

Relazione
di
Sistemi per la Progettazione Automatica

Sintetizzatore
basato sul modello di
KLATT

Realizzato da
Dinardo Saverio
Nardelli Graziano

Lo scopo del progetto realizzato è l'implementazione con strumenti informatici (nel caso specifico Matlab) del modello di Klatt per la generazione sintetica di segnali vocali. La caratteristica fondamentale del modello di Klatt è la generazione di questi segnali attraverso la concatenazione di singoli fonemi ciascuno dei quali viene realizzato "passando" al sintetizzatore opportuni parametri. Il sintetizzatore è essenzialmente composto da elementi di sorgente di segnali (che ricalcano le funzioni dei pomlioni nella produzione reale di suoni e rumori) e da filtri risuonatori (introdotti per tenere conto del fatto che nella realtà il suono è l'effetto di una serie di risonanze di impulsi d'aria all'interno delle cavità dell'apparato fonatorio umano e di opportune strozzature). La taratura dei vari parametri è un'operazione molto delicata per diverse ragioni la cui principale è senza dubbio l'impossibilità di individuare una precisa regola nel parlato umano legato sia alle varietà tra le lingue che nelle lingue stesse.

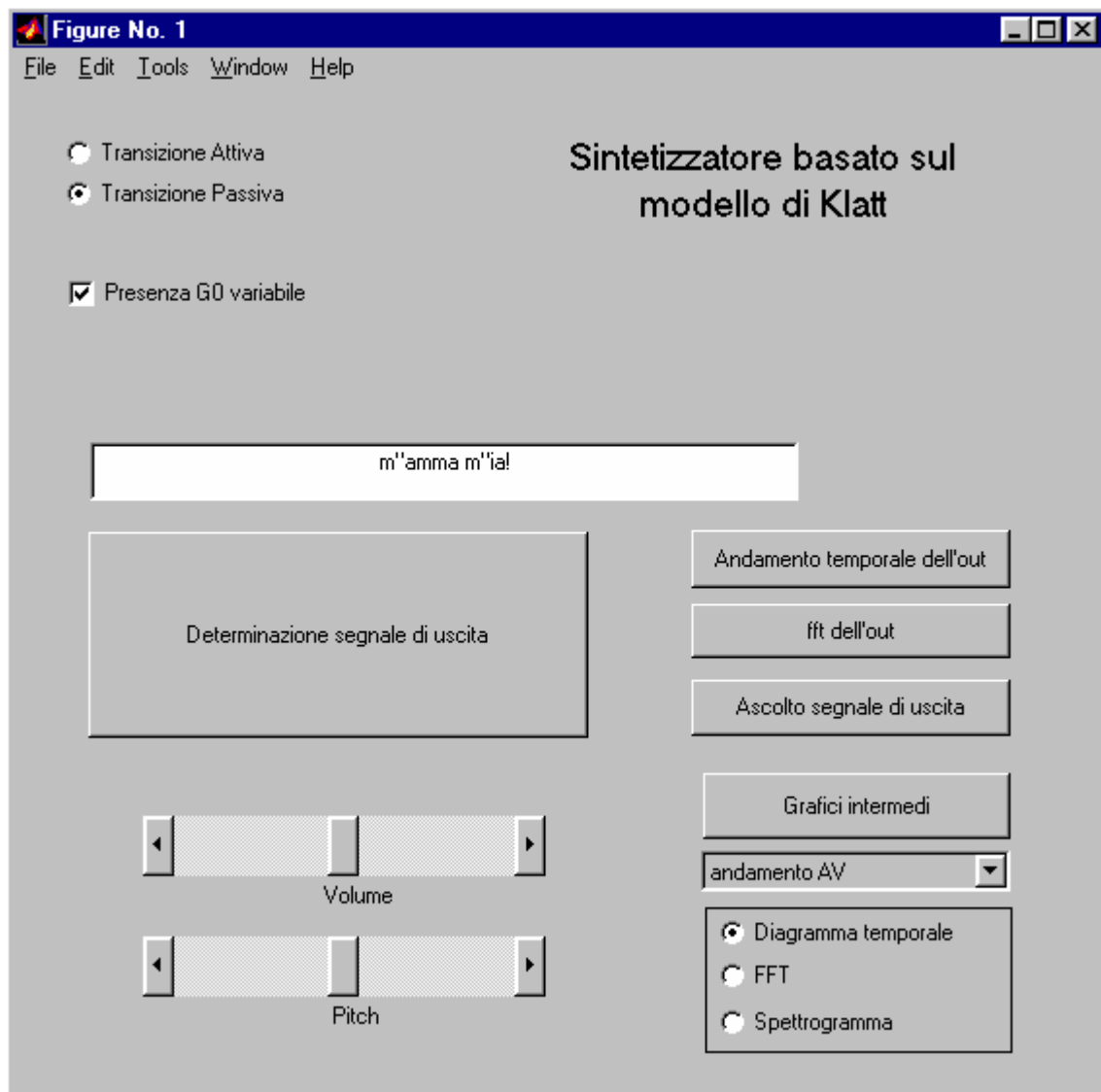
Sia pur con approssimazioni i parametri del modello di Klatt sono stati tabellati per ciascun fonema tipico della lingua italiana. E' a questa tabella che fa riferimento il nostro sintetizzatore.

