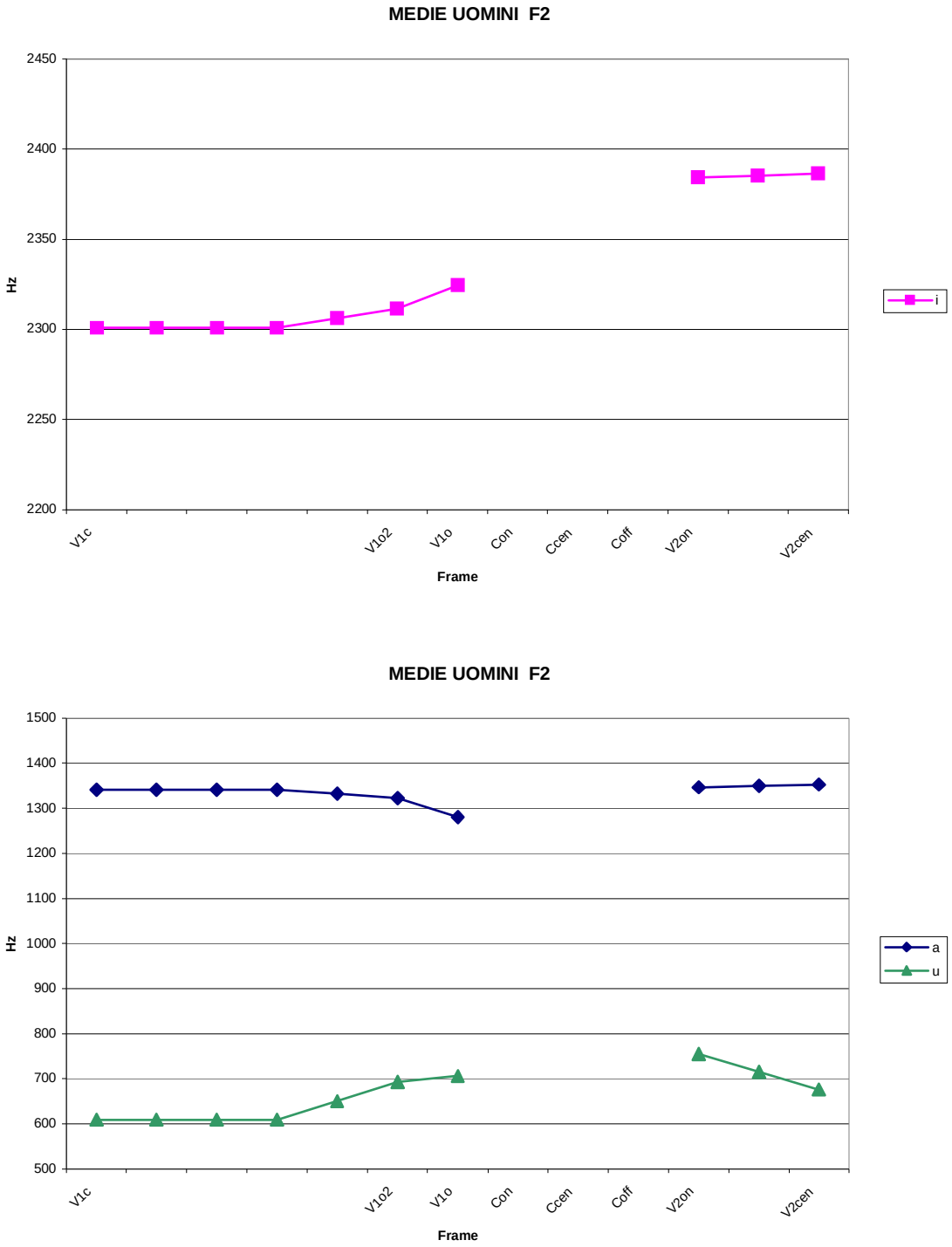


Fig. 4.7 Grafici degli andamenti medi della formante F2 nelle pronunce dei parlatori maschili. In ascissa è riportato il frame mentre in ordinata la frequenza in Hz. Notare che il range di frequenze rappresentato in ordinata è diverso nei due grafici.

Analisi della nasalizzazione

Altri test di Anova sono stati effettuati infine sulla formante di nasalizzazione e sulla sua ampiezza. In particolare, per ogni frame in cui Fn e An sono state misurate, si è condotta un'analisi multifattoriale, considerando, al solito, sesso, tipo, vocale e consonante come fattori di variabilità.

Per quanto riguarda la formante di nasalizzazione, il test di Anova evidenzia che questa è influenzata solo dal **sesso** (in tre dei cinque frame in cui è stata misurata). In particolare si nota come in tutti i frame della prima vocale la formante di nasalizzazione sia più alta per gli uomini mentre per tutti i frame della seconda vocale Fn è maggiore per le donne. Questo risultato è in controtendenza rispetto a quello rilevato per tutte le altre formanti che risultano essere sempre significativamente maggiori per le donne.

Decisamente più interessanti sono i risultati dei test Anova eseguiti sulle ampiezze delle formanti di nasalizzazione . Infatti, ci sono vari fattori che influenzano An. Vediamoli nel dettaglio. Il **sesso** influenza significativamente An in tutti e cinque i frame in cui essa è stata misurata. In particolare per le donne l'ampiezza della formante di nasalizzazione è sempre maggiore. Sempre in tutti e cinque i frame considerati An è influenzata dalla **vocale**. In particolare le An sono maggiori per la vocale u che non per la vocale i. Infine il **tipo** risulta influire significativamente su An in quattro dei cinque frame presi in considerazione. Solo al centro di V1, infatti, il valore p è maggiore di 0,05. In particolare le geminate risultano avere una media sempre maggiore delle singole. Questo ultimo risultato può essere interpretato come un aumento del carry-over di nasalizzazione, quando la pronuncia è geminata. In altre parole l'accoppiamento bocca-naso è maggiore se la pronuncia è geminata.

Il test di Anova per verificare se Fn e An dipendano dalla posizione all'interno della pronuncia in cui sono state misurate ha dato risultati positivi solo per An. In particolare An decresce andando verso la fine della pronuncia. Questo risultato è in linea con quanto detto riguardo alle ampiezze di tutte le altre formanti. Si noti che il valore p non è comunque molto più basso di 0,05.

