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 f1 subisce il massimo della diminuzione in corrispondenza della fine della costrizione e ritorni in breve tempo prossima al suo valore nominale.

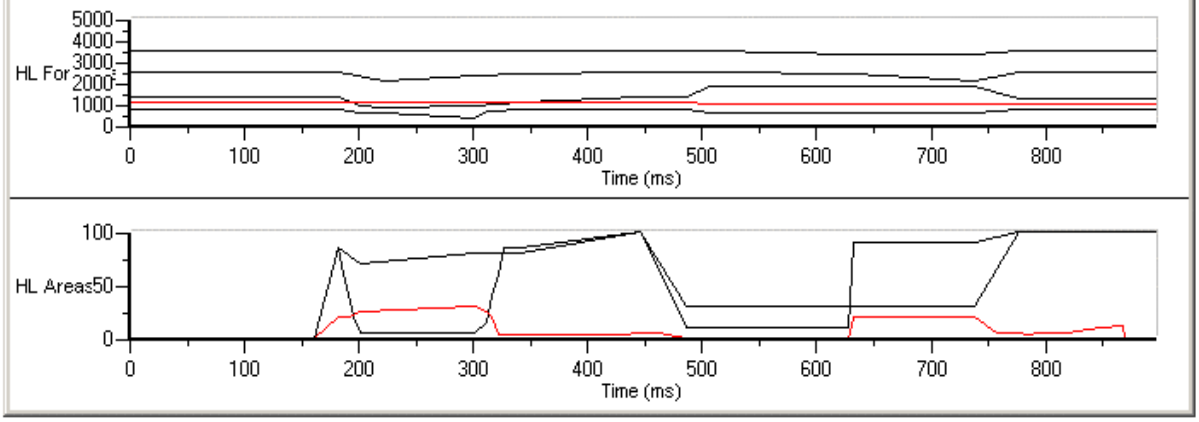
Ps

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.

faccia - HL Parameters

	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

faccia - HL Graphic



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 occlusone 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 bbassamento rispetto al valore nominale.

Formanti

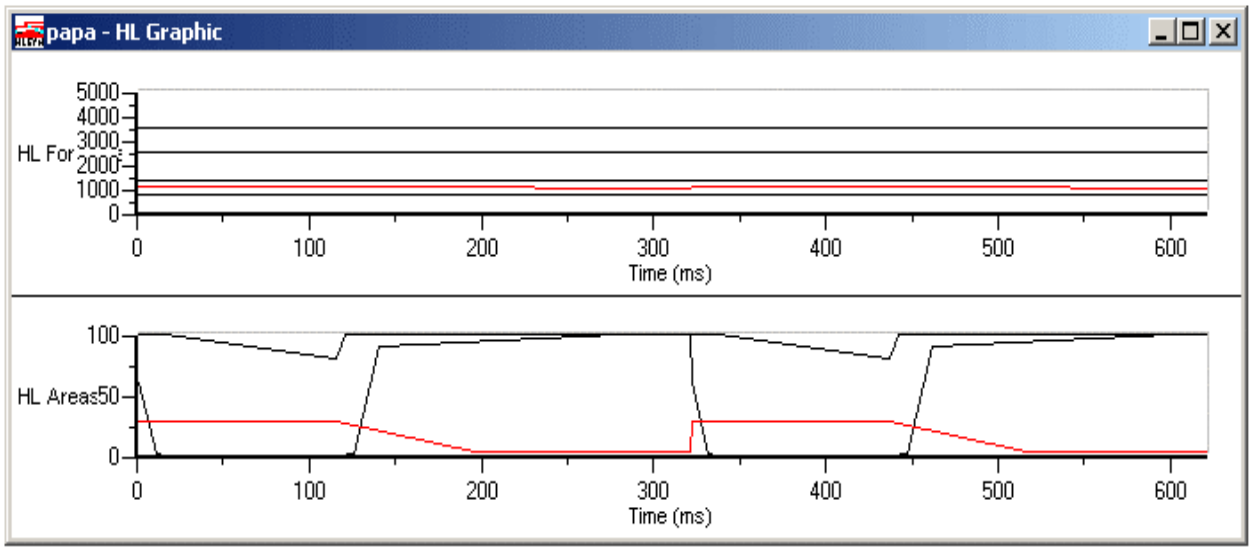
I valori delle formanti non hanno richiesto una particolare modifica, non influendo in maniera significativa alla sintesi del fonema.

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.

papa - HL Parameters

	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 coarticolarli. 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.