Altro problema che si incontra quando si realizza un sintetizzatore è l'effetto dell'intonazione nella pronuncia delle parole che permette una migliore comprensione del significato dell'intera frase.

La punteggiatura, ad esempio, determina modifiche nei parametri di cui bisogna tener conto, così come un altro elemento importante da valutare è la durata e l'intensità dei fonemi che influenzano molto la comprensione. Tutti questi problemi sono stati risolti brillantemente da Klatt per un modello sostanzialmente basato sul parlato inglese ma non sempre si adattano bene al parlato italiano (senza considerare poi le miriadi di varietà che l'italiano presenta). Il risultato finale è comunque accettabile pur essendo abbastanza riconoscibile l'origine sintetica del segnale ascoltato. Un altro problema che si presenta nella concatenazione dei fonemi è quello di evitare che tra un fonema e l'altro si presenti una brusca variazione del segnale che dal punto di vista audio determina un suono indesiderato (sparo). Ciò è da evitare in quanto in natura i segnali evolvono con continuità e non vi sono grosse discontinuità. Nel sintetizzatore invece queste brusche variazioni dipendono dal fatto che i parametri dei filtri di un fonema sono diversi dal fonema successivo. Per evitare questi spari bisogna fare in modo che questi parametri varino con continuità tra un fonema all'altro, (questa situazione viene valutata come zona transitiva).

Lo scopo del nostro progetto è quello di realizzare un software che permette di implementare tutte le problematiche del modello di Klatt. Il concetto di base del software è quello di implementare lo schema del modello, realizzando i rispettivi segnali di ingresso e la giusta locazione dei filtri e degli amplificatori. Per fare questo bisogna lavorare molto con le operazioni matematiche. Questa valutazione ha determinato la scelta di utilizzare il Matlab come linguaggio di programmazione. Questo linguaggio interprete ha la capacità di lavorare in maniera immediata su matrici e vettori ed è molto veloce da utilizzare anche perché non occorre dichiarare le variabili. Altre caratteristiche di questo linguaggio che ci hanno portato a questa scelta sono la grande varietà di rappresentazioni grafiche e la presenza di una grande libreria incorporata (toolbox) dalla quale si possono prelevare importanti funzioni.

Il nostro progetto ha la seguente schermata iniziale:



Descriviamo adesso (anche se è molto intuitiva) le caratteristiche essenziali di questa interfaccia: Nell'interfaccia è presente un campo testo, nel quale è possibile inserire la frase che vogliamo sintetizzare.

Questa frase deve essere composta da simboli appartenenti alla lista dei grafemi (non ci siamo occupati, infatti, della traduzione testo-fonemi).

