

**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 2000/2001

**Sintesi Articolatoria di Parole Isolate  
Trascritte Foneticamente**

Relatore:

**Prof. Maria Gabriella Di Benedetto**

Laureando:

**Sparro Massimiliano**

Matr. N. 09090331

# INDICE

## VOLUME I

### **INTRODUZIONE** **I**

---

#### **CAPITOLO 1**

#### **Nozioni di grammatica e fonetica propedeutiche**

---

<b>INTRODUZIONE</b>	<b>1</b>
<b>1.1 CENNI DI ITALIANISTICA</b>	<b>1</b>
1.1.1 La grammatica e la fonetica	1
1.1.2 La fonetica articolatoria	
1.1.3 Cenni di elementi prosodici	
1.1.4 Il ritmo di parola, l'accento e la durata	
<b>1.2 LA PREDIZIONE DELL'ACCENTO PRIMARIO</b>	<b>9</b>
1.2.1 La classificazione delle parole in funzione dell'accento	9
1.2.2 Regole grammaticali per la determinazione della posizione dell'accento	11
1.2.3 Regole fonologiche per la predizione della posizione dell'accento	12
1.2.4 Regole di trascrizione fonetica	22
1.2.5 Regole di corretta pronuncia ricavate dal D.O.P.	23
1.2.6 Conclusioni di italianistica	30

#### **CAPITOLO 2**

#### **ACUSTICA DEL SEGNALE VOCALE E SINTETIZZATORE HLSYN**

---

##### **INTRODUZIONE**

<b>2.1 CENNI DI FISIOLOGIA</b>	<b>31</b>
2.1.1 L'organo dell'udito	31
2.1.2 Gli apparati di produzione della voce	34

<b>2.2 IL SUONO E L'ACUSTICA DEL SEGNALE VOCALE</b>	<b>40</b>
2.2.1 Lo spettro acustico	41
2.2.2 Suoni sordi e sonori	42
2.2.3 La frequenza fondamentale o pitch	44
2.2.4 Frequenze formanti	46
2.2.5 Caratteristiche generali del segnale emesso	47
2.2.6 Caratteristiche acustiche della sensazione uditiva	48
<b>2.3 IL SINTETIZZATORE HLSYN</b>	<b>51</b>
2.3.1 Caratteristiche generali e parametri di controllo	51
2.3.2 Il software del sintetizzatore	56

## CAPITOLO 3

### **SOFTWARE E LINGUAGGI DI PROGRAMMAZIONE UTILIZZATI** **59**

---

#### INTRODUZIONE

<b>3.1 PERCHE' PHP</b>	<b>60</b>
3.1.1 Operazioni su stringhe	61
3.1.2 Operazioni su vettori	62
3.1.3 I dati nel protocollo http	63
3.1.4 Operazioni su file e funzioni	63
3.1.5 Istruzioni di sessione	64
3.1.6 Interazione con Mysql	64
3.1.7 Espressioni regolari	65
<b>3.2 GLI ALTRI SOFTWARE UTILIZZATI</b>	<b>67</b>
3.2.1 Apache	67
3.2.2 Mysql	67
3.2.3 Spyguru	68
3.2.4 Javascript	69
3.2.5 PHP MyAdmin 2.2.0.	70

## CAPITOLO 4

### **SINTESI DEI FONEMI** **71**

---

#### INTRODUZIONE

<b>4.1 SINTESI DEI CONTOIDI</b>	<b>71</b>
4.1.1 Il contoide "f"	71
4.1.2 Il contoide "p"	74
<b>4.2 LA SINTESI DEI FONI: LA COARTICOLAZIONE</b>	<b>76</b>

## **CAPITOLO 5**

### **CONCLUSIONI**

---

#### **INTRODUZIONE**

**5.1 RIEPILOGO DEI RISULTATI SULLA TRASCRIZIONE FONETICA 77**

**5.2 LA SINTESI DEI FONEMI 78**

**5.3 L'INTERAZIONE CON L'HLSYN E SPUNTI PER APPLICAZIONI FUTURE 79**

**BIBLIOGRAFIA 80**

# CAPITOLO 1

## NOZIONI DI GRAMMATICA E FONETICA PROPEDEUTICHE

Nel sistema di comunicazione umano c'è una sostanziale distinzione fra la lingua scritta e quella parlata tale da non avere un'informazione immediata della pronuncia di una parola dalla sua trascrizione grafica. Nasce così l'esigenza di usare un alfabeto in cui ogni simbolo corrisponde ad un suono. Il carattere scritto viene chiamato grafema mentre il suono fonema. Introduciamo delle nozioni di italianistica relative alla grammatica ed alla fonetica, necessarie per la comprensione del lavoro sviluppato.

### 1.1 CENNI DI ITALIANISTICA

---

#### 1.1.1 La grammatica e la fonetica

La **grammatica** è definibile come l'insieme e la descrizione sistematica delle regole riguardanti gli elementi costitutivi di una lingua quali i suoni, le forme, le parole e le strutture sintattiche.

La **fonetica** è invece 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 è poi 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/ pronunciata da un settentrionale, da un toscano o da un meridionale); tuttavia i fonemi nell'italiano sono soltanto 28 e poiché le lettere del nostro alfabeto soltanto ventuno i segni di trascrizione non coincidono con i fonemi. Una lettera può servire per più di un fonema o, viceversa, uno stesso fonema è trascritto con più grafemi come ad esempio i **digrammi** e i **trigrammi**.

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/.

I fonemi si distinguono per essere sordi o sonori: sono sordi quelli che non richiedono l'intervento delle corde vocali per la loro generazione a differenza di quelli sonori. Per distinguerli, si può ricorrere ad una semplicissima prova, che consiste nell'accostare la mano alla gola durante la produzione del suono: nel caso di suono sonoro si avvertirà la vibrazione delle corde vocali all'interno della gola. Un maggiore approfondimento si ha nella parte relativa alla fisiologia della generazione del suono.

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 avvertite nei vari tipi di italiano regionale in maniera evidente così pure le opposizioni fra vocali aperte e chiuse come /é/ chiusa ed /è/ aperta ed /ó/ chiusa e /ò/ aperta. Comunque queste differenze esistono e sono specifiche della lingua: nei casi di omografia, cioè parole aventi stessa grafia ma diversa pronuncia, queste differenze risultano fondamentali per la corretta comprensione della parola.

Come accennato all'inizio del capitolo, 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 simbolo 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.). Il font utilizzato in questo lavoro avente queste caratteristiche, e che troviamo anche nel programma di scrittura Word, è il Lucida Sans Unicode.

## 1.1.2 La fonetica articolatoria

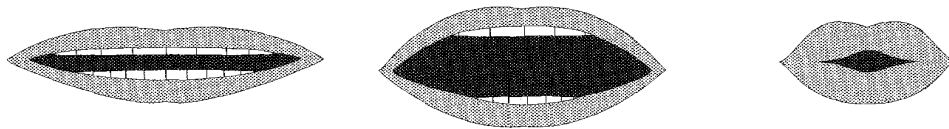
Ogni suono linguistico è compreso in una delle due classi principali chiamate tradizionalmente vocali e consonanti. In questo contesto si useranno i termini vocoidi e contoidi rispettivamente per le vocali e per le consonanti. Per lo studio dell'articolazione di tutti i fonemi ci si serve di diagrammi che

mostrano la posizione dei vari organi coinvolti. In particolare, per i vocoidi si usa il **trapezio fonetico**, e per i contoidi lo **spaccato sagittale** (sezione di profilo) dell'apparato fonatorio<sup>1</sup>.

### Articolazione dei vocoidi

Si possono definire **vocoidi** (in termini articolatori) quei suoni sonori, 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, centrale, posteriore). La figura 1.1 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<sup>2</sup>, 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. 1.1 Posizione delle labbra nelle tre articolazioni vocaliche estreme dell'italiano neutro:**

<b>Labbra non arrotondate</b>	<b>Labbra non arrotondate</b>	<b>Labbra arrotondate</b>
<b>vocale alta anteriore[i]</b>	<b>vocale bassa centrale[a]</b>	<b>vocale alta posteriore[u]</b>

(Canepari, 1992).

---

<sup>1</sup> 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).

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

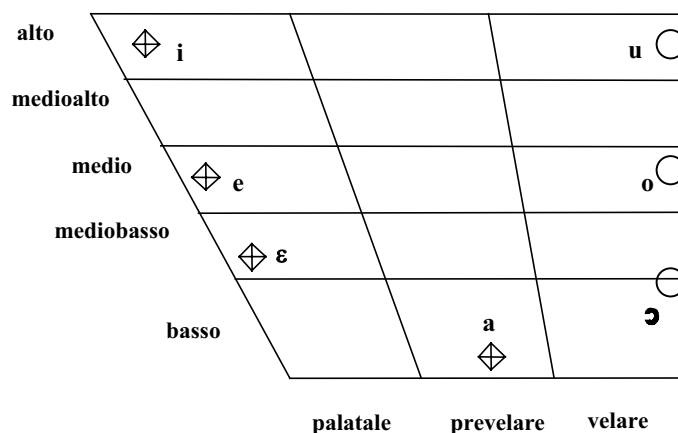


Fig. 1.2 Trapezio fonetico dell'Italiano (Canepari, 1979).

### 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 1.3. 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 **uomo**);
- **Bilabiale**, in cui, per realizzare il modo di articolazione, vengono usate entrambe le labbra (p. es. i fonemi /p/ di **papa**, /b/ di **biro**, /m/ di **mano**);
- **Labiodentale**, che prevede l'uso del labbro inferiore e dei denti superiori (p. es. i fonemi /f/ e /v/ **favo**);
- **Dentale**, in cui sono interessati la punta della lingua e i denti superiori (p. es. i fonemi /s/, /ts/, /d/ e /t/ di **senza di te**, /dz/ di **zero**, /z/ di **osare**);
- **Alveolare**, realizzato con la punta della lingua e gli alveoli che prendono parte all'articolazione (p. es. i fonemi /r/ di **rane**, /l/ di **lana**, /n/ di **nana**);
- **Alveopalatale**, con la lingua alta e con la punta in zona intermedia tra alveoli e palato (p. es. i fonemi /tʃ/ di **cinta**, /dʒ/ di **giro** e /ʃ/ di **scimmia**);
- **Palatale**, con il dorso della lingua ed il palato (p. es. i fonemi /j/ di **ieri**, /ɲ/ di **gli**, /ŋ/ di

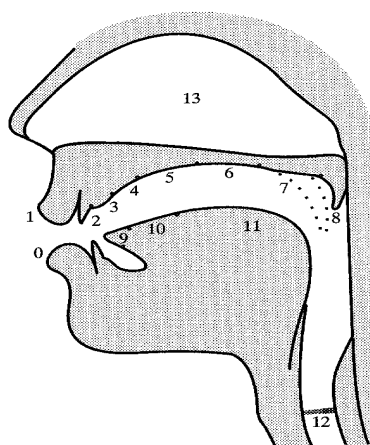
legno);



- **Velare**, con il dorso della lingua ed il velo (p. es. i fonemi /k/ e /g/ di **canguro**).

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 **china** e **ghiro**). Rispetto al punto d'articolazione velare, in tal caso, la parte interessata risulta più spostata verso il palato.

0	labbro (inferiore)
1	labbro (superiore)
2	denti (superiori)
3	alvéoli
4	post-alveoli
3-4	pre-palato
5	palato
6	pre-velo
7	velo (palatino)
8	uvula
9	apice (o punta, della lingua)
10	lamina (della lingua)
11	dorso (della lingua)
12	glottide (o laringe):
	1- ≡ corde (o pliche) vocali
	-2 ≡ aritenoidi
13	cavità nasale.



**Fig. 1.3 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 **tipo**, /k/ e /d/ di **dico**);
- **Fricativi**, prodotti operando un'occlusione non completa, che causa una particolare frizione dell'aria uscente (p. es. i fonemi /f/ e /v/ di **favo**, /s/ di **sano**, /z/ di **osare**, /ʃ/ di **scena**);
- **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 **zucchero**, /dz/ di **zaino**, /tʃ/ di **cima**, /dʒ/ di **giugno**);
- **Nasali**, prodotti ocludendo 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 **mano**);
- **Laterali**, realizzati bloccando il flusso d'aria al centro della bocca ma lasciandola fluire lateralmente (p. es. i fonemi /l/ e /ʎ/ di **luglio**);
- **Vibranti** (mono o poli vibranti), prodotti mediante la vibrazione di un organo mobile (p. es. il fonema /r/ di **rosa**).
- **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 **ieri** e il fonema /w/ di **uomo**);

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 1.1.

MODO DI ARTICOLAZI ONE	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			ɲ	k, g
Vibranti					r			
Laterali					l		λ	
Nasali		m			n			

**Tab. 1.1 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), il modo di articolazione e il punto di articolazione. Questo tipo di classificazione non è l'unico possibile, esistendone un'altra operata tramite la fonetica binarista ma che noi non esporremo, risultando significativa per la sintesi delle parole, le nozioni fin qui introdotte. E' possibile infatti trovare una stretta correlazione fra i parametri introdotti nella tabella 1.1 e quelli usati dall'HSyn per la sintesi dei fonemi e dei foni.

### 1.1.3 Cenni di 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 ed il ritmo, che tratteremo più approfonditamente di seguito 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. 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 vi è 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 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.

### 1.1.4 Il ritmo di parola, l'accento, la durata.

Il **ritmo** è quell'aspetto del linguaggio relativo all'andamento delle sillabe accentate ed atone all'interno della parola. Definiamo la sillaba come un fonema, od un insieme di essi, che costituiscono un gruppo stabile e ricorrente nella catena parlata. Le modalità relative a questa combinazione però, variano da lingua a lingua. Relativamente all'italiano le sillabe sono classificate in libere, cioè terminanti per vocale e implicate cioè terminante per consonante. Nella maggior parte dei casi non c'è contraddizione tra la sillabazione grammaticale e quella fonetica benché vi siano alcune eccezioni: ad esempio la parola chiosco, viene sillabata nella scrittura come chio-sco, ma fonologicamente è 'kjos-ko.

Le sillabe vengono organizzate in sequenze di piedi metrici secondo un algoritmo di metrificazione. Nell'italiano l'andamento del ritmo è trocaico, cioè di tipo forte-debole, a differenza del francese che è debole-forte. Nei casi possibili, cioè di sillabe in numero pari, i piedi metrici vengono organizzati con andamenti di tipo binario, alternando accenti forti con quelli deboli. Vediamo un esempio per spiegare meglio il concetto. La parola improvvisamente avrà il seguente andamento: 'im- prov -'vi-sa -'men-te. Come si può sentire ad orecchio, il piede metrico formato dalle sillabe men-te è quello di maggior rilievo: l'accento portato da questo piede è detto **primario**. L'**accento primario**, quindi, è l'enfasi che viene data ad una particolare sillaba di una parola ed in fonologia è indicato con il simbolo " ' " fatto precedere alla sillaba portatrice dello stesso. Gli altri accenti verranno assegnati cercando sempre di rispettare l'alternanza binaria. Nel nostro esempio si avrà un accento secondario sulle sillabe im-prov ed uno terziario su vi-sa. Il criterio di assegnazione degli accenti è ciclico mentre più complesso è il discorso per l'assegnazione dell'accento primario. Da sottolineare che sia il ritmo che l'andamento dell'accento sono fortemente influenzati da un fenomeno tipicamente italiano, quello dell'allungamento della durata di alcuni fonemi in relazione alla loro posizione od alla loro natura. L'allungamento può interessare sia le consonanti, per le quali si parla di geminazione, sia le vocali, per le quali si parla di allungamento vocalico.

Per la geminazione possiamo distinguere i seguenti casi:

- geminazione consonantica contrastiva
- consonanti intrinsecamente geminate
- parole ad iniziale intrinsecamente lunga

Il primo caso è quello che si incontra in parole come cane-canne, polo-pollo, casa-cassa ecc.. e non è ricavabile tramite una regola ma deve essere appreso caso per caso. Si noti, analogamente alla prova di commutazione per i fonemi, che la geminazione comporta in questo caso due diversi significanti.

Il secondo caso è relativo alle consonanti  $\eta$ ,  $\int$ ,  $\lambda$ , che hanno una durata maggiore delle altre indipendentemente dal contesto in cui si trovano.

Il terzo caso riguarda le parole, anch'esse non assoggettabili a regole, che presentano la consonante iniziale intrinsecamente geminata tipo Dio, più, qui ed altre.

Il caso di allungamento vocalico si ha nel caso di vocale in sillaba libera portatrice di accento che non sia finale di parola. Ad esempio in ca-ro o pa-pa si ha allungamento vocalico sulla a presente

nella prima sillaba ma non in car-ro o sulla prima a di pa-pà.

Nel caso di sillaba implicata portatrice di accento, la consonante che chiude la sillaba subisce allungamento. Questo fenomeno è detto allungamento della coda della sillaba ma non ha ragioni specifiche che ne giustifichino la presenza. E' un fenomeno tipicamente italiano.

Gli ultimi due casi di allungamento, il raddoppiamento sintattico e quello inverso, riguardano le consonanti di due parole adiacenti : nel primo caso si raddoppia la consonante della parola seguente,

ma va → ma 'vva, a noi → a nnoi

viceversa nel secondo

sud ovest → sudd ovest.

Non entreremo nello specifico di questi aspetti non essendo stati implementati nel presente lavoro, ma ne abbiamo segnalato la presenza per eventuali sviluppi futuri. Esaminiamo invece la questione della predizione dell'accento primario, che rappresenta il punto di riferimento per tutta la struttura metrica della parola, trascurando in questa sede l'assegnazione degli accenti di ordine superiore.

---

## 1.2 L'ACCENTO PRIMARIO E LA SUA PREDIZIONE

### 1.2.1 La classificazione delle parole in funzione della posizione dell'accento

L'accento primario interessa la vocale della sillaba sulla quale cade o, nel caso ce ne sia più di una, solamente una di esse. Le parole possono essere classificate relativamente al numero di sillabe che le compongono, monosillabi se sono costituiti da una sillaba, bisillabi da due ecc., ed in base alla sillaba interessata dall'accento: se è sull'ultima sillaba si ha una parola tronca, sulla penultima piana ecc.. Nello schema seguente sono riportate le due classificazioni, evidenziando il numero minimo di sillabe per realizzare la condizione relativa all'accento.

