CAPITOLO 2

ACUSTICA DEL SEGNALE VOCALE E SINTETIZZATORE HLSYN

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 sono dati cenni di fisica e nel terzo paragrafo acustica è descritto il sintetizzatore HLSyn e la generazione sonora in relazione agli apparati fonatori descritti.

2.1 CENNI DI FISIOLOGIA

2.1.1 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 all'asse della rocca petrosa (l'osso temporale in cui è scavato il labirinto), l'altro, quello posteriore, di 18 millimetri parallelamente a quest'ultima; il terzo canale, quello esterno, di 12 millimetri, è orizzontale.

La chiocciola consiste di tre sezioni: un nucleo, detto **columnella** 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**.

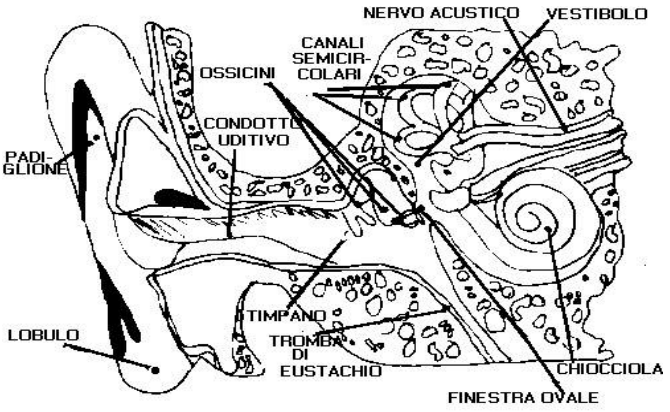


Fig. 2.1 Schematizzazione dell'organo dell'udito.

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.

2.1.2 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.

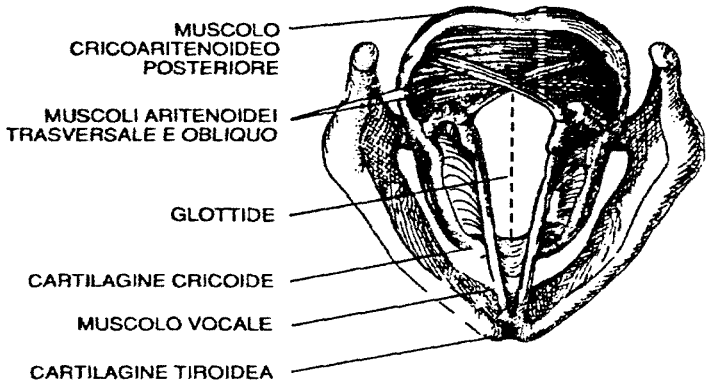


Fig. 2.2 Sezione longitudinale della laringe.

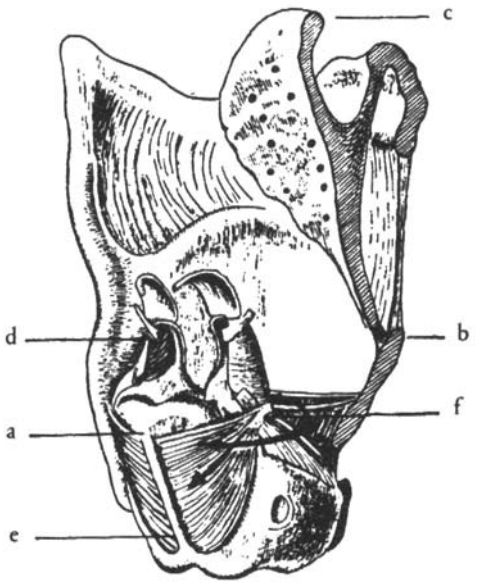


Fig. 2.3 La laringe vista da dietro. a: cartilagine cricoidea; b: cartilagine tiroidea; c: epiglottide; d: aritenoide (sinistra); e, f: muscoli (le frecce indicano le direzioni di movimento).

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.

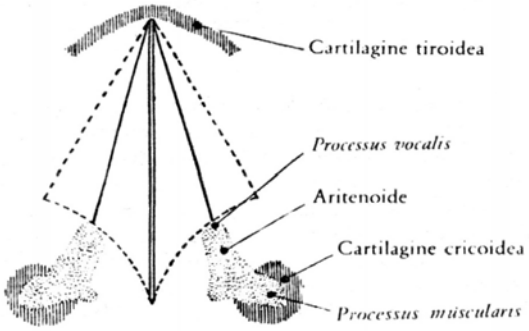


Fig. 2.4. Disegno schematico del meccanismo di apertura e chiusura della glottide. Le due linee più grosse indicano la posizione delle corde vocali durante la respirazione normale, le linee tratteggiate più grosse la posizione durante la respirazione profonda. Le due linee verticali sottili indicano la posizione di fonazione. Le linee tratteggiate sottili indicano la direzione del movimento delle aritenoidi quando la glottide cambia forma. (Da I.Tarneaud).