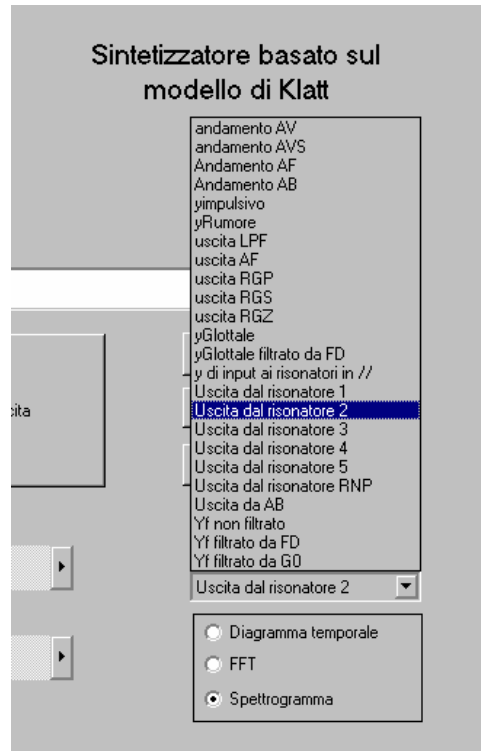
Inoltre sono previsti tutti i simboli di punteggiatura. Per valutare inoltre bene la tonalità, in ogni parola (che nel nostro caso rappresenta un gruppo di respiro) ci deve essere un accento.

L'accento in questo programma è simboleggiato dall'introduzione del carattere “ ” da mettere prima della vocale accentata.

Una volta inserita la frase bisogna elaborare il nostro segnale che rappresenta il parlato della frase presente nel campo testo. Per fare questo si clicca sul tasto: “Determinazione segnale di uscita”.

Una volta fatta questa operazione possiamo valutare tutti gli output. È Possibile ascoltare il segnale di uscita, e vedere inoltre il rispettivo andamento nel tempo e in frequenza.

È possibile inoltre valutare tutti i grafici intermedi che si presentano nei vari punti del modello di Klatt. Questo è possibile farlo cliccando su “Grafici Intermedi” una volta selezionato da quale punto si vuole estrapolare il segnale e quale tipo di grafico si vuole.



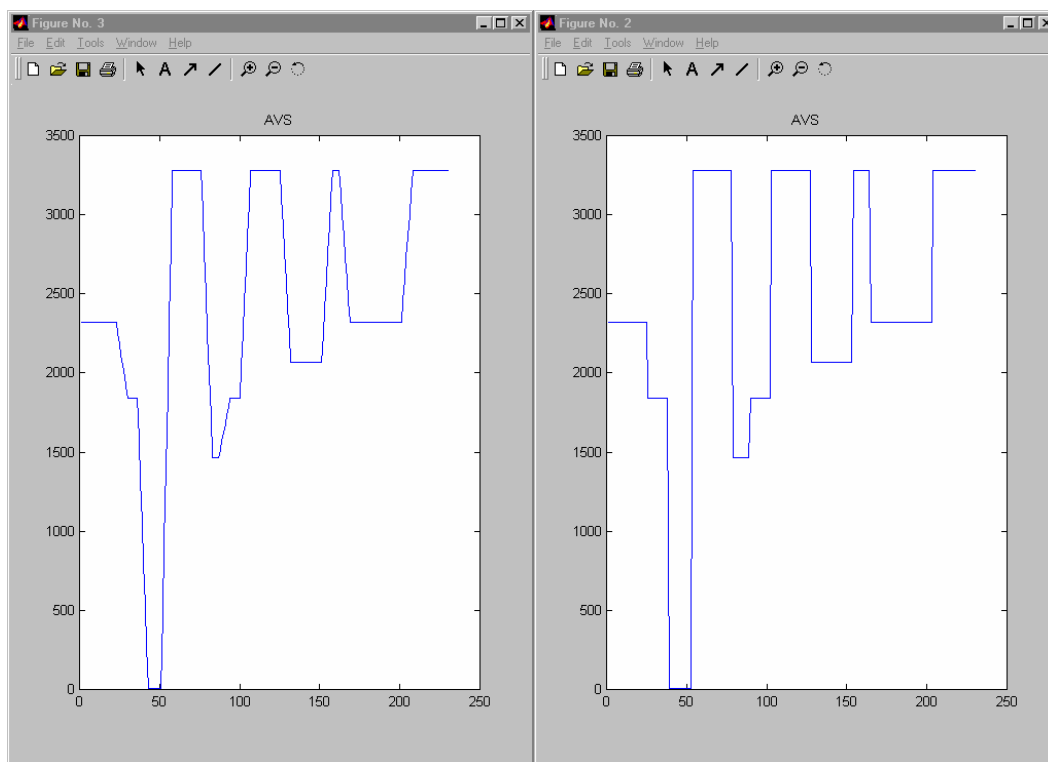
Altre funzionalità introdotte dall'interfaccia sono la possibilità di variare la velocità e il volume della riproduzione della voce.

