



MINISTERO DELLO SVILUPPO ECONOMICO
DIREZIONE GENERALE PER LA TUTELA DELLA PROPRIETÀ INDUSTRIALE
UFFICIO ITALIANO BREVETTI E MARCHI

UIBM

DOMANDA NUMERO	101994900404423
Data Deposito	23/11/1994
Data Pubblicazione	23/05/1996

Priorità	9303902-2
Nazione Priorità	SE
Data Deposito Priorità	

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
G	10	L		

Titolo

DISPOSIZIONE E METODO DI SINTESI VOCALE

RM94 A 000763

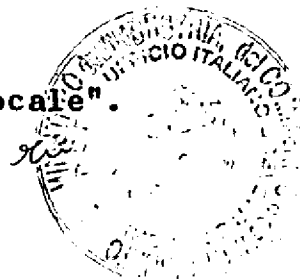
Descrizione

dell'invenzione avente per titolo:

"Disposizione e metodo di sintesi vocale".

a nome: TELIA AB - FARSTA (SVEZIA)

inventore: Svensson Tomas



Campo tecnico

La presente invenzione si riferisce alla sintesi della parola. Nella sintesi della parola, o sintesi vocale, vengono individuate parole che vengono suddivise in un certo numero di suoni caratteristici chiamati fonemi. Nella individuazione di sequenze parlate, è essenziale che i detti fonemi siano correttamente individuati. I fonemi vengono anche utilizzati per generare sequenze parlate con mezzi artificiali.

Stato della tecnica

Quando la parola viene generata artificialmente, si utilizza normalmente una biblioteca con fonemi fondamentali. Quando questi fonemi vengono assemblati in parole, essi devono in molti casi essere trasformati entro periodi più lunghi o più corti che sono rappresentati dai fonemi di base. E' noto a questo riguardo che si individuano i fonemi in un certo numero di punti. Quando si trasforma il fonema origina-



le in una differente scala di tempo, il che può significare un allungamento o un accorciamento della scala di tempo, è noto che si esegue la trasformazione in un certo numero di punti prescelti. Quando la scala di tempo viene allungata, ciò coinvolge certi punti nel fonema originale rappresentando un certo numero di punti nel nuovo fonema. Quando la scala di tempo viene accorciata, un certo numero di punti nel fonema originale viene combinato per formare un unico punto nel nuovo fonema. Quando il fonema originale viene trasferito ad una scala di tempo che, per esempio, è del 25% più lunga del fonema nella biblioteca, in certo numero di punti nel fonema della biblioteca viene selezionato. Nel nuovo fonema, che è formato dalla trasformazione, il 25% di punti in più viene inserito rispetto al fonema della biblioteca. In seguito alla trasformazione, il nuovo fonema conterrà dunque un numero di punti che non sono definiti le fonema della biblioteca. In seguito alla trasformazione viene selezionato ciascun quarto punto nel fonema della biblioteca. Queste parti del fonema vengono duplicate e trasferite a due punti nel fonema allungato. I punti restanti vengono trasferiti dal fonema della biblioteca al fonema allungato, punto per punto. Ciò consente un



allungamento del tempo del fonema originale mediante un allungamento uniforme del tempo sull'intero fonema. Nei casi in cui il fonema della biblioteca è più lungo del fonema che deve essere formato ciascun quarto punto viene scelto nella stessa maniera come sopra, supponendo che l'accorciamento del tempo sia del 25%. Quando il fonema accorciato nel tempo è stato formato, questi punti vengono rimossi nella trasformazione. Nel Brevetto EP 252544 è descritta la modifica della scala della parola di un nuovo punto segnale. Ciò si basa, tra l'altro, sulla constatazione che la compressione della scala di tempo riduce il contenuto d'informazioni e che l'espansione della scala di tempo aumenta il contenuto d'informazioni. Si possono così rimuovere "periodi di tono" o inserirli rispettivamente su un segmento. L'invenzione consiste in un metodo per migliorare il metodo SOLA mediante sovrapposizione di blocchi parzialmente sovrapposti.

Il Brevetto US 4435832 mostra una sintesi vocale con allungamento e compressione della scala di tempo senza cambiare il tono della parola sintetica. I parametri LPC vengono campionati a partire da forme d'onda segmentate prelevate dalla parola normale in un dato intervallo di tempo, da informazioni riguardan-



ti fonemi vocali/non vocali, tono e informazioni sul volume. L'LPC viene interpolato e l'intervallo della scala di tempo per l'interpolazione viene migliorato.