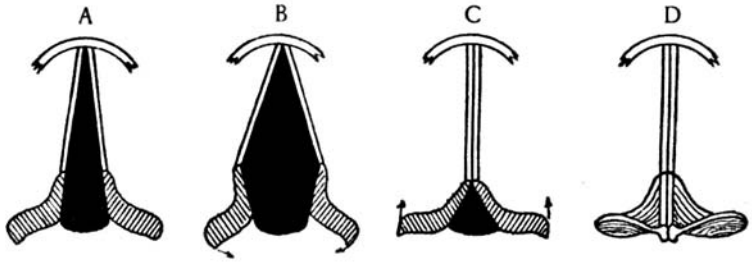


Fig. 2.5. Disegno schematico delle differenti posizioni della glottide: A respirazione normale, B respirazione profonda, C bisbiglio (le corde vocali sono chiuse ma il passaggio fra le aritenoidi resta libero), D fonazione. (Da J. Forchhammer).

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:

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

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 sollevando 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 e 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

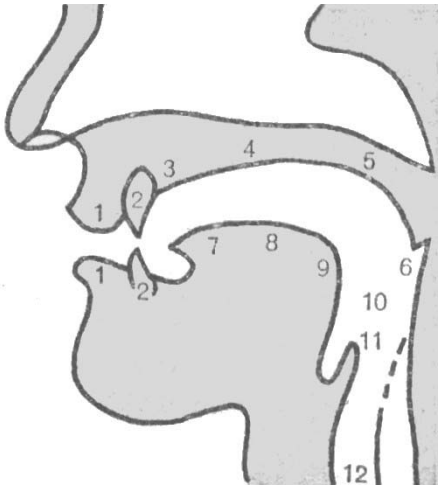


Fig. 1.6 .

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.

2.2 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[1], 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

[1] 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.

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.

2.2.1 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)[2]. 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.

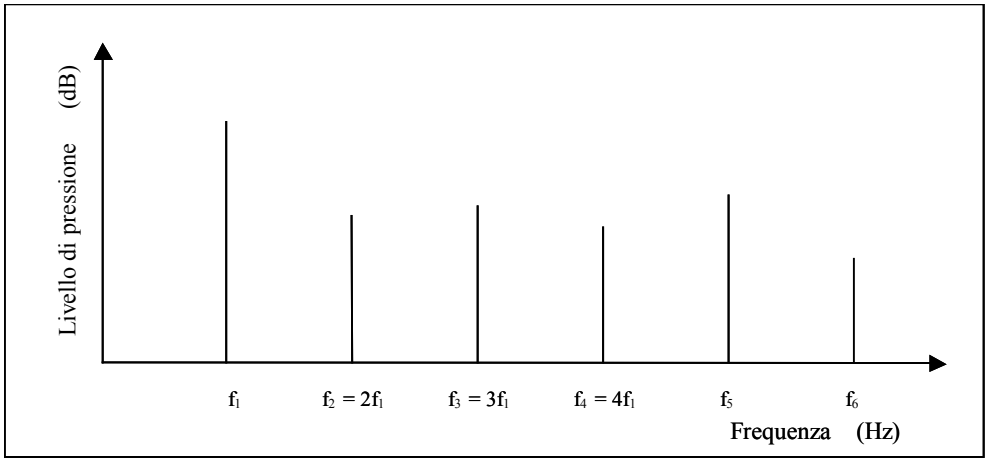


Fig. 2.1 Spettro acustico di un suono complesso.

[2] 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.

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.[3]

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).

2.2.2 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

[3] 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.])

posizione della costrizione o dell'improvvisa apertura del condotto).

Come esempio di suoni sordi riportiamo le consonanti [p t k f s ⏝] 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[4] 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.

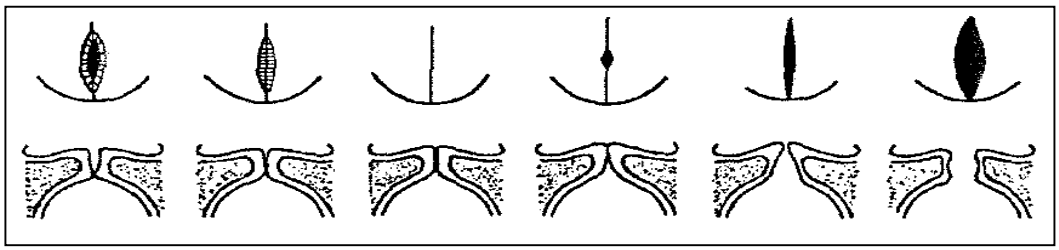


Fig. 2.2 Rappresentazione schematica dello stato di affrontamento delle corde vocali. Parte superiore: sezione longitudinale delle corde vocali (la mancanza del contatto è evidenziata in colore nero); parte inferiore: sezione trasversale.

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[5]. A questo punto l'aria

[4]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.

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

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[6].

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] di *bene, due, gara ,vetta ,usi*; inoltre in italiano sono sempre sonore [m, n, r, l] come in *mese, anfora, notte, àncora, rosa, lupo*. Le vocali sono tutte suoni sonori.

