

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 821 345 A1

(12)

## DEMANDE DE BREVET EUROPEEN

(43) Date de publication:

28.01.1998 Bulletin 1998/05

(51) Int Cl.<sup>6</sup>: G10L 3/00

(21) Numéro de dépôt: 97401752.7

(22) Date de dépôt: 21.07.1997

(84) Etats contractants désignés:

AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE

(30) Priorité: 24.07.1996 FR 9609313

(71) Demandeurs:

• FRANCE TELECOM  
75015 Paris (FR)

• LA POSTE

F-92777 Boulogne Billancourt Cédex (FR)

(72) Inventeur: Schwartzmann, Jean-Jacques  
14760 Bretteville sur Odon (FR)(74) Mandataire: Fréchède, Michel  
Cabinet Plasseraud  
84, rue d'Amsterdam  
F-75440 Paris Cédex 09 (FR)

## (54) Procédé d'extraction de la fréquence fondamentale d'un signal de parole

(57) L'invention concerne un procédé et un dispositif d'extraction de la fréquence fondamentale d'un signal de parole numérique.

Le signal de parole  $sp$  est préaccentué a). Le signal de parole préaccentué  $spp$  est b) subdivisé en trames successives de  $N$  échantillons, en recouvrement de  $N/2$  échantillons et un premier ensemble de valeurs  $X_1(k)$  du logarithme du spectre d'énergie est calculé sur  $M_1$  points. Un nombre  $p$  de coefficients cepstraux  $C(m)$  est calculé c) par transformée en cosinus discrète. Le signal de parole préaccentué  $spp$  est filtré et sous-échantillonné d) et un deuxième ensemble de valeurs  $X_2(k)$  du logarithme du spectre d'énergie du signal  $spf$  et les valeurs  $H(k)$  du spectre de fréquence lissé sont calculées e). Le maximum d'une fonction

$$P(k) = \sum_{l=1}^L X_2(k) - H(k)$$

est extrait pour  $k = F_0$  fréquence fondamentale du signal de parole.

Application à la reconnaissance vocale et à

l'authentification/identification du locuteur.

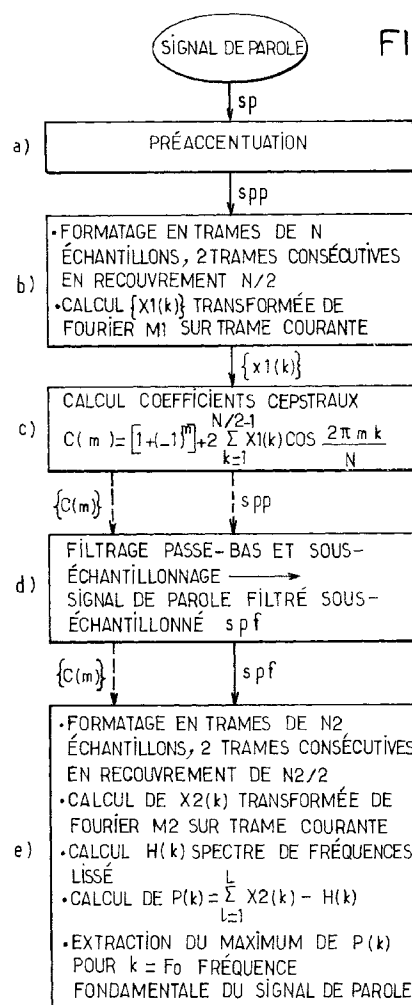


FIG.1a.

EP 0 821 345 A1

## Description

L'invention concerne un procédé d'extraction de la fréquence fondamentale d'un signal de parole.

Les techniques actuelles de traitement des signaux numériques de parole ont pour objet essentiel d'en extraire les paramètres fondamentaux, en vue d'en améliorer la qualité, par amélioration du rapport signal à bruit, et, le cas échéant, de déterminer l'origine du locuteur, en vue par exemple d'une authentification de ce dernier.

Parmi les paramètres fondamentaux précités, la fréquence fondamentale est l'un des paramètres qui caractérisent le mieux la voix d'un locuteur donné et qui permet donc de contribuer à l'authentification certaine de celui-ci.

De nombreux processus d'extraction de la fréquence fondamentale d'un signal de parole ont été proposés. Pour un panorama général des techniques proposées, on pourra utilement se reporter à l'ouvrage publié par W. HESS intitulé *"Pitch determination of speech signals : algorithms and methods"*, Springer-Verlag, New-York 1983.

Les techniques ou méthodes précitées peuvent être classées en deux grandes familles.

- Les méthodes temporelles telles que celles mettant en oeuvre un processus d'autocorrélation avec écrêtement central et comparaison des pics à une valeur de seuil ou celles désignées par AMDF, ces dernières ayant été décrites par R. BOITE et M. KUNT dans l'ouvrage intitulé *"Traitement de la parole"*, pages 193-195, Presses polytechniques romandes, Lausanne 1987, sont relativement peu coûteuses en temps de calcul car elles ne nécessitent pas la mise en oeuvre d'opérations arithmétiques de multiplication. Toutefois, elles manquent de précision et il est nécessaire, en conséquence, de procéder à un suréchantillonnage du signal de parole, afin d'obtenir une précision convenable, ce qui, bien entendu, entraîne une augmentation notable du temps de calcul effectif.

- Les méthodes fréquentielles sont, au contraire, basées sur l'analyse de la structure harmonique du spectre d'énergie en fonction de la fréquence du signal de parole. Parmi celles-ci, la méthode dite du peigne, décrite par P. MARTIN dans l'article intitulé *"Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne"*, publiée aux Journées d'Etude Parole 12, pp. 221-232, 1981, consiste à calculer la fonction d'intercorrélation entre le spectre du signal numérique de parole et une fonction en peigne, pour différentes valeurs de la distance entre les dents du peigne. Le maximum de la fonction d'intercorrélation est obtenu pour une distance entre deux dents consécutives du peigne, égale à la fréquence fondamentale du signal à analyser. Cette méthode présente une bonne fiabilité mais elle est relativement complexe, dans la mesure où elle nécessite un prélèvement fréquentiel consistant à ne retenir que les maxima du spectre et les valeurs adjacentes. En outre, il est nécessaire d'effectuer une interpolation afin d'augmenter la précision du résultat.

Une autre méthode, désignée par méthode de compression spectrale, a été publiée par NOLL (1970), confer l'ouvrage de W. HESS précédemment cité pages 414-417. Cette méthode, basée sur une analyse de la structure harmonique du spectre d'énergie en fonction de la fréquence du signal de parole, consiste à comprimer le spectre d'énergie du signal de parole le long de l'axe des fréquences, par des facteurs entiers successifs, puis à additionner les spectres comprimés obtenus au spectre initial. Ces opérations permettent, en principe, d'obtenir un maximum significatif, lequel résulte de la contribution cohérente des harmoniques de la fréquence fondamentale après compression. L'extraction de la fréquence fondamentale consiste alors à chercher le maximum du logarithme du produit harmonique défini par :

$$P(k) = \sum_{l=1}^L X(l)$$

où

$L = M/k$ ,  $M$  désignant le nombre de points du spectre  $X(l)$  désigne le logarithme du spectre d'énergie.

L'inconvénient de cette méthode réside dans le fait que l'amplitude des pics harmoniques décroît en fonction de la fréquence, avec une pente de l'ordre de -12 dB/octave. Bien qu'un processus de pré-accentuation permette de relever le niveau des harmoniques de fréquence élevée, certains pics harmoniques présentent un niveau d'énergie plus faible que d'autres en raison de la contribution des formants, ce qui provoque des erreurs fréquentes dans l'estimation de la valeur de la fréquence fondamentale.

La présente invention a pour objet la mise en oeuvre d'un procédé d'extraction de la fréquence fondamentale d'un signal de parole dans lequel l'extraction de la fréquence fondamentale est obtenue avec une fiabilité accrue.

Un autre objet de la présente invention est la mise en oeuvre d'un procédé d'extraction de la fréquence fondamentale d'un signal de parole dans lequel le processus d'extraction proprement dit de la fréquence fondamentale peut être conditionnel à la détection du voisement ou de l'absence de voisement des sons constitutifs du signal de parole.

Un autre objet de la présente invention est enfin la mise en oeuvre d'un procédé d'extraction de la fréquence

fondamentale d'un signal de parole dans lequel la valeur de fréquence fondamentale extraite est en outre soumise à un processus de post-traitement, du type par apprentissage, afin d'éliminer toute valeur improbable ou aberrante.