Numero sillabe	Classificazione accento	Tipo parola
1	Tronca	monosillabo
2	Tronca,piana	Bisillabo
3	Tronca,piana,sdrucchiola	Trisillabo
4	Tronca,piana,sdrucchiola,bisdrucchiola	Quadrisillabo
5	Tronca,piana,sdrucchiola,bisdrucchiola,trisdrucchiola	Polisillabo

La sillabazione di una parola è realizzabile in maniera univoca tramite le seguenti regole:

- Una vocale iniziale di parola, seguita da consonante, forma sillaba a sé
- Una consonante semplice fa sillaba con la vocale seguente
- I digrammi, i trigrammi ed i gruppi solo grafici con la vocale seguente formano sillaba a sé ch, gh, gl, gn sc, cia, cio, ciu
- Non si dividono i gruppi consonantici del tipo (b,c,d,f,g,p,t,v) + (l,r)
- Non si divide un gruppo di una o più consonanti inizianti per s
- Due consonanti uguali consecutive si dividono
- Qualsiasi altro gruppo di due consonanti deve essere diviso e comunque non è ammessa una configurazione che non possa essere presente come iniziale di parola
- Nei gruppi di tre o più consonanti la divisione avviene fra la prima e la seconda. Se fra la seconda e la terza si riscontra un nesso consonantico non consentito la divisione avviene fra queste due
- Si dividono solo le vocali consecutive che formano iato

Esponiamo i concetti di iato e dittongo, che avranno un notevole peso nella struttura del programma.

In italiano esistono dei suoni che sembrano avere caratteristiche intermedie fra le vocali e le consonanti: sono le semiconsonanti j,w e si impostano come la i e la u nel caso precedano una vocale. La durata del suono è però minore rispetto alle vocali semplici poiché l'articolazione si sposta velocemente sulla vocale seguente. Nel caso la i e la u seguano un elemento vocalico si parla invece di semivocali. Un dittongo si realizza nell'incontro di una semiconsonante seguita da vocale o da vocale seguita da semivocale. Riportiamo i casi possibili:

- semiconsonante j  
ia, ie, io, iu
- semiconsonante w  
ua, ue, ui, uo
- semivocale i  
ai, ei, oi, ui
- semivocale u  
au, eu

Nel caso si incontrino una semiconsonante, una vocale ed una semivocale, oppure due semiconsonanti ed una vocale si ha un trittongo. I casi riscontrabili in italiano sono.

- /j/ + vocale + semivocale
- /w/ + vocale + semivocale
- /j/ + /w/ + vocale
- /w/ + /j/ + vocale

Nel caso due vocali consecutive non formino dittongo si ha uno iato. Si ha sempre nei casi in cui nessuna delle vocali è u oppure i. La u e la i formano iato se sono toniche, cosa normalmente non indicata nella grafia, od appartengono a prefissi come ri-avere e su-esposto. Nel caso di tre vocali consecutive, è possibile anche avere un dittongo in iato con la vocale seguente o viceversa.

L'allocatione dell'accento nelle varie sillabe non è meccanica: l'italiano infatti, insieme all'inglese, il tedesco ed altre lingue, è considerata una lingua ad accento libero, dove cioè la sua posizione non è fissata a priori. Esempi di lingue ad accento fisso sono il francese dove le parole sono tutte tronche od il polacco dove le parole sono tutte piane. Inoltre non vi sono elementi grafici all'interno della nostra lingua che permettano in maniera univoca di prevedere la posizione dell'accento a differenza invece dello spagnolo, che pur essendo ad accento libero indica tipo e posizione dello stesso affinché la pronuncia sia immediatamente ricavabile dalla grafia. In questi termini, creare un automatismo che consenta di prevedere la posizione dell'accento e quindi la corretta trascrizione fonetica della parola in esame, sembra impossibile, se non ricorrendo ad una lista contenente le coordinate dell'accento per ogni parola.

In realtà la locuzione "lingua ad accento libero" presenta un'imprecisione, dando luogo ad ambiguità: il fatto che l'accento possa assumere più posizioni, non significa che non sia possibile prevedere dove cada ed inoltre, essendo l'italiano una lingua romanza derivata dal latino, ne eredita alcune proprietà, tra le quali quella di avere parole sostanzialmente piane oppure sdrucchiole. Ciò può sembrare una contraddizione con lo schema sopra riportato ma in realtà la possibilità di avere parole bi-tri- o quadrisdrucchiole, è ristretto ad un numero di casi limitato, soggetti a regole specifiche. Le parole bisdrucchiole in italiano ricorrono nei verbi del modo indicativo e congiuntivo terza persona plurale al presente come ad esempio capitano/ capitino telefonano/ telefonino e che naturalmente abbiano almeno quattro sillabe. Si noti che nel caso di capitano, si potrebbe altrettanto correttamente pronunciare capitano, intendendo con esso il comandante di una nave un aereo ecc. Questi casi non poco frequenti in italiano, di parole che hanno un eguale scrittura ma una diversa pronuncia si chiamano omografi, e creano una certa difficoltà nella elaborazione automatica della trascrizione fonetica poiché, ponendo l'accento in una posizione diversa da quella voluta si cambia radicalmente, oltre al loro significato, la loro pronuncia. L'unica possibilità di discernere due omografi è usare un analizzatore contestuale che esamini le parole seguenti e precedenti e ricavi informazioni sulla frase in cui l'omografo è inserito.

Un'altra categoria grammaticale che può presentare parole di tipo bisdrucchiole, trisdrucchiole e

quadrisdrucchiole, è quella dei verbi con enclitico. In italiano esistono delle particelle grammaticali, i clitici, in grado di modificare la posizione dell'accento della parola: se spostano l'accento sulla sillaba seguente della parola o addirittura su quella della parola seguente sono dette proclitiche, viceversa se lo anticipano, vengono dette enclitiche. Generalmente le particelle enclitiche sono pronomi che servono a snellire il discorso evitando ripetizioni o sottintendendo dei complementi:

Prendi quella → prendi-la

Dai quello a lui → da-glie-lo

Si noti che l'aggiunta dell'enclitico non modifica nella sostanza la posizione dell'accento, infatti, il verbo pren-di è un bisillabo avente l'accento sulla prima sillaba e pren-di-lo continua ad avere l'accento sulla prima sillaba. La differenza risiede unicamente nell'aver unito graficamente la particella al verbo trasformando così la parola in un trisillabo e cambiandone la classificazione. In base a quest'osservazione però, è possibile isolando le varie particelle eventualmente aggiunte in coda, risalire all'accentazione del verbo senza enclitico, individuando così la sillaba originariamente influenzata dall'accento e segnandola poi nella parola in esame. A questo punto, il problema è ricondotto al caso binario di scelta di accentazione piana o sdrucciola relativamente al verbo.

Rimane comunque il problema di discernere fra parole piane o sdrucciole. Per raggiungere lo scopo è possibile fare ricorso a strategie differenti, non escludente l'altra: la prima è di tipo grammaticale, nella quale considereremo anche i suffissi relativi alla trascrizione fonetica mentre la seconda è di tipo fonologico.

## **1.2.2 Regole grammaticali per la determinazione della posizione dell'accento**

Come già detto l'indicazione dell'accento in italiano non è mai obbligatorio tranne nei casi di polisillabi tronchi o monosillabi bivocalici: perché, già, caffè etc..

Questa prima restrizione, sebbene possa sembrare minima, permette di eliminare l'ambiguità sull'assegnazione della sede dell'accento in tutti i casi di bisillabi, in quanto, se troviamo la finale di parola accentata saremo in presenza di una parola tronca, viceversa l'accento cadrà sulla prima sillaba.

Vi sono poi degli affissi, cioè sillabe od insiemi di sillabe, che possono occupare delle posizioni prestabilite all'interno di una parola: in testa alla parola si chiamano prefissi mentre in coda suffissi. I prefissi non influenzano la posizione dell'accento, viceversa i suffissi hanno la proprietà di attirarlo su di esso o la sillaba che lo precede. Esempi di prefissi sono super, mega, anti ecc... che troviamo nelle parole composte mentre per i suffissi abbiamo one, etto, ezza ecc... Individuare queste categorie di parole permette perciò di individuare immediatamente la posizione dell'accento oppure di scartare la porzione di parola che sicuramente non lo contiene, continuando l'elaborazione sulla porzione restante. La parola anticarro verrà esaminata nel seguente modo:

- individuazione del prefisso anti

- eliminazione dello stesso
- sillabazione della parte rimanente
- essendo carro un bisillabo senza accento in coda sarà una parola piana
- si riunisce il prefisso in testa alla parola così elaborata

La parola *attenzione* invece, presentando la terminazione *one*, e sarà conseguentemente accentata piana.

Alcuni suffissi contenenti le vocali *e,o*, inoltre, permettono di determinare il grado di apertura delle vocali, dando la possibilità di scegliere fra i fonemi *è,é* ed *ò,ó*. Ciò porta un duplice vantaggio in quanto solo nel caso in cui l'elemento vocalico sia accentato può variare il suo grado di apertura, risultando invece sempre chiuso in caso di assenza: questi suffissi consentono in un sol colpo la determinazione del tipo di fonema e della posizione dell'accento, evitando tutta l'elaborazione relativa alla sua predizione. Alcuni suffissi sono specifici di alcune forme verbali ed in questo caso si può controllare se la parola analizzata sia o meno una di queste. La loro lista è riportata nel paragrafo della trascrizione fonetica.

Tra le regole grammaticali non ve n'è nessuna in grado di risolvere il caso in cui vi siano più vocali consecutive all'interno di una parola: solo nell'incontro di (*a e o*) vi è una separazione automatica ma non se una delle vocali sia *u* oppure *i*. La separazione (iato) risulterebbe possibile nell'eventualità che sulla vocale fosse indicato l'accento grafico, condizione non obbligatoria per la lingua italiana, altrimenti le vocali rimarrebbero unite (dittongo nel caso di due vocali trittongo in caso di tre). Nemmeno le regole fonologiche danno una soluzione a questo problema che può essere risolto solo tramite l'uso di una lista parole al cui interno è riportata l'eventuale presenza e la posizione di uno iato. Logicamente, poiché l'accento primario in una parola è unico, ci può essere al più uno iato nel caso vi siano più vocali consecutive.

Tutte le parole che non sono interessate da queste regole, o per le quali con esse si ha solo un'analisi parziale, saranno esaminate con le seguenti regole fonologiche.

### 1.2.3 Regole fonologiche per la predizione della posizione dell'accento

Abbiamo visto che l'accento primario è univocamente assegnabile fino alla categoria dei bisillabi. Per quanto esposto nella descrizione del ritmo di parola, la penultima sillaba, nei casi di parole con almeno tre sillabe, rappresenta la candidata solo più probabile a ricevere l'accento primario, senza averne, però, la certezza. Analizzando la struttura sillabica della parola è però possibile far emergere delle connessioni fra l'assegnazione dell'accento ed i contesti esaminati ed applicare conseguentemente le relative regole. Con la dizione contesto intenderemo la combinazione fra il fonema esaminato in una sillaba ed i fonemi della sillaba seguente e le proprietà eventualmente necessarie per attuare la regola: nel contesto /i/ in penultima sillaba seguita da /l m n t λ tʃ / in ultima sillaba e la parola è un verbo con enclitico si assegna l'accento in posizione sdrucciola.

Le seguenti regole, tratte dall'articolo di Rodolfo Del Monte "L'ACCENTO DI PAROLA NELLA PROSODIA DELL'ITALIANO STANDARD", riguardano la categoria dei trisillabi. Si può enunciare un gruppo di regole per i trisillabi, con varie eccezioni, aventi in penultima sillaba una consonante geminata, includendo anche quelle intrinsecamente geminate. L'altro gruppo riguarda i casi rimanenti. I gruppi di regole operano in maniera **disgiuntiva**, per cui viene applicata solo una regola alla volta a seconda del contesto individuato.



### Primo gruppo di regole

**piana** se la penultima sillaba è chiusa da una consonante forte escluso -GLI O Un gruppo consonantico diverso da -BR -TR -PL  
-ST -NT -GR -DR -RL;

per questi ultimi gruppi consonantici valgono le seguenti regole: -

**sdrucchiola** se la vocale è

- seguita da -PL;
- /e u/ ed è seguita da -BR è con qualche eccezione;
- /a i e/ ed è seguita da -TR è con qualche eccezione;

**piana** se la vocale non è

- /a/ seguita da -NT in alcuni casi;
- /i/ seguita da -ST in un caso; e non è /o/ seguito da -RL in un caso;
- /e/ seguita da -GR in un solo caso, oppure da -DR in un caso, o è un verbo in -ERE seguito da enclitico.

### Secondo gruppo di regole

Nel seguito intenderemo con – CONSONANTE v- il contesto CONSONANTE seguita da vocale.

#### **Con la vocale /i/ in penultima sillaba la parola si accenta:**

**sdrucchiola** se seguita da /t d l m k tʃ/ e la parola non è una delle eccezioni;

**sdrucchiola** se seguita da /l m n t λ tʃ / e la parola è un verbo seguito da un enclitico; ;

**sdrucchiola** se seguita da /g r n t/ e la parola è verbo della I coniugazione in -INARE -IGARE -ITARE oppure uno dei sostantivi e aggettivi elencati come eccezioni;  
piana in tutti gli altri casi.

**nel contesto –G v- è accentata piana;**

Fanno eccezione:

- verbi della I coniugazione in -IGARE alla 1 e 3 pers. sing. pres. ind./cong.  
LEVIGO MITIGO OBBLIGO IRRIGO LITIGO

**nel contesto –M v- è accentata sdrucchiola;**

Fanno eccezione:

-verbi della II in -IMERE alla 1 e 3 pers. sing. pres. ind./cong.  
IMPRIMO OPPRIMO COMPRIMO DEPRIMO

-alcuni aggettivi e sostantivi: REGIME MANGIME OPIMO  
SUBLIME ENZIMA

**nel contesto –D v- è accentata sdrucchiola;**

Fanno eccezione:

- verbi della I e II coniugazione 1 e 3 pers. sing. pres. ind./cong. DECIDO  
INCIDO UCCIDO DERIDO DIVIDO DIFFIDO CONFIDO AFFIDO RECIDO ARRIDO ANNIDO  
SORRIDO

- alcuni aggettivi sostantivi e nomi propri: DISFIDA CORRIDA  
CUPIDO DIFFIDA MALFIDO INFIDO FASTIDI SUSSIDI DISSIDI ECCIDI ALGIDE ALIDA  
SIGFRIDO SUICIDA

**nel contesto –L v- è accentata sdrucciola;**

Fanno eccezione:

- verbi della I in -ILARE alla 1 e 3 pers. sing. pres. ind./cong.  
INFILO DEFILO DEPILO AFFILO PROFILO

- aggettivi: EDILE CIVILE GENTILE MENSILE SENILE  
MASCHILE VIRILE SCURRILE SOTTILE

- sostantivi nomi propri e geografici: FUCILE BARILE CORTILE  
CANILE APRILE MONILE OVILE PROFILO PONTILE PORCILE VINILE METILE BACILE  
ETILE SACILE BRASILE TRAFILA ASILO DANILO BADILE SEDILE

**nel contesto –N v- è accentata sdrucciola;**

Fanno eccezione:

- verbi della I coniugazione in -INARE alla 1 e 3 pers. sing. pres, ind./cong.  
NOMINO LESINO MACINO PETTINO ORDINO

- sostantivi e nomi geografici: ACINO ASINO DOMINO ARGINE CARDINE CRIMINE FEMMINA  
FRASSINO CULMINE VIMINI ORDINE TERMINE FULMINE TURBINE TENDINE FASCINO  
GLICINE GLUTINE GRANDINE LAMINA LESINA MACINA MACCHINA MARGINE NOMINA  
PAGINA PETTINE PAMPINO PATTINI PLATINO POLLINE RONDINE RESINA RETINA  
REDINI RICINO CECINA FELSINA

**nel contesto –C v- è accentata sdrucciola;**

Fanno eccezione:

verbi della I coniugazione in -ICARE alla 1 e 3 pers. sing. pres. ind./cong.  
INDICO INTRICO DISDICO PREDICO

sostantivi aggettivi e nomi propri: FATICA CERVICE NEMICOMOLLICA MOTRICE ORTICA  
PUDICO -PENDICE NARICE CORNICE VARICE FELICE FENICE NUTRICE ANTICO  
FORMICA PERNICE LUBRICO MENDICO MATRICE RUBRICA RADICE ALICE VERNICE  
UFFICI VESCICA AMICO LOMBRICO ORTICA

**nel contesto –T v- è accentata sdrucciola;**

Fanno eccezione:

- participi passati dei verbi della III coniugazione SCUCITO GRADITO TRADITO  
MARCITO CUCITO SUBITO AMBITO

- alcuni verbi alla 1 e 3 pers. sing. pres. ind./cong. AVVITO INVITO  
ADDITO

- sostantivi aggettivi e nomi propri: LEVITA INVITO SMENTITA  
MATITA PARTITA SORTITA ITTITA MASTITE CISTITE AVITO NEFRITE BARRITO  
GARRITO FERRITE TURRITO PRURITO SQUISITO SEMITA CAMITA GRANITA LIGNITE  
SANNITA PEPITA SCIPITO MARITO PERITO BANDITO UDITO GRAFITE SOLFITO VAGITO  
ACCHITO SALITA COLITE LOLITA FLEBITE USCITA RIUSCITA

**nel contesto –Rv- è accentata piana;**

Fanno eccezione:

- MARTIRE ZEFIRO SATIRA SATIRO

**nel contesto –BRv- è accentata piana;**

Fanno eccezione:

- CALIBRO

**nel contesto –GLv- è accentata piana;**

Fanno eccezione:

- verbi con enclitico -GLI

**nel contesto –STv- è accentata piana;**

Fanno eccezione:

- ARISTA

**nel contesto –TR v- è accentata sdrucciola;**

Fanno eccezione:

- SALNITRO

**nel contesto –Bv- è accentata piana;**

Fanno eccezione:

- ALIBI CELIBE

**nel contesto –Pv- è accentata piana;**

Fanno eccezione:

- ADIPE DISSIPO POLIPO PRINCIPE

**nel contesto –Sv- è accentata piana;**

Fanno eccezione:

- BRINDISI

**Con la vocale /e/ in penultima sillaba la parola si accenta:**

**piana** se seguita da /r/ in un numero di forme irregolari di verbi, nomi propri, sostantivi, aggettivi e avverbi;

**sdrucciola** se seguita da /l m n v s λ tʃ / e la parola è un verbo seguito da enclitico;

**sdrucciola** se seguita da /l k tʃ / e la parola non è un verbo o uno dei sostantivi elencati nelle eccezioni;

**sdrucciola** se seguita da /s t/ in alcuni sostantivi e nei verbi della l' in -ETARE;

**piana** in tutti gli altri casi.

**nel contesto –R v- è accentata sdrucciola;**

- vengono elencate qui tutte le parole utilizzate per produrre delle radici chiave e che coprono anche parole prefissate o flesse: LEGGERO MADERA MEGERA MISTERO MESSERE PANCERA DISPERO PIACERE PARERE PANTERA PODERE POTERE OMERO ALTERO INVERO INTERO IMPERO DOVERE GODERE GALERA DOLERE CRATERE DAVVERO CLISTERE BARBERA CADERE BUFERA COLERA CHIMERA AVERE SAPERE TENERE GIACERE TEMERE RIMANERE SEDERE VEDERE VOLERE TACERE VALERE MISTERO CIMITERO STASERA SINCERO SEVERO AVVERO OVVERO PRIMAVERA MATERA ALGERI ALGHERO NOCERA AUSTERO BATTISTERO DICASTERO MENZOGNERO EMISFERO TORERO TRINCERA INGEGNERE MISERERE ROSTICCERE SFINTERE SOLERE STATERE SUADERE URETERE TAGLIERE SOMBRERO PACERE LUTERO

**nel contesto -Sv- è accentata piana;**

Fanno eccezione:

- NEMESI PROTESI GENESI SINTESI EFESO DIOCESI  
- verbo con enclitico –si

**nel contesto –B v- è accentata piana;**

Fanno eccezione:

- EFEBO

**nel contesto –TR v- è accentata sdrucciola;**

Fanno eccezione:

- FARETRA URETRA

**nel contesto –DRv- è accentata piana;**

Fanno eccezione:

- CATTEDRA

**nel contesto –Gv- è accentata piana;**

Fanno eccezione:

- DELEGA ALLEGRE

**nel contesto –Tv- è accentata piana;**

Fanno eccezione:

- EBETE VENETO IMPETO  
- verbi della I coniugazione in -ETARE alla 1 e 3 pers. sing. pres. ind./cong.  
VEGETO INTERPRETO ecc.