2.2.3 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_0) 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 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

[6]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.

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*[7]. Data la natura periodica, il suo spettro è a righe, le cui 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).

[7]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.

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

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.

2.2.4 Frequenze Formanti

I suoni sonori sono caratterizzati, oltre che dalla F_0 anche dalle frequenze formanti. Vediamo, come abbiamo fatto nel precedente paragrafo per la F_0 , 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[8]. 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

(1.2)

[8]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.

frequenze di risonanza v delle onde stazionarie che vi si generano assumono i valori dati dall'espressione:

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

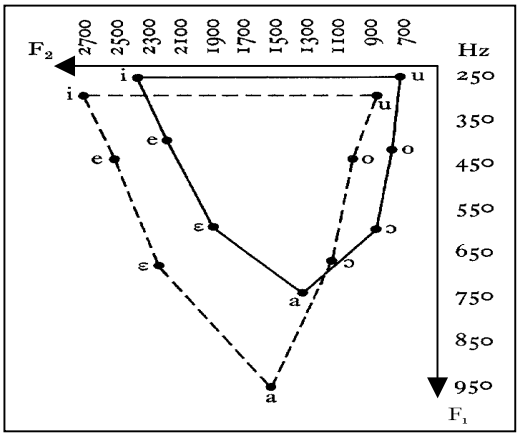


Fig. 2.3 Medie delle prime due formanti dei sette vocoidi tonici italiani: voci maschili (linea continua) e femminili (linea tratteggiata) sovrapposte. (Canepari, 1979).

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.

2.2.5 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[9] 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.

2.2.6 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:

[9] 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.

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

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.

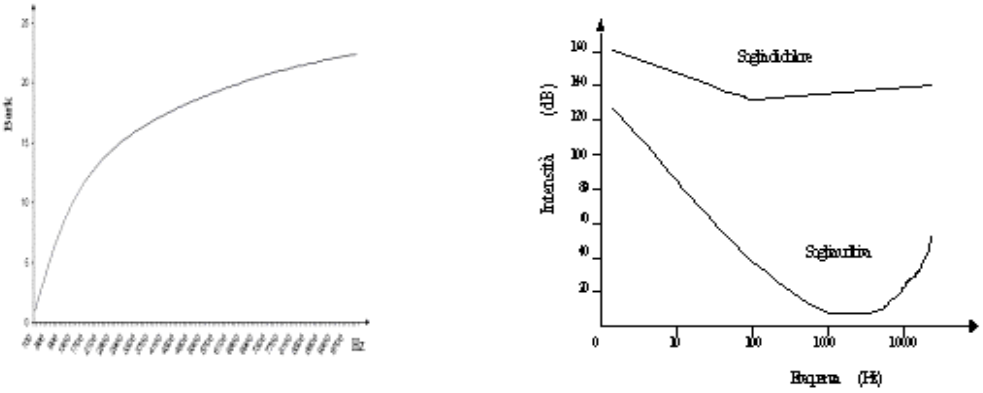


Fig. 2.4 a) Conversione di scala Hz/Bark b)Il campo di sensibilità dell'orecchio umano alle vibrazioni.

2.3 IL SINTETIZZATORE HLSYN

In questo paragrafo verranno descritte le principali caratteristiche e funzionalità del sintetizzatore articolatorio HLsyn. Si daranno soltanto le informazioni necessarie a comprendere il lavoro svolto per ovvi motivi di spazio, rimandando al manuale per una descrizione più approfondita e completa del del sintetizzatore.

2.3.1 Caratteristiche generali e parametri di controllo

Il sintetizzatore articolatorio HLsyn si basa sul precedente sintetizzatore per formanti KLsyn (Scarlino, 1993). In pratica si può dire che il sintetizzatore HL utilizza il precedente KL tramite delle relazioni matematiche che convertono i valori dei parametri impostati nell'HL nei valori del KL. Tale approccio è basato sull'osservazione che esistono dei legami e dei vincoli tra gli oltre quaranta parametri di controllo (formanti, loro ampiezze e larghezze di banda, ampiezze delle eccitazioni fricative e sonore ecc.), del sintetizzatore KLsyn. Questi vincoli esistono perché il processo fisico della produzione del parlato impone dei limiti sulle combinazioni dei parametri di sintesi che ci possono essere in ogni particolare istante della fonazione e in come questi parametri possono variare nel tempo. In accordo a questi limiti, è stato proposto un insieme di 10 (poi ampliato a 13) parametri ad un più alto livello (HL, higer level) di quelli del sintetizzatore per formanti KL. Questi parametri HL sono legati più direttamente allo stato e ai movimenti del tratto vocale di quanto non lo fossero i parametri del KLsyn. Un insieme di relazioni, implementate nell'HLsyn, trasforma i parametri HL in parametri KL che si occupano di controllare il sintetizzatore KLsyn88. Oltre a questi 13 parametri che possono essere variati a proprio piacimento (sempre entro i limiti previsti) durante la pronuncia, ce ne sono altri 24 che possono essere impostati dall'utente ma che restano costanti per tutta la durata della pronuncia sintetizzata e alcune altre decine invisibili all'utente e che non possono essere modificate.