Le procédé d'extraction de la fréquence fondamentale d'un signal de parole, succession d'échantillons numériques, objet de la présente invention, est remarquable en ce qu'il comprend au moins les étapes consistant à soumettre ce signal de parole à un processus de préaccentuation, pour engendrer un signal de parole préaccentué, calculer, à partir du signal de parole préaccentué, pour chaque trame courante d'une succession de trames correspondant chacune en durée à un nombre déterminé N d'échantillons, deux trames consécutives présentant chacune un recouvrement de durée en nombre d'échantillons consécutifs communs au plus égal à 50/100 du nombre N d'échantillons, un premier ensemble de valeurs  $X_1(k)$  du logarithme du spectre d'énergie par transformée de Fourier sur un nombre  $M_1$  de points, calculer, à partir de ce premier ensemble de valeurs, un nombre p déterminé de premiers coefficients cepstraux  $C(m)$ , par application d'une transformée en cosinus discrète auxdites valeurs  $X_1(k)$  sur un nombre de ces valeurs au moins égal à la moitié du nombre N d'échantillons constitutifs de la trame courante, cette transformée vérifiant la relation :

$$C(m) = [1 + (-1)^m] X(0) + 2 \sum_{k=1}^{N/2-1} X_1(k) \cdot \cos\left(\frac{2\pi mk}{N}\right)$$

avec  $m = [1, 2, \dots, p]$ , soumettre le signal de parole préaccentué à un filtrage de type passe-bas et à un sous-échantillonnage, pour engendrer un signal de parole filtré sous-échantillonné, calculer, par compression spectrale, à partir du signal de parole filtré sous-échantillonné et à partir des coefficients cepstraux pour chaque trame courante d'une succession de trames de même recouvrement de durée, la fréquence fondamentale, maximum de rang k, d'une fonction  $P(k)$  représentative de la différence entre un deuxième ensemble des valeurs  $X_2(k)$  du logarithme du spectre d'énergie et l'ensemble des valeurs  $H(k)$  du spectre de fréquences lissé, ladite fonction vérifiant la relation :

$$P(k) = \sum_{l=1}^L X_2(k) - H(k)$$

avec  $L = M_2/k$ , k variant entre une première et une deuxième valeur représentatives d'une bande de fréquences basses comprises entre 70 et 450 Hz, ladite fonction  $P(k)$  présentant un maximum pour  $k=F_0$ , valeur extraite de la fréquence fondamentale du signal de parole.

Le procédé objet de la présente invention trouve en particulier application à la reconnaissance vocale et à l'identification de locuteurs à partir de signatures sonores.

Il sera mieux compris à la lecture de la description et à l'observation des dessins ci-après dans lesquels :

- la figure la représente un organigramme illustratif de l'ensemble des étapes permettant la mise en oeuvre du procédé objet de la présente invention ;
- la figure 1b représente un organigramme illustratif d'une variante de mise en oeuvre avantageuse du procédé objet de la présente invention, dans laquelle certaines étapes sont conduites en parallèle ou, le cas échéant, sous système d'exploitation multitâche afin de permettre un mode opératoire en temps réel, sans toutefois nécessiter une puissance de calcul très importante ;
- la figure 2a représente un détail de réalisation d'une succession d'étapes élémentaires permettant une mise en oeuvre optimale de l'étape terminale de calcul par compression spectrale de la fréquence fondamentale du signal de parole du procédé, objet de la présente invention, illustré conformément à la figure 1a ou 1b ;
- la figure 2b représente une série de signaux obtenus dans le domaine fréquentiel suite à la mise en oeuvre des étapes élémentaires illustrées en figure 2a ;
- les figures 3a, 3b, 3c et 3d représentent un mode opératoire de formatage de trames d'échantillons, constitutifs du signal de parole, un processus de discrimination des trames courantes en fonction d'un critère relatif au caractère voisé ou non voisé de chaque trame courante, un mode d'établissement de ce critère et un abaque d'attribution d'un indice de voisement de segments temporels constitutifs de chaque trame respectivement ;
- la figure 4 représente un schéma synoptique de l'architecture d'un dispositif permettant la mise en oeuvre du procédé, objet de la présente invention, à partir d'un micro-ordinateur hôte et d'un processeur de signal numérique spécialisé ou dédié connectés par une liaison de type BUS.

Une description plus détaillée du procédé d'extraction de la fréquence fondamentale d'un signal de parole, objet de la présente invention, sera maintenant donnée en liaison avec les figures 1a et 1b.

Ainsi qu'on l'observera sur la figure 1a, le signal de parole sur lequel on souhaite procéder à l'extraction de la fréquence fondamentale, conformément au procédé objet de la présente invention, est par exemple un signal analogique représentatif de mots et de syllabes distincts, ce signal analogique étant transformé en une succession d'échantillons numériques, le signal de parole, dans sa forme numérique, étant désigné par sp sur la figure 1a.

Ainsi qu'il apparaît en outre sur la figure précitée, le signal de parole sp est alors soumis à un processus de préaccentuation permettant d'engendrer un signal de parole préaccentué, noté spp. Le processus de préaccentuation est un processus de type classique, lequel, à ce titre, ne sera pas décrit de manière détaillée. Ce processus consiste en une préaccentuation globale, laquelle consiste en fait à appliquer une valeur de gain croissante avec la fréquence pour compenser l'atténuation des harmoniques de rang supérieur. A titre d'exemple non limitatif, on indique que le processus de préaccentuation globale peut consister à appliquer au signal de parole sp une fonction de transfert du type :

$$G(z) = 1 - z^{-1}.$$

Dans la relation précitée, on indique que  $z = e^{j\omega}$  où  $\omega = 2\pi f$ , f désignant la fréquence instantanée du signal de parole.

Le procédé objet de la présente invention, ainsi que représenté en figure 1a, consiste ensuite, en une étape b), à effectuer un formatage du signal de parole préaccentué spp. Cette opération de formatage consiste en fait à constituer le signal de parole préaccentué spp en trames successives comportant chacune N échantillons et correspondant à une durée de ces N échantillons, deux trames consécutives présentant chacune un recouvrement de durée en nombre d'échantillons consécutifs communs au plus égal à 50/100 du nombre N d'échantillons constitutifs de chaque trame.

L'étape b) précitée consiste également à calculer, sur chaque trame courante désignée par  $T_q$ , un premier ensemble de valeurs, noté  $X_1(k)$  du logarithme du spectre d'énergie pour la trame considérée par application d'une transformée de Fourier sur un nombre  $M_1$  de points.

D'une manière pratique, on indique que le nombre  $M_1$  de points sur lequel la transformée de Fourier est appliquée est choisi de façon que le théorème de Shannon soit satisfait. A titre d'exemple non limitatif, on indique que pour des trames constituées par 256 échantillons successifs et pour une durée de chaque trame courante égale à 32 ms, le nombre  $M_1$  de points peut être pris égal à 128.

L'étape b) précitée, représentée en figure 1a, permet alors de disposer du premier ensemble de valeurs, noté  $\{X_1(k)\}$ .

Ainsi que représenté sur la figure la précitée, le procédé objet de la présente invention consiste ensuite à effectuer en une étape c) le calcul, à partir du premier ensemble de valeurs  $\{X_1(k)\}$ , un nombre p déterminé de premiers coefficients cepstraux notés  $C(m)$  du logarithme du spectre d'énergie défini par le premier ensemble de valeurs  $\{X_1(k)\}$ .

Les coefficients cepstraux précités vérifient la relation :

$$C(m) = [1 + (-1)^m] X(0) + 2 \sum_{k=1}^{N/2-1} X_1(k) \cdot \cos\left(\frac{2\pi m k}{N}\right)$$

Dans cette relation, on indique que m est un entier prenant les valeurs = [1, 2, ..., p], p désignant le nombre de premiers coefficients cepstraux calculé et retenu pour la mise en oeuvre du procédé objet de la présente invention. A titre d'exemple non limitatif, on indique que p peut être limité à 16.

A la fin de l'étape c), on dispose ainsi des coefficients cepstraux précités, lesquels vont permettre la mise en oeuvre des étapes suivantes du procédé objet de l'invention, tel que représenté en figure 1a.

Suite à l'étape c) précitée, le procédé objet de la présente invention consiste, en une étape d), à soumettre le signal de parole préaccentué spp à un filtrage de type passe-bas et à un sous-échantillonnage pour engendrer un signal de parole filtré sous-échantillonné, noté spf.