Dall'interfaccia si può dare anche la possibilità di valutare il segnale di uscita sia con l'amplificatore variabile di uscita (G0) del modello di Klatt sia senza questo amplificatore.

Un'altra cosa importante che è possibile fare è quella di decidere che tipo di transizione applicare tra i fonemi. Infatti è possibile scegliere tra un tipo di transizione detta “transizione passiva” e un tipo di transizione detta “transizione attiva”.

Descriviamo adesso i due tipi di transizione: la transizione passiva permette un passaggio tra 2 fonemi facendo variare alla fine della durata del fonema i parametri dei filtri in modo da farli arrivare con gradualità ai parametri dei filtri del fonema successivo.

Questa situazione è possibile capirla meglio con i seguenti grafici:



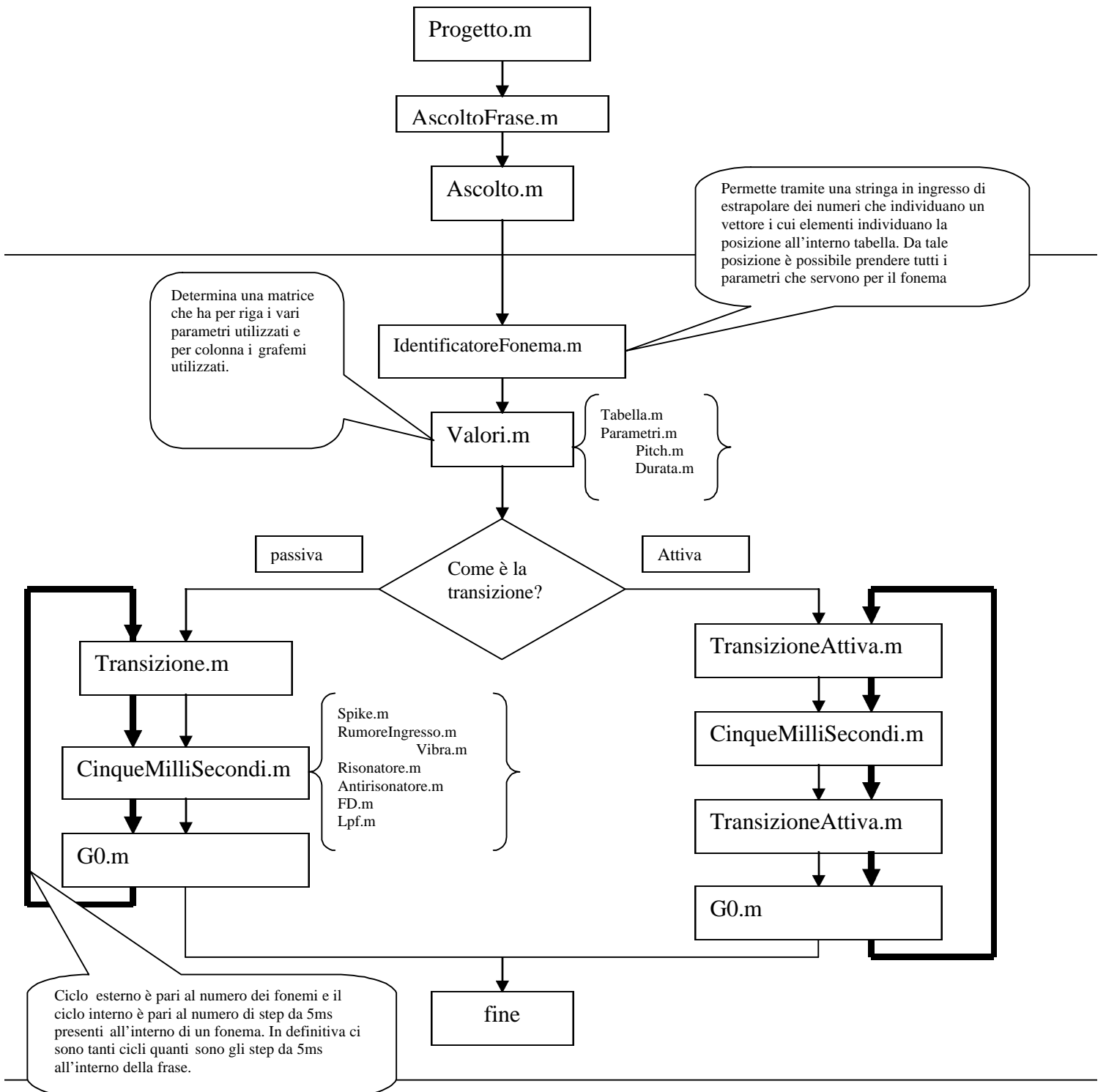
La transizione attiva permette di analizzare la situazione.