Analizziamo ora quali sono i parametri di controllo e come essi sono legati alle caratteristiche che l'apparato vocale assume durante la fonazione. In Tabella 5.1 sono illustrati i parametri di controllo con una loro breve descrizione mentre in Figura 5.3 si può vedere come essi agiscono sulle caratteristiche dell'apparato fonatorio umano.

I primi cinque parametri del sintetizzatore HLsyn sono molto simili (e in alcuni casi uguali) ai parametri del KLsyn. Questi sono la frequenza fondamentale **f0** e le quattro frequenze formanti **f1**, **f2**, **f3** e **f4** che specificano le frequenze naturali del tratto vocale assumendo che non ci siano accoppiamenti acustici con la trachea o con la cavità nasale e che non ci siano costrizioni localizzate causate dalla punta della lingua e dalle labbra. Le frequenze formanti specificano come la forma del tratto vocale cambia durante la produzione del parlato (si pensi, ad esempio, alle differenti forme che assume la bocca pronunciando una [a] o una [u] e a come si ripercuotono sulla posizione ed ampiezza delle formanti). Se ci sono accoppiamenti con la trachea o con il naso o se c'è una costrizione localizzata (come specificato dai parametri **an**, **ag**,

al e **ab**) le relazioni di mappatura modificano i parametri del sintetizzatore KLsyn. I parametri **f1**, **f2**, **f3** e **f4** descrivono gli aspetti del tratto vocale che sono determinati dalla posizione del corpo della lingua, dalla posizione della mascella, dalla forma della faringe e dall'eventuale arrotondamento delle labbra.

Parametro	Descrizione
f1, f2, f3, f4	Prime quattro frequenze naturali del tratto vocale. Queste sono le frequenze naturali quando la faringe è chiusa, non c'è accoppiamento acustico con la trachea e non ci sono occlusioni, anche parziali, davanti al tratto vocale formate dalla lingua o dalle labbra..
f0	Frequenza fondamentale di vibrazione delle corde vocali. E' data un decimi di Hz.
ag	Area dell'apertura della glottide. Il range di variazione normale è tra 0 e 40 mm^2. Il valore medio per suoni sonori è di circa 3 - 5 mm^2.
al	Area trasversale della costrizione formata dalle labbra durante la produzione delle consonanti. Il range di variazione è tra 0 e 100 mm^2. Il valore 100 mm^2 corrisponde alla configurazione senza costrizione.
ab	Area trasversale della costrizione formata dalla lingua durante la produzione delle consonanti. Il range di variazione è tra 0 e 100 mm^2. Il valore 100 mm^2 corrisponde alla configurazione senza costrizione
an	Area trasversale della costrizione del velo faringeo. Il range di variazione è tra 0 e 100 mm^2.
ue	Rapidità di aumento del volume del tratto vocale durante l'intervallo di occlusione di una consonante occlusiva sonora. Un valore positivo di ue corrisponde ad una espansione della cavità dietro al punto di occlusione, un valore negativo ad una contrazione. L'integrale di ue calcolato sull'intervallo di costrizione è l'aumento o la diminuzione totale del volume.
ps	Pressione subglottale. Permette di aumentare o diminuire l'intensità del segnale prodotto. L'unità di misura è in cm di H_2O.
dc	Variazione percentuale dell'elasticità delle pareti dell'apparato fonatorio durante la pronuncia.
ap	Area dell'interstizio glottale posteriore che persiste attraverso un ciclo glottale. L'unità di misura è mm^2.

Tabella 2.2 Elenco completo dei parametri di controllo del sintetizzatore HLsyn. Gli ultimi 3 (ps, dc e ap) sono stati introdotti sulla attuale versione del sintetizzatore (Versione 2.2).

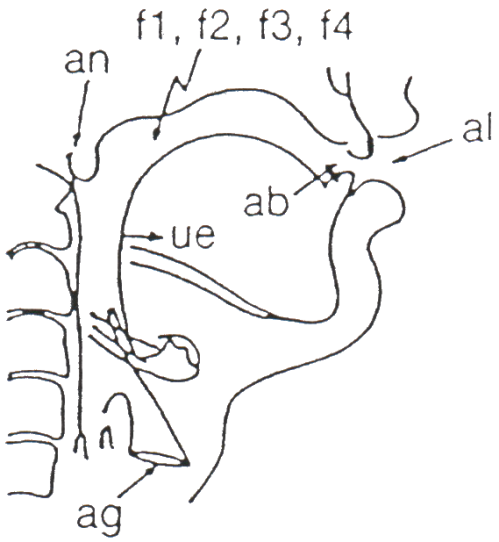


Figura 2.5 Legame tra i parametri del sintetizzatore e le caratteristiche del tratto vocale.