Le medie di Fn e An sia globali che suddivise a seconda di tipo, sesso, consonante e vocale possono essere trovate nell'appendice C. Da questi dati emerge anche un altro risultato interessante: la percentuale di pronunce sulle quali è stato possibile misurare la formante di nasalizzazione. Si vede innanzi tutto che mentre la prima vocale è nasalizzata solo in meno della metà delle pronunce, la seconda vocale è nasalizzata nel 66/70% delle pronunce. Scendendo più nel dettaglio di cosa succede per la seconda vocale che è quella più nasalizzata, si nota che, come percentuale di pronunce nasalizzate, non c'è grande differenza tra [i] e [u] mentre le geminate sono nasalizzate in circa il 10% in più dei casi rispetto alle singole, le pronunce di parlatori femminili presentano la formante di nasalizzazione nel 10/20% in più dei casi rispetto alle pronunce dei parlatori maschili ed infine, il risultato più evidente: le n sono nasalizzate nel 20/30% in più dei casi rispetto alle m.

4.2.3 Elaborazioni statistiche e risultati dell'analisi energetica

Scopo dell'analisi energetica è:

- Indagare su quali siano i fattori che influenzano i parametri energetici considerati, con particolare attenzione al fenomeno della geminazione.
- Indagare sul fenomeno della nasalizzazione.

Medie aritmetiche e deviazioni standard

Anche le medie e le deviazioni standard sui dati relativi alle misure energetiche sono state condotte rispetto alle ripetizioni, ai parlatori, al sesso ed infine sulla totalità dei dati. La raccolta di tutti i dati elaborati si trova nell'appendice B. Dalla tabella 4.8 si nota che l'unico parametro che sembra variare tra le singole e le geminate, è l'energia totale della consonante. Passiamo quindi all'analisi della varianza per vedere se effettivamente queste medie sono significativamente diverse e se emerge qualche altra dipendenza.

	EtotV1	**PmV1**	**EtotC**	**PmC**	**EiV1cent.**	**EiV1-C**	**EiCcent.**	**EiCoffset**
Singole	96.0	63.4	88.5	58.9	88.0	84.8	82.8	83.1
(StD)	*5.7*	*5.8*	*5.8*	*6.1*	*6.0*	*6.1*	*6.2*	*6.2*
Geminate	95.2	64.4	92.4	59.2	89.2	85.0	82.9	83.2
(StD)	*4.7*	*4.3*	*5.4*	*5.6*	*4.8*	*4.7*	*6.0*	*6.5*

Tab. 4.8 Medie e deviazioni standard (StD) rispetto a tutti i parlatori [6], le ripetizioni [3], le vocali [a, i, u] e le consonanti [m, n] per il gruppo delle singole (108 pronunce) e per quello delle geminate (108 pronunce). Tutte le misure sono in dB.

Analisi della varianza

E' stata condotta un'analisi multifattoriale su tutti i parametri energetici considerando come fattori di variabilità il sesso, la vocale, la consonante e il tipo. Riportiamo di seguito i risultati ottenuti.

Tutti e otto i parametri energetici sono risultati significativamente maggiori per le **donne** che non per gli uomini. Considerando che durante le registrazioni del database si è fatta particolare attenzione a mantenere costante, la distanza tra microfono ed il parlatore, questo denoterebbe una maggiore intensità del parlato femminile. Sicuramente questa affermazione necessita di una indagine più approfondita, essendo molteplici i fattori che possono influenzare l'intensità del parlato

Per quanto riguarda i quattro parametri relativi alla consonante si è trovata una forte dipendenza dal **tipo** dell'energia totale della consonante (p=0,0000). Poiché la potenza media della consonante non

risulta essere influenzata dal tipo si può dedurre che la maggiore energia della consonante è semplicemente un effetto della maggiore durata della consonante stessa quando essa è geminata.

Stessa cosa non accade per la prima vocale. Infatti la sua energia totale non risulta essere statisticamente dipendente dal tipo. Sappiamo invece che V1d dipende in maniera molto forte dal tipo e quindi ci aspettiamo che questa volta sia la potenza media della vocale a dipendere dal tipo. In particolare ci aspettiamo una potenza media maggiore per le geminate, essendo le singole mediamente più lunghe. Questo accade ma in maniera poco evidente (p=0,0406). Accade inoltre che l'energia istantanea al centro della vocale dipenda dalla geminazione anche se in maniera non molto forte (p=0,0169). Tutti questi risultati ci inducono a pensare che vi siano diversi effetti di compensazione e che per quanto riguarda i parametri energetici relativi alla prima vocale non ci sia una reale dipendenza dalla geminazione.

Osserviamo infine, come era logico prevedere, una dipendenza dei parametri energetici relativi alla prima vocale dalla **vocale** pronunciata. In particolare sia l'energia totale che quella istantanea che le energie istantanee sono significativamente maggiori per la [a] che non per le altre vocali.

Test di correlazione di Spearman

E' stato condotto un test di correlazione di Spearman tra tutti i parametri energetici. Si sono considerate prima le singole e le geminate separatamente e poi è stato fatto un terzo test sulla totalità delle pronunce. Non vale la pena riportare i risultati nel dettaglio non essendo emersa alcuna correlazione tra i parametri energetici della consonante e quelli della vocale e non essendoci nessuna differenza tra i risultati forniti dai tre test. Aggiungiamo, per maggiore chiarezza, che, ovviamente, si sono invece trovati valori molto alti tra i parametri "naturalmente" correlati tra loro come, ad esempio, energia e potenza dello stesso fonema.

Classificazione delle pronunce

Avendo trovato che l'energia totale della consonante è l'unico parametro significativamente influenzato dal tipo, tentiamo di classificare le pronunce sulla base di questo parametro.

Abbiamo utilizzato, come per la classificazione operata sulla base dei parametri temporali, il Maximum Likelihood Criterion. I risultati sono illustrati nella tabella 4.9. Risulta evidente che non è possibile una classificazione soddisfacente delle singole e delle geminate sulla base dell'energia della consonante.