In tal caso non si lavora all'interno della durata dei fonemi ma si lavora all'esterno. In questa situazione pertanto rimangono costanti all'interno della durata dei fonemi i parametri dei filtri del modello di klatt. La transizione attiva prevede di aggiungere tra un fonema e l'altro una piccola quantità di segnale che permette di non avere una brusca discontinuità tra i segnali di due fonemi. Nel nostro caso questo intervallo è di 30ms. In questo intervallo si cambiano con continuità i parametri del modello di klatt e si determina l'uscita utilizzando la funzione CinqueMilliSecondi.m che ci permette in base ai valori dei parametri, di valutare l'uscita del modello di klatt per 5ms. E' chiaro che questa funzione sarà applicata in modo da soddisfare tutta la durata della transizione.

La schermata analizzata è un'interfaccia che nasconde il software utilizzato e permette una facile interpretazione da parte di un utente qualsiasi.
 I vari bottoni della schermata richiamano le funzioni che permettono di eseguire le varie procedure che implementano il modello.

Descriviamo adesso come è strutturato il progetto con il seguente schema a blocchi:

Il nostro progetto è formato nel complesso dalle seguenti funzioni:



Progetto.m

Rappresenta l'interfaccia che comunica con l'utente

AscoltoFrase.m

```
[GRAFICI, YUSCITA]=AscoltoFrase(FraseConPunteggiatura)
```

È La prima funzione che viene richiamata nel momento in cui si vuole elaborare il segnale.

Questa funzione permette di suddividere la frase nei vari gruppi di respiro. Questi gruppi di respiro vanno interpretati dalla funzione ascolto.m che viene descritta nel punto successivo. A questa funzione gli viene passato come parametro FraseConPunteggiatura che rappresenta una stringa il cui valore arriva dal campo testo dell'interfaccia. Questa funzione determina YUSCITA e GRAFICI. GRAFICI rappresenta una matrice le cui righe individuano l'andamento di alcuni parametri dei filtri utilizzati: AV AVS AF AB mentre YUSCITA è sempre una matrice le cui righe rappresentano i segnali nei vari punti del modello di klatt. Da questi punti è possibile valutare il loro andamento grafico sia rispetto al tempo che rispetto alla frequenza. La matrice in uscita viene ottenuta come composizione delle matrici che si ottengono dalla funzione ascolto.m

Ascolto.m

```
[grafici, yUscita]=ascolto(ParolaConPunteggiatura)
```

```
global Pind Ind FlagPrimoPassaggio
```

```
global f0 blocca_oscill xr AF AV AVS AB fc nn h Parametri Tempoaccento N
```

```
global FlagTransizioneAttiva FlagPresenzaG0
```

```
global xr0 yGlott0 yRGP0 yRGS0 yRGZ0 y10 y20 y30 y40 y50 yRNPO
```

Questa funzione viene richiamata dalla funzione descritta precedentemente. A tale funzione gli viene passata la ParolaConPunteggiatura. Questo parametro rappresenta una stringa. Questa stringa è una frazione della frase completa e rappresenta un gruppo di respiro cioè una parola che non presenta spazi all'interno. In questo gruppo di respiro è presente un solo accento. Questa funzione determina grafici e yUscita. Entrambi sono delle matrici e servono per comporre le matrici descritte nel punto precedente.