I parametri HL includono le aree di quattro costrizioni che si possono avere nella bocca e che sono:

- **an**, sezione di apertura della cavità nasale, data dal maggiore o minore abbassamento del velo palatino
- **ag**, area media dell'apertura della glottide
- **al**, area della costrizione formata dalle labbra
- **ab**, area della costrizione formata dalla punta della lingua

C'è da dire che **an** interviene solo per le consonanti nasali o, più in generale, quando c'è una nasalizzazione di qualche fonema, mentre **al** e **ab** intervengono solo durante la produzione delle consonanti.

Nella produzione di consonanti occlusive sonore si ha il passaggio di aria attraverso le corde vocali (per la produzione della sonorità) che però non può fuoriuscire all'esterno fino al momento del rilascio a causa dell'occlusione formata per produrre la consonante stessa. Si ha allora all'interno della bocca un aumento del volume compreso tra le corde vocali e il punto di occlusione. Di ciò tiene conto il parametro **ue**, che rappresenta la rapidità con cui questo volume varia e può essere sia positivo (per permettere la vibrazione delle corde vocali durante le consonanti occlusive) che negativo. Il suo integrale rappresenta ovviamente l'aumento o la diminuzione totale del volume all'interno della bocca.

Gli ultimi tre parametri, introdotti su questa ultima versione del sintetizzatore HL, sono **ps**, **dc** e **ap**. Il primo, **ps**, rappresenta la pressione sub-glottale e permette di variare l'intensità della sorgente sonora. Si può utilizzare, per esempio, per aumentare l'ampiezza di una vocale per le sillabe accentate. Per quanto riguarda **dc** c'è da fare una premessa: è stato dimostrato che la tensione delle pareti del tratto vocale, quando sottoposto ad una forza periodica, come ad esempio l'eccitazione dovuta alla vibrazione delle corde vocali, può variare significativamente

durante una pronuncia (Svirsky et al., 1997). Il parametro **dc** (delta compliance) tiene conto di ciò, rappresentando la variazione percentuale che l'elasticità delle pareti dell'apparato fonatorio subisce durante la pronuncia. Infine **ap** rappresenta l'area dell'interstizio glottale posteriore che persiste durante un ciclo glottale. Grazie ad esso ora si può, per esempio, avere un miglior controllo del flusso d'aria per sintetizzare fricative sonore e si possono sintetizzare occlusive sonore aspirate.

Parametro	Descrizione	Val. Default	Parametro	Descrizione	Val. Default
TLm	tilt	5 dB	**Cwm**	elasticità pareti tratto vocale	0.001 cm^5/dina
OQm	quoziente di apertura	50%	**Rw**	resistenza pareti tratto vocale	10 dina*s*cm^-5
B1m	largh. di banda 1° formante	80 Hz	**Cgm**	elasticità corde vocali	8E-6 cm^5/dina
B2m	largh. di banda 2° formante	90 Hz	**Lg**	lungh. orizzontale glottide	1 cm
B3m	largh. di banda 3° formante	150 Hz	**LabialAB**	guadagno per il filtro parallelo	55 dB
B4m	largh. di banda 4° formante	350 Hz	**PalVelarA2f**	A2F per fricaz. palatovelare	55 dB
B5m	largh. di banda 5° formante	500 Hz	**PalVelarA3f**	A3F per fricaz. palatovelare	60 dB
B2f	largh. di banda per F2 in parall.	250 Hz	**PalVelarA5f**	A5F per fricaz. palatovelare	50 dB
B3f	largh. di banda per F3 in parall.	320 Hz	**RetroflexA3f**	A3F per fricaz. retroflessa	50 dB
B4f	largh. di banda per F4 in parall.	350 Hz	**LateralA3f**	A3F per fricaz. laterale	40 dB
B5f	largh. di banda per F5 in parall.	500 Hz	**F5**	quinta formante	4500 Hz
Psm	pressione subglottale	8 cm H_2O	**F6**	sesta formante	4990 Hz

Tabella 2.3 Elenco dei parametri caratteristici del singolo parlatore

Come già detto oltre a questi 13 parametri che variano durante la pronuncia ce ne sono altri 24 che possono essere impostati dall'utente ma che si mantengono costanti per tutta la durata della parola sintetizzata (si può pensare ad esse come delle grandezze caratteristiche di ciascun parlatore). Questi sono elencati in Tabella 5.2 con una brevissima descrizione. Si fa presente che per la sintesi delle pronunce di questa tesi sono stati utilizzati i valori di default, validi per un generico parlatore maschile (per maggiori dettagli si veda la documentazione del sintetizzatore).

Analizziamo ora un po' più in dettaglio come i parametri di controllo appena descritti possono essere utilizzati nella sintesi di una pronuncia. Verranno descritti solo gli aspetti principali per ovvi motivi di spazio, lasciando al lettore interessato lo studio del manuale del sintetizzatore.