Nel Brevetto US 4864620 è descritto un metodo per modificare la scala di tempo di informazioni vocali o segnali vocali per riprodurre la parola registrata ad una differente velocità senza cambiamento del tono. Campioni del settore tempo vengono prelevati in inquadrature dove il numero dei campionamenti per inquadratura è funzione del desiderato fattore di cambiamento della parola. Blocchi vengono formati a partire dalle inquadrature. Si realizzano transizioni relativamente morbide mediante ponderazione graduata.

La modifica della scala di tempo di segnali vocali è ancora indicata nel Brevetto US 5216744. Viene determinato il numero di campionamento che costituiscono un "periodo di tone". Inoltre un gruppo campione combinato formato di un primo gruppo campione e da un secondo gruppo campione viene formato. Il numero di campioni in ciascun gruppo è uguale al numero di campioni che costituiscono un periodo di tono.

Descrizione dell'invenzione

Problema tecnico



Nella sintesi vocale è essenziale che parole e frasi prodotte artificialmente siano riprodotte naturalmente. E' anche essenziale che la parola prodotta da una persona sia identificata correttamente. A questo riguardo è possibile identificare un certo numero di suoni, fonemi, caratteristici per differenti lingue. Questi fonemi sono disposti in differenti forme di biblioteche. Detto fonemi costituiscono un nucleo di base. I fonemi possono estendersi su un tempo più lungo o più corto che gli intervalli di tempo che sono rappresentati dal fonema di base a seconda del contesto e della parole i cui sono inclusi. Ciò comporta che i fonemi rappresentati nella biblioteca devono essere trasformati in periodi di tempo più lunghi o più corti. In questo contesto è essenziale in una trasformazione del genere che la caratteristica del fonema non venga cambiata. Ciò implica che le parti portanti informazioni del fonema non vengano cambiate. E' dunque desiderabile che cambiamenti di tempo avvengano nelle parti del fonema che portano meno informazioni. Nell'assemblaggio di un certo numero di fonemi in parole e frasi, è anche essenziale che le transizioni tra fonemi abbiano luogo in maniera tale che le parti portanti informazioni di un rispettivo fonema non siano cambiate.



Nella parola naturale, il tono fondamentale viene cambiato con il medesimo fonema nello svolgimento della parola. Le soluzioni presentate finora non hanno tenuto conto di questo fenomeno. E' quindi desiderabile che il cambiamento nel tono fondamentale, di maggiore o minor frequenza, sia preso in considerazione quando si trasformano fonemi.

L'invenzione caratterizzata si propone di indicare una soluzione al problema caratterizzato.

Soluzione

La presente invenzione riguarda un metodo per la sintesi vocale. Viene identificato un fonema in un certo numero di punti nella eccitazione della corrispondente corda vocale dell'oratore. Il fonema deve essere trasformato ad un altro tempo rispetto a quello che è rappresentato dal fonema originale. Dopo che i punti sono stati scelti, i punti nel fonema che sono portatori di informazioni vengono identificati. Per trasporto d'informazioni s'intendono a questo riguardo le parti del fonema necessarie affinché il fonema sia correttamente compreso. Le parti del fonema che portano meno informazioni vengono anch'esse identificate. Parti che portano meno informazioni possono essere cambiate senza che la caratteristica del fonema venga cambiata



nella sua parte più essenziale. Quando vengono adoperati fonemi, per esempio, per generare parola artificiale, è desiderabile che si possa adoperare un certo numero di fonemi di base i quali vengono trasformati in valori desiderati in differenti occasioni. L'invenzione tiene conto di questa situazione e sposta le transizioni tra differenti fonemi verso le parti che portano meno informazioni. Nella trasformazione in una nuova scala di tempo, la compressione o rispettivamente lo stiramento hanno essenzialmente luogo nelle parti del fonema che portano meno informazioni. In questo modo le parti portanti informazioni del fonema sono mantenute essenzialmente intatte.

La disposizione comprende un elemento che sceglie un fonema da una sequenza vocale o da un elemento memorizzatore. L'elemento identifica un certo numero di punti nel fonema. Vengono poi identificate le parti portanti informazioni del fonema o, rispettivamente le parti del fonema portanti meno informazioni.