All'interno di questa funzione di presentano le funzioni che si presentano in seguito.

- **IdentificatoreFonema.m**

```
VettoreFonema=IdentificatoreFonema(ParolaSenzaPunteggiatura)
```

Questa funzione viene richiamata dalla funzione descritta precedentemente. A tale funzione gli viene passata la ParolaSenzaPunteggiatura. Questo parametro rappresenta una stringa determinata all'interno della funzione precedente ed è la stringa che rappresenta il gruppo di respiro ripulito da accento e segni di punteggiatura. Questa funzione determina un vettore: VettoreFonema i cui parametri individuano nella tabella in cui sono presenti i parametri dei filtri, il corrispondente valore della riga. Da questa riga si prendono le e si manipolano gli opportuni valori e si mettono in ordine utilizzando la funzione successiva: Valori.m

- **Valori.m**

```
EstrazioneTabella=valori(VettoreNumeriFonemi)
```

Questa funzione segue quella precedente, utilizzando come parametro il risultato della funzione precedente. Si ottiene una matrice le cui righe rappresentano in ordine i vari fonemi che formano il gruppo di respiro e i parametri delle righe sono così ordinati:

```
F1 F2 F3 F4 B1 B2 B3 B4 A2 A3 A4 A5 AB AV AVS A1 AF dur1 dur2 PresenzAccento AN FNP  
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
```

- **Tabella.m**

```
y=tabella(riga, colonna)
```

Questa funzione viene richiamata dalla funzione valori.m e serve per prelevare i parametri fissi utili per comporre poi tramite la funzione parametri.m la matrice ordinata in base ai fonemi al cui interno sono presenti sia i valori sempre fissi che i valori che dipendono dalla punteggiatura.

- **Parametri.m**

```
[PitchDurata, VettoreNumeriFonemiPulito, PosizioneAccento]=parametri(VettoreNumeriFonemi)
```

Questa funzione è richiamata dalla funzione precedente e serve per determinare:

VettoreNumeriFonemiPulito che rappresenta lo stesso vettore passato alla funzione come parametro, meno la presenza dell'elemento che rappresenta l'accento (100) e il valore che rappresenta il punto di interpunzione (un valore superiore a 100), se sono presenti.

PosizioneAccento individua la posizione della vocale nella parola nella quale è presente l'accento. PitchDurata determina una matrice fatta da 2 colonne dove la prima colonna individua per ogni fonema il corrispondente pitch e la seconda colonna individua sempre per ogni fonema il valore di prcnt. Questo valore rappresenta un valore percentuale calcolato in base alla punteggiatura (nella funzione valore.m) e serve per determinare lo scostamento della durata dal valore minimo. Le

colonne della matrice ottenuta vengono calcolate separatamente all'interno delle funzioni descritte successivamente: *Pitch.m* e *Durata.m*

○ **Pitch.m**

`y=pitch(PosizioneAccento,PosizionePretonica,dimensione,gruppo)`

Il valore del pitch dipende da dove è posizionato l'accento e anche da dove è posizionata la vocale che precede l'accento. Questo viene calcolata della funzione descritta precedentemente e viene passato come parametro a questa funzione con il nome *PosizionePretonica*. Dipende anche dalla dimensione del vettore ripulito dai segni di punteggiatura e dal gruppo. Il gruppo caratterizza l'andamento del pitch in base al segno di interpunzione presente alla fine della parola.

○ **Durata.m**

`y=durata(VettoreNumeriFonemi)`

Questa funzione determina per ogni fonema la rispettivo valore percentuale che è utilizzato per determinare l'effettivo valore della durata del fonema stesso. Le durate dei vari fonemi dipendono da dove sono disposti nella parola, dal accento e dai segni di interpunzione.