MLC			
	Ectot		
Contesto	E.P.P.	Errori	Err.%
Totale	89.90	71/216	32.87
Uomini	87.58	36/108	33.33
Donne	92.43	50/108	46.30
[a]	90.82	29/72	40.28
[i]	90.62	23/72	31.94
[u]	88.50	27/72	37.50
[m]	89.90	36/108	33.33
[n]	89.94	35/108	32.40

Tab. 4.9 Criterio MLC per la classificazione del tipo, condotto sulla base del parametro Ectot. E.P.P. rappresenta il punto di equiprobabilità o di separazione delle due gaussiane. E.P.P. è espresso in dB.

Analisi della nasalizzazione

Vogliamo concludere questo capitolo parlando dei tentativi effettuati per studiare la nasalizzazione anche dal punto di vista energetico. Alla fine delle misure delle formanti di nasalizzazione si è osservato che esse erano comprese nel range tra 429Hz e 585 Hz. Ci si è posti quindi la domanda se fosse stato possibile, andando a calcolare l'energia della seconda vocale (la più nasalizzata), in un range simile, trovare differenze rispetto all'energia calcolata nello stesso range per consonanti non nasali.

Sono stati effettuati vari tentativi, calcolando l'energia totale della vocale nel range tra 400Hz e 600Hz, l'energia del frame centrale della vocale nello stesso range e questi parametri in intervalli leggermente diversi. Si è anche calcolato il rapporto tra l'energia totale e quella nel range tra 400Hz e 600Hz sia per la totalità dei frame della vocale, sia per il frame centrale. Anche in questo caso si sono fatte prove anche con intervalli in frequenza leggermente diversi. Tutte queste prove hanno tuttavia fornito risultati con varianze enormi e questo non ha reso possibile alcun confronto con altre classi consonantiche. Una ragione di questa altissima variabilità dei parametri che si sono calcolati per questa analisi può essere sicuramente individuata nella grande ricchezza spettrale alle basse frequenze delle vocali. La presenza di vari picchi di ampiezze confrontabili alla formante di nasalizzazione che talvolta rientrano e talvolta non rientrano nel range frequenziale scelto hanno quindi reso impossibile questa strada. Del resto questo è stato anche il motivo che ha reso impossibile la misura della formante di nasalizzazione per le [a].

CAPITOLO 5

CONFRONTI E CONCLUSIONI

INTRODUZIONE

Nel capitolo quattro sono stati descritti in maniera dettagliata i risultati dell'analisi acustica condotta sulle consonanti nasali italiane, con particolare attenzione al fenomeno della geminazione. In questo capitolo finale si riprenderanno i risultati più importanti di questo lavoro per poi confrontarli con quelli degli studi svolti per altre classi di consonanti nell'ambito del progetto GEMMA. Inoltre si confronteranno i risultati ottenuti anche con quelli di studi sulla geminazione in lingue diverse dall'Italiano. Infine saranno dati alcuni spunti per ulteriori ricerche

Come allegato a questa tesi, dopo le appendici, è riportato l'articolo, scritto insieme alla prof. Maria Gabriella Di Benedetto, intitolato "Acoustic analysis of singleton and geminate nasals in Italian". Questo articolo illustra i risultati relativi allo studio della geminazione e, al momento della stesura della tesi, è in corso di pubblicazione nel journal "The European Student Journal of Language and Speech" (WEB-SLS).

5.1 RIEPILOGO DEI RISULTATI DELL'ANALISI SULLA GEMINAZIONE DELLE CONSONANTI NASALI

I risultati fondamentali dall'analisi sulla geminazione delle consonanti nasali in Italiano, ampiamente descritta nel capitolo quattro, possono essere riassunti come segue:

- Nell'analisi nel domino del ***tempo***, si è trovato che sia la durata della prima vocale che quella della consonante sono significativamente influenzate dal fenomeno della geminazione. In particolare, mentre la vocale si accorcia nelle geminate, la consonante si allunga. Queste due durate sono dunque legate in maniera inversa (all'aumentare di una decresce l'altra) con un coefficiente di correlazione di Spearman pari a –0,77. Anche la durata dell'intera pronuncia è un parametro che si è rivelato dipendere in maniera statisticamente significativa dalla geminazione. Questa dipendenza è meno forte che non per i precedenti parametri. I risultati appena riepilogati hanno portato all'ipotesi che vi sia un effetto di compensazione che, però non appare completo.
- In ***frequenza***, solo il pitch e la prima formante, calcolati in frame particolari (i frame tra V1 e C per F0 ed il frame V2 onset per F1) variano a seconda della presenza della geminazione. Mentre per F0 le variazioni (circa 12Hz in media) possono essere considerate percettivamente significative, la stessa cosa non si può dire per F1, la cui variazione media di 15Hz è molto vicina al limite minimo percettibile (Kewley-Port e Watson, 1994).
- ***Energeticamente***, non sembra esserci differenza alcuna tra i parametri misurati in presenza o assenza di geminazione tranne che per l'energia totale della consonante, che risulta significativamente maggiore per le geminate. Questo risultato può essere messo in relazione con la maggiore durata delle geminate, anche in considerazione del fatto che la potenza della consonante non risulta differente tra singole e geminate.
- Per quanto riguarda gli effetti della **nasalizzazione** sulla geminazione si è trovato che l'ampiezza della formante di nasalizzazione nella seconda vocale è significativamente maggiore nelle geminate. Questo risultato può essere interpretato in termini di un maggior prolungamento dell'effetto di nasalizzazione quando la pronuncia è geminata.

I risultati della classificazione sulla base dei parametri significativamente dipendenti dalla geminazione verranno ripresi nel prossimo paragrafo in relazione ai risultati ottenuti per altre classi di consonanti finora analizzate nell'ambito del progetto GEMMA. In relazione a questi studi verranno commentati anche i risultati relativi ai test di correlazione e verranno formulate ulteriori ipotesi sul fenomeno della geminazione.

5.2 CONFRONTO TRA GLI EFFETTI DELLA GEMINAZIONE NELLE CONSONANTI NASALI, FRICATIVE, OCCLUSIVE E LIQUIDE.

In uno studio estensivo sul fenomeno della geminazione risulta di fondamentale importanza capire quali effetti della geminazione sono generali e quali legati ad una particolare classe di consonanti.