Sur la figure 1a, on a représenté une liaison en trait mixte entre l'étape c) et l'étape d), cette liaison en trait mixte indiquant une opération réalisée sur le signal de parole préaccentué spp disponible postérieurement à l'étape a) de préaccentuation globale. On comprend en particulier que le signal de parole sous forme numérique sp, consistant en fait en une salve de mots successifs par exemple, le signal de parole préaccentué spp peut être mémorisé postérieurement à l'étape de préaccentuation réalisée à l'étape a), et que, bien entendu, l'étape d) peut être réalisée à partir du signal de parole préaccentué spp précédemment cité.

D'une manière générale, on indique que le filtrage de type passe-bas peut être réalisé grâce à un filtre passe-bas de fréquence de coupure égale à 2 kHz au moyen d'un filtre à réponse impulsionnelle finie, dit filtre RIF, à 47 coefficients. Le signal filtré issu du filtrage précité peut alors être soumis à un sous-échantillonnage, le sous-échantillonnage pouvant être réalisé par décimation, pour délivrer le signal de parole filtré sous-échantillonné noté spf.

L'étape d) précitée est alors suivie, ainsi que représenté en figure 1a, d'une étape e) consistant à calculer par compression spectrale la fréquence fondamentale maximum de rang k d'une fonction P(k) représentative de la différence entre un deuxième ensemble de valeurs  $X_2(k)$  du logarithme du spectre d'énergie du signal de parole filtré sous-échantillonné spf, et de l'ensemble des valeurs H(k) du spectre de fréquences lissé obtenu à partir des coefficients cepstraux disponibles à la fin de l'étape c) précédemment mentionnée dans la description.

La fonction P(k) vérifie la relation :

$$P(k) = \sum_{l=1}^L X_2(k) - H(k)$$

D'une manière générale, l'étape e) représentée en figure 1a consiste également en une étape de formatage en trames de  $N_2$  échantillons, avec  $N_2 = N/2$ , deux trames consécutives étant en recouvrement de  $N_2/2$  échantillons du signal de parole filtré sous-échantillonné spf, le formatage étant bien entendu semblable au formatage appliqué au début de l'étape b) sur le signal de parole préaccentué spp.

L'étape de formatage réalisée à l'étape e) est alors suivie d'une étape effective de calcul du deuxième ensemble des valeurs  $\{X_2(k)\}$  du logarithme du spectre d'énergie, ce calcul étant effectué par application d'une transformée de Fourier sur un nombre  $M_2$  de points pour chaque trame courante obtenue à l'issue du formatage réalisé. Le deuxième ensemble de valeurs  $\{X_2(k)\}$  est avantageusement calculé par l'intermédiaire d'une transformée de Fourier rapide FFT appliquée sur  $M_2 = 2048$  points en utilisant la méthode de remplissage par des zéros.

L'étape de calcul du deuxième ensemble de valeurs  $\{X_2(k)\}$  est alors suivie d'une étape de calcul du spectre de fréquences lissé H(k) à partir des coefficients cepstraux C(m) disponibles dès la fin de l'étape c), la liaison entre l'étape c) et l'étape e) sur la figure 1a étant représentée en trait mixte pour cette raison. Le spectre lissé H(k) est calculé par l'application d'une transformée en cosinus sur les p coefficients cepstraux disponibles.

L'étape de calcul du spectre de fréquences lissé est alors suivie d'une étape de calcul de la fonction P(k) vérifiant la relation précédemment citée dans la description. Dans cette relation, on indique que L est égal à  $M_2/k$  pour k variant entre une première et une deuxième valeur représentatives d'une bande de fréquences basses comprises entre 70 et 450 Hz. La fonction P(k) présente alors un maximum pour  $p = F_0$ , valeur extraite de la fréquence fondamentale du signal de parole.

Le procédé d'extraction de la fréquence fondamentale d'un signal de parole, objet de la présente invention, permet, par compression spectrale, par le calcul du produit harmonique de la différence entre le spectre d'énergie du signal de parole et le spectre du signal lissé, d'éliminer la contribution des formants et d'extraire la structure harmonique de la fréquence fondamentale du signal de parole.

Dans le mode de réalisation de la figure 1a, on a représenté, à titre d'exemple non limitatif, une réalisation de type séquentiel, les étapes a) à e) pouvant être exécutées successivement. On comprend en particulier que, d'une part, le signal de parole préaccentué spp, et que, d'autre part, les coefficients cepstraux, en particulier les p coefficients cepstraux utilisés, peuvent être mémorisés à l'issue de l'étape c) respectivement postérieurement à l'étape a) pour permettre la mise en oeuvre séquentielle des étapes b) à e) précédemment mentionnées.

Toutefois, et afin de ne pas surcharger inutilement le processeur de calcul utilisé pour la mise en oeuvre des étapes a) à e) précitées, mais afin toutefois de faciliter l'exécution des étapes précitées en temps réel, le procédé objet de la présente invention peut être mis en oeuvre, dans une variante d'exécution telle que représentée en figure 1b, en parallèle, les étapes, b), c) étant réalisées séquentiellement, en parallèle avec les étapes d) et e) à partir du signal de parole préaccentué spp. Ce mode de réalisation tel que représenté en figure 1b, est rendu possible en raison du fait que les étapes b) et c) sont qualitativement indépendantes des étapes d) et e) et peuvent être réalisées en parallèle sur le signal de parole préaccentué spp.

En ce qui concerne les sous-étapes de formatage réalisées aux étapes b) et e) sur le signal de parole préaccentué spp, respectivement sur le signal de parole filtré sous-échantillonné spf, on indique que ces étapes de formatage peuvent être réalisées par un adressage approprié sur le signal de parole préaccentué spp, respectivement le signal de parole filtré sous-échantillonné spf. Bien entendu, la réalisation de la sous-étape de calcul du spectre de fréquences lissé H(k) de l'étape e) est conditionnelle à la disponibilité des p coefficients cepstraux C(m) en fin de l'étape c).

La mise en oeuvre de la variante de réalisation du procédé objet de la présente invention telle que représentée en figure 1b ne préjuge aucunement de la structure mono ou multiprocesseur du dispositif permettant la mise en oeuvre du procédé objet de la présente invention, une structure monoprocesseur avec système d'exploitation multitâche pouvant bien entendu être envisagée, ainsi qu'il sera décrit ultérieurement dans la description.

En outre, on indique que, dans une autre variante de réalisation, le procédé, objet de la présente invention, peut consister à calculer un seul ensemble de valeurs, noté X(k), du spectre d'énergie du signal de parole à l'étape c) sur un nombre M de points égal par exemple à 2048, c'est-à-dire à la valeur  $M = M_2$  la plus grande précédemment décrite

dans la description, et à mémoriser cet ensemble de valeurs. Le nombre  $M_1 = 128$  de valeurs utilisées pour le calcul des coefficients cepstraux à l'étape c) peut alors être obtenu par décimation à partir de l'ensemble de valeurs  $X(k)$ . Toutefois, on indique que cet autre mode de réalisation, bien qu'équivalent au mode de réalisation décrit avec calcul du premier ensemble de valeurs  $X_1(k)$  puis du deuxième ensemble de valeurs  $X_2(k)$ , présente l'inconvénient de nécessiter le maintien en mémoire de l'ensemble des valeurs  $X(k)$  pendant la totalité du temps d'exécution du processus de calcul pour chacune des trames courantes, ce qui provoque un encombrement de mémoire néfaste à la gestion de l'ensemble des ressources de calcul.

Une description plus détaillée du processus de mise en oeuvre de l'étape e) du procédé, objet de la présente invention, telle que représentée en figures 1a et 1b, sera maintenant donnée en liaison avec la figure 2a.

Selon la figure précitée, l'étape e) de calcul par compression spectrale consiste, ainsi que mentionné précédemment dans la description, à réaliser une étape  $e_1$ ) comprenant le formatage en trames de  $N_2$  échantillons à partir du signal de parole filtré sous-échantillonné  $spf$  et de calcul du deuxième ensemble de valeurs  $X_2(k)$  du logarithme du spectre d'énergie par application d'une transformée de Fourier sur un nombre  $M_2$  de points sur une bande de fréquences comprises entre 0 et 2 KHz.

La sous-étape  $e_1$ ) est suivie d'une sous-étape  $e_2$ ) consistant à calculer l'enveloppe spectrale  $H(k)$  ou spectre de fréquences lissé de la trame courante sur la bande de fréquences comprises entre 0 et 2 kHz sur un même nombre  $M_2$  de points, par application sur les  $p-1$  premiers coefficients cepstraux d'une transformée en cosinus vérifiant la relation :

$$H(k) = C(0) + 2 \sum_{m=1}^{p-1} C(m) \cdot \cos\left(2\pi \frac{mk}{Q}\right)$$

Dans cette relation,  $k$  prend les valeurs  $[0, 1, 2, \dots, M_2]$  et  $M_2$  est égal à  $Q/4$  avec  $Q = 8192$ .