L'elemento provvede allora a che la trasformazione del fonema in un tempo più lungo/più corto abbia luogo mediante compressione o rispettivamente mediante stiratura nella parte del fonema che porta meno informazioni. In questo modo il carattere del



fonema viene essenzialmente conservato. Inoltre viene data la possibilità di ottenere transizioni tra differenti fonemi che danno un'impressione naturale.

L'invenzione permette di memorizzare una serie di fonemi di biblioteca rappresentanti un certo numero di suoni che si trovano nel linguaggio. Questi fonemi di biblioteca possono essere utilizzati per trasformazioni in un tempo più lungo o più corto di quanto rappresentato dal fonema di biblioteca. Con la soluzione indicata, il fonema trasformato viene corretto in misura minima rispetto al fonema di biblioteca. Ciò si deve al fatto che le parti del fonema essenziali per l'interpretazione del fonema sono invariate o variate in minor misura. L'invenzione permette anche di tener conto di cambiamenti nel tono fondamentale nel fonema. Si possono introdurre così variazioni nel tono fondamentale nel fonema trasformato, in relazione al fonema di biblioteca. Il significato è che a sequenze vocali createsi può dare un carattere concordante con la parola naturale. Ciò è essenziale, sia per comprendere la parola sia per ottenere una intonazione naturale nel suono creato.

Descrizione delle figure

La Figura 1 mostra esempi di cartografia lineare della scala di tempo.



La Figura 2 mostra la formazione della scala di tempo secondo l'invenzione;

La Figura 3 mostra l'invenzione in forma di schema a blocchi.

La Figura 4 mostra un fonema nel quale una finestra A esclude asimmetricamente un impulso.

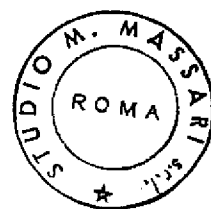
Forma di attuazione preferita

Nel testo che segue, l'invenzione è descritta con riguardo alle figure. Nel creare una parola artificiale, un testo arriva in 1 nella Figura 1. Il testo viene analizzato da 1 e suddiviso nei suoi componenti fondamentali. I fonemi vengono ora scelti dalla biblioteca. Il fonema nella biblioteca rappresenta un valore campione. Ciò comporta che il fonema si è visto conferire un valore campione rispetto alla durata, al tono ecc. Quando il fonema deve essere quindi inserito nel testo che è arrivato, occorre di norma una qualche forma di modifica del fonema, il che significa che l'estensione del fonema nel tempo deve essere cambiata. Ciò è rappresentato per esempio da tempi lunghi, corti o medi durante i quali deve essere rappresentata per esempio una vocale. Per trasformare il fonema di biblioteca, esso viene identificato in un certo numero di punti. Il fonema viene ora analizzato da 1. Nell'analisi,



vengono determinate parti portanti informazioni e parti portanti meno informazioni.. Le parti portanti meno informazioni vengono ora scelte per la trasformazione. E' stato osservato che le transizioni tra differenti fonemi sono di maggior significato che le parti più stabili nell'interno dei fonemi. Il processo di sviluppo , che contiene informazioni decisive riguardanti l'interpretazione del fonema, è di particolare importanza in questo contesto. I punti portanti meno informazioni vengono poi copiati per dare un certo numero di punti equivalenti nella nuova scala di tempo quando si prolunga il tempo. Ciò è illustrato nella Figura 2 in cui certi punti sono trasferiti dalla scala di tempo più corta ad un certo numero di punti nella scala di tempo più lunga. In questo modo le parti portanti informazioni del fonema sono conservate nella stiratura della scala ^{che} di tempo senza/la caratteristica del fonema venga cambiata.

La scala di tempo viene accorciata in maniera corrispondente. In questo caso, due o più punti nella parte del fonema che non porta informazioni sono combinati per formare un singolo punto. In questo modo le parti portanti informazioni sono anch'esse mantenute in gran parte intatte quando la scala di



tempo nel fonema viene accorciata.