Faremo riferimento per questi confronti ai lavori sulle consonanti occlusive [p, b, t, d, c, g] (A. Vannucci 1993; R. Rossetti, 1993), sulle consonati liquide [l, r] (F. Argiolas, 1995; F. Macrì 1995) e sulle consonanti fricative [f, v, s, z, ʃ] (M.Giovanardi, 1998).

La prima osservazione da fare è che l'effetto principale della geminazione rilevato nel presente studio, cioè l'allungamento di Cd e l'accorciamento di V1d nelle pronunce geminate, è stato riscontrato anche in tutte le altre classi di consonanti. Le tabelle 5.1 e 5.2 forniscono a questo proposito interessanti informazioni.

	FRICATIVE			LIQUIDE			OCCLUSIVE		
	V1d	**Cd**	**Cd/V1d**	**V1d**	**Cd**	**Cd/V1d**	**V1d**	**Cd**	**Cd/V1d**
Singole	176	135	0.80	171	61	0.37	168	91	0.57
Geminate	127	233	1.97	122	174	1.52	125	182	1.56

Tab. 5.1 Medie dei parametri temporali V1d, Cd e Cd/V1d per le consonanti fricative, liquide ed occlusive per il gruppo delle singole (648 totali) e per quello delle geminate (648 totali). Cd/V1 è il rapporto tra le durate di C e di V1 ed è un numero puro; tutti gli altri parametri sono ms.

	V1d	**Cd**	**Cd/V1d**
Singole	184	91	0.51
Geminate	125	212	1.78

Tab. 5.2 Medie dei parametri temporali V1d, Cd e Cd/V1d per le consonanti nasali [m, n] per il gruppo delle singole (108 pronunce) e per quello delle geminate (108 pronunce). Cd/V1 è il rapporto tra le durate di C e di V1 ed è un numero puro; tutti gli altri parametri sono ms.

Per le nasali, la differenza tra la durata media di V1 nelle singole e nelle geminate è di 59ms (-32% per la geminate), mentre la differenza tra la durata media di C nei due gruppi è di 121ms (+134% per le geminate). Per le fricative queste stesse differenze sono di 49ms (-28% per le geminate) per V1d e di 98 ms (+73% per le geminate) per Cd. Ancora, per quanto riguarda le occlusive troviamo per V1d una differenza di 43ms (-26% per le geminate) e di 92ms per il tempo di occlusione (+101% per le geminate). Infine nelle liquide si è trovata per V1d una differenza di 49ms (-29% per le geminate) e per Cd una differenza di 113ms (+185% per le geminate) tra singole e geminate. Questi risultati evidenziano che nelle nasali l'effetto di accorciamento di V1d è più evidente che per tutte le altre classi di consonanti e lo è in valore assoluto anche per l'allungamento di Cd. Questo fa già prevedere che le nasali possano essere più facilmente classificate rispetto alle altre classi di consonanti. Vediamo nel dettaglio i risultati della classificazione tramite il MLC sulla base di Cd e Cd/V1d per le nasali, le fricative e le occlusive, non disponendo di questo dato per le liquide.

Nelle nasali la classificazione ha portato allo 0,47% di errori sia utilizzando Cd che Cd/V1d, per le fricative si è ottenuto il 12% di errori sia sulla base di Cd sia del rapporto Cd/V1d; infine per le occlusive gli errori sono risultati pari al 4% e all'8% utilizzando Cd e V1d/Cd rispettivamente.

Inoltre il criterio di classificazione ha fissato il punto di equiprobabilità (EPP) ai seguenti valori: nelle nasali 130ms per Cd e 0,80 per Cd/V1d, nelle fricative 182ms per Cd e 1.30 per Cd/V1d e nelle occlusive 128ms per Cd (durata dell'occlusione) e 0.93 per Cd/V1d.

Questi risultati indicano che le nasali e le occlusive si comportano in maniera simile in termini di Cd e Cd/V1d ma in maniera differente rispetto alle fricative. Questa differenza, già osservata da Bertinetto e Vivalda (1978) troverebbe una giustificazione nelle caratteristica [-continua] delle nasali e delle occlusive rispetto alle fricative che invece hanno il TDI [+continua]. Vogliamo aggiungere che la caratteristica [-continua] è attribuita alle nasali italiane, secondo quanto deduce Muljacic (1972), da i dati parziali di Di Pietro (1967). Tuttavia Muljacic stesso non caratterizza le nasali italiane con il TDI di continua come anche D.Brozovic (1967). Al contrario Saltarelli (1970) attribuisce alle nasali il TDI [+continua]. Come si vede, quindi nell'affermazione che il comportamento delle nasali è simile a quello delle occlusive per il loro attributo comune di [+continua], il condizionale è d'obbligo.

Per avere un'idea più immediata della maggiore difficoltà di classificazione ad esempio delle fricative rispetto alle nasali riportiamo di seguito i grafici a dispersione nel piano bidimensionale V1d-Cd per entrambe le classi di consonanti. Si nota che la separazione tra singole e geminate è molto più netta nelle nasali che non nelle fricative.