**nel contesto –Lv- è accentata sdrucciola;**

Fanno eccezione:

- aggettivi e sostantivi: FEDELE CANDELA BABELE CRUDELE VANGELO MISCELA SFACELO CAUTELA CARMELO PAMELA DISGELO

- verbi della I coniugazione in -ELARE alla 1 e 3 pers. sing. pres. ind./cong.

ANELO CONGELO TRAPELO RIVELLO ecc.

**nel contesto –Cv- è accentata sdrucciola;**

Fanno eccezione:

-verbi della I coniugazione in -ECARE alla 1 e 3 pers. sing. pres. ind./cong. IMPRECA ACCECA ARRECO ecc.

**nel contesto -Vv- è accentata piana:**

Fanno eccezione:

- verbi seguiti da enclitico –vi

**nel contesto -Mv - è accentata piana;**

Fanno eccezione:

- EREMO CINEMA

- verbi seguiti da enclitico –MI

**nel contesto –Nv- è accentata piana;**

Fanno eccezione:

- MODENA DARSENA ELENA

- verbi seguiti da enclitico –NE

**nel contesto -GRv- è accentata piana;**

Fanno eccezione:

- INTEGRO

**nel contesto –GLv- è accentata piana;**

Fanno eccezione:

- verbi seguiti da enclitico –GLI

**Con la vocale /a/ in penultima sillaba la parola si accenta:**

**sdrucciola** se seguita da /l m n s t λ tʃ v/ e la parola è verbo seguito da enclitico;

**sdrucciola** se seguita da /r m n s l k tʃ / e la parola è un so stantivo o verbo elencato come eccezione;

**piana** in tutti gli altri casi.

**Nel contesto –BRv- è accentata piana;**

Fanno eccezione:

- MACABRO

**nel contesto –NTv- è accentata piana;**

Fanno eccezione:

- TARANTO OTRANTO LEPANTO LEVANTO

**nel contesto –Bv- è accentata piana;**

Fanno eccezione:

- ARABO SILLABA SILLA

**nel contesto –TRv- è accentata sdrucciola;**

Fanno eccezione:

- ARATROTEATRO

**nel contesto –Cv- è accentata sdrucciola;**

Fanno eccezione:

- OPACO MORDACE PROCACE MENDACE VERACE FUGACE

TENACE SAGACE VIVACE CAPACE LOQUACE RAPACE

- Sostantivi: LUMACA BENACO AMACA MACACO

**nel contesto -Dv - è accentata piana;**

Fanno eccezione:

- MONADE DECADE GONADE RAGADE LAMPADA

**nel contesto -Rv - è accentata piana;**

Fanno eccezione:

- aggettivi: ILARE DISPARI IMPARI BULGARO BARBARO

- sostantivi: NETTARE ETTARO BACARO TARTARO ZINGARO SIGARO PICARO DOLLARO

ICARO FIGARO GASPARE LAZZARO CAGLIARI LIPARI MARMARA PESARO SASSARI

CHIAVARI

**nel contesto -Fv - è accentata piana;**

Fanno eccezione:

- ORAFO

**nel contesto -Vv - è accentata piana;**

Fanno eccezione:

- AGAVE CONCAVO INCAVO

- verbi seguiti da enclitico –vi

**nel contesto –Pv- è accentata piana;**

Fanno eccezione:

- CANAPA SENAPE

**nel contesto –Mv- è accentata piana;**

Fanno eccezione:

- BIGAMO BALSAMO BERGAMO DINAMO DITTAMO SESAMO TALAMO

- verbi seguiti da enclitico –mi

**nel contesto -Gv - è accentata piana;**

Fanno eccezione:

- FOLAGA PELAGO MALAGA SARAGO

**nel contesto –Nv- è accentata piana;**

Fanno eccezione:

- MANGANO GIOVANE EBANO SCORFANO SEDANO SATANA TIMPANO TRAPANO  
TETANO COFANO PLATANO ABANO LIBANO TRAPANI

- verbi della I e III coniugazione alla 3 pers. pl. pres. ind./cong.

- verbi seguiti da enclitico –ne

**nel contesto –Tv- è accentata piana;**

Fanno eccezione:

- PROSTATA AGATA DALMATA FEGATO STIGMATE SABATO SOCRATE

- verbi seguiti da enclitico –ti

**nel contesto –Lv- è accentata piana;**

- VANDALO ITALO SOMALO SEGALE SANDALO SCANDALO PETALO CABALA ALCALE

- verbi seguiti da enclitico -Lv

**nel contesto -Sv- è accentata piana;**

- ENFASI OASI ESTASI PEGASO CAUCASO

- verbi seguiti da enclitico –si

### **Con la vocale /o/ in penultima sillaba la parola si accenta:**

**sdrucchiola** se seguita da /l m n t s λ tʃ v/ e la parola è un verbo seguito da enclitico;

**piana** se è seguita da /k n l d/ e la parola è un verbo in -OCARE -OLARE oppure uno dei sostantivi elencati come eccezioni;

**sdrucchiola** se seguita da /d n m g v r l/ e la parola è uno dei sostantivi elencati;

**piana** in tutti gli altri casi.

**nel contesto –Sv-; è accentata piana;**

Fanno eccezione:

- PROGNOSI DIAGNOSI

- verbi seguiti da enclitico –si

**nel contesto –Vv- è accentata piana;**

Fanno eccezione:

- PADOVA MANTOVA VEDOVA VESCOVO

- verbi seguiti da enclitico –vi

**nel contesto –Bv- è accentata piana;**

Fanno eccezione:

- MICROBO REPROBO

**nel contesto -Tv- è accentata piana;**

Fanno eccezione:

- DESPOTA

- verbi seguiti da enclitico –ti

**nel contesto –Gv- è accentata piana;**

Fanno eccezione:

- DEROGA PROROGA ECLOGA TEOLOGO SILLOGE

**nel contesto -Pv- è accentata piana;**

Fanno eccezione:

- SINCOPE

**nel contesto -Cv- è accentata piana;**

Fanno eccezione:

- PARROCO EPOCA REVOCA

- verbi della I coniugazione in -OCARE alla 1 e 3 pers. sing. pres. ind./cong.  
PROVOCO REVOCO EVOCO CONVOCO INVOCO

- verbi seguiti da enclitico –ci

**nel contesto –Lv- è accentata sdrucciola;**

Fanno eccezione:

- sostantivi e nomi propri: CANNOLO CASSOLA PINOLO PIGNOLO STAGNOLA PAROLA  
ASSOLO TRITOLO ANGOLA BENZOLO NICOLA IOLE VIGNOLA POGGIOLO MENTOLO  
MANDOLA FAGIOLO FENOLO PIANOLA SPAGNOLO DIPOLO TAGLIOLA

- verbi della I coniugazione -OLARE alla 1 e 3 pers. sing. pres. ind./cong.  
SORVOLO CONVOLO CONSOLO IMMOLO ecc.

**nel contesto –Dv- è accentata piana;**

Fanno eccezione:

- ANODO CATODO COMODO ESODO METODO SCOMODO SINODO

**nel contesto –Mv- è accentata piana;**

Fanno eccezione:

- ATOMO PRODROMO SAGOMA SINDROME SINTOMO ZIGOMO SODO MA EPITOME

- verbi seguiti da enclitico –mi

**nel contesto -Nv- è accentata piana;**

Fanno eccezione:

- AFONO CANONE ATONO DEMONE BRETONE CONSONO ESTONE LET TONE SASSONE



- verbi della II coniugazione alla 3 pers. pl. pres. ind.

**nel contesto –RLv- è accentata piana;**

Fanno eccezione:

- MANDORLA

**nel contesto –Rv- è accentata piana;**

Fanno eccezione:

- sostantivi: BIFORA FEMORE ANCORA ANGORA BOSFORO CANFORA FORFORA FOLGORE  
LOGÓRO MEMORE MENTORE MARTORA MORMORO PORPORA RETORE REMORA  
SCORPORO TORTORA TRIFORA PECORA PLETORA ETTORE

**nel contesto -GLv - è accentata piana;**

- verbi seguiti da enclitico –GLI

**Con la vocale /u/ in penultima sillaba la parola si accenta:**

**sdrucchiola** se seguita da /t/ e la parola è un verbo in -UTARE;

**sdrucchiola** se seguita da /p l/;

**piana** in tutti gli- \_altri\_ casi.

**nel contesto -Tv- è accentata piana;**

- verbi della I coniugazione in -UTARE alla 1 e 3 pers. sing. pres. ind./cong.

AMPUTO COMPUTO CONFUTO DEPUTO DISPUTO REPUTO SVALUTO VALUTO

**nel contesto –Lv- è accentata sdrucchiola;**

Fanno eccezione:

- RINCULO BATTICULO

**nel contesto -Rv- è accentata piana;**

Fanno eccezione:

- LIGURE MURMURE SATURO BISTURI

**nel contesto –Mv- è accentata piana;**

Fanno eccezione:

- POSTUMO CUCCUMA

**nel contesto -Gv- è accentata piana;**

Fanno eccezione:

- PROFUGO

**nel contesto –Pv- è accentata sdrucchiola;**

Fanno eccezione:

- DIRUPO

**nel contesto –Cv- è accentata piana;**

Fanno eccezione:

- REDUCE EDUCO ALLUCE

**nel contesto –BRv- è accentata sdrucciola;**

Fanno eccezione:

- MANUBRI

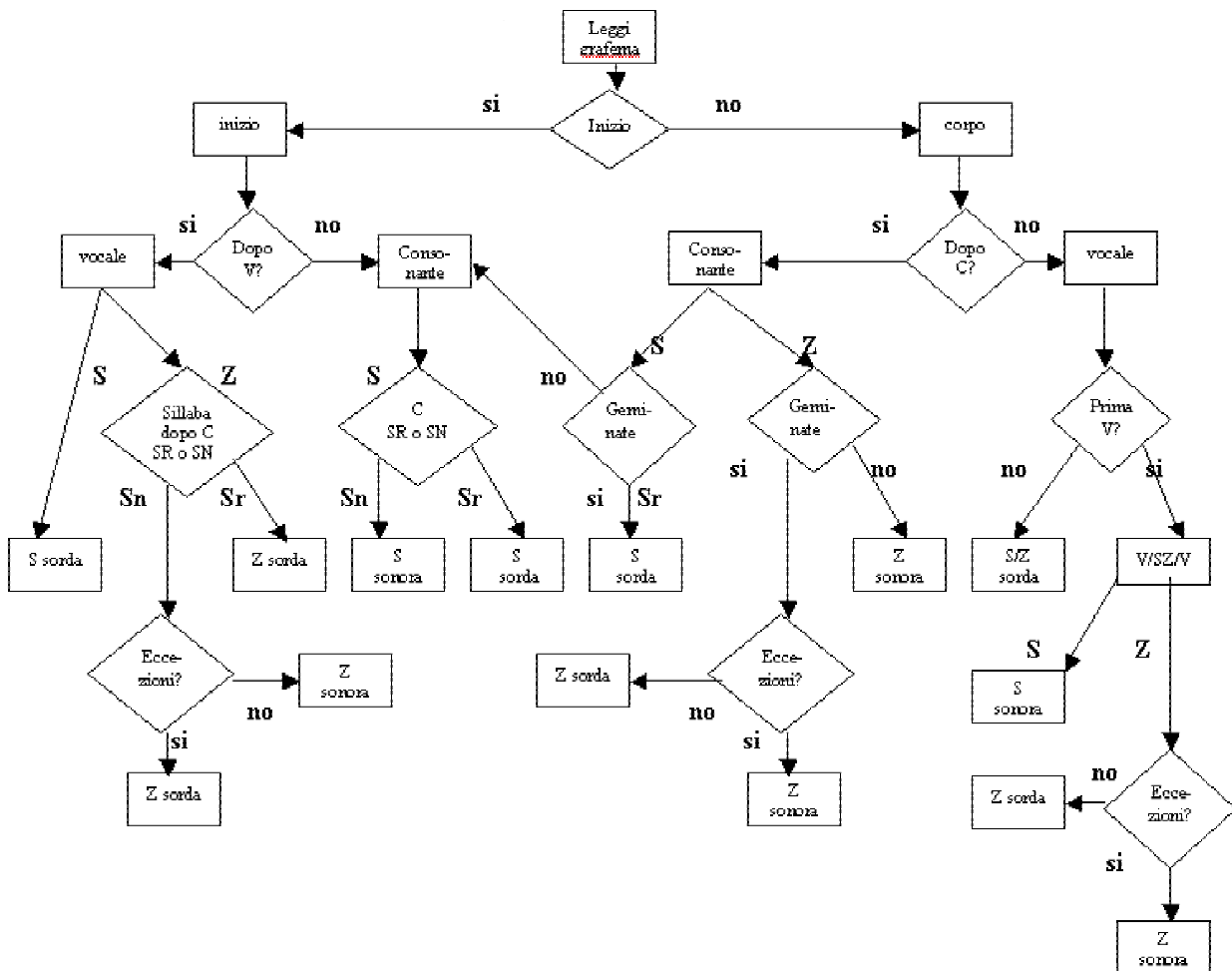
## 1.2.4 Regole di trascrizione fonetica

Avendo individuato il tipo di fonema con le regole precedenti, si può procedere alla codificazione fonetica secondo lo schema riportato nella tabella a fianco.

Fonemi	Lettere
/a/ . . . . .	a
/b/ . . . . .	b
/č/ (cero) . . . . .	c (digramma <i>ci</i> e <i>ch</i> )
/k/ (casa) . . . . .	c
/d/ . . . . .	d
/é/ (néro) . . . . .	e
/è/ (bène) . . . . .	e
/f/ . . . . .	f
/ǵ/ (gita) . . . . .	g (digramma <i>gi</i> e <i>gh</i> )
/ɣ/ (gara) . . . . .	g
— . . . . .	h
/i/ . . . . .	i
/l/ . . . . .	l
/λ/ (foglio) . . . . .	— (digramma <i>gl</i> e trigramma <i>gli</i> )
/m/ . . . . .	m
/n/ . . . . .	n
/ɲ/ (gnomo) . . . . .	— (digramma <i>gn</i> )
/ó/ (pólio) . . . . .	o
/ò/ (póco) . . . . .	o
/p/ . . . . .	p
— . . . . .	q
/r/ . . . . .	r
/s/ (suono) . . . . .	s
/ʒ/ (caso) . . . . .	s
/ʃ/ (scemo) . . . . .	— (digramma <i>sc</i> e trigramma <i>sci</i> )
/t/ . . . . .	t
/u/ . . . . .	u
/v/ . . . . .	v
/z/ (pazzo) . . . . .	z
/z/ (zona) . . . . .	z

Tab. 1.2 Tabella dell'associazione grafemi fonemi

Per i grafemi s e z che non hanno una codificazione biunivoca, si può utilizzare il seguente algoritmo ricavato dall'articolo di Rodolfo del Monte, integrato con le regole ricavate dal D.O.P ed adattato al nostro scopo.



## 1.2.5 Regole di corretta pronuncia ricavate dal D.O.P.

### La lettera S

La lettera s ha il suono *sordo* [s] nei casi che seguono:

in -ese, suffisso di nomi per lo più etnici (es. cinese, malese, marsigliese), compresi i derivati (es. cineseria), con qualche eccezione (es. borghesia, francese, Malesia, marchese);

in -esi, -ese, -esero, -eso, desinenze del passato remoto e del participio passato (es. resi, rese, resero, reso; accesi, difesi, presi, ripresi, scesi, stesi, ecc.), e loro derivati (es. discesista, impresario), con qualche eccezione (es. lesi, lesivo);

in -osi , -ose , -osero , -oso , desinenze del passato remoto e del participio passato (es. rosi , rose , rosero , roso ; nascosi, risposi , ecc.), e loro derivati (es. rosicchiare ), con qualche eccezione (es. corrosi, corrosione );

in -oso , -osa , suffissi di aggettivi e di sostantivi (es. bisognoso , cellulosa, generoso , vanitoso , ventosa ), compresi i derivati (es. curiosità, vanitosetto ), questi però con qualche eccezione (es. celluloso);

in alcune parole isolate:

casa , chiusi , cosa, così , mese, peso , Pisa , posare , rasi , risi , socchiusi , sorrisi ).