La durata delle vocali dipendono dal fatto che sono appartenenti a sillabe aperte (finiscono in vocale) e in sillabe chiuse (finiscono in consonanti) o anche dal fatto che sono seguite da altre vocali.

Per le consonanti il valore percentuale è sempre fisso tranne se rappresenta il primo fonema della parola. C'è da dire inoltre che l'ultimo fonema della parola ha un volare maggiore se rappresenta una vocale.

Transizione.m

`A=Transizioni(Parametri,num)`

`global FlagTransizioneAttiva`

Questa funzione elimina le variazioni brusche presenti nei valori di AF,AV,AVS,AB.

Per fare questo viene determinato prima il loro andamento per tutto il gruppo di respiro e poi viene realizzata per loro la linearizzazione.

È presente come variabile globale *FlagTransizioneAttiva* perché la linearizzazione viene eseguita solo nel caso della transizione passiva cioè quando il valore di tale flag è 0.

TransizioneAttiva.m

`yUscita=TransizioneAttiva(Fonema,FonemaSuccessivo,ParametriFonema,ParametriFonemaSuccessivo,... DurataTransizione)`

`global Pind Ind FlagPrimoPassaggio`

`global f0 blocca_oscill xr`

`global AF AV AVS AB fc nn Tempoaccento N`

`global xr0 yGlott0 yRGP0 yRGS0 yRGZ0 y10 y20 y30 y40 y50 yRNP0`

Questa funzione viene eseguita solo se il frag *FragTransizioneAttiva=1* (questa interrogazione viene fatta in ascolto.m). Questa funzione valuta tutte le situazioni che si possono presentare tra i vari fonemi, valutando le transizioni fonema-silenzio, fonema-plosiva, fonema-fonema. Questo determina una transizione di un fonema e l'altro con un aggiunta di segnale pari al valore presente come parametro: *DurataTransizione*. Il

FlagPrimoPassaggio è utile perché questa funzione deve essere utilizzata all'inizio del gruppo di respiro come transizione silenzio-fonema. Negli altri casi prima bisogna calcolare il fonema e poi valutare la transizione.

CinqueMilliSecondi.m

`function YUscitaFonema=CinqueMilliSecondi(ParametriFonema)`

`global Pind Ind FlagPrimoPassaggio`

`global f0 blocca_oscill xr`

`global AF AV AVS AB fc nn Tempoaccento N`

`global xr0 yGlott0 yRGP0 yRGS0 yRGZ0 y10 y20 y30 y40 y50 yRNP0`

Questa funzione implementa il modello di klatt. Questa funzione in base ai parametri dei filtri del relativo fonema determina l'andamento dell'uscita per 5ms. In questa funzione sono determinati i vari segnali di ingresso e sono determinati i segnali uscenti dai vari filtri. Per determinare questi segnali vengono utilizzate le funzioni che seguono.

● **ImpulsoIngresso.m**

`h=ImpulsoIngresso(f,blocca_oscill)`

`global Ind Pind`

Produco 5 ms di segnale impulsivo alla frequenza *f*. In 5 ms ho 50 campioni se la frequenza di campionamento è di 10 kHz.

Questa funzione ha memoria. Infatti si ricorda nel ciclo precedente quando è stata eseguita la stessa funzione, dove era situato l'impulso precedente. Per questo motivo sono stati messi global i parametri *Ind* e *Pind*.

Il parametri *Ind* di incrementa di 50 ogni volta che si ripete la funzione e il parametro *Pind* si incrementa del Numero di campioni tra impulsi (10000/f) nel momento in cui si verifica l'impulso.

- **RumoreIngresso.m**

`[rum,UltimoValore]=RumoreIngresso(f,blocca_oscill,Ind,rum)`

Questa funzione presenta come ingresso la frequenza *f* derivante dalle caratteristiche del fonema. Sono presenti anche i campioni del rumore dei 5ms precedenti da cui si estrae l'ultimo valore e si presenta anche *blocc_oscill*. Questo è un flag utile per vedere se il rumore deve essere modulato o meno da un'onda quadra. Questo flag è controllato dai parametri AV e AVS del fonema in questione nel senso che il flag è nullo se entrambi i valori sono nulli.