Per ridurre l'effetto dell'eccitazione di una corda vocale precedente, è stata scelta una finestra che è stata ritagliata asimmetricamente. Ciò è illustrato nella Figura 4. La finestra viene così tagliata ripidamente all'inizio, registrando così il periodo iniziale dell'impulso e una parte minima della parte finale degli impulsi precedenti. E' anche conveniente il fatto che venga tagliata una parte dell'impulso tanto grande che si ottiene il suo valore massimo ed una proporzione dell'impulso smorzato. Questa soluzione offre la possibilità di spostare le transizioni tra gli impulsi di eccitazione delle corde vocali alle zone dove gli impulsi sono smorzati e non contengono informazioni significative. Un ritaglio a finestra di questo tipo dà inoltre come conseguenza di poter identificare la significanza dei singoli impulsi per comprendere i fonemi.

L'invenzione permette inoltre di ponderare differenti punti nel fonema di biblioteca in relazione agli elementi portanti informazioni. La ponderazione viene utilizzata nella trasformazione del fonema in maniera tale che i punti ai quali è stata conferita una minor ponderazione siano trasformati in un periodo di tempo più lungo delle parti che hanno ri-



cevuto una maggior ponderazione. Punti con bassa ponderazione vengono dunque assegnati per esempio a tre punti in una scala di tempo più lunga, mentre punti che rappresentano una ponderazione media sono trasformati per esempio in due punti nella nuova scala di tempo, e punti con ponderazione massima vengono trasferiti invariati nella nuova scala.

All'atto della trasformazione in una scala di tempo più corta di quella rappresentata nel fonema di base, tre punti che rappresentano per esempio la ponderazione minore sono combinati in un unico punto in maniera simile, e punti rappresentanti una ponderazione media sono combinati a due a due in un unico punto nel fonema accorciato nel tempo. Punti con la ponderazione la più alta sono trasferiti invariati nella nuova scala di tempo.

In questo modo l'invenzione rende possibile la scalata nel tempo di fonemi da effettuare senza che le parti del fonema portanti informazioni siano cambiate in una qualsiasi parte essenziale. Il metodo permette anche di legare l'uno all'altro differenti fonemi in maniera tale che che importanti informazioni nel fonema non siano distrutte nelle transizioni di fonemi. Ciò è causato dalla transizione tra i fonemi che avviene in parti non portanti alcuna informazione.



In questo modo l'invenzione permette che parole ed espressioni create tramite la sintesi vocale diventino quasi naturali.

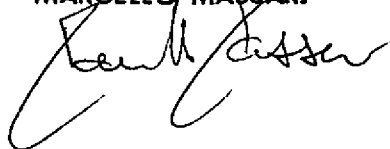
Poichè i punti scelti nel fonema rappresentano eccitazioni delle corde vocali nella parola, è possibile cambiare il tono fondamentale. Questo è necessario, per esempio, per dare il giusto carattere al fonema che viene creato. Il cambiamento del tono fondamentale viene ottenuto dalle eccitazioni delle corde vocali nel fonema creato che viene riprodotto in punti che vengono cambiati in relazione al fonema originale. Si supponga, per esempio, che il fonema di base rappresenti un suono con tono fondamentale immutato. Ciò comporta che le eccitazioni delle corde vocali avvengano con lo stesso reciproco distanziamento. In un fonema trasformato, tuttavia, il tono fondamentale viene cambiato durante il periodo di vita del fonema. Conoscendo il cambiamento nel tono fondamentale caratteristico, si deve tener conto di ciò nella trasformazione nel fonema nuovo, che in questo caso può essere un fonema che è invariato nel tempo o che viene trasformato in un tempo più lungo o più corto, gli intervalli di tempo sono determinati tra ciascuna eccitazione delle corde vocali che comparirà nel fonema. Così, per esempio, l'intervallo



di tempo tra la prima e la seconda eccitazione delle corde vocali è T_1 e l'intervallo tra l'ultima e la penultima eccitazione delle corde vocali è T_2 . Se in questo caso avviene che l'alterazione nel tono fondamentale cambia uniformemente con il passare del tempo, le eccitazioni intermedie delle corde vocali devono essere distribuite prendendo ciò in considerazione. Detta distribuzione viene opportunamente effettuata mediante modelli matematici conosciuti. Rispettive eccitazioni delle corde vocali nel fonema di base vengono poi trasferite a rispettivi punti nel fonema trasferito. Ciò consente una variazione nel tono fondamentale che corrisponde alla parola naturale.

L'invenzione non è limitata alla forma di realizzazione mostrata sopra ma può essere sottoposta a modifiche entro l'ambito di applicazione delle rivendicazioni che seguono ed entra il concetto dell'invenzione.