La sous-étape  $e_2$ ) est suivie d'une sous-étape  $e_3$ ) consistant à calculer la différence, notée  $D(k) = X_2(k) H(k)$ .

La sous-étape  $e_3$ ) est elle-même suivie d'une sous-étape  $e_4$ ) consistant à calculer la fonction  $P(k)$  par compression spectrale de la différence  $D(k)$  sur la bande de fréquences basses comprises entre 70 et 450 Hz. La fonction  $P(k)$  n'est autre que le produit harmonique de la différence  $D(k)$ . Ce calcul est effectué pour  $L = M_2/k$ ,  $k$  variant pour des valeurs représentatives de 70 à 450 Hz, c'est-à-dire dans la bande de fréquences basses précédemment citée.

Enfin, la sous-étape  $e_4$ ) est elle-même suivie d'une sous-étape  $e_5$ ) réalisant l'extraction du maximum de la fonction  $P(k)$  pour la valeur de  $k$  représentative de la valeur  $F_0$ , fréquence fondamentale du signal de parole.

La sous-étape  $e_5$ ) peut être réalisée à partir d'un programme de tri des valeurs successives de la fonction  $P(k)$  dans la bande de fréquences basses précitée. Le programme de tri est un programme de type classique de recherche de valeur maximum parmi plusieurs valeurs.

Sur la figure 2b, on a représenté successivement des diagrammes dans un espace énergie  $W$ -fréquence relatifs successivement au spectre à court terme entre 0 et 2 KHz d'une trame d'un signal de parole, la trame ayant une durée de 32 ms sur 2048 points, ce diagramme pouvant correspondre à une trame obtenue suite à la sous-étape de formatage réalisée en la sous-étape  $e_1$ ) de la figure 2a, l'enveloppe spectrale obtenue par transformée en cosinus appliquée sur les 16 premiers coefficients cepstraux, cette enveloppe représentant uniquement la contribution des formants, c'est-à-dire le spectre lissé  $H(k)$  obtenu à l'issue de la sous-étape  $e_2$ ) de la figure 2a par exemple, la différence  $D(k)$  entre les deux spectres précédents, différence dans laquelle il ne subsiste que la structure de fréquence fondamentale du signal de parole, la contribution des formants étant éliminée, ce diagramme correspondant aux valeurs  $D(k)$  de la différence obtenue à l'issue de la sous-étape  $e_3$ ) de la figure 2a, puis, enfin, la courbe obtenue par compression spectrale de la structure de fréquence fondamentale du signal de parole entre 70 et 450 Hz, cette fonction présentant une valeur maximum ou pic significative pour la fréquence  $F_0$ , ce dernier diagramme correspondant à la mise en oeuvre des sous-étapes  $e_4$ ) et  $e_5$ ) de la figure 2a.

Le procédé objet de la présente invention peut normalement être mis en oeuvre sur un flot continu ou pseudo-continu de mots ou syllabes constitutifs d'un signal de parole.

Toutefois, des investigations poussées ont montré l'intérêt de la mise en oeuvre d'un processus de discrimination entre trames voisées et trames non voisées, car l'échantillonnage de trames non voisées est susceptible d'entraîner des erreurs dans l'évaluation de la fréquence fondamentale du signal de parole en raison du fait que, pour les trames non voisées, les sons ne résultent pas d'une vibration périodique des cordes vocales, ces trames non voisées n'étant pas significatives de la fréquence fondamentale de ce signal de parole.

Dans ce but, et suite à la sous-étape consistant à soumettre le signal de parole préaccentué  $spp$ , respectivement le signal de parole filtré sous-échantillonné  $spf$  à la sous-étape de formatage en trames, le procédé objet de la présente invention peut consister avantageusement, en outre, à discriminer, parmi l'ensemble des trames successives, les tra-

mes voisées et les trames non voisées puis à éliminer chaque trame non voisée. En fait, les trames non voisées ne sont pas éliminées physiquement de la succession des trames courantes. Ces trames non voisées sont discriminées par affectation à celles-ci d'une valeur de fréquence fondamentale arbitraire, valeur nulle, ainsi qu'il sera décrit ultérieurement dans la description.

Ainsi, comme représenté en figure 3a, la constitution de ces signaux en trames successives de N respectivement N<sub>2</sub> échantillons peut être réalisée de manière classique par réception et mémorisation de ces échantillons à des adresses spécifiques d'une mémoire vive par exemple, puis lecture séquentielle, ainsi que représenté en figure 3a, des trames successives, avec lecture par exemple de la trame de rang q-1 par lecture simultanée des N échantillons correspondants, puis lecture au bout de la durée de trame, soit 32 ms, de la trame de rang q ultérieure correspondant à N échantillons en recouvrement de N/2 échantillons par rapport à la trame antérieure de rang q-1, et ainsi de suite pour la trame de rang q+1 et les trames suivantes. Ce processus de lecture peut être réalisé avantageusement par simple adressage en lecture de la mémoire contenant les échantillons du signal de parole. Ainsi que représenté en figure 3b, le formatage en trames ayant été effectué sur l'un ou l'autre signal ainsi que décrit en relation avec la figure 3a, le processus de discrimination entre trames voisées et trames non voisées peut consister, à partir de la trame courante T<sub>q</sub>, en une étape 100, à appliquer un critère 101 de discrimination entre trames courantes voisées ou non voisées. Sur réponse négative au critère 101 précité, à la trame courante T<sub>q</sub> est affectée une valeur arbitraire de fréquence fondamentale, valeur zéro par exemple, en une étape 102, alors qu'au contraire, sur réponse positive au critère 101, la trame courante est conservée à l'étape 103 pour traitement selon le processus de calcul pour réaliser l'extraction de la fréquence fondamentale du signal de parole. La succession des trames courantes conservées à l'étape 103 est alors soumise, en fonction du signal considéré spp, respectivement spf, au calcul du premier ensemble de valeurs X<sub>1</sub>(k) ou X<sub>2</sub>(k) respectivement, dans le cadre de la mise en oeuvre de l'étape b) ou de l'étape e), ou sous-étape e<sub>1</sub>), des figures 1a, 1b ou 2a.

En ce qui concerne la discrimination proprement dite des trames voisées et non voisées, on indique que celle-ci peut consister, ainsi que représenté en liaison avec la figure 3c, à subdiviser chaque trame courante T<sub>q</sub> en un nombre ST de segments de trames contigus successifs, puis à établir, pour chacun des segments de trame, un critère de discrimination de voisement. Sur la figure 3c, on a représenté quatre segments de trame contigus, notés S<sub>1</sub> à S<sub>4</sub>, chaque segment de trame comportant donc 64 échantillons et occupant une durée de 8 ms.

Selon un mode de réalisation particulièrement avantageux non limitatif, on indique que le critère de discrimination de voisement peut consister à affecter à chaque segment de trame considéré un indice de voisement dont la valeur est comprise entre 0 et 1. Chaque indice de voisement est noté Vs(1) à Vs(4) et est représentatif du niveau d'énergie basse fréquence du segment de trame S<sub>1</sub> à S<sub>4</sub> considéré, selon une loi sensiblement linéaire. Enfin, chaque trame courante T<sub>q</sub> est classée comme trame non voisée par comparaison d'une combinaison linéaire des indices de voisement de chaque segment à une valeur de seuil déterminée. A titre d'exemple non limitatif, on indique que la combinaison linéaire précitée des indices de voisement peut consister à calculer la moyenne arithmétique de ces indices et à comparer cette moyenne arithmétique à la valeur de seuil ε précitée, le critère de comparaison de la combinaison linéaire s'écrivant :

$$\sum_{k=1}^4 V_s(k)/4 < \epsilon .$$

Enfin, ainsi que représenté en figure 3d, la valeur de chaque indice de voisement peut être affectée en fonction de l'énergie basse fréquence de chaque segment selon l'abaque représenté sur la figure précitée. Dans le mode de réalisation étudié pour la mise en oeuvre du procédé objet de la présente invention, on indique que la valeur d'indice de voisement affectée est linéaire entre les valeurs 0 et 1 pour des valeurs d'énergie basse fréquence de chaque segment comprises entre -35 et -15 dB. Ces valeurs peuvent bien entendu être modifiées.