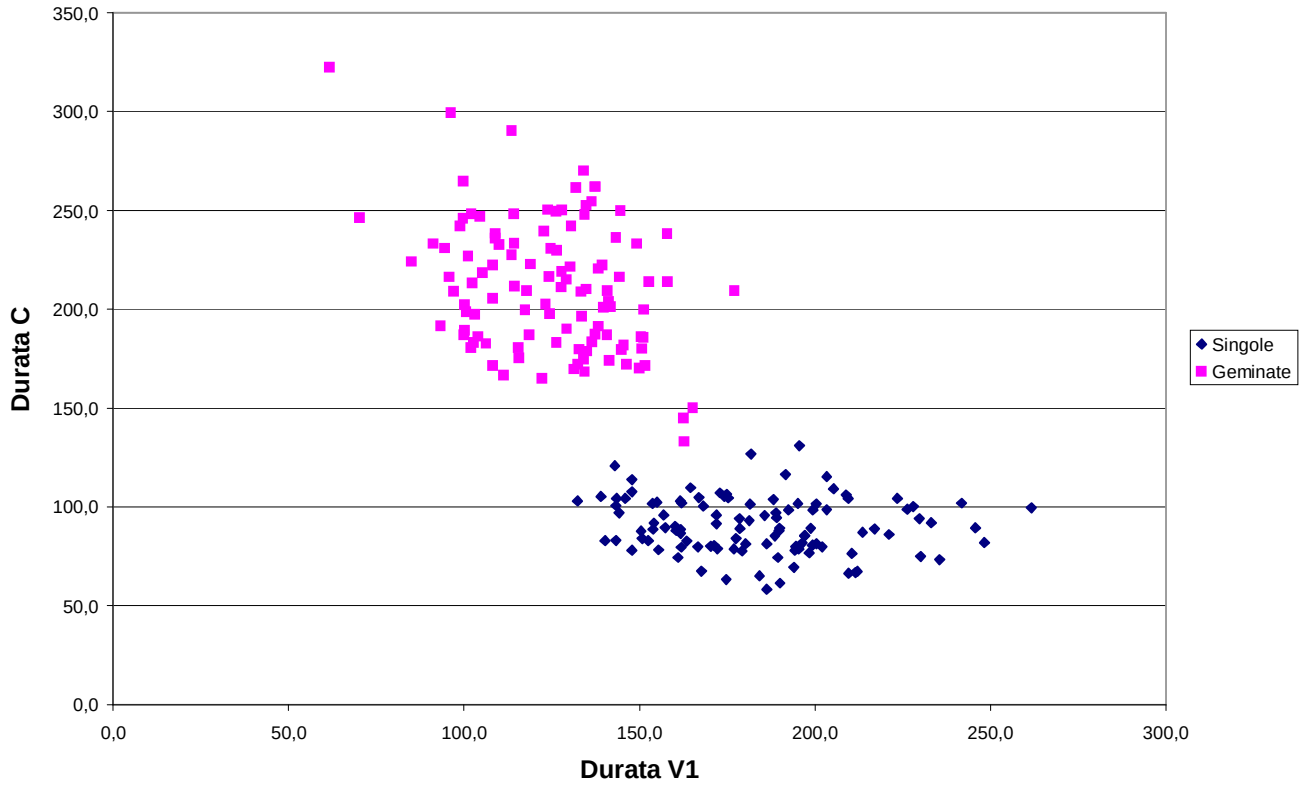


Fig. 5.1 Grafico a dispersione nel piano bidimensionale V1d-Cd per le consonanti nasali (108 singole e 108 geminate). Le durate sono in ms.

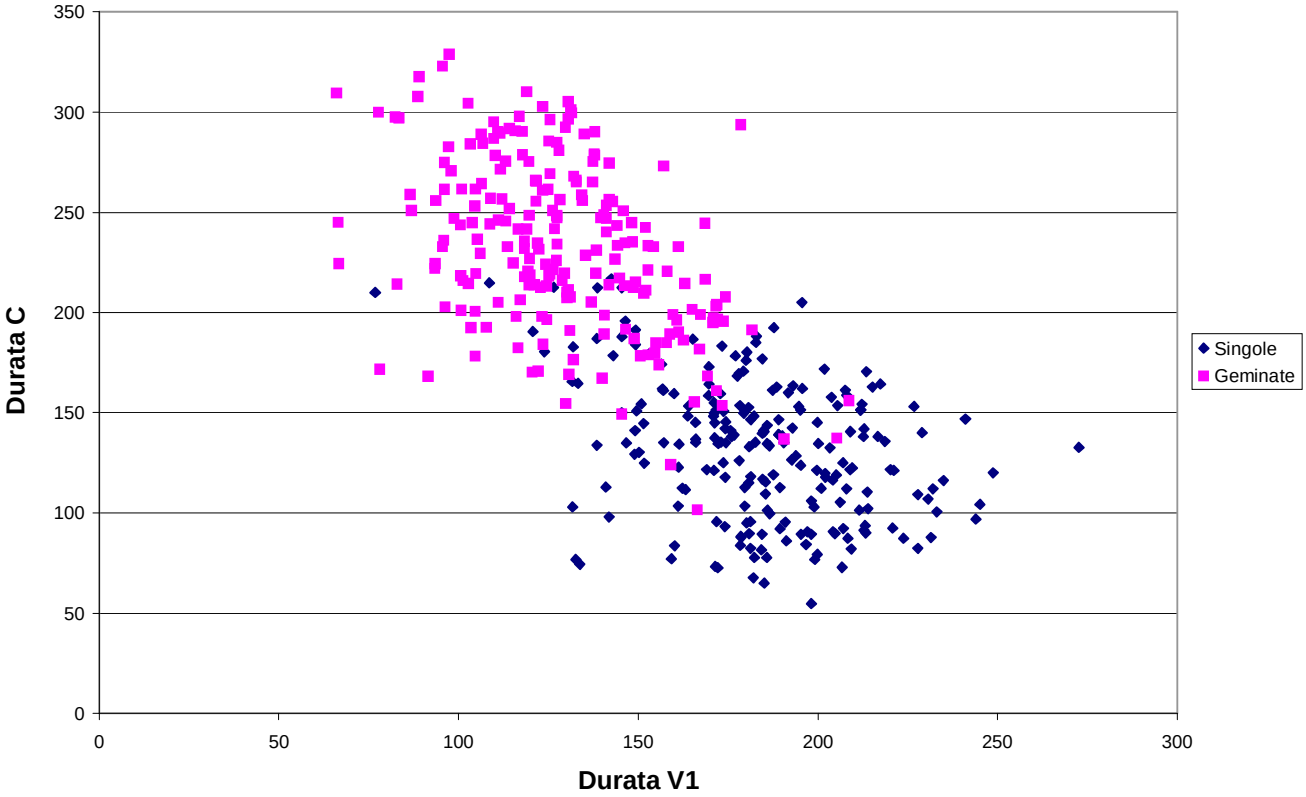


Fig. 5.2 Grafico a dispersione nel piano bidimensionale V1d-Cd per le consonanti fricative (216 singole e 216 geminate). Le durate sono in ms.

Dall'osservazione delle tabelle 5.1 e 5.2 si trae un'altra interessante considerazione: il rapporto medio Cd/V1d è per tutte le classi di consonanti inferiore ad 1 nel caso delle singole e superiore ad 1 nel caso delle geminate.