Il Mandatario
MARCELLO MASSARI




RIVENDICAZIONI

1. Metodo di sintesi vocale per trasformare un dato fonema da una prima scala di tempo ad una seconda scala di tempo, caratterizzato dal fatto che vengono determinati punti con un intervallo di tempo circostante, rappresentanti una parte di una curva del fonema; che vengono identificate le parti del fonema portanti più e rispettivamente, meno informazioni e che le parti del fonema portanti meno informazioni vengono trasformate nella seconda scala di tempo entro un periodo più lungo/più corto nella seconda scala di tempo, e che le parti del fonema portanti più informazioni vengono trasformate nella seconda scala di tempo essenzialmente senza essere cambiate nel tempo, per cui il carattere originale del fonema viene essenzialmente conservato.

2. Metodo secondo la rivendicazione 1, caratterizzato dal fatto che sono identificati i differenti punti nei fonemi ed essi ricevono differente ponderazione rispetto al grado d'informazione che rappresentano.

3. Metodo secondo le rivendicazioni 1 o 2, caratterizzato dal fatto che i punti con una minor ponderazione vengono trasformati entro un periodo



più lungo/più corto dei punti con maggior ponderazione, e che la trasformazione avviene mediante duplicazione o rimozione di punti con la minor ponderazione.

4. Metodo secondo la rivendicazione 1, caratterizzato dal fatto che transizioni di fonemi avvengono nelle parti del fonema che non portano alcuna informazione.

5. Metodo secondo la rivendicazione 1, caratterizzato dal fatto che i punti prescelti nella seconda scala di tempo sono scelti con lo stesso o con un altro intervallo di tempo rispetto alla prima scala di tempo per cui il tono fondamentale viene mantenuto o viene cambiato in relazione al dato fonema all'atto della trasformazione del fonema.

6. Disposizione nella sintesi vocale che comprende un elemento di selezione che sceglie un fonema da una sequenza vocale o da un elemento di memoria, per trasferire il fonema da una prima scala di tempo ad una seconda scala di tempo, caratterizzata dal fatto che l'elemento di selezione identifica un certo numero di punti con un intervallo di tempo circostante che rappresenta una parte della curva del fonema, per cui vengono



identificate le parti del fonema portanti informazioni, e rispettivamente portanti meno informazioni; che l'elemento trasforma le parti del fonema portanti meno informazioni entro un tempo più lungo/più corto nella trasformazione del fonema in un'altra scala di tempo rispetto alla scala di tempo originale rappresentata dal fonema; e che il carattere originale del fonema viene essenzialmente conservato.

7. Disposizione secondo la rivendicazione 6, caratterizzata dal fatto che l'elemento di selezione identifica e pondera differenti punti a seconda del contenuto d'informazione dei detti punti rispetto alla identificabilità del fonema.

8. Disposizione secondo le rivendicazioni 6 o 7, caratterizzata dal fatto che l'elemento di selezione trasforma punti con minor ponderazione entro una scala di tempo più lunga dei punti che rappresentano una ponderazione media, e che punti ai quali è stata attribuita una ponderazione superiore vengono trasformati in condizione invariata.

9. Disposizione secondo le rivendicazioni 6 o 7, caratterizzata dal fatto che tre o più punti con ponderazione bassa vengono combinati, e che punti con ponderazione media vengono combina-



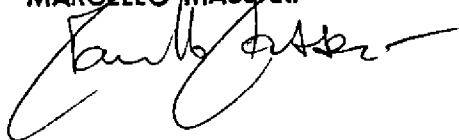
ti in un numero di punti minore che non i punti con bassa ponderazione; e che punti con ponderazione alta vengono trasformati in condizione invariata.

10. Disposizione secondo la rivendicazione 6, caratterizzata dal fatto che l'elemento di selezione cambia il tono fondamentale nel fonema durante il trasferimento alla seconda scala di tempo; e che i punti prescelti nel fonema rappresentano eccitazioni delle corde vocali nella parola.