Enfin, des erreurs peuvent survenir dans l'estimation de la valeur de la fréquence fondamentale du signal de parole, ces erreurs pouvant être dues à la présence dans une même trame de segments voisés et de segments non voisés ou de silences. Ces types d'erreurs sont désignés par erreurs de transition. De telles erreurs peuvent également survenir dans les trames voisées ou mixtes de faible énergie. Dans certaines conditions, il est alors possible de corriger ces erreurs alors que, lorsque la correction n'est pas possible, la valeur de la fréquence fondamentale du signal de parole est prise égale arbitrairement à une valeur fictive, la valeur zéro, par convention par exemple, de manière semblable à la valeur attribuée aux trames non voisées ou aux trames de silence.

Le procédé objet de la présente invention peut consister alors, en outre, à effectuer un post-traitement de la valeur extraite de fréquence fondamentale du signal de parole.

Cette étape de post-traitement peut consister par exemple à établir un histogramme des fréquences fondamentales, afin de déterminer la plage de valeurs de fréquences les plus probables ainsi que les bornes de valeurs inférieure

et supérieure de ces valeurs. Suite à l'établissement de l'histogramme des fréquences fondamentales, le processus de post-traitement peut consister à soumettre chaque valeur extraite de fréquence fondamentale à un critère de tri par rapport aux bornes de valeurs inférieure et supérieure, pour obtenir des valeurs triées représentatives de l'évolution des valeurs extraites de fréquence fondamentale.

5 Ces valeurs triées peuvent ensuite être soumises à un filtrage non linéaire pour supprimer les valeurs aberrantes.

Ainsi, pour une bande de fréquences la plus probable comprise entre des valeurs notées B.Sup respectivement B.Inf, valeur supérieure et valeur inférieure de la bande de fréquences, et pour des valeurs de fréquence fondamentale successives notées  $F_0(i)$ , le processus de correction peut être réalisé selon les étapes de calcul ci-après :

si  $F_0(i) > B.Sup$

10

$$F_0(i) = F_0(i)/2$$

si  $F_0(i) > B.Sup$  ou  $F_0(i) < B.Inf$

15

$$F_0(i) = 0$$

sinon si

20

$$|F_0(i) - F_0(i-1)| > \gamma$$

25

$$F_0(i) = 0$$

sinon si  $F_0(i) < B.Inf$

30

$$F_0(i) = F_0(i)*2$$

si  $F_0(i) > B.Sup$  ou  $F_0(i) < B.Inf$

35

$$F_0(i) = 0$$

sinon si  $|F_0(i) - F_0(i-1)| > \gamma$

40

$$F_0(i) = 0.$$

Dans le processus de calcul précité, l'indice  $i$  affecté aux valeurs de fréquence fondamentale désigne l'ordre successif des valeurs extraites,  $\gamma$  représente une valeur de seuil arbitraire à laquelle est comparée la différence entre deux valeurs de fréquence fondamentale successives de rang  $i$  et  $i-1$ .

45 Suite au filtrage non linéaire, les valeurs nulles isolées sont ensuite recalculées par interpolation linéaire, alors que les valeurs non nulles isolées au milieu d'une suite de zéros sont affectées à la valeur 0 par convention. Enfin, des paramètres statistiques tels que les valeurs  $F_0$  maximum et minimum ainsi que la valeur moyenne peuvent être calculés.

50 Une description d'un dispositif permettant la mise en oeuvre du procédé, objet de la présente invention, sera maintenant donnée en liaison avec la figure 4.

Le dispositif représenté sur la figure précitée permet la mise en oeuvre du procédé, objet de la présente invention, précédemment décrit dans la description. Ce dispositif présente une architecture adaptée à la mise en oeuvre de ce procédé.

55 Ainsi que représenté sur la figure précitée, il comprend un circuit 1 d'échantillonnage et de conversion analogique-numérique d'un signal de parole analogique d'entrée en une suite d'échantillons numériques. En outre, un ordinateur hôte 2 est prévu afin de permettre la conduite de la succession des étapes a) à e) du procédé objet de la présente invention, ainsi que la gestion et la commande d'organes périphériques tels que notamment le circuit d'échantillonnage 1 et de conversion analogique-numérique, ainsi qu'il sera décrit ultérieurement dans la description.



Le dispositif représenté en figure 4 comporte en outre un processeur de signal numérique dédié 3 interconnecté, d'une part, par une liaison par BUS au micro-ordinateur hôte 2, et, d'autre part, par une liaison spécifique au circuit de conversion analogique-numérique 1, ce processeur de signal numérique 3 permettant d'effectuer les opérations de calcul du premier ensemble de valeurs  $X_1(k)$  du logarithme du spectre d'énergie du signal de parole par transformée de Fourier sur un nombre  $M_1$  de points, le calcul des premiers coefficients cepstraux, le filtrage passe-bas et le sous-échantillonnage du signal de parole  $sp$  ainsi que le calcul du deuxième ensemble de valeurs  $X_2(k)$  du logarithme du spectre d'énergie, le calcul de l'ensemble des valeurs  $H(k)$  du spectre de fréquences lissé, le calcul de la fonction  $P(k)$  et l'opération d'extraction du maximum de la fonction  $P(k)$  pour  $k = F_0$ , valeur extraite de la fréquence fondamentale du signal de parole. L'acquisition des échantillons constitutifs du signal de parole  $sp$  est conduite par l'ordinateur hôte 2, par l'intermédiaire du processeur de signal 3.

Dans un mode de réalisation non limitatif, on indique que le processeur de signal numérique dédié 3 peut être constitué par un processeur de signal MOTOROLA, référencé DSP56001, cadencé à la fréquence d'horloge de 33 MHz. Le micro-ordinateur hôte 2 peut avantageusement être constitué par un micro-ordinateur de type PC-PENTIUM, cadencé à une fréquence d'horloge de 90 MHz et doté d'un système d'exploitation tel qu'un système d'exploitation multitâche MS-WINDOWS. Le processeur de signal numérique dédié 3 est un processeur à 24 bits en virgule fixe, ce type de processeur permettant d'effectuer les calculs précédemment cités, pour la mise en oeuvre des étapes a) à e) du procédé objet de la présente invention de manière optimale. Ce processeur de signal 3 est en fait constitué par une unité centrale de traitement 30, notée DSP-CPU, à laquelle est associé un espace de mémoire de programme noté P, référencé 31, et deux espaces de mémoire de données, notés X et Y, de capacité de 512 mots chacun et référencés 32. Les espaces de mémoire P, X et Y sont accessibles chacun par trois BUS indépendants de 24 bits, l'adressage étant effectué par trois BUS de 16 bits permettant d'adresser séparément chaque espace mémoire qui peut donc être étendu à 64 k-mots.

Pour des raisons de rapidité, les programmes et sous-programmes de calcul sont exécutés dans les 512 mots de la mémoire interne P, ces programmes ou sous-programmes étant préalablement chargés dans les 8 k-mots de la mémoire P externe. Sur instruction du micro-ordinateur hôte 2, un programme ou un sous-programme peut être transféré de la mémoire externe à la mémoire interne pour y être exécuté. Les données à traiter, données relatives au signal de parole, ainsi que les tables de calcul nécessaires au calcul des coefficients cepstraux par exemple et les résultats intermédiaires sont mémorisés dans les espaces X et Y 32 étendus à  $2 \times 64$  k-mots.

Le micro-ordinateur hôte 2 dispose de programmes et sous-programmes permettant d'assurer un dialogue avec le processeur de signal numérique dédié 3 en vue d'effectuer le chargement de code et de données, la lecture de données, le transfert de code, l'exécution d'un ou plusieurs programmes ainsi que l'initialisation du module de conversion analogique-numérique 1 pour assurer l'acquisition et la reproduction du signal de parole.

L'ensemble constitué par le circuit de conversion analogique-numérique 1 et le processeur de signal numérique dédié 3 est implanté sur une carte additionnelle, telle qu'une carte commercialisée par la Société DIGIMETRIE, sous la référence PC-DSP56k/AD/MEM. Cette carte, outre le processeur de signal numérique DSP56001, comporte un convertisseur analogique-numérique / numérique-analogique commercialisé par la Société TEXAS INSTRUMENTS, sous la référence TCL32040CN permettant d'assurer l'acquisition des signaux de parole, ce convertisseur portant la référence 10 sur la figure 4.

Compte tenu d'une telle architecture, on indique que le temps de calcul de la fréquence fondamentale, pour 100 trames de parole de durée 32 ms, est d'environ 2,7 secondes, soit 27 ms par trame de 32 ms. Le calcul de logarithme du spectre d'énergie, soit le deuxième ensemble de valeurs  $\{X_2(k)\}$  sur  $M_2 = 2048$  points nécessite un temps de calcul de 14 ms. Compte tenu de la complexité des calculs effectués, les temps de calcul apparaissent remarquablement courts. On indique d'ailleurs qu'il est possible d'effectuer ces calculs en temps réel, puisque le temps de calcul effectif de 27 ms par trame est inférieur à la durée de chaque trame.