### **La lettera s ha il suono *sonoro* [z] nei casi che seguono**

quand'è finale di prefisso (es. bisunto ,cispino , disadatto , esautorare , trasandato , trisavolo );

in -asi , -ase , -asero , -aso , desinenze del passato remoto e del participio passato (es. invasi, invase , invasero, invaso , persuasi , ecc.), con qualche eccezione (es. rasi )

in -esimo , suffisso dei numerali ordinali (es. cinquecentesimo ,ventiquattresimo );

in -esimo , suffisso di sostantivi astratti (es. cristianesimo , urbanesimo );

in -isi , -ise , -isero , -iso , desinenze del passato remoto e del participio passato (es. incisi , incise , incisero , inciso, misi , preciso , ecc.), con qualche eccezione (es. risi , sorrisi );

in -usi , -use , -usero , -uso , desinenze del passato remoto e del participio passato (es. infusi , infuse , infusero , infuso, esclusi , protruso , ecc.), con qualche eccezione (es. chiusi , socchiusi ).

## **La lettera Z**

### **La lettera z ha il suono *sordo* [ts] nei casi che seguono:**

iniziale di parola, quando la seconda sillaba comincia per consonante sorda, cioè per c f, p, t (es. zampa , zoccolo ), con alcune eccezioni (es. zaffiro) ;

preceduta da l (es. alzare , filza ), con poche eccezioni (es. elzeviro ) ; seguita da un i tonico o atono, seguito a sua volta da un'altra vocale (es. agenzia , polizia , razzia ; grazia, Lazio ), ospizio , pronunzia , silenzio , spezie ; anziano , conversazione , fazioso ), con poche eccezioni (es. azienda ), a cui vanno aggiunti i casi di derivazione da parole con z sonora (es. romanziere );

in -anza , suffisso di sostantivi (es. speranza , usanza );

in -azzare , suffisso di verbi (es. scorrazzare , spiegazzare ); in -enza , suffisso di sostantivi (es. assenza, supplezza ) o terminazione in genere (es. Cosenza , lenza , scienza , senza );

in -ezza , suffisso di sostantivi (es. grandezza , saggezza ); in -onzolo , suffisso di sostantivi (es. ballonzolo , lattonzolo );

in -ozzo , -ozza , suffisso di sostantivi (es. barilotto , predicozzo ; carrozza , Carrozza );

in -uzzo , -uzza , suffisso di sostantivi (es. labbruzzo , palazzo ; pagliuzza , pietrucce );

in molte parole isolate:

(es. dinanzi , Firenze , forza , innanzi , pezzo , piazza , ragazzo , sostanza , spezzare , stanza , terzo ).

### **La lettera z ha il suono *sonoro* [dz] nei casi che seguono:**

iniziale di parola, quand'è seguita da due vocali (es. zaino, zuavo), con poche eccezioni (es. zio);

iniziale di parola, quando la seconda sillaba comincia per z (es. zanzara, zizzania ), con poche eccezioni (es. zazzera);

iniziale di parola, quando la seconda sillaba comincia per consonante sonora, cioè per b, d, g, l, m, n, r, v (es. Zara, zebra, zibibbo, Zodiaco), con alcune eccezioni (es. zanna);

quand'è scritta scempia in mezzo a due vocali semplici (es. azalea, bizantino), con alcune eccezioni (es. nazismo);

in -izzare, suffisso di verbi (es. armonizzare, organizzare);

in alcune parole isolate (es. azzurro, mezzo inteso a metà, medio, romanzo).

### **Le lettere E ed O**

Per i grafemi e ed o si utilizzeranno i seguenti criteri.

#### **La lettera e ha il suono *aperto* [è] nei casi che seguono:**

nel dittongo ie (es. bandiera, cavaliere, chiedere , chiesa, dieci, diedi etc)

nelle parole che finiscono in consonante (es. negus, semel, vademecum)

quand'è seguita da vocale (es. colei, costei, feudo, idea etc);

quand'è seguita da una consonante scritta scempia, seguita a sua volta da due vocali vere o apparenti (es. assedio, criterio, serio, tenue), salvo che nelle voci in -eguo e simili (es. dileguo, seguito); regola che vale anche se la consonante che segue è una z, scempia nella scrittura ma doppia nella pronuncia (es. facezia, screzio), o anche se è un c o un g dolce e delle due vocali la prima è un i muto (es. specie, spregio), in quest'ultimo caso con qualche eccezione (es. fregio, impecio);

in -edine, suffisso di sostantivi (es. pinguedine, salsedine);

in -rei, -rebbe, -rebbero, desinenze del condizionale (es. crederei, crederebbe, crederebbero );  
in -elio, suffisso di diminutivi o vezzeggiativi (es. fossatello , porcello);

in -endo , desinenza del gerundio (es. aprendo , leggendo ), o terminazione in genere (es. ammenda , tremendo , attendo, rendo ), qui però con qualche eccezione (es. scendo , vendo);

in -enne , suffisso di sostantivi e aggettivi derivati da numerali (es. decenne , settantenne );

in -ennio , suffisso di sostantivi derivati da numerali (es. biennio , millennio );

in -eno , suffisso di nomi etnici (es. madrilenno , nazareno);

in -ense , suffisso di aggettivi (es. forense , portuense) o terminazione in genere (es. amanuense , intense );

in -ente , desinenza del participio presente (es. perdente , scendente) o terminazione in genere (es. accidente, gente );

in -enza , suffisso di sostantivi (es. partenza , urgenza ) o terminazione in genere (es. lenza , Faenza , senza );

in -errimo , suffisso di certi superlativi (es. asperrimo , integerrimo);

in -esi, terminazione di termini dottrinali (es. catacresi, mimesi) ;

in -esimo ,suffisso dei numerali ordinali (es. trentesimo, milionesimo);

in -estre, suffisso di aggettivi (es. terrestre, rupestre) o terminazione in genere (es. bimestre, Mestre);

in -etti, -ette, -ettero, desinenze del passato remoto (es. credetti, credette, credettero);

in molte parole isolate:

accendere, accento, accetto, aperto, argento, aspetto, attento, bello, bene, breve, celeste, cento, certo, cipresso, contento, coperto, debbo, decimo, devo, diciassette, divento, è , ebbene , ebbi, ecco, ero, esco, essere, eterno, febbre, festa, finestra, Firenze , gesto , getto , greco , guerra , intendere , leggere , lettera, letto , levo , maestro , meglio , merce , mezzo , novella , offerto, osservo , peggio , penso , perdere, petto , pezzo , prego prendere , presento del verbo presentare, presso , presto , rendere , resto , riesco, ripetere , riprendere , scoperto , sempre , senso , sento, seppi , servo , sesto , sette , silenzio , sorella ,

spero , spezzo , stendere , temo , tempo , tengo , terra , terzo , testa , vecchio , vengo , vento , verso ,  
vesto.

**La lettera e ha il suono *chiuso [è]* nei casi che seguono:**

in -eccio , suffisso di aggettivi e di sostantivi (es. cicaleccio , villereccio );

in -efice , suffisso di sostantivi (es. artefice , orefice );

in -eggio , suffisso di sostantivi e di verbi (es. campeggio , posteggio );

in -ei , -esti , -é , -emmo , -este , -erono , desinenze del passato remoto (es. credei , credesti, crede ,  
credemmo, credeste , crederono );

in -remo , -rete , desinenze del futuro (es. crederemo, crederete );

in -mente , suffisso di avverbi (es. amabilmente , sinceramente ) ;

in -mento , suffisso di sostantivi (es. accoglimento , sentimento ) o terminazione in genere (es.  
momento , palmento , sacramento , strumento );

in -ere , desinenza dell'infinito dei verbi (es. accadere , avere , bere , cadere , dovere , godere , parere ,  
piacere , potere , rimanere , rivedere , sapere, sedere , tacere , temere , tenere , valere , vedere , volere );

in -esco ,suffisso di aggettivi (es. fanciullesco);

in -ese suffisso di nomi perlopiù etnici (es. albanese ,valdese ) o terminazione in genere (es. arnese ,  
mese , paese , palese);

in -esimo , suffisso di sostantivi astratti (es. feudalesimo , umanesimo ) ;

in -essa , suffisso di sostantivi femminili (es. contessa , ostessa );

in -essi , -esse , -essimo , -este , -ersero ,desinenze dell'imperfetto congiuntivo (es. credessi ,credesse ,  
credessimo, credeste , credessero ).

in -resti , -remmo , -reste ,desinenze del condizionale (es. crederesti , crederemmo , credereste );

in -ete , desinenza del presente indicativo e dell'imperativo (es. credete , prendete );

in -eto , -eta , suffisso di sostantivi (es. agrumeto , diavoletto , cipresseto , sughereti );

in -etto , suffisso di diminutivi e di collettivi (es. agretto , pezzetto , terzetto , ottetto );

-evo , -evi , -eva , -evano , desinenze dell'imperfetto indicativo (es. credevo , credevi , credeva ,  
credevano ) ;

in .-evole , suffisso di aggettivi (es. amichevole , meritevole);

in -ezza , suffisso di sostantivi (es. debolezza , fanciullezza);

in molte parole isolate :

acceso , allegro , almeno , appena , bevo , bevvi , capello , cerco , che congiunzione e pronome, ché a perché , codesto e celeste , credere ,, davvero , dentro , detto , dimentico , egli , ella, esso , feci del verbo fare, fermo , freddo , fresco , fretta , invece , maestro , me , medesimo ,meno , mente mettere , né , nemmeno , nero , orecchio , perché , permettere , poiché , preso , quello , questo , reso , ricevere, ripreso , rivedo , scendere , se congiunzione, sé pronome, sedici , segno , seguo , sembro , sera, sereno, spesso , steso , stesso , stretto , te , temo, tenni , tre , , trenta , vedo, velo , venni , venti, verde , vero.

**La lettera o ha il suono *aperto* [ò] nei casi che seguono:**

nel dittongo uo (es. buono , cuore , fuoco , fuori etc);

nelle parole che finiscono in consonante (es. colon , factotum , nord , stop );

quand'è seguita da una consonante scritta scempia, seguita a sua volta da due vocali vere o apparenti (es. custodia , Microbio , memoria , petrolio , roseo , stentoreo , storia ), anche se la consonante che segue è una z, scempia nella scrittura ma doppia nella pronunzia (es. Coste , negozio ), o anche se è un c o un g dolce e delle duevocali la prima è un i muto (es. cervogia , consocio , mogio in quest'ultimo caso con qualche eccezione (es. incrocio , vocio);

in -ò , desinenza del passato remoto (es. amò , gridò ); in -rò , desinenza del futuro (es. mangerò , scherzerò );

in -ò terminazione in genere (es. ciò , do , falò , fio , ho , no , però , po' , so , sto );

in -occio , -occia , suffisso di aggettivi e di sostantivi (es. belloccio , cartoccio , saccoccia );

in -oide , suffisso di aggettivi e di sostantivi (es. albuminoide , cilindroide , mattoide , tiroide );

in -oldo , terminazione di nomi propri (es. Aroldo , Leopoldo );

in -olfo , terminazione di nomi propri (es. Adolfo , Castel Gandolfo );

in -olo , -ola , suffisso di aggettivi e di sostantivi (es. castagnola, fenolo , stagnola , tritolo ), compreso -nolo , -uola , -aiolo , -aiola , -aiuolo , -aiuola (es. barcaiolo e barcaiuolo , mazzuolo , museruola, trecciaiola e trecciaiuola );

in -osi , suffisso di termini medici (es. artrosi, tubercolosi );

in -otto , suffisso di aggettivi e di sostantivi (es. bambolotto , chioggiotto , semplicitto );

in -ottolo , -ottola , suffisso di sostantivi (es. collottola , nanerottolo , pianerottolo , viottola );

in -ozzo , -ozza , suffisso di sostantivi (es. bacherozzo , maritozzo , tavolozza , tinozza );



in molte parole isolate :

accorgersi , cogliere, copro ,corpo , cosa, diciannove , diciotto , donn, forte , forza , , gioia , godo ,  
grosso , importo , modo , morte, morto , nonno , nono, nostro , nove , occhio , odo, offro , oggi , oh ,  
opera , oso, , otto , parola , piuttosto , poco , poi , porta, , porto , posso , povero, proprio , provo ,  
ricordo , scopro , scorgere , solito , togliere , troppo , trovo, voglio, volgere, volli , volta , volto del  
verbo voltare e del verbo volgere, vostro .

**La lettera o ha il suono *chiuso* [ó] nei casi che seguono:**

in -ognolo , suffisso di aggettivi (es. amarognolo, verdognolo );

in -oio , -oia , suffisso di aggettivi e di sostantivi (es. cottolo , levatoio, mattatoio, smoccolatoio ;  
mangiatoia , scorciatoia );

in -ondo , -onda , terminazione di aggettivi e di sostantivi (es. fondo , giocondo , mondo , profondo ,  
secondo ; baraonda, fionda , sonda );

in -one , -ona , suffisso di sostantivi maschili e femminili (es. cascai , librone ; bambolona , ragazzona,  
espressione , fienagione , partizione, stratificazione , trasfusione, trasmissione e terminazione in genere  
(es. carbone , padrone , ragione, termosifone, Ancona , persona , rifocillano, con qualche eccezione es.  
testimone , zona );

-onte , terminazione di aggettivi e di sostantivi (es. arconte , fronte , Rodomonte );

in -onzolo, suffisso di sostantivi (es. mediconzolo , raponzolo );;

in -ore , suffisso di sostantivi (es. amore , coltivatore , creditore , dolore , fondatore , rumore );

in -oso , -osa , suffisso di aggettivi e di sostantivi (es. afoso , erboso , piccoso ; tuberosa );

in molte parole isolate:

abbandono, allora , ancora, ascolto , bisogno , bocca , coda , colore , coloro , colpa , colpo , come,  
condotto , conoscere , conto , contro , correre , costoro ,dodici , dolce , dopo , dove , fiore , forma ,  
forse , giorno , giovane , gola , incontro , interrompere , intorno , lavoro , loro , maggiore, migliore ,  
moglie , molto , mostro , nascondere , nipote , noi , nome , ogni , oltre , ombra , ora , porre , posto ,  
pronto , quattordici , racconto, riconoscere, rispondere , Roma rompere , rosso , signore , sogno , sole ,  
solo , sono , sopra , sotto , tocco del verbo toccare, torno , voce , voi .

Le seguenti desinenze grammaticali e suffissi prevedono accentazione sdrucciola o piana e non rientrano nei casi esposti.

## **Accentazione sdrucciola**

### **Suffissi**

abile (venerabile), evole (lodevole), ibile (credibile), issimo (fortissimo), istico (borsistico).

### **Desinenze grammaticali**

Ano (dormano, lodano, temano, vendano), arono (lodarono) , assero (lodassero), assimo(lodassimo), avano (lodavano), crebbero( loderebbero, temerebbero , venderebbero) erono(temerono , venderono), essero (temessero , vendessero), essimo (temessimo , vendessimo), ettero(temettero , vendettero) , evano(temevano , vendevano), irebbero (dormirebbero) ,irono (dormirono), issero (dormissero), issimo (dormissimo), ivano (dormivano).

## **Accentazione piana**

### **Desinenze grammaticali**

Avamo( lodavamo), avate(lodavate), erano (loderanno temeranno, venderanno), eremmo(loderemmo, venderemmo, temeremmo), eremo(loderemo, venderemo, temeremo), ereste ( lodereste, vendereste, temereste), eresti ( loderesti, venderesti, temeresti),erete( loderete, venderete, temerete), erò ( loderò, venderò, temerò), evamo (temevamo , vendevamo), evate (temevate , vendevate), iranno (dormiranno), iremmo(dormiremmo), iremo(dormiremo), ireste (dormireste), iresti (dormiresti), irete (dormirete) , ivamo (dormivamo), ivate (dormivate).

## **1.2.6 CONCLUSIONI DI ITALIANISTICA**

Non essendo state trovate relativamente ai polisillabi regole simili, sono state usate sinergicamente le due strategie sopraesposte per cercare di ricondurre quante più parole possibili al caso dei trisillabi. Per i quadrisillabi sono state trovate quattro regole fonologiche in maniera autonoma, usando il medesimo approccio usato per i trisillabi. Tutte le parole che non rientrano in nessuna delle analisi effettuate viene considerata automaticamente piana in virtù dei concetti esposti. Qualora non fossero piane, occorre inserire le stesse in una lista di eccezioni.

# 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 vibratorii 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. È 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.

È 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 semicircolari** 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 **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**.

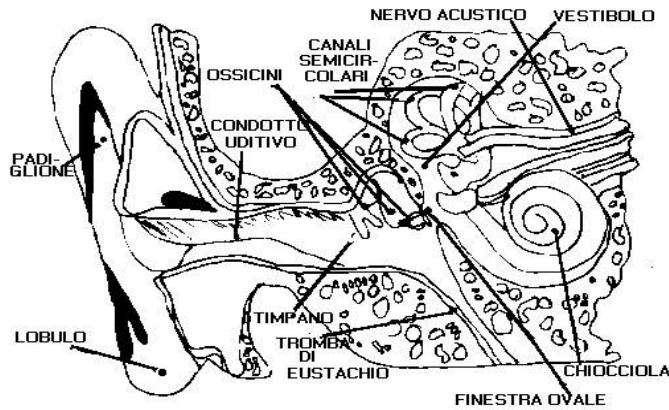


Fig. 2.1 Schematizzazione dell'organo dell'udito.

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 **perilinf**a, e sono pieni di un altro liquido, l'**endolinf**a. 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 Reissner** 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 Reissner 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 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.