Questa considerazione porta all'ipotesi che l'intenzione astratta che il parlatore ha nel produrre una geminata si traduca nella realizzazione di un fonema che sia almeno più lungo di quello che lo precede. Per la verifica di questa ipotesi sarebbe necessaria un'analisi percettiva; tuttavia un primo riscontro su quanto si è appena affermato può essere ottenuto da due considerazioni.

La prima riguarda il valore del punto di equiprobabilità trovato nelle nasali per il rapporto Cd/V1d. Questo valore è risultato pari a 0,80 ma è sicuramente non del tutto significativo. Infatti, il bassissimo numero di errori commessi (1/216) può aver spostato in maniera lieve ma casuale il valore dell'EPP. Se proviamo a spostare la frontiera di separazione ad un valore di Cd/V1d=1,00 si trova che il numero di errori cresce di pochissimo (3/216) e quindi si capisce che il valore 0.80 trovato con il MLC non contraddice l'ipotesi che stiamo testando.

La seconda considerazione che vogliamo fare si basa sui test effettuati alla ricerca di un criterio generale di classificazione per tutte le classi di consonanti italiane. A tale proposito si è condotta una classificazione "per tentativi" su 1512 pronunce appartenenti alle sopraindicate classi di consonanti sulla base di Cd e di Cd/V1d. Questo procedimento consiste nello spostare il punto di separazione per la classificazione delle singole e delle geminate finché non si minimizza il numero di errori.

Si è trovato che il minimo di errori commessi nel classificare singole e geminate, si ottiene per un valore di Cd/V1d pari a *1.03*, con una percentuale d'errore pari al 7,2%[1].

Questo numero di separazione così prossimo all'unità avvalora sicuramente l'ipotesi formulata che l'intenzione del parlatore nella produzione di una geminata si traduca nella realizzazione di un fonema che sia almeno più lungo di quello che lo precede.

Aggiungiamo che la classificazione per tentativi sulla base di Cd ha portato ad una percentuale di errori pari al 9,66% con un punto di separazione pari a 136 ms. Quindi sembra che complessivamente il parametro Cd/V1d sia il più indicato per operare una classificazione delle singole e delle geminate. Inoltre, mentre la durata assoluta della consonante è sicuramente influenzata dallo "speaking rate" (velocità di elocuzione), Cd/V1d, essendo un parametro relativo, potrebbe risultarne meno dipendente (Blumstein, 1998).

Concludiamo con un'ultima considerazione. Come ci si poteva attendere da quanto detto finora, osservando le matrici di correlazione, costruite per le occlusive (A. Vannucci, 1993; R. Rossetti, 1993) e le fricative (M.Giovanardi, 1998) nello stesso modo in cui è stata costruita quella per le nasali (tab 4.5), si vede che i risultati sono perfettamente analoghi in tutte e tre le classi di consonanti. In particolare il valore del coefficiente di correlazione tra Cd e V1d oscilla tra –0,71 e –0,78.

5.3 CONFRONTO TRA GLI EFFETTI DELLA GEMINAZIONE NELL'ITALIANO E IN ALTRE LINGUE

Come è stato detto nel secondo capitolo, il fenomeno della geminazione è caratteristico della lingua italiana. Tuttavia la geminazione risulta un argomento di particolare interesse anche per studiosi non italiani. Un motivo di ciò può essere individuato nel fatto che la geminazione è forse l'unico carattere distintivo legato soprattutto ad aspetti prosodici. Sta di fatto che sono molti gli studi condotti sul fenomeno nelle più disparate lingue e dialetti e da molti punti di vista.

Ad esempio citiamo lo studio condotto da Shrotriya et al. (1995), sulle consonanti occlusive dell'Hindi. Anche in questo lavoro è stato trovato un significativo allungamento della consonante nelle pronunce geminate. Citiamo, inoltre, altri studi sulla geminazione in lingue diverse dall'Italiano: quello di Blumstein et al., (1998) e quello di Rochet e Rochet, (1995).

E' inoltre doveroso in questa sede ricordare che si è tenuto recentemente (Agosto 1999) un simposio sulla geminazione nelle lingue presso l'International Conference of Phonetic Sciences a San Francisco.

Gli articoli presentati alla conferenza, si riferivano a tre dialetti indonesiani (Cohn et al., 1999), al Pattani Malay (Abramson, 1999), un linguaggio austronesiano, al Malayalam (Local e Simpson, 1999), un dialetto dravidiano, al Greco cipriota (Arvaniti, 1999) e al Berbero (Louali e Maddieson, 1999).

[1] Si è notato che più del 50% degli errori erano commessi nelle fricative erano sul parlatore GD.

Molti dei risultati presentati negli articoli appena citati sono in accordo con quelli ottenuti per l'Italiano; in particolare si è trovato che, sia per i dialetti indonesiani,sia per il Greco cipriota, la durata è il principale correlato acustico per la classificazione delle pronunce singole e geminate. Lo studio condotto sul Pattani Malay (Abramson, 1999) focalizza la propria attenzione sulle variazioni di F0 in relazione a pronunce che presentano la geminazione della consonante iniziale. Il risultato di questo studio indica che c'è una variazione significativa della F0 in dipendenza della geminazione ma non per tutte le classi di consonanti. In particolare si è trovato che, nelle consonanti nasali, F0 non è influenzata dalla geminazione. Lo studio sul Malayalam (Local e Simpson, 1999) si discosta leggermente dai risultati degli altri studi contraddicendo l'affermazione che la durata è il principale correlato della geminazione. In particolare per il Malayalam sono risultati significativi aspetti legati sia al tempo che alla frequenza. Infine, lo studio sul Berbero si è interessato del problema della classificazione delle occlusive geminate anche quando, in alcuni dialetti, non esistono più le corrispondenti singole che nei secoli sono diventate aspirate. I risultati di questo studio indicano che è appropriato considerare queste consonanti ancora come geminate e che esse sono effettivamente caratterizzate da una durata dell'occlusione superiore a quella delle occlusive singole che ancora esistono in altri dialetti berberi.