Dans le but d'améliorer les performances du système et en vue d'assurer un traitement en parallèle des étapes b), c) et d), e) du procédé, objet de la présente invention, tel que représenté par exemple en figure 1b, on indique que le micro-ordinateur hôte peut être configuré à partir du système d'exploitation MS-Windows de façon à fonctionner en mode multitâche, ce qui permet d'effectuer la conduite des opérations en parallèle dans le mode multitâche précité. Un tel mode opératoire n'est pas indispensable mais il permet d'optimiser l'utilisation des ressources de calcul.

On comprend enfin qu'en ce qui concerne les opérations de post-traitement, celles-ci peuvent être réalisées au niveau du micro-ordinateur hôte 2 dans la mesure où le processus de post-traitement, tel que décrit précédemment dans la description selon l'algorithme défini précédemment, peut être réalisé grâce à un programme écrit au moyen d'un langage tel que le langage C par exemple, permettant une rapidité suffisante de traitement pour assurer la correction des valeurs et fréquences fondamentales successives extraites  $F_0(i)$ .

Compte tenu de l'architecture précitée, on indique en particulier que le procédé et le dispositif, objets de la présente invention, peuvent avantageusement être utilisés de façon à réaliser un système d'authentification du locuteur avec une grande probabilité de réussite. En effet, on comprend en particulier que la construction de l'histogramme des fréquences peut être réalisée, soit de manière générale pour un nombre déterminé de locuteurs, soit, au contraire,

pour un locuteur particulier pour lequel l'histogramme des fréquences est effectivement représentatif de ce locuteur. Il en est bien entendu de même en ce qui concerne la valeur des bornes inférieure et supérieure, ainsi que, le cas échéant, des paramètres statistiques tels que les valeurs  $F_{0\max}$  et  $F_{0\min}$  et valeur moyenne de la fréquence fondamentale du signal de parole de ce locuteur. Bien entendu, l'histogramme des fréquences précité, pour un locuteur déterminé, peut alors être réactualisé dans le temps en fonction de l'évolution de la voix du locuteur.

## Revendications

1. Procédé d'extraction de la fréquence fondamentale d'un signal de parole, succession d'échantillons numériques, caractérisé en ce que ce procédé comprend au moins les étapes consistant à :

a) soumettre ledit signal de parole à un processus de préaccentuation, pour engendrer un signal de parole préaccentué ;

b) calculer, à partir du signal de parole préaccentué, pour chaque trame courante d'une succession de trames correspondant chacune en durée à un nombre déterminé N d'échantillons, deux trames consécutives présentant chacune un recouvrement de durée en nombre d'échantillons consécutifs communs au plus égal à 50/100 du nombre N d'échantillons, un premier ensemble de valeurs  $X_1(k)$  du logarithme du spectre d'énergie par transformée de Fourier sur un nombre  $M_1$  de points ;

c) calculer, à partir dudit ensemble de valeurs, un nombre p déterminé de premiers coefficients cepstraux  $C(m)$ , par application d'une transformée en cosinus discrète auxdites valeurs  $X_1(k)$  sur un nombre de ces valeurs au moins égal à la moitié du nombre N d'échantillons constitutifs de ladite trame courante, ladite transformée vérifiant la relation :

$$C(m) = [1 + (-1)^m] X(0) + 2 \sum_{k=1}^{N/2-1} X_1(k) \cdot \cos\left(\frac{2\pi mk}{N}\right)$$

avec  $m = [1, 2, \dots, p]$ ,  $C(m)$  désignant lesdits coefficients cepstraux ;

d) soumettre ledit signal de parole préaccentué à un filtrage de type passe-bas et à un sous-échantillonnage, pour engendrer un signal de parole filtré sous-échantillonné ;

e) calculer, par compression spectrale, à partir dudit signal de parole filtré sous-échantillonné et à partir desdits coefficients cepstraux pour chaque trame courante d'une succession de trames de même recouvrement de durée, la fréquence fondamentale maximum de rang k d'une fonction  $P(k)$  représentative de la différence entre un deuxième ensemble des valeurs  $X_2(k)$  du logarithme du spectre d'énergie et l'ensemble des valeurs  $H(k)$  du spectre de fréquences lissé, ladite fonction vérifiant la relation :

$$P(k) = \sum_{l=1}^L X_2(k) - H(k)$$

avec  $L = M_2/k$ , k variant entre une première et une deuxième valeur représentatives d'une bande de fréquences basses comprises entre 70 et 450 Hz, ladite fonction  $P(k)$  présentant un maximum pour  $k=F_0$ , valeur extraite de la fréquence fondamentale du signal de parole.

2. Procédé selon la revendication 1, caractérisé en ce que ladite étape de calcul par compression spectrale consiste successivement à :

- calculer sur ledit signal de parole filtré sous-échantillonné, pour chaque trame courante, ledit deuxième ensemble de valeurs  $X_2(k)$  du logarithme du spectre d'énergie par transformée de Fourier sur un nombre  $M_2$  de points sur une bande de fréquences comprises entre 0 et 2 kHz ;

- calculer l'enveloppe spectrale  $H(k)$ , spectre de fréquences lissé de ladite trame courante sur ladite bande de fréquences comprises entre 0 et 2 kHz sur un même nombre  $M_2$  de points, par application sur lesdits p-1 premiers coefficients cepstraux d'une transformée en cosinus vérifiant la relation :

$$H(k) = C(0) + 2 \sum_{m=1}^{P-1} C(m) \cdot \cos\left(2\pi \frac{mk}{Q}\right)$$

5

avec  $k = [0, 1, 2, \dots, M_2]$  et  $M_2 = Q/4$  ;

- calculer la différence  $D(k) = X_2(k) - H(k)$  ;
- calculer le produit harmonique représentatif de la fonction  $P(k)$  par compression spectrale de ladite différence  $D(k)$  sur ladite bande de fréquences basses comprises entre 70 et 450 Hz ;
- déterminer par un processus de tri le maximum de la fonction  $P(k)$  et le rang  $k=F_0$  correspondant, valeur extraite de la fréquence fondamentale.

10

3. Procédé selon la revendication 1 ou 2, caractérisé en ce que, suite à l'étape consistant à soumettre le signal de parole préaccentué respectivement filtré sous-échantillonné à un formatage en trames, celui-ci consiste en outre à discriminer, parmi l'ensemble des trames, les trames voisées et les trames non voisées, le processus d'extraction de la fréquence fondamentale étant conduit sur les trames voisées.

15

4. Procédé selon la revendication 3, caractérisé en ce que l'étape consistant à discriminer les trames voisées et les trames non voisées consiste :

20

- à subdiviser chaque trame en un nombre ST de segments de trames contigus successifs ;
- à établir pour chacun desdits segments de trame un critère de discrimination de voisement, à partir d'un indice de voisement, compris entre 0 et 1 représentatif du niveau d'énergie basse fréquence du segment de trame considéré selon une loi sensiblement linéaire ;
- à classifier chaque trame comme trame non voisée par comparaison d'une combinaison linéaire des indices de voisement de chaque segment à une valeur de seuil déterminée.

25

5. Procédé selon l'une des revendications 1 à 4, caractérisé en ce que suite à l'étape de détermination du maximum de rang  $k$  de ladite fonction  $P(k)$ ,  $k=F_0$  représentant la valeur de la fréquence fondamentale du signal de parole, et en vue d'éliminer toute valeur de fréquence fondamentale aberrante et supprimer les risques d'erreur dues à la présence d'erreurs de transitions engendrées par l'existence, dans une même trame, de segments voisés, non voisés ou de silences ainsi que par l'existence de trames voisées ou mixtes de faible niveau d'énergie, ledit procédé consiste en outre à effectuer un post-traitement de ladite valeur extraite de fréquence fondamentale dudit signal de parole, cette étape de post-traitement consistant à :

30

- établir un histogramme des fréquences fondamentales, afin de déterminer la plage de valeurs de fréquences les plus probables et les bornes de valeurs inférieure et supérieure de ces valeurs ;
- soumettre chaque valeur extraite de fréquence fondamentale à un critère de tri par rapport auxdites bornes de valeurs inférieure et supérieure, pour obtenir des valeurs triées représentatives de l'évolution des valeurs extraites de fréquence fondamentale ;
- soumettre ces valeurs triées à un filtrage non linéaire pour supprimer les valeurs aberrantes.