Roma, 23 NOV. 1994

p. TELIA AB

Il Mandatario
MARCELLO MASSARI




RM 94 A 000763

P. TELIA AB

Il Mandatario
MARCELLO MASSARI

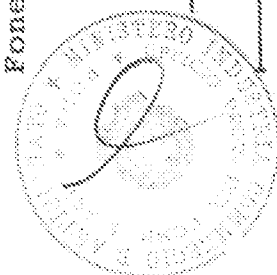
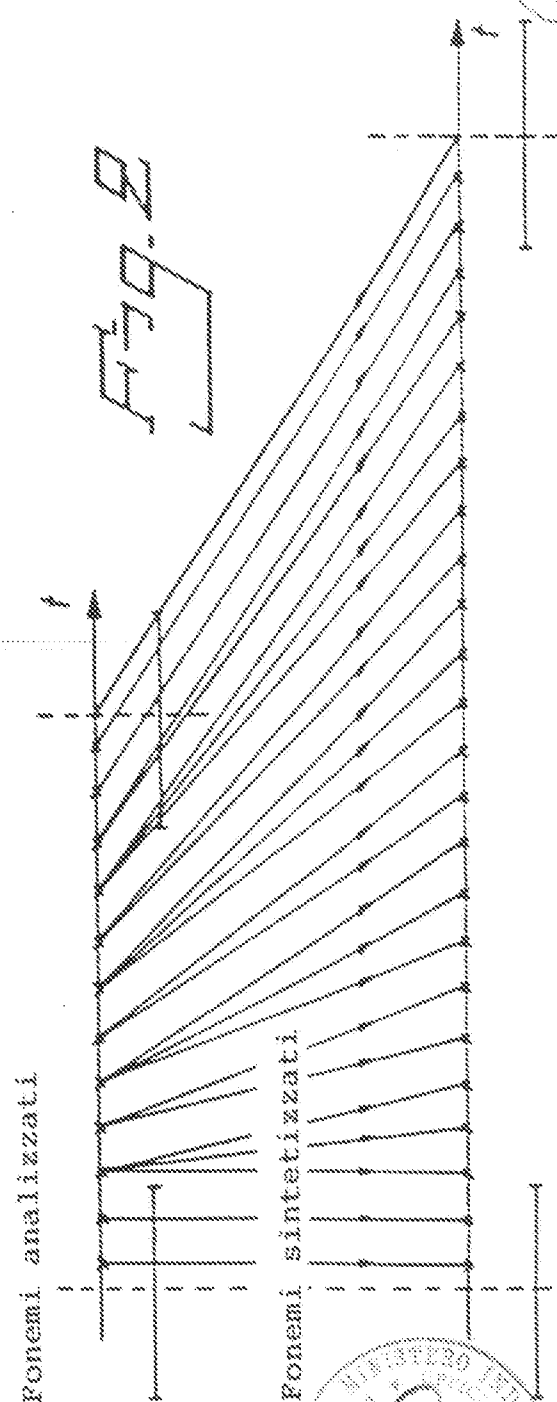
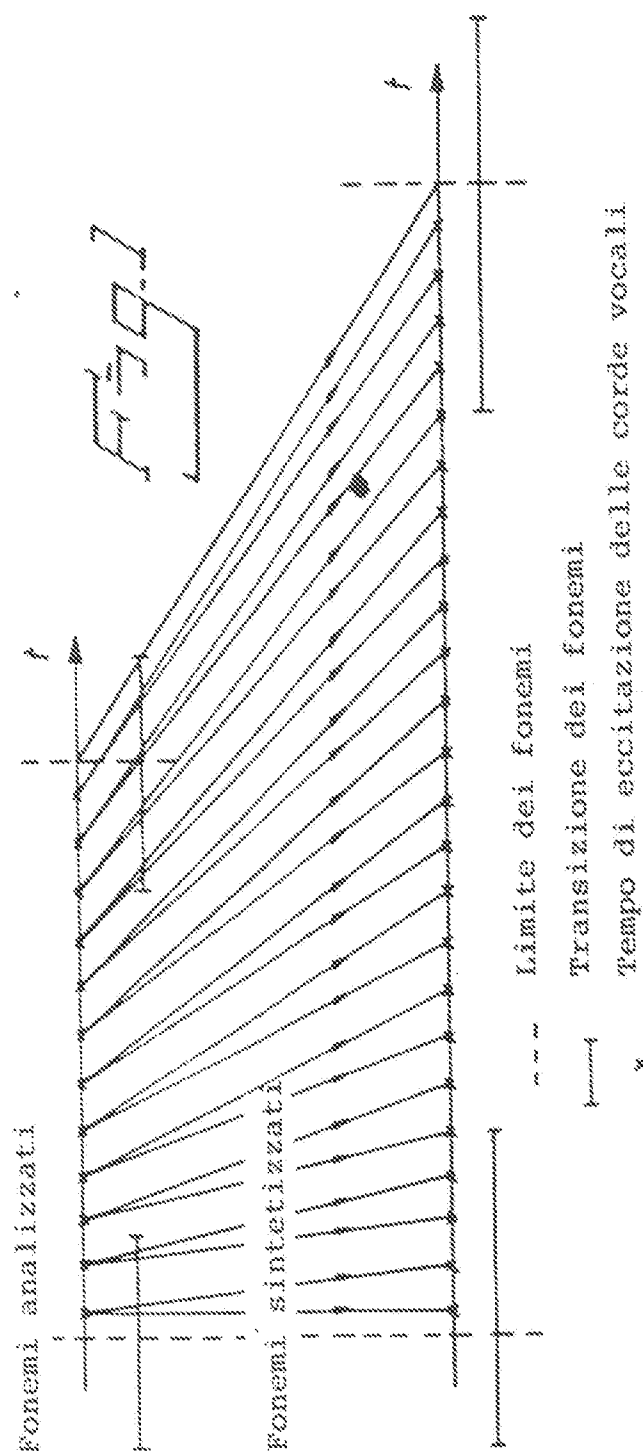