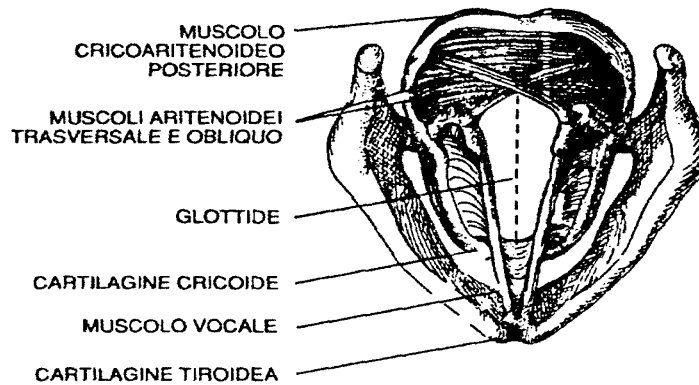


Fig. 2.2 Sezione longitudinale della laringe.

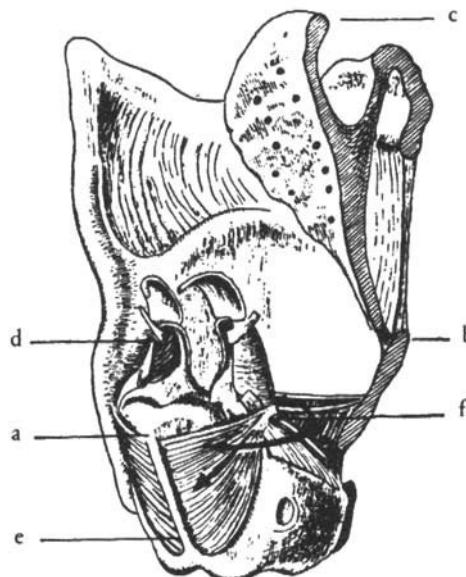


Fig. 2.3 La laringe vista da dietro. a: cartilagine cricoidea; b: cartilagine tiroidea; c: epiglottide; d: aritenoidi (sinistra); e, f: muscoli (le frecce indicano le direzioni di movimento).

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<sup>2</sup> con picchi tipici dell'ordine di 15 mm<sup>2</sup>.

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.



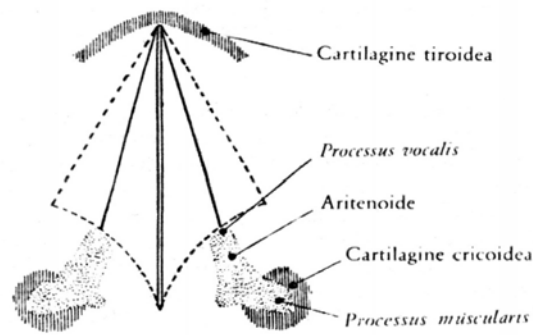


Fig. 2.4. Disegno schematico del meccanismo di apertura e chiusura della glottide. Le due linee più grosse indicano la posizione delle corde vocali durante la respirazione normale, le linee tratteggiate più grosse la posizione durante la respirazione profonda. Le due linee verticali sottili indicano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

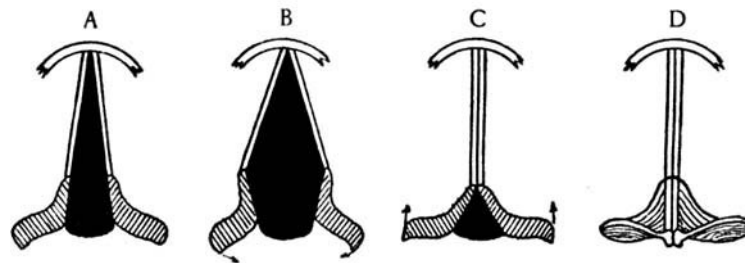


Fig. 2.5. Disegno schematico delle differenti posizioni della glottide: A respirazione normale, B respirazione profonda, C bisbiglio (le corde vocali sono chiuse ma il passaggio fra le aritenoidi resta libero), D fonazione. (Da J. Forchhammer).

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:

$$v = \frac{n}{2l} \sqrt{\frac{T}{\mu}} \quad n = 1, 2, 3, \dots \quad (1.1)$$

ove  $\mu$  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<sup>3</sup>.

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

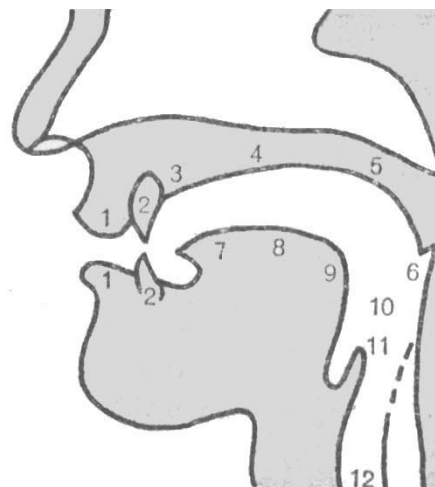


Fig. 1.6 .

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 toccare 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<sup>1</sup>, 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

---

<sup>1</sup> 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 di sotto 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)<sup>2</sup>. 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  $f_1$ ); le altre righe corrispondono alle frequenze  $f_2 = 2f_1$  (seconda armonica),  $f_3 = 3f_1$  (terza armonica) ecc.

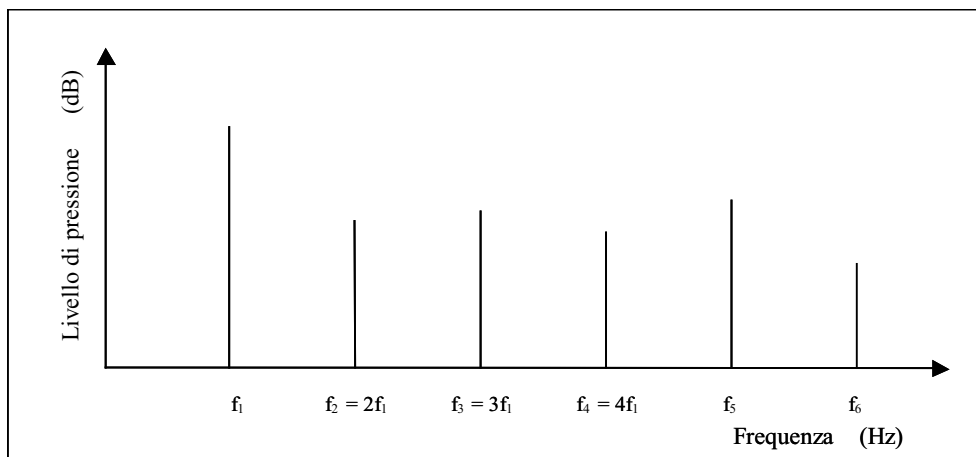


Fig. 2.1 Spettro acustico di un suono complesso.

---

<sup>2</sup> 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.<sup>3</sup>

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

---

<sup>3</sup> 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.)

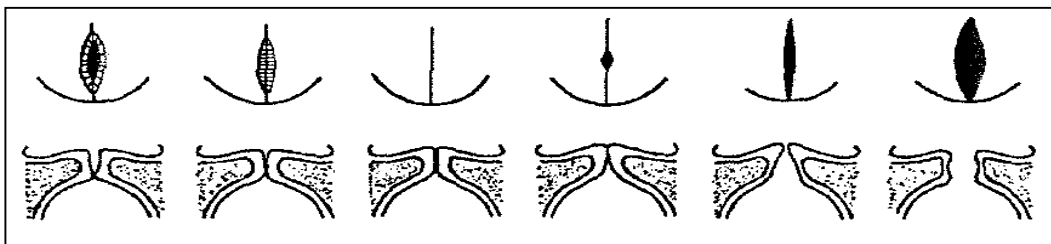
posizione della costrizione o dell'improvvisa apertura del condotto).

Come esempio di suoni sordi riportiamo le consonanti [p t k f s ʃ] 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<sup>4</sup> 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<sup>5</sup>. A questo punto l'aria

<sup>4</sup>Generalmente il valore della pressione dell'aria proveniente dai polmoni al livello glottale è di 7 cm H<sub>2</sub>O per il parlato normale, 2 cm H<sub>2</sub>O per un parlato appena percettibile, e di 20 cm H<sub>2</sub>O per un parlato a voce molto alta.

<sup>5</sup>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<sup>6</sup>.

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, ancora, 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  $1 \text{ cm}^3/\text{ciclo}$  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

---

<sup>6</sup>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*<sup>7</sup>. Data la natura periodica, il suo spettro è a righe, le cui componenti periodiche sono multipli interi della frequenza fondamentale. L'involuppo 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).

---

<sup>7</sup>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 <sub>0</sub> minima (Hz)	F <sub>0</sub> media (Hz)	F <sub>0</sub> 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<sub>0</sub>=500 Hz e del 20% per F<sub>0</sub>=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<sub>0</sub> anche dalle frequenze formanti. Vediamo, come abbiamo fatto nel precedente paragrafo per la F<sub>0</sub>, 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<sup>8</sup>. 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)

<sup>8</sup>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 della lunghezza del condotto vocale.

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

$$\nu = \frac{c}{4l}(2n+1) \quad n = 1, 2, 3, \dots$$

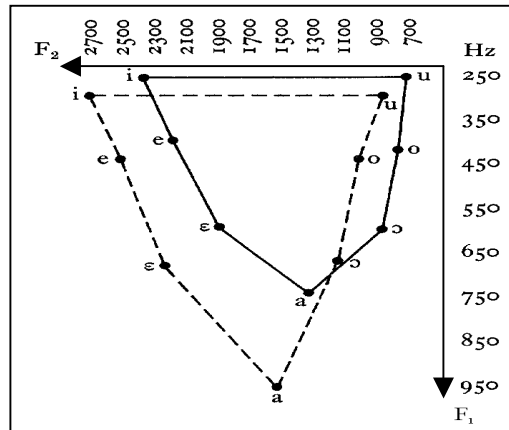


Fig. 2.3 Medie delle prime due formanti dei sette vocoidi tonici italiani: voci maschili (linea continua) e femminili (linea tratteggiata) sovrapposte. (Canepari, 1979).

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  $\nu$ : 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<sup>9</sup> di un parlatore è dell'ordine di 20  $\mu$ W 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  $\mu$ W (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:

---

<sup>9</sup> 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.

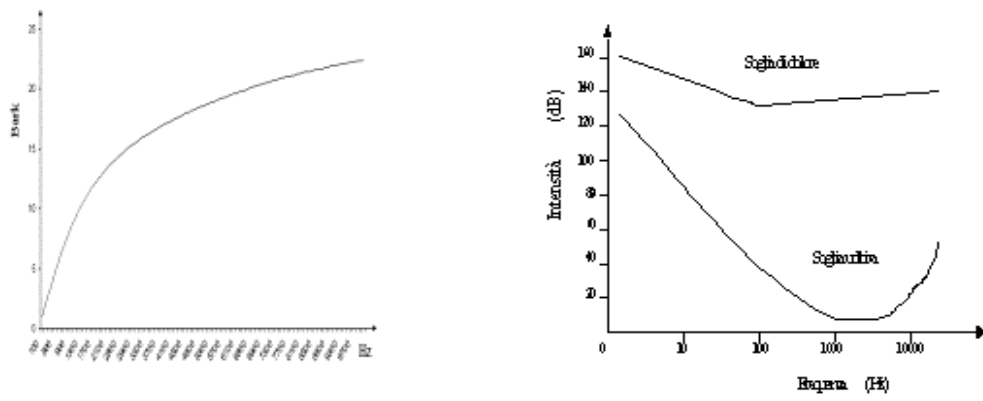
$$Bark = 13 \cdot \arctg(0.76 \cdot f_{kHz}) + 3.5 \cdot \arctg\left(\frac{f_{kHz}}{7.5}\right)^2 \quad (1.3)$$

L'effetto della trasformazione è una compressione dei valori in frequenza ( $5\text{kHz} = 18.54\text{B}$ ), 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  $\tau$  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, higher 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.

<b>Parametro</b>	<b>Descrizione</b>
<b>f1, f2, f3, f4</b>	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..
<b>f0</b>	Frequenza fondamentale di vibrazione delle corde vocali. E' data un decimi di Hz.
<b>ag</b>	Area dell'apertura della glottide. Il range di variazione normale è tra 0 e 40 mm <sup>2</sup> . Il valore medio per suoni sonori è di circa 3 - 5 mm <sup>2</sup> .
<b>al</b>	Area trasversale della costrizione formata dalle labbra durante la produzione delle consonanti. Il range di variazione è tra 0 e 100 mm <sup>2</sup> . Il valore 100 mm <sup>2</sup> corrisponde alla configurazione senza costrizione.
<b>ab</b>	Area trasversale della costrizione formata dalla lingua durante la produzione delle consonanti. Il range di variazione è tra 0 e 100 mm <sup>2</sup> . Il valore 100 mm <sup>2</sup> corrisponde alla configurazione senza costrizione
<b>an</b>	Area trasversale della costrizione del velo faringeo. Il range di variazione è tra 0 e 100 mm <sup>2</sup> .
<b>ue</b>	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.
<b>ps</b>	Pressione subglottale. Permette di aumentare o diminuire l'intensità del segnale prodotto. L'unità di misura è in cm di H <sub>2</sub> O.
<b>dc</b>	Variazione percentuale dell'elasticità delle pareti dell'apparato fonatorio durante la pronuncia.
<b>ap</b>	Area dell'interstizio glottale posteriore che persiste attraverso un ciclo glottale. L'unità di misura è mm <sup>2</sup> .

**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).**



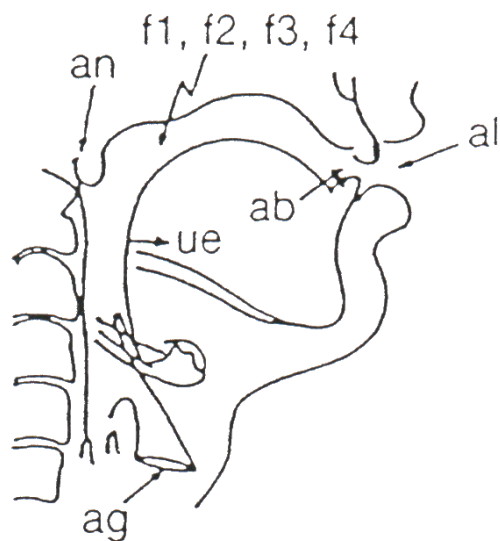


Figura 2.5 Legame tra i parametri del sintetizzatore e le caratteristiche del tratto vocale.

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

	ag	al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	4.000	100.0	100.0	0.0	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
250.0	4.000	100.0	100.0	0.0	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
270.0	4.000	100.0	100.0	15.00	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
290.0	4.000	100.0	5.000	30.00	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
465.0	4.000	100.0	5.000	30.00	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
485.0	4.000	100.0	100.0	27.14	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
505.0	4.500	100.0	100.0	24.29	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
520.0	5.000	100.0	100.0	22.14	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
545.0	5.500	100.0	100.0	18.57	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
555.0	6.000	100.0	100.0	17.14	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
570.0	7.000	100.0	100.0	15.00	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
585.0	9.000	100.0	100.0	12.86	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
630.0	10.00	100.0	100.0	6.429	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
631.0	0.0	100.0	100.0	6.286	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0
675.0	0.0	100.0	100.0	0.0	0.0	1070	750.0	1300	2500	3500	8.000	0.0	0.0

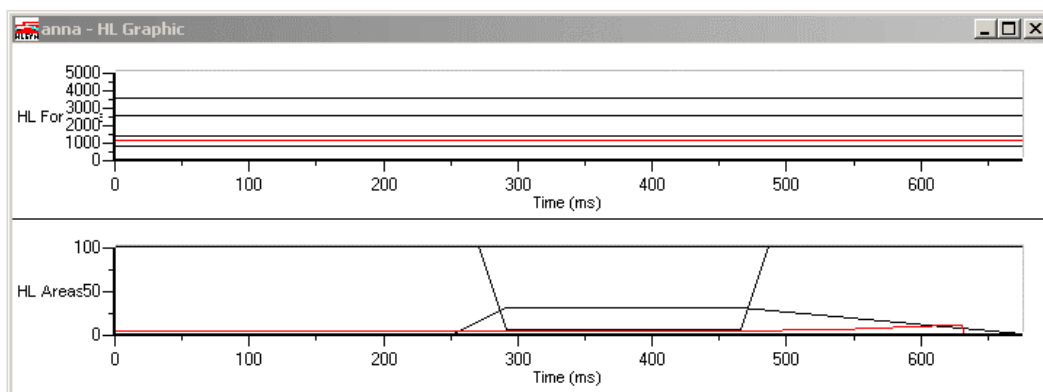
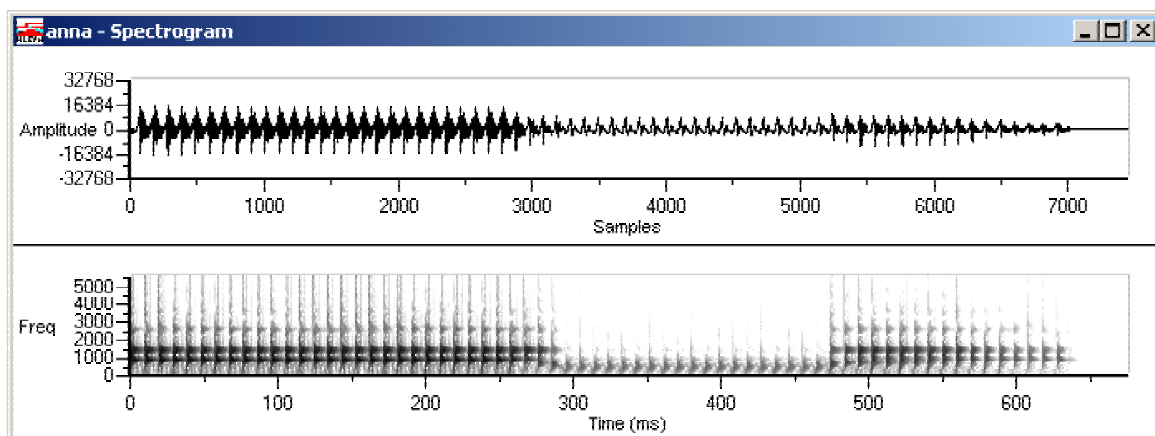


Figura 2.6 Finestra dei parametri HL di una parte di pronuncia. Si ricorda che il parametro  $f_0$  è in decimi di Hz mentre le frequenze formanti sono in Hz.