35

6. Procédé selon la revendication 1, caractérisé en ce que les étapes a) à e) sont réalisées séquentiellement.

45

7. Procédé selon la revendication 6, caractérisé en ce que les étapes b) et c), respectivement d) et e) sont réalisées sous système d'exploitation multi-tâches, ce qui permet d'effectuer l'extraction de la fréquence fondamentale en temps réel.

50

8. Dispositif d'extraction de la fréquence fondamentale d'un signal de parole, conformément au procédé selon l'une des revendications 1 à 7, caractérisé en ce que ce dispositif comprend :

- des moyens d'échantillonnage et de conversion analogique-numérique d'un signal de parole en une suite d'échantillons numériques ;
- un micro-ordinateur hôte permettant la conduite de la succession des étapes a) à e) du procédé et la gestion et la commande d'organes périphériques, notamment lesdits moyens d'échantillonnage et de conversion analogique-numérique ;
- un processeur de signal numérique interconnecté par une liaison par BUS audit micro-ordinateur hôte et permettant d'effectuer les opérations de calcul du premier ensemble de valeurs  $X_1(k)$  du logarithme du spectre

55

d'énergie par transformée de Fourier sur un nombre  $M_1$  de points, des  $p$  premiers coefficients cepstraux, de filtrage passe-bas et de sous-échantillonnage, du deuxième ensemble de valeurs  $X_2(k)$  du logarithme du spectre d'énergie, de l'ensemble des valeurs  $H(k)$  du spectre de fréquences lissé, de la fonction

5

$$P(k) = \sum_{l=1}^L X_2(k) - H(k) ,$$

10

de l'extraction du maximum  $P(k)$  pour  $k = F_0$  valeur extraite de la fréquence fondamentale du signal de parole.

9. Utilisation du procédé et du dispositif d'extraction de la fréquence fondamentale d'un signal de parole selon l'une des revendications 1 à 8, pour l'authentification d'un ou plusieurs locuteurs.

15

20

25

30

35

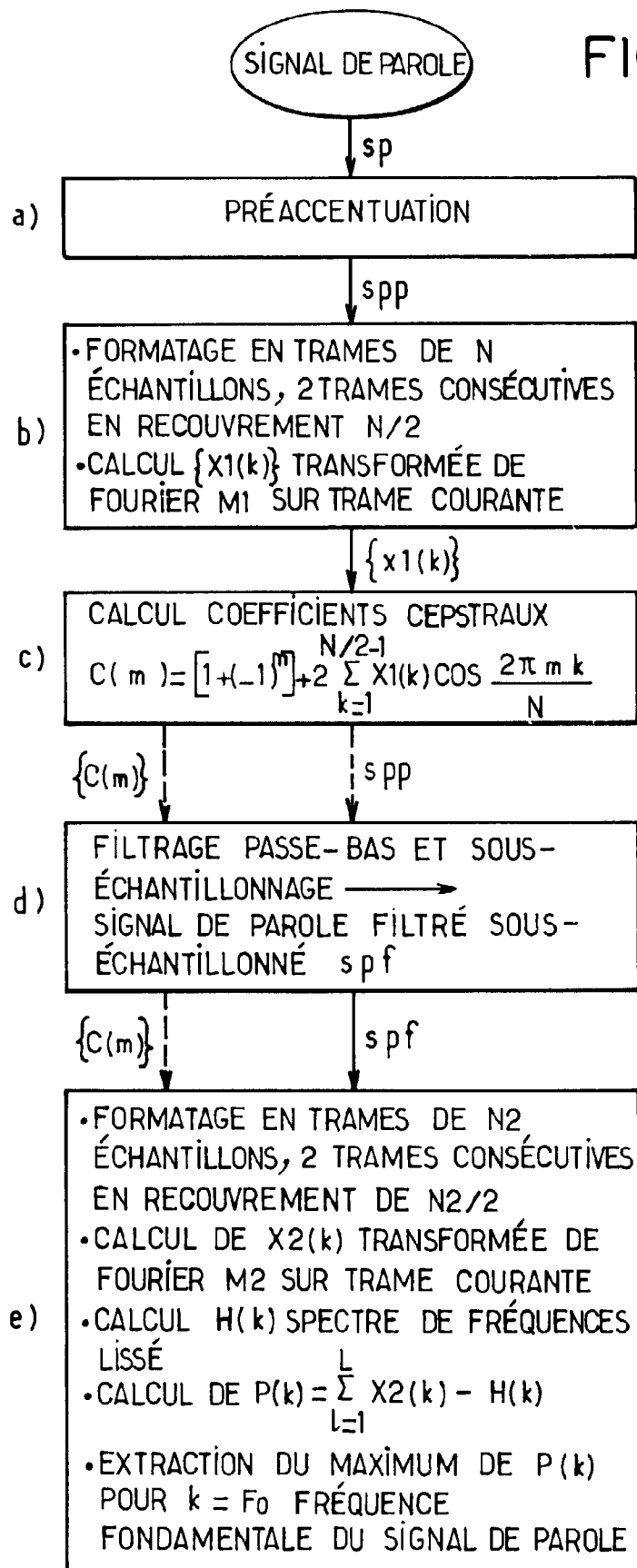
40

45

50

55

FIG.1a.