Fig. 3

RM 94 A 000763

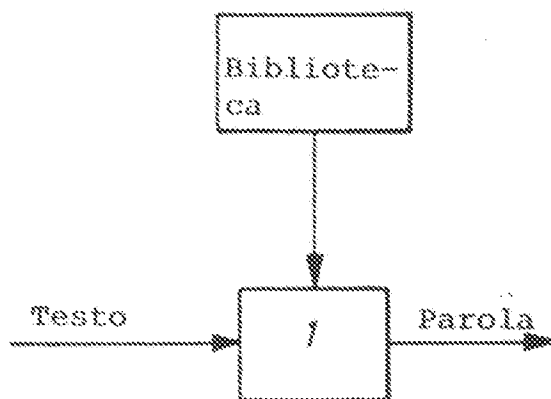
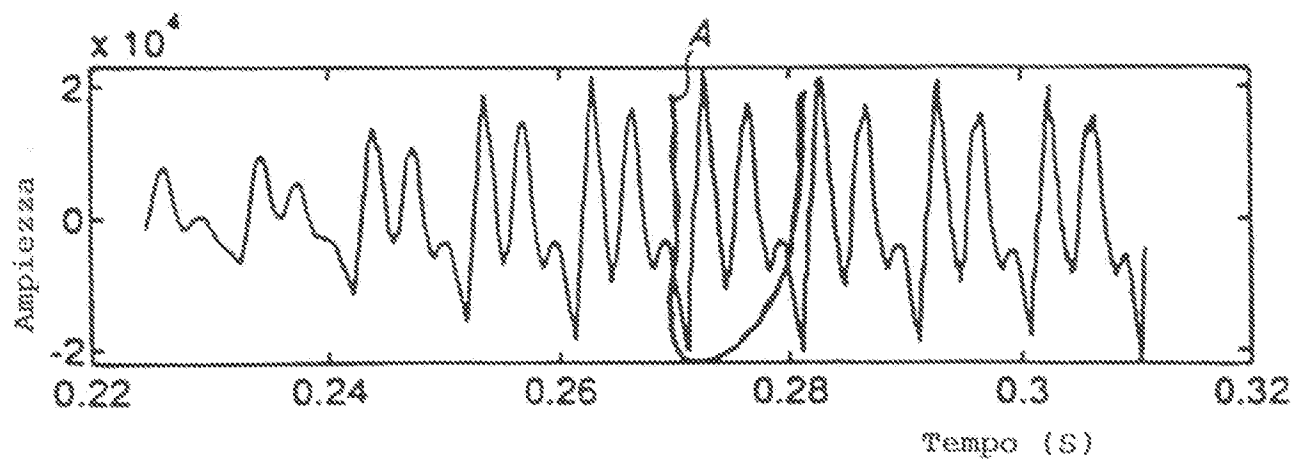


Fig. 4



p. TELIA AB

Il Mandatario
MARCELLO MASSARI

[Handwritten signature]