Costrizioni del tratto vocale e ampiezza delle sorgenti

Le proprietà della sorgente sono determinate dai flussi e dalle cadute di pressione attraverso le costrizioni nella glottide e nelle vie superiori. Ci sono tre tipi di orifizi che possono influenzare i flussi e le pressioni:

1. l'area trasversale dell'apertura del velo faringeo
2. l'area trasversale dell'apertura della glottide
3. la minima area trasversale presente nel tratto vocale sopra la laringe

Il primo di questi è dato semplicemente dal parametro **an** ed è diverso da zero solo per le pronunce nasalizzate (limitate a [m, n, ŋ] nell'italiano ma molto frequenti nella lingua inglese). Il secondo è dato dal parametro **ag**, escluso il caso in cui la pressione aumenti nel tratto sopra la glottide. In questo caso viene imposta sulla superficie delle corde vocali un aumento di pressione che può portare ad un aumento dell'area di apertura della glottide. In questo caso il sintetizzatore

utilizza, per calcolare i flussi e le pressioni, un parametro modificato chiamato **agx**, che ottiene in base a calcoli ed algoritmi implementati sul software stesso. Il terzo tipo di strettoia che si può avere nel tratto vocale può essere formata con le labbra, con la punta della lingua o con il corpo della lingua. Se la costrizione è formata dalle labbra o dalla punta della lingua, l'area della sezione così formata è data rispettivamente da **al** o **ab**. Quando invece è l'intera lingua a formare il restringimento alzandosi verso il palato, la lunghezza della costrizione è maggiore rispetto alle due precedenti. Ciò provoca un effetto globale sulla forma del tratto vocale. In questo caso la sezione del restringimento non è data da un semplice parametro del sintetizzatore ma viene calcolata in base ad altre grandezze, soprattutto la prima formante. L'innalzamento della lingua provoca infatti un abbassamento della frequenza di f1. Quando allora si è di fronte a una occlusione formata da tutto il corpo della lingua (come avviene ad esempio nella pronuncia della [ ]) si deve modificare la grandezza **f1** per sintetizzare correttamente tale fenomeno.

Filtraggio delle sorgenti per la produzione di consonanti sonore e vocali

Per le vocali non nasalizzate (**an**=0) la funzione di trasferimento tra velocità del flusso d'aria nella glottide e velocità sulle labbra è una funzione a tutti poli. Assumendo che, durante un ciclo di vibrazione delle corde vocali, non ci siano cambiamenti significativi nella frequenza o nella larghezza di banda delle formanti, la sintesi di una vocale si può ottenere con la sorgente glottale standard (controllata, lo ricordiamo, dal parametro **ag** compreso tra 3 e 5 mm^2) filtrata da una cascata di cinque frequenze formanti. Le quattro frequenze formanti **f1**, **f2**, **f3** e **f4** possono essere variate a piacere durante la pronuncia mentre la quinta va impostata come costante per ogni parlatore. Tali formanti dovranno essere quelle caratteristiche della vocale che si sta sintetizzando, potendo subire delle variazioni in base alle caratteristiche del singolo parlatore (ad esempio se si sta sintetizzando una voce maschile o femminile). In questa versione del sintetizzatore le larghezze di banda nominali delle diverse formanti sono fissate per tutta la pronuncia e i valori di default sono quelli in Tabella 5.2. Queste sono le larghezze di banda utilizzate quando la sorgente glottale è impostata per la produzione di suoni sonori (tipicamente ag=4 mm^2). Le effettive larghezze di banda dipendono dalla vocale (ossia dalla frequenza delle formanti e da quanto esse sono vicine l'una all'altra) e dalla lunghezza del tratto vocale del parlatore. Attualmente tali variazioni non sono incluse nelle relazioni di mappatura del software e la larghezza di banda delle formanti è un parametro fisso.

Filtraggio delle sorgenti di rumore (sorgenti fricative)

Dalle costrizioni che si possono avere nell'apparato fonatorio (labbra, punta o corpo della lingua) si può ottenere, per ogni istante, quella che ci dà la più piccola sezione di passaggio dell'aria. Si può pensare che il flusso d'aria che attraversa l'apparato boccale sia controllato da tale sezione minima e che la turbolenza dell'aria sia generata nelle vicinanze di tale costrizione. Il rumore così prodotto attraversa un insieme di filtri in parallelo che hanno il compito di modellizzare il comportamento dell'apparato fonatorio umano. Dato che le quattro frequenze formanti sono conseguenza della forma del tratto vocale, può essere possibile dedurre la posizione della costrizione da queste frequenze. La posizione e forma della costrizione determina quali formanti sono eccitate dal rumore di fricazione.