Se il flag non è nullo ai campioni del rumore ottenuto viene eseguito la funzione che segue: *Vibra.m*

- **Vibra.m**

`x=vibra(ing,f,Ind)`

Ind rappresenta l'indirizzo da cui iniziare la modulazione. *f* è la frequenza del rumore e *ing* rappresenta il rumore che deve essere modulato. L'uscita della funzione rappresenta il rumore modulato (è sempre un vettore da 50 elementi cioè un segnale di 5ms)

- **Lpf.m**

`y=lpf(x,b)`

è un filtro dove *x* è il vettore di ingresso e *b* è il suo valore iniziale.

La funzione da implementare è : $y(nT)=x(nT)+y((n-1)T)$

- **Risonatore.m**

`y=risonatore(b,f,fc,x,yPrecedente)`

è un filtro dove *x* è il vettore di ingresso, *yPrecedente* sono i valori iniziali, *fc* è la frequenza di campionamento ed *f* è la frequenza *f0* del fonema in questione con *b* che rappresenta l'ampiezza di picco del risonatore.

La funzione da implementare è : $y(nT)=A*x(nT)+B*y((n-1)T)+C*y((n-2)*T)$

Dove A B C vengono calcolati noti b ed f

- **Antirisonatore.m**

`y=antirisonatore(b,f,fc,x,yPrecedente)`

è un filtro dove *x* è il vettore di ingresso, *yPrecedente* sono i valori iniziali, *fc* è la frequenza di campionamento ed *f* è la frequenza *f0* del fonema in questione con *b* che rappresenta l'ampiezza di picco del risonatore.

La funzione da implementare è : $y(nT)=A1*x(nT)+B1*x((n-1)T)+C1*x((n-2)*T)$

Dove A1 B1 C vengono calcolati noti b ed f

- **FD.m**

`y=FD(x,x0)`

è un filtro dove *x* è il vettore di ingresso, *x0* è il valore iniziale.

La funzione da implementare è : $y(nT)=x(nT)-x((n-1)T)$

G0.m

`y=G0(fc,Tempoaccento,yf)`

è un filtro che lavora su tutti i campioni del segnale (pertanto non viene richiamato dalla funzione *CinqueMilliSecondi.m*). I parametri di questa funzione sono: *fc* che rappresenta la frequenza di campionamento, *yf* che rappresenta il segnale di ingresso al filtro e *Tempoaccento*....

Descrizione generale del funzionamento: il doppio ciclo

Una volta commentate tutte le caratteristiche delle funzioni utilizzate, descriviamo il funzionamento generale del progetto. Una volta frazionati i gruppi di respiro e analizzato dove si trova l'accento si inizia ad ottenere attraverso la funzione *CinqueMilliSecondi.m* il segnale in uscita.

Questa funzione determina 5ms del segnale per volta . Pertanto viene eseguita per ogni fonema tante volte quanto basta per determinare l'intera sua durata. Per implementare questo discorso pertanto si è dovuto fare un doppio ciclo: il ciclo esterno viene eseguito per ogni fonema del gruppo di respiro mentre il ciclo interno viene eseguito per un numero di volte pari al numero di step di 5ms che formano il fonema.

Le caratteristiche dei parametri dei filtri nel ciclo interno sono fissi mentre variano, sempre nello stesso fonema, nel ciclo esterno: questo permette di migliorare le caratteristiche del segnale del fonema stesso rendendolo molto più simile al segnale del parlato umano.

Nel ciclo esterno si passa da un fonema all'altro. In tal caso pertanto si presentano grandi variazioni dei parametri dei filtri. Queste variazioni portano nella transizione a dei grandi picchi del segnale di uscita (spari). Per evitare questo bisogna fare in modo che nelle regioni di transizione tra un fonema e l'altro non ci sia una brusca variazione.

E' per ovviare a questo che si sono applicati i 2 metodi visti in precedenza: transizione attiva e transizione passiva.