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  $f_c=10000$  Hz con 50 campioni per frame o  $f_c=11025$  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'.

# Capitolo 3

## SOFTWARE E LINGUAGGI DI PROGRAMMAZIONE UTILIZZATI

In questo capitolo vedremo i programmi ed i linguaggi utilizzati per implementare i concetti esposti nel primo capitolo, evidenziandone le caratteristiche ed accennando, ove necessario, la sintassi delle varie istruzioni. Per approfondimenti ulteriori invece si rimanda ai manuali di riferimento disponibili on-line. Ricordiamo che uno degli scopi del presente lavoro è quello di rendere disponibile su **Web** i risultati ottenuti, sia per la consultazione dei procedimenti utilizzati, sia come elaborazione in tempo reale delle parole immesse nel sistema. Come visto nel capitolo precedente, il sintetizzatore utilizzato per questo lavoro richiede una piattaforma **Windows** compatibile per poter girare. Poiché non è automatizzabile l'apertura del programma da un altro applicativo, è comunque possibile far girare il programma su di un'altra piattaforma ed importare solo il file da sintetizzare. Per comodità operativa nel nostro caso si è utilizzato **Windows 2000**. I criteri guida nella scelta degli applicativi sono stati molteplici:

- 1) software in grado di integrarsi in maniera semplice con le pagine **Web**
- 2) grande flessibilità nella gestione delle stringhe
- 3) realizzazione di un programma multiplatforma il cui funzionamento fosse indipendente dal sistema operativo
- 4) contenimento della spesa per l'uso dei software necessari

Il primo punto riguarda la possibilità di realizzare un sistema user friendly, basato sulle note e largamente diffuse pagine **Web**, permettendo l'accesso tramite rete **Internet** a tutto il lavoro sviluppato.

Il secondo riguarda la presenza nel linguaggio di programmazione di funzioni potenti e versatili nella elaborazione delle stringhe, che non si limitino alla semplice concatenazione o estrazione di sottostringhe.

Il terzo assicura l'adattabilità del progetto a qualsiasi tipo di sistema operativo.

L'ultimo criterio viene soddisfatto utilizzando software di tipo freeware, rendendo assolutamente svincolato il progetto da eventuali obblighi economici per l'acquisizione di licenze e quindi a spesa minima, al di là del costo marginale del lavoro di programmazione.

Ottemperare a tutte queste esigenze è stato possibile tramite l'impiego di un pacchetto di programmi

gratuito che va sotto l'acronimo di **L.A.M.P.** Ogni lettera è l'iniziale del nome di un software: **L** sta per **Linux**, il sistema operativo che in questo caso non è stato utilizzato per quanto detto sul sintetizzatore **HLsyn**; **A** sta per **Apache**, ed è un **Web** server che copre il 67% del mercato ed è completamente gratuito, **M** sta per **Mysql**, ed è un database; **P** è l'iniziale di **PHP**, un linguaggio di programmazione estremamente versatile che richiama molte strutture del **PERL** e del **C++**. Oltre questi software, sono stati utilizzati il programma **Spyguru**, per la decodifica del file di testo necessario per caricare i dati all'interno del sintetizzatore; il **JavaScript**, per operazioni di verifica ed elaborazione lato client, il **DISC 95**, un vocabolario in formato elettronico che ha permesso di rilevare proprietà statistiche della lingua italiana e ricavare un nutrito database di parole utilizzate per il corretto funzionamento dell'intero progetto. Per finire citiamo il programma di scrittura **Word 2000** con il quale è stata redatta la tesi.

## 3.1 PERCHE' PHP?

---

Oltre i motivi sopra riportati ci sono altri aspetti che hanno fortemente influito nell'adozione di questo linguaggio di programmazione: ogni pagina **Web** è scritta con un linguaggio denominato **HTML**, che permette di gestire la formattazione di una pagina con l'inserimento di testo, di links, immagini, sfondi etc.. in maniera rigida e determinata. Infatti, potremmo dire che l'**HTML** non rappresenta un vero e proprio linguaggio di programmazione, ma più una sorta di specifiche relative ad i vari attributi, in quanto è possibile con qualsiasi editor di pagine **Web**, posizionare gli elementi desiderati nella pagina ed ottenere il codice relativo, cosa impossibile per qualsiasi linguaggio di programmazione vero e proprio: la categoria dei programmatori inciderebbe in maniera drammatica sul tasso di disoccupazione mondiale! Si intuisce come una siffatta pagina **Web** non possa interagire con un eventuale utente o comunque con un sistema remoto, in quanto rappresenta una semplice visualizzazione di più elementi assemblati insieme: è, cioè, una pagina **statica**.

Per far sì che la pagina **Web** diventi dinamica e quindi attiva, occorre poter prelevare i dati inseriti, analizzarli ed inviarli ad un'altra pagina deputata all'elaborazione tramite un protocollo specifico e riceverne il risultato.

Esistono due tipi di programmi nel mondo **Web**: uno detto lato **client**, cioè residenti sull'**host** richiedente il servizio, come ad esempio il **Javascript**, ed uno detto lato **server**, cioè residenti sull'**host** server. Per questi ultimi, la connettività può essere realizzata con due modalità: tramite l'uso di **CGI** (common gateway interface), cioè interfacce appositamente studiate per dialogare con il **Web server** ed usate ad esempio nel **Perl**, o tramite l'uso di linguaggi "immersi" nel codice **HTML** come l'**ASP**.

Il **PHP** fa parte di questa famiglia e può essere tranquillamente incorporato nelle varie parti di codice **HTML**. Il **PHP** è un linguaggio interpretato ed il suo output rappresenta proprio codice **HTML**. La sua sintassi deriva direttamente dal **C/C++** e richiede un brevissimo periodo di addestramento per chi ha già esperienza di programmazione con il suddetto linguaggio. Essendo poi un codice di tipo immerso, la generazione di codice **HTML** risulta semplificata in confronto all'uso di **CGI**. L'aspetto fondamentale del suo uso in questo contesto, risiede nella vastissima libreria di funzioni dedicate alle stringhe, estremamente versatile, ed alla implementazione di una tecnica di programmazione che va sotto il nome di "espressioni regolari" o **REGEXP**. Inoltre, presenta piena compatibilità con le **PCRE**, le espressioni regolari del **Perl**, forse il linguaggio più potente per l'elaborazione delle stringhe.

Come abbiamo visto nel capitolo I, dovremo fare un numero elevato di verifiche su parole e contesti fonologici e, quindi, controlli di elevata complessità su combinazioni di caratteri e stringhe per



riscontrare la presenza o meno di determinate proprietà. Queste operazioni risultano particolarmente semplici ed immediate con l'uso delle espressioni regolari permettendo di realizzare un codice compatto ed elegante a vantaggio sia della leggibilità che delle prestazioni. Descriviamo adesso le istruzioni, le funzioni e le espressioni regolari utilizzate per la stesura dei programmi, descrivendone le caratteristiche e la sintassi.

Nel **PHP** a differenza del **C**, non è necessario dichiarare le variabili prima del loro utilizzo ma basta semplicemente scriverne il nome per renderla disponibile alla parte di codice seguente. Al nome che contrassegna la variabile deve essere anteposto il simbolo "\$". Analogamente agli altri linguaggi il simbolo "=", assegna il valore riportato alla sua destra alla variabile scritta alla sua sinistra. Se il valore da assegnare ad una variabile è racchiuso fra virgolette, viene interpretato come una sequenza di caratteri o più propriamente una stringa. Facciamo un esempio:

```
$parola = "gioco";
```

con questa istruzione si è creata una variabile di nome parola alla quale viene assegnato un valore di tipo stringa – *gioco* -. Il simbolo di ";" indica la fine dell'istruzione. Si noti inoltre come non sia necessario specificare il tipo di variabili, che risulta implicito nell'operazione di assegnazione. Questo aspetto, benché comodo, potrebbe portare difficoltà di lettura o addirittura ad errori di programmazione nel caso si dichiarino operazioni di conversione di tipo (**casting**): occorre fare attenzione ad utilizzare nomi diversi per le variabili oppure tenere memoria dell'ultima conversione affinché possano essere correttamente svolte le operazioni successive.

I tipi di dato presente nel **PHP** sono quelli che comunemente troviamo in qualsiasi linguaggio: interi, reali, stringhe, vettori, vettori associativi. Ricordiamo che un vettore associativo è un vettore che invece di utilizzare un indice numerico per la sua scansione, utilizza delle stringhe denominate parole chiave.

Le operazioni di tipo aritmetico, logico, per l'esecuzione di cicli (for o while) o verifiche di condizione (if else) sono quelle implementate in qualsiasi linguaggio di programmazione e la sintassi è simile a quella del linguaggio **C**. In questo linguaggio non è obbligatoria l'indentazione, benché utilizzata in questo lavoro per rendere più leggibile e chiaro il programma. Saranno invece esaminate con un certo dettaglio le operazioni e le funzioni applicabili alle stringhe, corredandole con esempi per capirne sia la funzionalità che la sintassi; le operazioni utilizzabili sui vettori; le istruzioni per prelevare i dati dal protocollo **HTTP** o per renderle disponibili in una sessione di lavoro, le righe di codice necessarie per potere interfacciare il **PHP** con il database **Mysql**, le funzioni per la scrittura e la lettura da file.

### 3.1.1 Operazioni su stringhe

```
$parola="gioco";
```

La variabile indicata fra virgolette è considerata di tipo stringa.

```
$lunghezza=strlen($appoggio);
```

La funzione **strlen** permette di misurare la lunghezza della stringa contenuta fra parentesi. Il valore viene assegnato alla variabile \$fine.

```
$stringa=substr($appoggio,0,$pos).substr($appoggio,$pos+1,$fine);
```

In questa espressione troviamo l'operatore di concatenazione “.” Usato per unire variabili stringhe e l'operatore **substr**, che genera ancora variabili di tipo stringa. Gli argomenti rappresentano nell'ordine: la stringa di cui si vuole prendere una parte, il punto iniziale della sottostringa scelta e la sua lunghezza. Questi valori possono essere numeri o anche variabili come nel nostro esempio.

### 3.1.2 Operazioni su vettori

```
$vocali=array("à"=>"a","è"=>"e","ò"=>"o","é"=>"e","ó"=>"o","ì"=>"i","ù"=>"u");
```

Questa espressione permette di creare un array inserendone i valori in un sol colpo. E' anche possibile aggiungerne in sequenza con l'espressione:

```
$vocali[]=$valoredaimmettere;
```

In questo esempio si è creato un array associativo, utilizzando come campi chiave le lettere indicate fra virgolette. Omettendo i campi chiave si sarebbe ottenuto un vettore con la classica indicizzazione.

```
$vettore=explode("-", $stringa);
```

Questa espressione permette di ottenere un vettore da una stringa utilizzando come simbolo separatore quello contenuto fra virgolette. Ad esempio la parola *ca-sa*, verrebbe memorizzata dopo l'**explode**, in un array bidimensionale il cui primo elemento conterrebbe *ca*, ed il secondo *sa*.

```
$stringa=implode("", $vettore);
```

Questa espressione attua esattamente la funzione inversa alla precedente. Anche in questo caso si può utilizzare un simbolo di unione. Si veda anche la funzione **join**.

```
count($vettore);
```

Questa funzione restituisce il numero di elementi contenuti in un vettore.

```
array_pop($vettore);
```

Questa funzione elimina l'ultimo elemento di un vettore.

```
array_push($vettore,$ultimoelemento);
```

Questa funzione inserisce come ultimo elemento di un vettore la variabile contenuta nel secondo argomento.

```
$vettore3=array_merge($vettore1,$vettore2);
```

Questa funzione restituisce un terzo vettore contenente gli elementi dei vettori passati come argomento.

### 3.1.3 I dati nel protocollo HTTP

```
$stringaapp=$HTTP_GET_VARS[caratteri];
```

Il valore nell'array associativo contenente i dati trasferiti su protocollo HTTP in modalità GET, è caricato nella variabile \$stringaapp.

### 3.1.4 Operazioni su file e funzioni

```
$file2=fopen("../convertitore_alfabeto/fileUTF/".$nomefile2.".UCT","w");
```

Questa funzione apre un file restituendo il puntatore ad esso. Nel primo argomento è specificato il percorso del file, nel secondo la modalità di apertura, se in lettura, scrittura od aggiunta.

```
fputs($file2,$HTTP_GET_VARS[fono]);
```

Questa funzione inserisce il valore contenuto nel secondo argomento, nel file il cui puntatore è passato nel primo argomento.

```
fclose($file2);
```

Questa funzione chiude il file il cui puntatore è passato come argomento.

```
include("funzioni_synth/caricavettore.php3");
```

Questa funzione carica in memoria la funzione richiamata dal programma principale il cui percorso è specificato nell'argomento.

```
<?if($dir=opendir("../convertitore_alfabeto/fileUTF")){  
    while(false!==($file=readdir($dir))){
```

In questo caso si è utilizzato come esempio una parte di ciclo, per mostrare direttamente un utilizzo pratico: se la directory, il cui percorso è specificato nell'argomento della **opendir**, è aperta con successo, vengono caricati tutti i file presenti con **readdir** finché questa funzione li trova al suo interno.

### 3.1.5 Istruzioni di sessione

```
session_start();
```

Questa espressione apre una sessione di lavoro.

```
session_register("vai","omo","vettoreomo","dueomo","vettoredistringhe","pas","login","user");
```

Questa espressione inserisce nel registro di sessione tutte le variabili che saranno utilizzate.

```
session_unregister("vettoreomo");
```

Questa espressione elimina dal registro di sessione la variabile specificata.

### 3.1.6 Interazione con mysql

```
$conn = mysql_connect() or die ("non c'e");
```

Questa espressione apre una connessione per l'uso del db. In caso di esito positivo restituisce l'indicativo di connessione altrimenti visualizza un messaggio di errore tramite la funzione **die**.

```
mysql_select_db("GEMMA",$conn) or die ("no db");
```

Questa espressione seleziona il db con l'indicativo di connessione ottenuto in precedenza. In caso di esito negativo visualizza un messaggio di errore tramite la funzione **die**.

```
$sql="SELECT * FROM prefissisuffissi WHERE tipo='s' ";  
$result = mysql_query($sql,$conn);
```

La prima espressione è un esempio di comando SQL mentre la seconda è la funzione che invia la query al db utilizzando l'indicativo di connessione. Ritorna una variabile booleana indicante l'esito dell'operazione.

```
while($suffisso=mysql_fetch_row($result)...
```

E' indicato un frammento di codice per indicare che se \$result è diversa da zero, esito positivo della query inviata, viene eseguita l'operazione di fetch, cioè caricamento, dei dati. Una modalità alternativa è **mysql\_fetch\_array**, che carica direttamente in un array i dati prelevati.

### 3.1.7 Espressioni regolari

Un'espressione regolare può essere considerata a tutti gli effetti un micro linguaggio inserito all'interno del linguaggio stesso, permettendo di effettuare una serie di controlli, anche molto complessi, con una stringa di codice estremamente compatto e leggibile. Come accennato troveremo due nomi per descrivere le espressioni regolari in **PHP**: le **REGEXP**, contrazione di regular expression, proprie del **PHP**, le cui istruzioni hanno la parola chiave **ereg**, e le **PCRE**, cioè **Perl Compatible REGEXP**, le cui istruzioni hanno la parola chiave **preg**. Queste ultime sono state le più usate nel presente lavoro poiché presentano un grado di flessibilità maggiore rispetto alle altre. Un'espressione regolare permette di individuare un carattere od una sequenza di caratteri specifica all'interno di una stringa, restituendo una variabile di tipo booleano in risposta. Ad esempio se volessimo verificare la presenza della sequenza di caratteri "io" all'interno della parola gioco scriveremmo:

```
$parola="gioco";  
  
ereg("io",$parola);  
preg_match("/io/",$parola);
```

In questo semplice caso l'istruzione verifica se all'interno della variabile parola sia presente la sequenza di caratteri io: poiché alla variabile è stato assegnato il valore gioco l'espressione regolare darà un esito positivo, viceversa si avrebbe nel caso parola fosse stata giacca. Nel caso sia necessario memorizzare la sequenza, o pattern, riscontrata all'interno della variabile in esame, è possibile introdurre una variabile di appoggio, che chiameremo solo per comodità **\$match**, nella quale viene memorizzato un vettore, contenente l'intera stringa nel primo elemento e nei successivi gli eventuali pattern trovati. Per attivare la funzione di memorizzazione è necessario però inserire fra parentesi tonde la parte di espressione che eventualmente interessa. Riprendendo l'esempio precedente, e non sapendo se a priori la parola che contenga io sia gioco o giocare, potremo scrivere la seguente espressione:

```
preg_match("/io(\w+)/",$parola,$match);
```

all'interno del primo elemento di **\$match** troveremo **co** o **care** a seconda del valore di parola. Oltre queste semplici operazioni di matching tra variabili è anche possibile modificare un'intera sequenza di caratteri con altre od addirittura, vettore di caratteri, tramite l'espressione **preg\_replace**. Di seguito riportiamo un esempio:

```
$parola=preg_replace("/io/","e",$parola);
```

questa istruzione sostituisce nella variabile parola la sequenza di caratteri io con la sequenza costituita dal solo carattere e: si noti che l'operazione non sarebbe visibile senza una preventiva assegnazione del risultato della stessa ad una variabile che, in questo caso, è la stessa sulla quale si opera il controllo. Nulla avrebbe vietato di riassegnare il risultato ad un'altra variabile lasciando quindi immutata quella originale. Tramite l'uso di operatori specializzati, è possibile controllare la presenza della sequenza di caratteri in testa alla variabile, in coda od in qualsiasi posizione; controllare ad esempio il numero di ripetizioni oppure la possibilità di combinarla con altri caratteri. Questi operatori sono utilizzabili con ciascuna delle espressioni regolari implementate nel linguaggio. Sono riportati nell'elenco seguenti le espressioni regolari utilizzate in questo lavoro, che rappresentano solo una parte di quelli implementati

dal linguaggio. Per ulteriori chiarimenti si rimanda al manuale.

```
preg_match("/(cg)([aeou])/i",$sequenza,$match);
```

In questa espressione si verifica se vi siano nella stringa `$sequenza`, caratteri *c*, *g* seguiti da una vocale: all'interno del vettore `$match` troveremo nell'elemento di indice uno il valore riscontrato nella prima parentesi ed analogamente per il secondo. Nel vettore di indice zero avremo tutta la sequenza riscontrata.

```
$stringa=preg_replace("/grafema/i","fonema",$stringa2);
```

Se nella variabile `$stringa2` è riscontrata la sequenza grafema, la si sostituisce con fonema ed il risultato è posto nella variabile `$stringa`.

```
$simboliappoggitre=array("/ó/", "/è/", "/é/", "/ò/");  
$simbolifoneticitre=array("o", "e", "e", "o");  
$tre=preg_replace($simboliappoggitre,$simbolifoneticitre,$tre);
```

Questa funzione è analoga alla precedente ma utilizza un vettore di variabili.