2.3.2 Il software del sintetizzatore

L'unità completa che contiene tutte le informazioni di un file sintetizzato è l'HL Document (file con estensione .hld). E' un file binario composto da sette gruppi di dati. Ogni gruppo può anche essere esportato separatamente in un file a sé stante con le seguenti estensioni:

1. file di descrizione del documento (.hli)
2. file di descrizione HL Speaker (.hls)
3. file di descrizione KL Speaker (.kls)
4. file con i parametri HL (.hl)
5. file con i valori di pressione dei flussi (.pf)
6. file con i parametri KL (.kl)
7. file in formato wave (.wav)

Anche un file nel formato del sintetizzatore KLsyn (.kld) può essere aperto e modificato con il programma HLsyn. La sintesi effettuata in questo modo corrisponde ad usare un sintetizzatore a formanti cascata-parallelo (Scarlino, 1993). Si può anche salvare un file di sintesi nel formato KL. In questo caso il file salvato (.kld) contiene quattro gruppi di dati, analogamente al formato .hld, e che contengono le seguenti informazioni:

1. file di descrizione del documento (.hli)
2. file di descrizione KL Speaker (.kls)
3. file con i parametri KL (.kl)
4. file in formato wave (.wav)

Tutte le operazioni sui file appena descritte si possono eseguire dal menù 'file' dell'interfaccia grafica del sintetizzatore. E' anche possibile importare file in formato wave per visualizzare forma d'onda, spettrogramma ecc. per poter fare dei confronti con le pronunce sintetizzate.

Il programma è in grado di visualizzare due tipi di finestre: finestre di testo e finestre grafiche. Le tre finestre di testo disponibili permettono di visualizzare, modificare e salvare i parametri HL e KL e di vedere i valori delle pressioni dei flussi (PF Values). Le quattro finestre grafiche permettono di visualizzare l'andamento dei parametri HL, KL, dei flussi PF e dello spettrogramma della pronuncia.

Il programma HLsyn implementa il metodo dei *punti di controllo* (control points) per l'inserimento dei valori dei parametri. Grazie a questo metodo si devono inserire i valori solo in corrispondenza di istanti di tempo scelti dall'utente. Il programma provvederà poi automaticamente a ricostruire con una interpolazione lineare i valori dei parametri tra due istanti precedentemente fissati. I punti di controllo possono essere fissati nelle finestre dei parametri HL e KL. La Figura 2.6 mostra appunto la finestra dei parametri HL e la relativa rappresentazione grafica. La prima colonna a sinistra contiene gli istanti temporali in msec, anche essi inseriti dall'utente secondo necessità. I caratteri più scuri indicano i valori fissati dall'utente mentre quelli più chiari sono i valori ricavati per interpolazione lineare dal programma stesso.

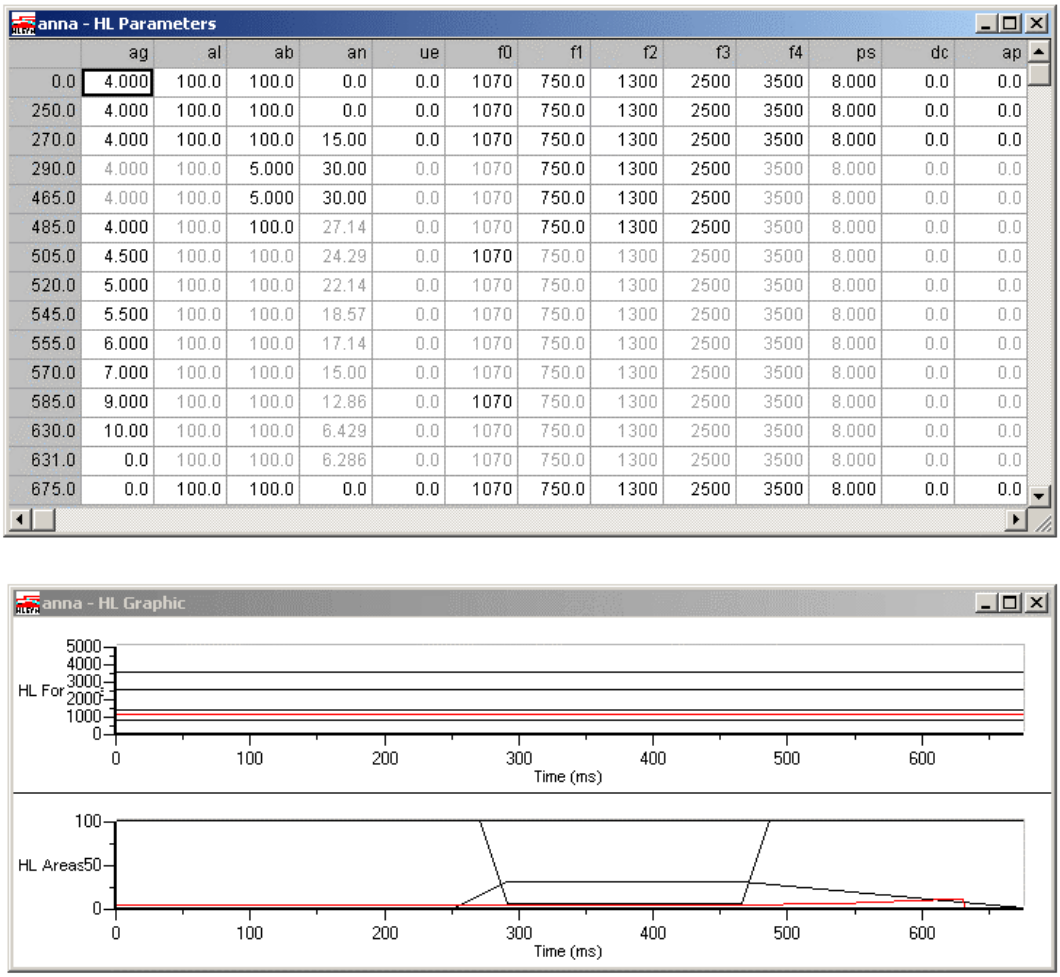
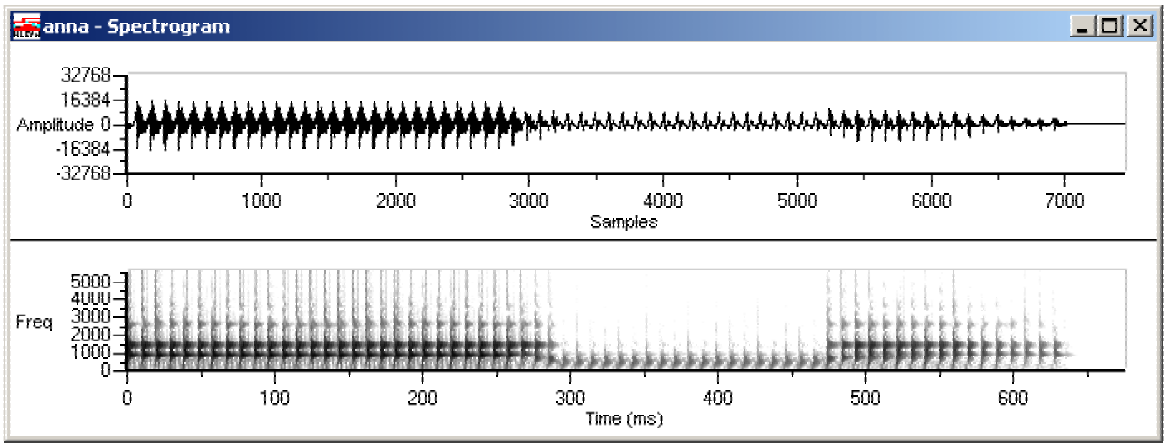