5.4 CONCLUSIONI

Alla luce di quanto emerso dal confronto tra gli effetti della geminazione nelle nasali italiane, nelle altre classi consonantiche italiane ed in altre lingue possiamo aggiungere ai risultati già riepilogati nel paragrafo 5.1 quanto segue:

- La classificazione delle nasali, sia sulla base della durata della consonante, sia sulla base del rapporto Cd/V1d, risulta più semplice che non per le altre classi di consonanti.
- Anche per le nasali continua a valere il valore distintivo del rapporto Cd/V1d molto prossimo all'unità.
- La dipendenza delle geminazione da parametri di durata è ricorrente in tutte le lingue citate nel paragrafo precedente.

Riguardo alla nasalizzazione questo studio ha evidenziato che una discreta percentuale (66/70%) delle vocali che seguono la consonante presentano una formante di nasalizzazione tra i 429 Hz e i 585 Hz.

Desidero, per concludere, ringraziare in modo particolare la professoressa Di Benedetto per il concreto aiuto nella risoluzione dei problemi incontrati durante questo lavoro, nonché per la preziosa collaborazione nella stesura dell'articolo "Acoustic analysis of singleton and geminate nasals in Italian", in corso di pubblicazione nel journal "The European Student Journal of Language and Speech".

5.5 SPUNTI PER RICERCHE FUTURE

Eventuali ricerche future potrebbero orientarsi secondo i seguenti punti:

- Condurre sulle nasali un esperimento percettivo per approfondire l'indagine sul valore ottimo del rapporto Cd/Vd1 che discrimina le singole dalle geminate.
- Analizzare come varia la funzione distintiva del rapporto Cd/V1d sulla geminazione in funzione dello *speaking rate*.
- Analizzare le restanti consonanti della base dati (in particolare le affricate) e vedere se continua a valere anche per quelle il valore distintivo del rapporto Cd/V1d molto prossimo all'unità.
- Analizzare le correlazioni temporali tra fonemi anche ad un livello di astrazione più alto, come in parole intere più lunghe dei bisillabi (come *affannato*, *rattoppato*, ecc.) o addirittura all'interno di intere frasi (come *sono andato al largo con la barca*, ecc.).
- Testare la rilevanza della nasalizzazione delle vocali nella lingua italiana con un esperimento percettivo.
- Infine, a livello di progetti ancora più ampi, si potrebbe cercare di sfruttare tutti i dati raccolti nell'ambito del progetto GEMMA per progettare e implementare un sistema di riconoscimento o ancora sfruttare gli stessi dati per implementare un sintetizzatore vocale per l'Italiano per scopi generali.

5.6 CONSIDERAZIONI FONOLOGICHE SULLA GEMINAZIONE

Pur non rientrando negli obiettivi di questa tesi, come molti tra coloro che hanno lavorato prima di me al progetto GEMMA, vorrei esprimere il mio parere riguardo alla disputa filosofico-fonologica sul concetto di geminazione. In realtà, questo paragrafo non nasce solo per omogeneità ai precedenti lavori ma soprattutto dal fatto che, dopo aver lavorato per mesi sul problema della geminazione, la presa di posizione sull'argomento nasce abbastanza spontanea.

Fatta questa premessa riassumiamo brevemente la disputa sulla geminazione alla quale partecipano da molti anni fonetisti e altri studiosi. Esistono infatti due modi contrastanti di interpretare la geminazione:

1. come una realizzazione alternativa al fonema consonantico scempio, articolato ora con aumentata durata e intensità. I *monofonematisti*, sostenitori di questa interpretazione, sostengono quindi l'esistenza di 15 fonemi lunghi, rafforzati o tesi;
2. come la ripetizione di uno stesso fonema consonantico, ossia come due occorrenze successive dello stesso fonema, realizzate eventualmente con diversi allofoni. I sostenitori di questa teoria sono invece i *bifonematisti*.

A favore della prima ipotesi si possono citare Esposito e Di Benedetto (1999); a favore della seconda Muljacic (1972), mentre Macrì (1995), ad esempio, prende una posizione neutrale considerando la [l:] come una ripetizione dello stesso fonema e la [r:] come un fonema lungo autonomo. Giovanardi (1998), infine, sostiene che entrambe le teorie siano valide secondo gli scopi.

Il mio parere è a favore della teoria monofonematista dato che, come sostiene lo stesso Giovanardi, le medesime caratteristiche spettrali ed energetiche possono essere trovate nelle due versioni singola e geminata che quindi possono distinguersi solo per la loro lunghezza temporale. Inoltre, anche dalla forma d'onda nel tempo non si è mai notata una minima differenza tra parte iniziale e parte finale della consonante. Per questi motivi, credo che il TDI *lungo* possa essere adeguato a rappresentare l'opposizione fonologica tra [m] e [m:] e tra [n] e [n:]. Il fatto poi, di considerare, ad esempio, le consonanti occlusive geminate come composte da due fonemi "uguali", di cui il primo senza la fase di esplosione e il secondo senza la fase di implosione, sembra già una contraddizione in termini.

BIBLIOGRAFIA

Arthur S. Abramson (1999) “Fundamental frequency as a cue to word-initial consonant length: Pattani Malay” ICPhS99 San Francisco pp 591-594

Francesca. Argiolas, “Analisi acustica e percettiva delle consonanti liquide [l, r] in italiano”, Tesi Univ. di Roma “La Sapienza”, 1995.

Francesca Argiolas, Federico Macrì, M.G. Di Benedetto "Acoustic analysis of Italian [r] and [l]," *Journal of the Acoustical Society of America* 97, no. 5, pt.2, pp.3418, 1995.

Amalia Arvaniti (1999) “Effects of speaking rate on the timing of single and geminate sonorants” ICPhS99 San Francisco pp 599-602

M. Bertinetto, E. Vivalda, "Recherches sur les oppositions des quantité en Italien", Journal of Italian Linguistics, No. 3, 1991, pagg. 97-119.

Blumstein S.E., Pickett E., Burton M., “Effects of speaking rate on Singleton/Geminate consonant contrast in Italian”, unpublished manuscript, 1998.

Brozovic D. “Sull’inventario dei fonemi serbocroati e i loro tratti distintivi” in “WSI”,XII (1967), pp161.172. (Contiene anche una matrice binaria dei fonemi italiani).