Questa panoramica sulle funzioni principalmente usate nel programma dovrebbe fornire gli strumenti necessari per la comprensione dei codici riportati nel manuale operativo del **CRISTAL**, dove sono esaminati tutte le funzioni ed i programmi del progetto.

## 3.2 GLI ALTRI SOFTWARES UTILIZZATI

---

### 3.2.1 APACHE

Come già accennato, il **PHP** è un linguaggio lato server, e come tale ha bisogno di un programma per il trasferimento dei dati tramite protocollo **http**, cioè un Web server e la possibilità, quindi, di inviare dati da una pagina all'altra durante la navigazione tramite il browser. Nello specifico è stato utilizzato il programma **Apache**. In rete è inoltre possibile reperire documentazione e faq per la sua installazione e configurazione.

### 3.2.2 MYSQL

La base di dati, è una struttura che permette di immagazzinare dati ordinati secondo un determinato criterio, come si farebbe ad esempio in un negozio, per realizzare l'inventario dei prodotti disponibili in un magazzino. All'interno di questa base di dati, è necessario creare delle tabelle, per poter catalogare dei dati aventi delle caratteristiche comuni: nell'inventario di un negozio di alimentari cercheremo di ordinare i prodotti in base alla loro deperibilità, o della provenienza, o del tipo. All'interno di ogni tabella poi, specificheremo, utilizzando come campo di ricerca il nome del prodotto, una serie di informazioni che potrebbero essere il prezzo, il peso ecc.

Una organizzazione logica e razionale del database permette di ottenere buone prestazioni generali in quanto la ricerca sarà effettuata nel minor tempo possibile. Per questa applicazione specifica, uno qualsiasi dei database disponibile, sarebbe andato bene. Fra i più utilizzati in coppia con il **PHP** troviamo il **Postgres** ed il **Mysql**. Fra i due è stato scelto quest'ultimo perché, pur essendo meno versatile a livello di operazioni su db, risulta estremamente veloce ed implementa comunque, tutte le funzionalità necessarie al progetto. Considerando che la base di dati potrebbe essere in continuo aggiornamento, e di conseguenza aumenterebbero i tempi di accesso all'intero sistema, si è preferito privilegiare quest'aspetto. Naturalmente inserire all'interno del codice dati presenti nel db avrebbe sicuramente velocizzato l'intero sistema ma non è pensabile poter inserire centinaia di vocaboli con le loro proprietà all'interno delle varie funzioni poiché qualsiasi aggiornamento della lista di vocaboli, avrebbe richiesto la modifica del codice. Con una base di dati esterna invece, è possibile aggiornare semplicemente quest'ultima indipendentemente dal programma, ed in più creare una procedura di correzione automatica, che permetta di inserire ogniqualvolta si riconosca un errore nella trascrizione della parola immessa, il dato corretto all'interno della lista.

Vengono di seguito riportate le istruzioni del linguaggio **SQL** utilizzate nel software rimandando invece all'organizzazione della base di dati al manuale operativo del **CRISTAL**. Si faccia attenzione all'uso dei due tipi di virgolette, necessari al corretto funzionamento della query.

```
SELECT * FROM `apertechiuse` WHERE grafema= '$parola'
```

Questa espressione permette di effettuare una selezione degli elementi nella tabella specificata dopo la parola **FROM**, dove l'elemento **grafema** assume il valore contenuto nella variabile **\$parola**. L'asterisco dopo **SELECT** indica su tutti gli elementi.

**DELETE FROM `apertechiuse` WHERE `grafema`= '\$sparola'**

Questa espressione permette di effettuare una cancellazione nella tabella specificata dopo la parola **FROM**, dove l'elemento **grafema** assume il valore contenuto nella variabile **\$sparola**.

**INSERT INTO `apertechiuse`(`grafema`,`tipo`,`posizione`)VALUES('\$sparola','\$tipo','\$posizione')**

Questa espressione permette di effettuare l'inserimento degli elementi contenuti dopo la parola **VALUE**, nella tabella specificata dopo la parola **INTO**, nei campi indicati dopo il nome della tabella.

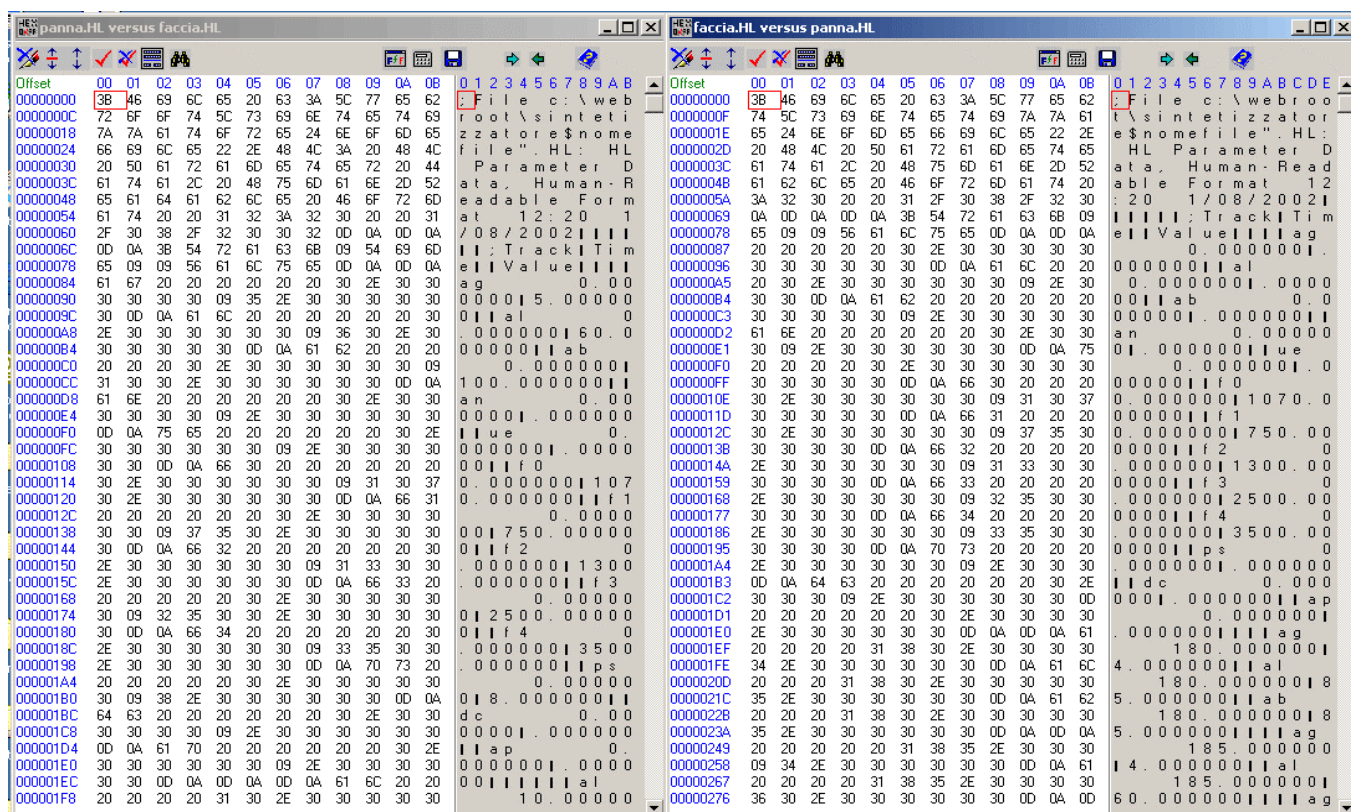
### 3.2.3 Spyguru

Questo piccolo programma, liberamente distribuito in rete, è un editor esadecimale che permette di visualizzare i codici relativi ad ogni simbolo all'interno di un file. Come visto in precedenza, il sintetizzatore **HLSyn** può comunicare con l'esterno solo attraverso dei files di testo opportunamente formattati: La procedura seguita per la decodifica è stata la seguente:

- 1) è stato generato un file di testo dal sintetizzatore
- 2) si è ricreato un file apparentemente uguale in base alla visualizzazione dal monitor
- 3) tramite la modalità versus di **Spyguru** sono stati aperti entrambi i files e si è effettuato il confronto simbolo per simbolo
- 4) ad ogni codice esadecimale è stato associato il relativo simbolo

La tabella con le corrispondenze è riportata nel **CRISTAL REFERENCE**.

Di seguito mostriamo un esempio della modalità versus.

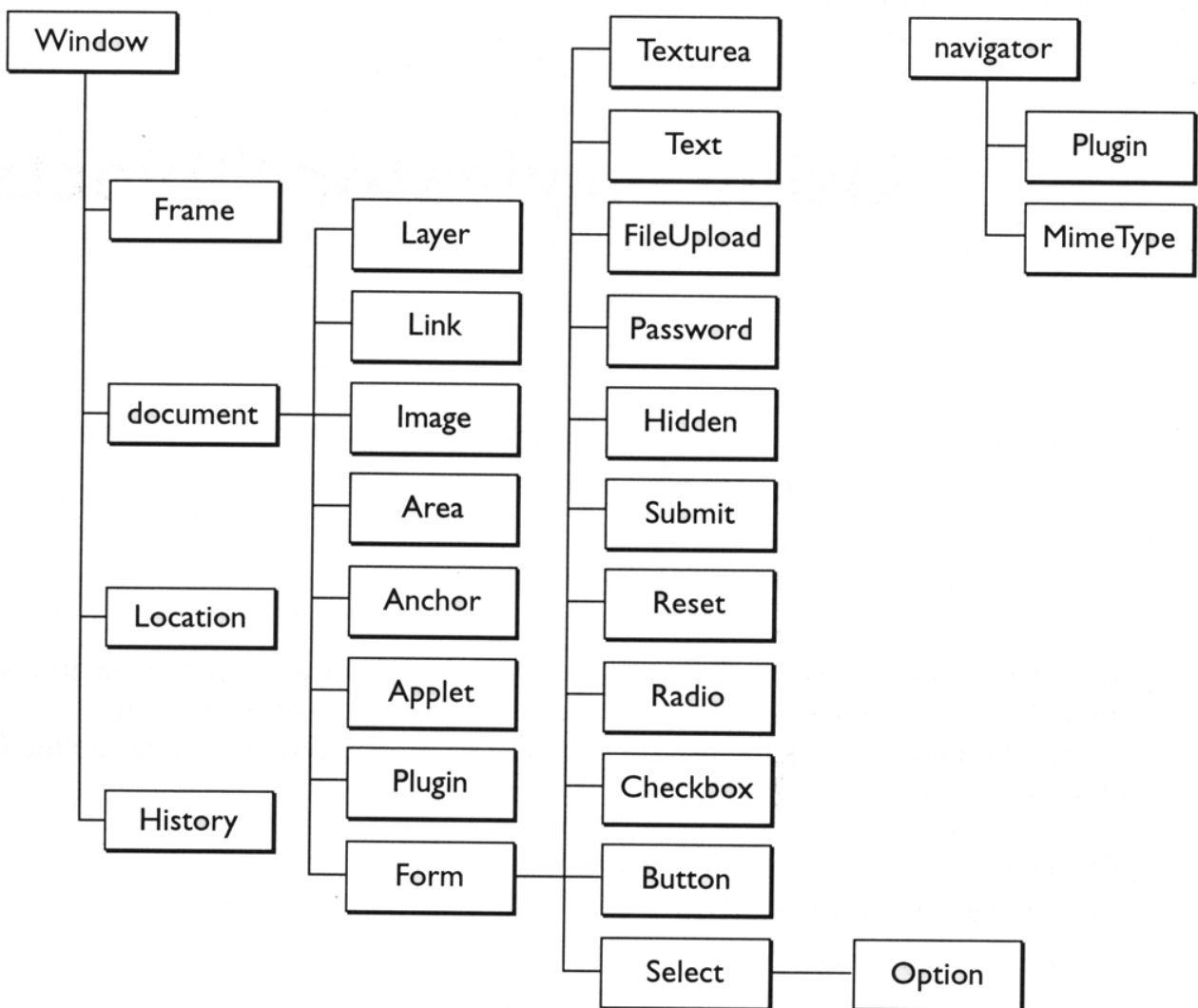




### 3.2.4 JavaScript

Il **JavaScript**, che come abbiamo visto è un applicativo lato client, è stato utilizzato per rendere interattive le pagine di immissione dati e poter inserire correttamente le informazioni necessarie alla corretta esecuzione del programma. E' possibile infatti attivare determinate funzioni di inserimento dati solo in conseguenza di altre azioni, come la pressione di un bottone o la selezione di una **checkbox**. Nel caso della procedura di correzione, infatti, è necessario indicare se la parola in esame appartenesse alla categoria omografi per poterla inserire opportunamente nella base di dati. Nel caso della sintesi, invece, occorre poter specificare il nome del file da creare solo dopo aver selezionato l'opzione corrispondente, per non generare files vuoti ad ogni invio dati.

Il **JavaScript** tratta tutti gli attributi **HTML** di una pagina come degli oggetti, associando loro delle proprietà e delle operazioni di modifica. Questa associazione avviene secondo il seguente schema ad albero:



Ogni proprietà è accessibile con la seguente sintassi:

```
document.correzione.elements["attiva_<? echo $stringa2[1];?>"].disabled=true;
```

Ad ogni punto corrisponde la selezione di un ramo dell'albero e l'ultima istruzione è la responsabile dell'assegnazione del valore desiderato. Nelle parentesi quadre è indicato il nome dell'elemento a cui ci si riferisce: in questo caso il nome è creato da codice **PHP**. E' possibile definire delle funzioni richiamabili da particolari eventi **HTML**, gli **event handlers** associati ad operazioni compiute dall'utente nell'interazione con il browser, come la selezione di una casella, la pressione su un bottone ed altro. Nel manuale di riferimento è possibile consultare l'elenco completo. Un esempio di uso di uno di questi **event handlers** è il seguente:

```
<INPUT type="checkbox" name="attiva_<? echo $stringa2[1];?>" onclick="stacca<?echo $indexdis; ?>('1','1');" >
```

Nel momento in cui viene selezionata la casella relativa della **checkbox**, quindi con un "click" sull'oggetto in esame, l'evento **onclick** è attivato ed invoca la funzione "stacca" con gli opportuni parametri. Questa funzione è utilizzata nelle pagine di correzione e si rimanda al **CRIiSTAL MANUAL** per vederne il funzionamento e gli effetti. Un altro uso dell'**onclick** si trova nella pagina dell'ascolto degli esempi **Wave** mentre nella pagina di scelta files **UTF** è usato l'evento **OnChange**. Per ulteriori approfondimenti si rimanda al manuale del linguaggio.

### 3.2.5 PHP-MYADMIN-2.2.0

Questo applicativo permette di utilizzare un'interfaccia di tipo **WEB** per interagire con il db senza dover utilizzare le righe di comando, permettendo una più veloce operatività ed una visione completa delle modifiche apportate.

# Capitolo 4

## SINTESI DEI FONEMI

In questo lavoro sono stati sintetizzati i fonemi consonantici sordi “**f**” e “**p**”, per poter realizzare in unione con i fonemi già presenti, dei lessemi completi. Essendo i fonemi di tipo sordo, non è stato utilizzato il parametro “**ue**”. Anche i valori “**dc**” e “**ap**” sono stati mantenuti a valori di default. Inoltre, non essendo fonemi che presentano fenomeni di nasalizzazione, il parametro “**an**” ha assunto sempre valore nullo. Per i restanti parametri illustriamo i valori scelti riportando i dati contenuti nelle due parole sintetizzate “*faccia*” e “*papà*”.

### 4.1 LA SINTESI DEI CONTOIDI

---

#### 4.1.1 Sintesi del contoide “**f**”

Il fonema “**f**” è di tipo costrittivo labio-dentale, in quanto c’è una restrizione della cavità orale a causa del contatto dei denti con il labbro inferiore. Ciò comporta la creazione di un rumore tipico nella fase espiratoria. Ricordiamo che “**f**” è sordo, per cui le corde vocali non intervengono nella produzione del suono. Questa fase è seguita da una rapida apertura della restrizione con la pronuncia della vocale seguente. Nella tabella seguente abbiamo la parte iniziale della parola “*faccia*” ottenuta con il programma: la sillaba “*fa*” inizia al tempo “0.0” e termina all’istante “484.0”. I valori di tempo sono espressi in ms. Esaminiamo singolarmente i parametri rilevanti per la sintesi.

##### **Ag**

Parte da un valore di 20 che permette il passaggio dell’aria senza provocare vibrazione delle corde vocali fino ad arrivare a 4, valore tipico dei suoni sonori, in corrispondenza della vocale **a**, al tempo 320. In corrispondenza della fase costrittiva, da 200 a 300, c’è un innalzamento del valore che provoca una diminuzione dell’ampiezza del segnale per la maggiore apertura della glottide.

##### **Al**

L’andamento di questo parametro è decrescente fino a raggiungere il minimo nell’istante iniziale della costrizione, per poi riaumentare fino al massimo in corrispondenza della produzione della vocale **a**. La fase di massima costrizione si estende per circa 100 ms da 200 a 300, mentre la posizione delle labbra varia per i primi 300 ms circa.

## **Ab**

I valori di **ab** sono stati raffinati per avere una pronuncia più naturale in combinazione con i valori di **al**, il responsabile della simulazione della costrizione, dato che anche la lingua modifica in parte la sua posizione, contribuendo alla modifica del cavo orale.