Figura 2.6 Finestra dei parametri HL di una parte di pronuncia. Si ricorda che il parametro f0 è in decimi di Hz mentre le frequenze formanti sono inHz.

Nella modalità grafica è sufficiente selezionare con il mouse un punto di una curva di interesse per vederne visualizzati i valori di ascissa (tempo) e di ordinata (valore del parametro).

Altre utili funzionalità del software di controllo riguardano gli spettrogrammi e gli spettri delle pronunce. Si possono visualizzare in finestre grafiche la forma d'onda del segnale sintetizzato, il suo spettrogramma e il suo spettro. In Figura 2.7 ne è illustrato un esempio.



Le quattro opzioni di calcolo e di visualizzazione possibili sono tutte attivabili cliccando con il tasto destro del mouse sulla finestra di interesse e selezionando una delle opzioni possibili dal menù che si apre. Tali opzioni sono:

- Pre-Emphasis: può essere abilitato o disabilitato il filtro di pre-enfasi nella visualizzazione dello spettro
- Window Size: si può impostare la dimensione (in numero di campioni) della finestra di Hamming per il calcolo dello spettro. Impostandolo a 64 campioni si ottiene uno spettro wide band mentre con una finestra di 512 si ha uno spettro narrow band
- Spectrum size: permette di scegliere il numero di campioni per il calcolo della FFT
- dB range: permette di aggiustare il livello di luminosità e contrasto dello spettrogramma per una visualizzazione ottimale

Tutti i valori caratteristici del singolo parlatore (elencati in Tabella 2.3) possono essere visualizzati e modificati aprendo l'apposita finestra con il comando 'KL Speaker' nel menù 'View'. Per impostare tutti i parametri di default del parlatore maschile o femminile è sufficiente selezionare il comando 'Generic Male Speaker' o 'Generic Female Speaker' dal menù 'Edit'.

Il software del sintetizzatore permette anche di selezionare la frequenza di campionamento e il numero di campioni per frame di analisi della pronuncia sintetizzata. Tali grandezze si possono modificare aprendo la finestra 'Document Info' nel menù 'View'. I valori usuali sono fc=10000 Hz con 50 campioni per frame o fc=11025 con 55 campioni per frame (sufficienti per l'analisi di un segnale vocale).

Una ultima considerazione riguarda la modalità di inserimento dei valori nelle finestre dei parametri. Purtroppo su questa versione non sono disponibili le familiari operazioni di 'taglia', 'copia' e 'incolla'.