L. Canepari, "Introduzione alla fonetica", Einaudi, 1979.

L. Canepari, “Manuale di pronuncia italiana”, Ed. Zanichelli, 1992.

Abigail C. Cohn, William H. Ham, Robert J. Podesva (1999) “The phonetic realization of singleton-geminate contrasts in three languages of Indonesia” ICPhS99 San Francisco pp 587-590

Giuseppe Cicchitelli “Probabilità e statistica” Maggioli editore, 1984.

Clifford, "Microphones (3rd edition)", Tab books inc., 1986.

W. R. Dillon, M. Goldstein, "Multivariate analysis", J. Wiley & Sons, 1984.

Di Pietro R.J. “Phonemics, Generative grammar and the italian sibilants” in “SL”, XXI (1967) pp. 96,106.

Esposito A., Di Benedetto M.G., “Acoustic and Perceptual Study of Gemination in Italian Stops”, Journal of the Acoustical Society of America, 1999.

G. Fant, “Acoustic theory of speech production”, Mounton and Company, Gravenhage, 1960.

Giovanni Flammia ”Classificazione statistica e neurale su base percettiva del riconoscimento delle vocali italiane.” Tesi Univ. di Roma “La Sapienza (1988).

J. L. Flanagan, A. Rosemberg, “Effecct of glottal pulse shape on the quality of natural vowels”, J.A.S.A. 53, 1971.

J. Flanagan, L. R. Rabiner, “Speech synthesis”, Stroudsburg, 1973.

Luisa Franchina, Piero Marietti, “Sistemi elettronici a banda frazionale stretta”, Masson Editore, 1994 (pag 239).

Fujimura O. e Lindqvist G (1971) “Sweep-tone measurements of vocal tract characteristics” *Journal of the Acoustical Society of America.* Vol 49,No. 2, pp 541-558

Giovanardi M. (1998).”Analisi Acustica e Sintesi delle consonanti fricative singole e geminate in Italiano” Tesi Univ. di Roma “La Sapienza.

Giovanardi M. (1998).” Acoustic analysis of singleton and geminate fricatives in Italian” European student journal of language and speech

Al Kelley Ira Pohl Didattica e programmazione C Addison-Wesley 1996

B. W. Kernighan e D. M. Ritchie, “Linguaggio C”, Jackson, 1990.

Kewley-Port D. and Watson C.S. (1994) "Formant-frequency discrimination for isolated English vowels", *Journal of the Acoustical Society of America.* Vol 95,No. 1, pp 485-496.

John Local and Adrian P. Simpson (1999) “Phonetic implementation of geminates in Malayalam nouns” ICPhS99 San Francisco pp 592-595

Naima Louali and Ian Maddieson (1999) “Phonological contrast and phonetic realization: the case of Berber stops” ICPhS99 San Francisco pp 603-606

Pierpaolo Luzzato Fegiz Appunti di Statistica Metodoligica Kappa Librerie editrice 1965-66

F. Macrì, “Raddoppiamento nelle liquide [l], [r]: acustica e percezione”, Tesi Univ. di Roma “La Sapienza”, 1995.

Shinji Maeda “Acoustics of vowel nasalization and articoulatory shifts in french nasal vowels” pubblicato in “Phonetic and Phonology” Volune 5 “Nasals,Nasalization, and the Velum” ACADEMIC PRESS INC. 1993

B. Malmberg, “Manuale di fonetica generale”, Ed. Il Mulino, Bologna, 1977.

P. Mandarini “Comunicazioni elettriche”, Ed. Ingegneria 2000, 1990.

Marchese Angelo, (1979) “Pratiche comunicative”, Principato Editore.

Z. Muljacic, “Fonologia della lingua italiana”, Ed. Il Mulino, Bologna, 1972.

V. Oppenheim, R. W. Schafer, “Digital signal processing”, Prentice Hall, 1975.

OROS AU21 CARD, manuali di riferimento”, OROS, 1991.

Athanasios Papoulis, “Probabilità, variabili aleatorie e Processi stocastici”, 1973, Boringhieri editore.

R. Rabiner, R. W. Schafer, “Digital processing of speech signals”, Prentice Hall, 1978.

Rochet,L.B., and Rochet,A.P. (1995) “The perception of the single-geminate consonant contrast by native speakers of Italian and Angliphones” in proceedings of ICPhS95, edited by K.Elenius and P.Branderud,Vol 3 (Arne Strombergs, Stockholm) pp. 616-619.

R. Rossetti “Gemination of Italian stops”, Journal of the Acoustical Society of America, 95, 2pSP25, pp.2874, 1994.

R. Rossetti, “Caratteristiche acustiche del fenomeno di geminazione nelle consonanti occlusive Italiane: applicazione all’adattamento automatico di pronunce straniere”, Tesi Università di Roma la Sapienza, 1993.

Saltarelli M. “A phonology of italian generative grammar” The Hague-Paris,1970

Shrotriya N., Siva Sarma A.S., Verma R., Agrawal S.S. “Acoustic and perceptual characteristics of geminate Hindi stop consonants”, in Proceedings of ICPhS95, edited by K.Elenius and P.Branderud, 4, (Arne Strombergs Grafiska Stockolm), pp.132-135, 1995.

M. Spiegel, “Statistica”, sec. Ed., McGraw-Hill, 1988.

Statgraphics Plus User Manual - Statistical graphics corp.1996.

Kenneth N. Stevens, Gunnar Fant, Sarah Hawkins “Some acoustical and perceptual correlates of nasal vowels”, 1987.

Kenneth N. Stevens Acoustic phonetics (1998) the MIT press

H. W.Strube, “Determination of the istant of glottal closure from the speech wav.”; J.A.S.A., vol. 56, n. 5, November 1974.

Turbo Pascal 6.0 Manuale di riferimento”, Borland, 1992.

A. Vannucci, “Correlati acustici di tratti distintivi: applicazione alla caratterizzazione del punto di articolazione delle consonanti occlusive dell’italiano e loro riconoscimento automatico”, Tesi Univ. di Roma “La Sapienza”, 1993.

Vecsys (1989). The Unice User Manual (Vecsys - Chemin du Chene rond - 91570 Bièvres, France).

T.H. Wonnacott- R.J Wonnacott (1972) Introduzione alla statistica Franco Angeli editore.