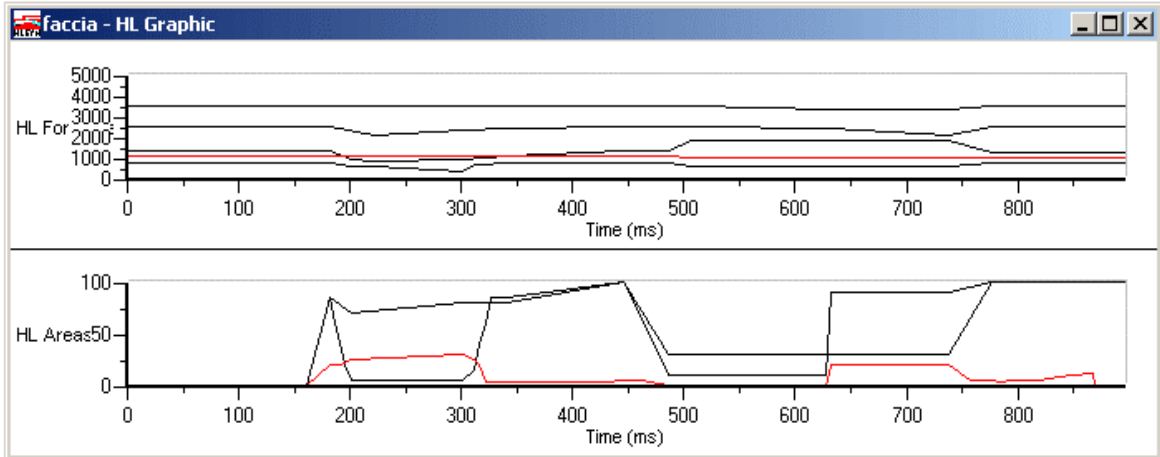
## **Formanti**

I valori delle formanti subiscono un decremento in corrispondenza della fase costrittiva parallelamente alla variazione del parametro **al**, a causa della modifica della cavità orale. Si noti come  $f_1$  subisce il massimo della diminuzione in corrispondenza della fine della costrizione e ritorni in breve tempo prossima al suo valore nominale.

## **Ps**

E' stato utilizzato un valore di 6.5, ritenuto adatto per l'ampiezza del segnale. Si confrontino i valori numerici dei parametri con l'andamento nel grafico sottostante.

	ag	al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc
180.0	20.00	85.00	85.00	0.0	0.0	1070	750.0	1300	2500	3500	6.500	0.0
185.0	20.00	60.00	81.25	0.0	0.0	1070	712.5	1200	2450	3500	6.500	0.0
190.0	20.00	40.00	77.50	0.0	0.0	1070	675.0	1100	2400	3500	6.500	0.0
195.0	23.00	15.00	73.75	0.0	0.0	1070	637.5	1000	2350	3500	6.500	0.0
200.0	25.00	5.000	70.00	0.0	0.0	1070	600.0	900.0	2300	3500	6.500	0.0
220.0	26.00	5.000	72.00	0.0	0.0	1070	600.0	800.0	2100	3500	6.500	0.0
300.0	30.00	5.000	80.00	0.0	0.0	1070	320.0	925.0	2300	3500	6.500	0.0
310.0	25.00	15.00	80.00	0.0	0.0	1070	670.0	961.2	2338	3500	6.500	0.0
315.0	20.00	40.00	80.00	0.0	0.0	1070	700.0	979.4	2356	3500	6.500	0.0
320.0	4.000	60.00	80.00	0.0	0.0	1070	710.0	997.5	2375	3500	6.500	0.0
325.0	4.000	85.00	80.00	0.0	0.0	1070	720.0	1016	2394	3500	6.500	0.0
340.0	4.000	85.00	80.00	0.0	0.0	1070	750.0	1070	2450	3500	6.500	0.0
410.0	4.000	95.00	93.33	0.0	0.0	1070	750.0	1223	2483	3500	6.867	0.0
445.0	4.350	100.0	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.051	0.0
460.0	4.500	73.08	65.38	0.0	0.0	1070	750.0	1300	2500	3500	7.129	0.0
480.0	0.7500	37.18	19.23	0.0	0.0	1070	750.0	1300	2500	3500	7.234	0.0
484.0	0.0	30.00	10.00	0.0	0.0	1070	750.0	1300	2500	3500	7.255	0.0
505.0	0.0	30.00	10.00	0.0	0.0	1040	600.0	1800	2500	3500	7.365	0.0
506.0	0.0	30.00	10.00	0.0	0.0	1040	600.0	1800	2499	3498	7.371	0.0



## 4.1.2 Sintesi del contoide “p”

Il fonema “p” è di tipo occlusivo bilabiale in quanto il passaggio dell’aria nella fase espiatoria è completamente bloccata dalle labbra. Ricordiamo che “p” è sordo, per cui le corde vocali non intervengono nella produzione del suono. Questa fase è seguita da una rapida apertura della occlusione con la pronuncia della vocale seguente.

Esaminiamo di seguito i parametri rilevanti.

### Ag

Parte da un valore di 29 che permette il passaggio dell’aria senza provocare vibrazione delle corde vocali fino ad arrivare a 4, valore tipico dei suoni sonori, in corrispondenza della vocale **a**, al tempo 195. In corrispondenza della fase occlusiva, da 15 a 115, rimane costante.

### Al

Questo parametro subisce una brusca diminuzione nei primi 15 ms arrivando al valore nullo, tenuto fino al tempo 115, corrispondente a tutta la fase occlusiva. L’aumento è speculare e termina a 140.

### Ab

I valori di **ab** sono stati raffinati per avere una pronuncia più naturale in combinazione con i valori di **al**, il responsabile della simulazione della occlusione, dato che anche la lingua modifica in parte la sua posizione, contribuendo alla modifica del cavo orale. Alla fine della fase occlusiva c’è un abbassamento rispetto al valore nominale.

### Formanti

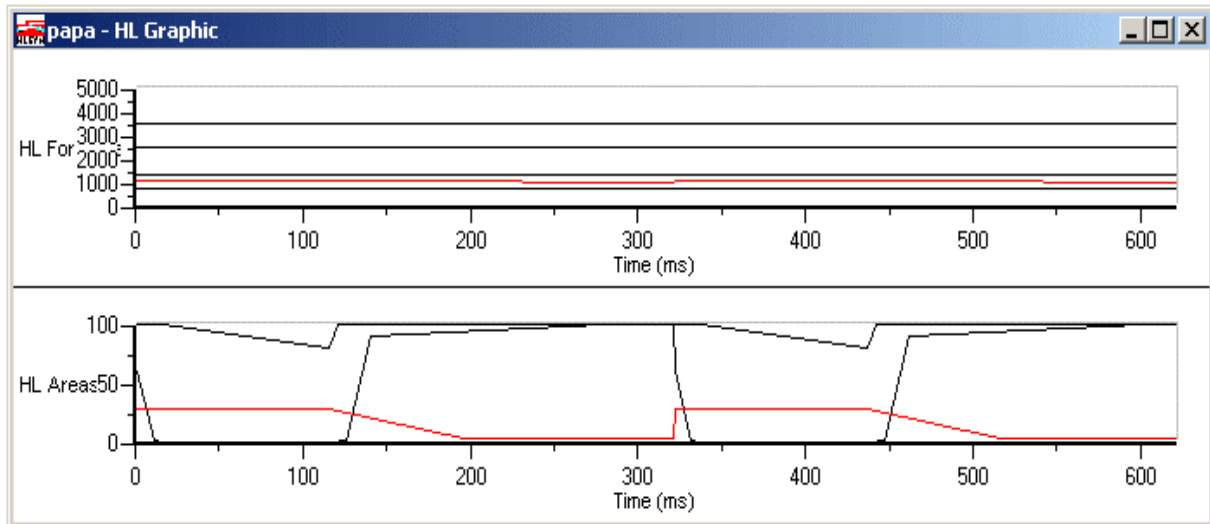
I valori delle formanti non hanno richiesto una particolare modifica, non influenzando in maniera significativa alla sintesi del fonema.

### Ps

E’ stato utilizzato un valore iniziale di 8, ridotto a 6.5 in corrispondenza della pronuncia della **a**, per diminuire l’ampiezza del segnale alla fine della sillaba.

Si confrontino i valori numerici dei parametri con l’andamento nel grafico sottostante.

	ag	al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc
0.0	29.00	60.00	100.0	0.0	0.0	1070	750.0	1300	2500	3500	8.000	0.0
10.00	29.00	2.000	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.964	0.0
15.00	29.00	0.0	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.946	0.0
115.0	29.00	0.0	80.00	0.0	0.0	1070	750.0	1300	2500	3500	7.589	0.0
120.0	27.44	1.000	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.571	0.0
125.0	25.88	2.000	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.554	0.0
140.0	21.19	90.00	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.500	0.0
195.0	4.000	93.93	100.0	0.0	0.0	1049	750.0	1300	2500	3500	7.304	0.0
260.0	4.000	98.57	100.0	0.0	0.0	1023	750.0	1300	2500	3500	7.071	0.0
280.0	4.000	100.0	100.0	0.0	0.0	1016	750.0	1300	2500	3500	7.000	0.0
320.0	4.000	100.0	100.0	0.0	0.0	1000	750.0	1300	2500	3500	6.500	0.0
321.0	29.00	60.00	100.0	0.0	0.0	1070	750.0	1300	2500	3500	8.000	0.0
331.0	29.00	2.000	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.964	0.0
336.0	29.00	0.0	100.0	0.0	0.0	1070	750.0	1300	2500	3500	7.946	0.0



## 4.2 LA SINTESI DEI FONI: LA COARTICOLAZIONE

---

Per realizzare la pronuncia di una parola completa è necessaria poter unire i fonemi fra loro: occorre cioè coarticularli. Finché questa procedura viene realizzata manualmente, è sempre possibile ottenere un risultato ottimale modificando opportunamente i parametri. Nel caso di trascrizione automatica, invece, bisogna studiare un meccanismo valido per ogni tipo di coarticolazione, affinché il programma sia funzionante in qualsiasi condizione di lavoro. Concretamente, sillabe formate da una sola vocale, libere od implicate devono comunque poter essere unite ottenendo una pronuncia più naturale possibile a prescindere dalla loro natura. Dalle prove e le analisi dei dati effettuate, è emerso che la condizione migliore di unione fra i fonemi, si ha mettendo in sequenza le sillabe componenti la parola. In effetti, anche per la sintesi della **p** e della **f**, più che una sintesi del fonema si è realizzata una sintesi della sillaba **fa**, **pa**. Modificando i valori delle formanti di queste pseudo-sillabe, è possibile ottenere anche le pronunce di **fi**, **fu**, **pi**, **pu**. Ciò si ottiene con una opportuna codifica dei dati all'interno del database. Per approfondimenti consultare il manuale del **CRIiSTAL** alla voce base di dati.

Per le geminate, si è ottenuto un ottimo risultato considerandole un fonema unico appartenente interamente alla sillaba seguente. Inoltre, le consonanti finali in sillaba implicata, sono ottenibili dalle relative versioni geminate, utilizzando una opportuna finestra di valori: più è dettagliato l'andamento dei valori a cavallo della geminazione, tanto meglio sarà la pronuncia dell'attacco della consonante. Il lavoro effettuato sui dati disponibili per l'HLSyn è stato il seguente:

- 1) determinazione dei punti di inizio e fine delle sillabe relative ad un fonema
- 2) determinazione della codifica dei dati per contenere tutte le informazioni necessarie ad una corretta sintesi (variazioni dei valori delle formanti per ogni vocale)
- 3) creazione della base di dati nel database **MySql** a partire dai valori dei fonemi codificati

I files sintetizzati venivano confrontati con quelli originali, raffinando i dati nel db per una migliore sintesi. Come procedura di sintesi per nuovi fonemi si potrebbe perciò seguire questa via:

- 1) creazione di una sequenza vocale – consonante - vocale della consonante da sintetizzare ad esempio *ara*, *ala*
- 2) realizzazione della sua versione geminata scegliendo i corretti istanti di sintesi ( i valori dei parametri rimangono pressoché uguali ma subiscono uno slittamento nel tempo)
- 3) determinazione degli istanti significativi per la creazione dell'attacco del fonema ad esempio *ar* in *ar-pa*.
- 4) sintesi della parola e confronto con la versione originale per un aggiustamento dei parametri

Avendo a disposizione un programma che genera automaticamente il file, la procedura risulta enormemente velocizzata, poiché l'unica possibilità per sintetizzare un suono, era compilare riga per riga un foglio di calcolo ed avviare la sintesi: ogni modifica temporale relativa ad una sola riga, comportava la riscrittura di tutte le seguenti.



# CAPITOLO 5

## CONCLUSIONI

Nei capitoli precedenti sono state descritte le problematiche inerenti alla realizzazione del progetto, presentando le scelte fatte per risolverle. In questo ultimo capitolo riassumiamo il lavoro svolto, indicandone le proprietà più importanti e le possibilità di sviluppo futuro.

### 5.1 RIEPILOGO DEI RISULTATI SULLA TRASCRIZIONE FONETICA

---

La trascrizione automatica delle parole nel caso dell'italiano ha presentato non poche difficoltà per ottenere nel complesso un buon risultato. Utilizzando sinergicamente i concetti grammaticali e fonologici è stato possibile realizzare un programma che operi una conversione con un errore bassissimo, al di sotto del 3%. Inoltre, gli eventuali errori di trascrizione, se dovuti ad una mancanza di dati nel db, sono correggibili in tempo reale, creando una sorta di procedura di apprendimento dinamica. Per migliorare le prestazioni del sistema, dopo un periodo di esercizio consistente, è possibile estrapolare regolarità fra i termini inseriti, soprattutto i polisillabi, così da implementare nuove regole da inserire nel programma.

Per quanto riguarda la lingua spagnola, invece, la trascrizione opera con un errore praticamente, in quanto abbiamo visto che la procedura di associazione grafema – fonema è rigida: le pochissime eccezioni sono costituite da alcune parole che presentano nel contesto **vocale - s – vocale** la versione sorda del fonema anziché quella sonora. Con la procedura di correzione questo problema è azzerabile. Possiamo affermare le due procedure tendono ad un errore uguale a zero ma con velocità di aggiornamento diverse.

La modularità del programma consente di implementare l'analisi di altre lingue, potendo estendere a livello europeo un progetto di Ateneo. Il programma, inoltre, può essere anche utilizzato in maniera svincolata dal sintetizzatore, come trascrittore automatico di files di testo, con tutte le applicazioni ad esso correlate.

## 5.2 LA SINTESI DEI FONEMI

---

Con il programma realizzato, considerando anche la possibilità di sintetizzare files già trascritti foneticamente, è possibile semplificare in parte la sintesi dei vari fonemi, in quanto la modifica dell'istante di emissione di un fonema non crea ripercussioni dirette alle righe adiacenti: si possono modificare a piacimento i valori nel db e la procedura non ne risente minimamente.

E' sempre possibile riprodurre esattamente con il programma, files già sintetizzati, semplicemente confrontando le due tabelle ed apportando le correzioni necessarie.

In linea di principio, si possono adattare i dati , con le codifiche introdotte, per ottenere la pronuncia migliore con qualsiasi combinazione fonetica: il tutto è limitato alla bravura di chi effettua la sintesi.

Si è visto infine, che una buona strategia di sintesi , per le conoscenze attuali, consiste nel creare dapprima una sequenza **vocale – fonema da sintetizzare – vocale**; utilizzare i valori ottenuti per realizzare la versione geminata traslandoli nel tempo; individuare l'istante più opportuno per separare il suono dell'attacco della consonante per riprodurre i fonemi in coda alle sillabe implicate.

Questo è solo un metodo di principio e andrebbero apportate le correzioni necessarie per far sì che la pronuncia risulti la più naturale possibile, ma consente comunque di ottenere tutte le possibili versioni di uno stesso fonema. Per i clusters, sia vocalici che consonantici, il problema non è stato affrontato, ma la sperimentazione unendo vari fonemi risulterà velocizzata dalla sintesi automatica: è possibile infatti utilizzarla come punto di partenza per ottenere il materiale da raffinare fino a trovare i parametri ottimali da inserire nel db.

Nel prossimo paragrafo sono esposti gli aspetti relativi all'interazione con l'**HLSyn**, mostrando due spunti per realizzazioni future che potrebbero migliorare sensibilmente sia la qualità della sintesi che i tempi realizzativi.

## 5.3 L'INTERAZIONE CON L'HLSYN E SPUNTI PER APPLICAZIONI FUTURE

---

L'aspetto fondamentale dell'intero lavoro era dimostrare la possibilità di interazione con il sintetizzatore, sebbene ciò non avvenga in maniera completamente automatica, a causa della struttura impostagli. Ciò permette di supporre che, in versioni successive del programma di sintesi, possano venire introdotte delle modifiche che consentano una più stretta operatività fra i due softwares, magari anche con suggerimenti ricavabili dal presente lavoro. I punti cruciali sono due:

- 1) non esiste per ora la possibilità di aprire il sintetizzatore e fargli caricare i dati tramite comando esterno
- 2) nel caso il numero di parole da sintetizzare sia maggiore di poche unità, i tempi di attesa per la sintesi diventano molto lunghi, creando seri problemi per una conversione in tempo reale o per una interazione uomo- macchina.

Questo secondo punto sarebbe aggirabile se, per ogni parola da tradurre, si potesse aprire un'istanza del programma riunendo poi i risultati parziali.

Se ciò fosse possibile, si potrebbe affrontare anche il problema della prosodia di frase più velocemente, in quanto inserendo una funzione specifica nel codice del programma che crea il file **.HL**, sarebbe già possibile implementarne i parametri caratteristici.

Inoltre anche a livello di singola parola, con la coppia *papa – papà*, si è dimostrato come siano accessibili tutti i parametri dei fonemi in tutti gli istanti di sintesi, cosicché un'eventuale ricerca sui valori assunti dai parametri della sintesi articolatoria in determinati contesti, potrebbero essere immediatamente implementati.

Per finire, sarebbe possibile realizzare due moduli per completare il lavoro appena sviluppato:

- 1) un modulo per trasformare in file **SQL** un file di tipo **HL**
- 2) un modulo di interfaccia grafica che consenta la conversione di un grafico in valori numerici compatibili al formato **HL**

Insieme, queste due procedure consentirebbero una maggiore velocità nella sintesi e nel confronto dei dati ottenuti, migliorando le prestazioni quantitative e qualitative dell'intero progetto.