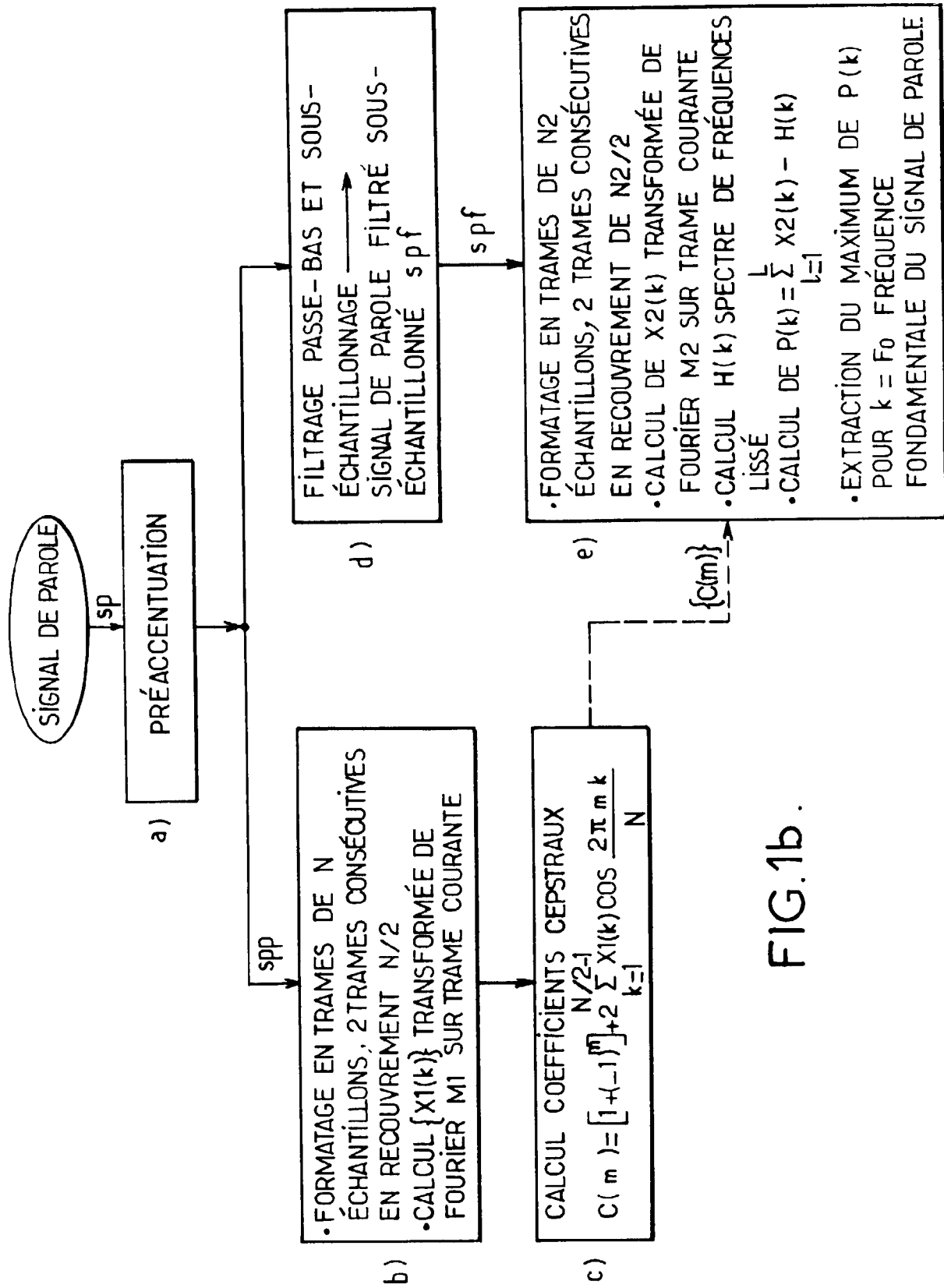
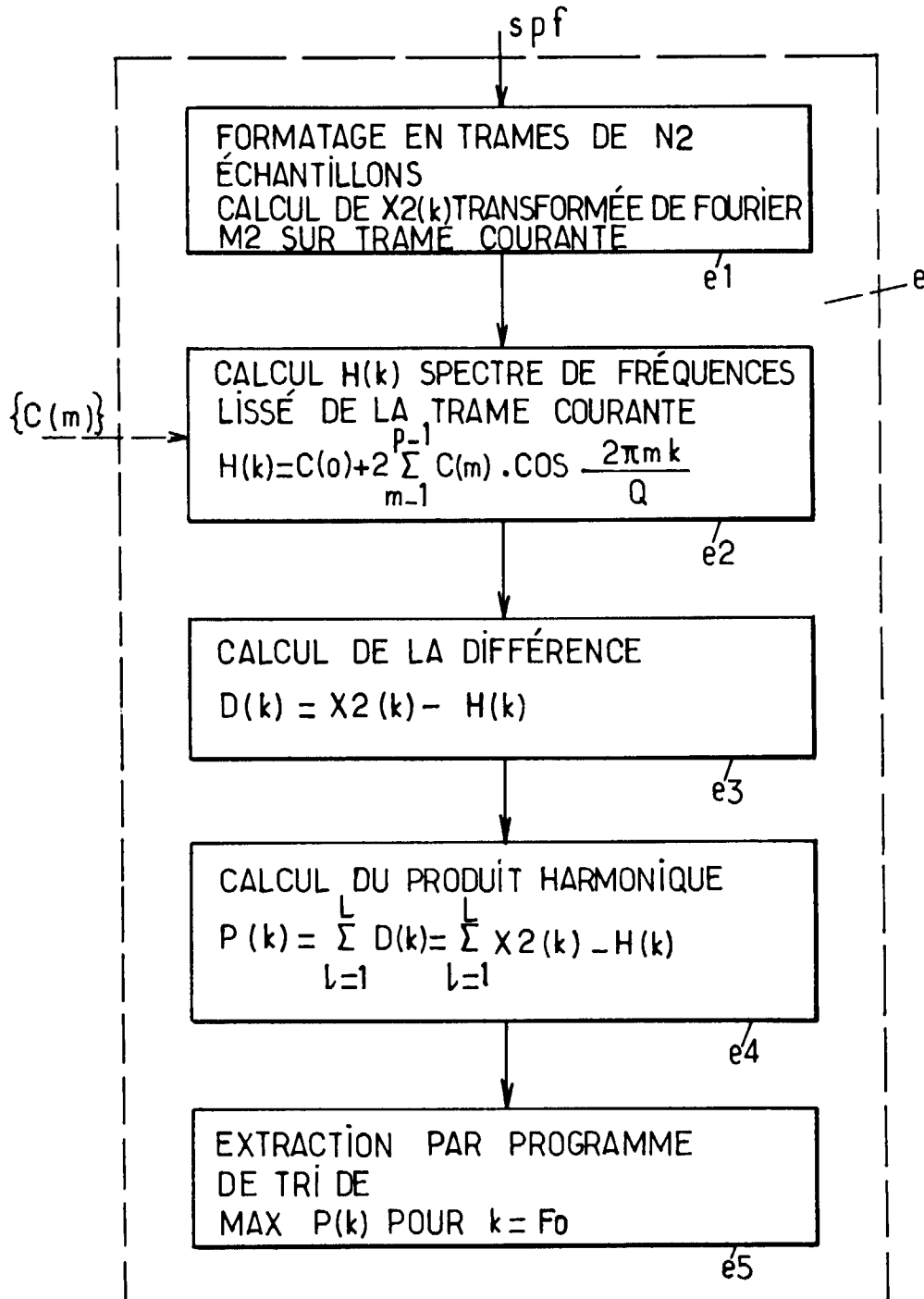


FIG. 2a.



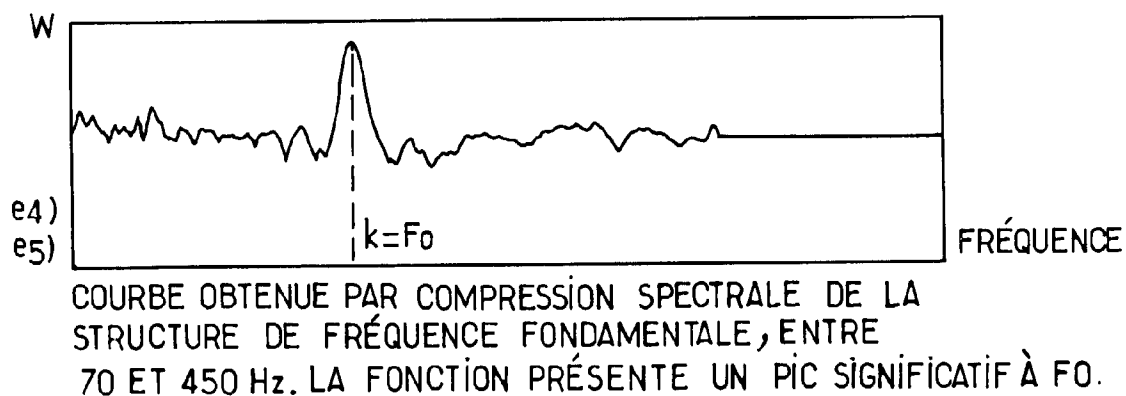
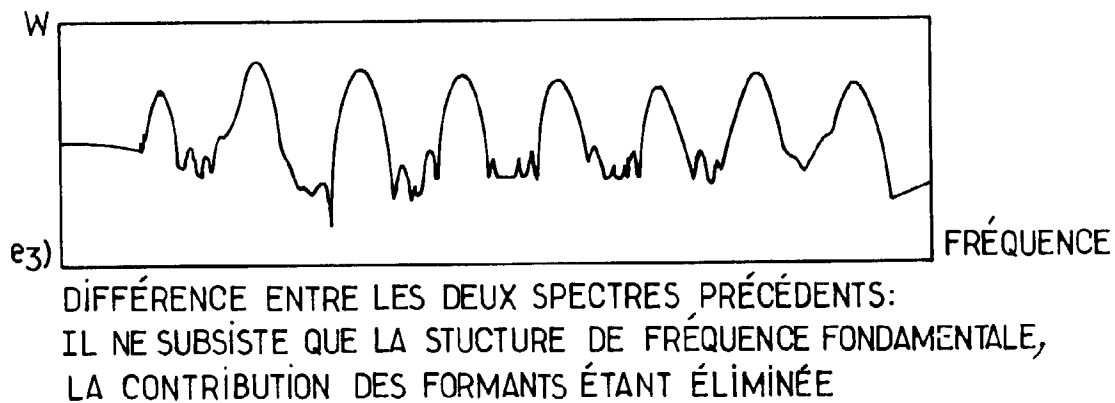
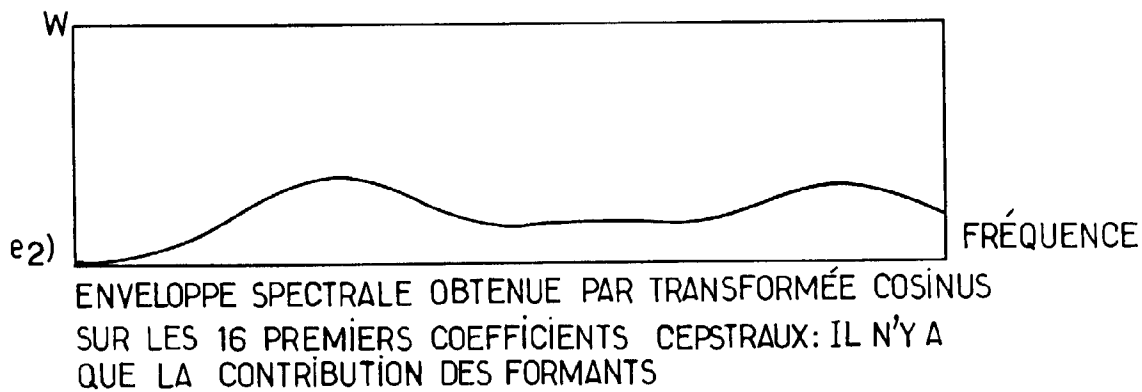
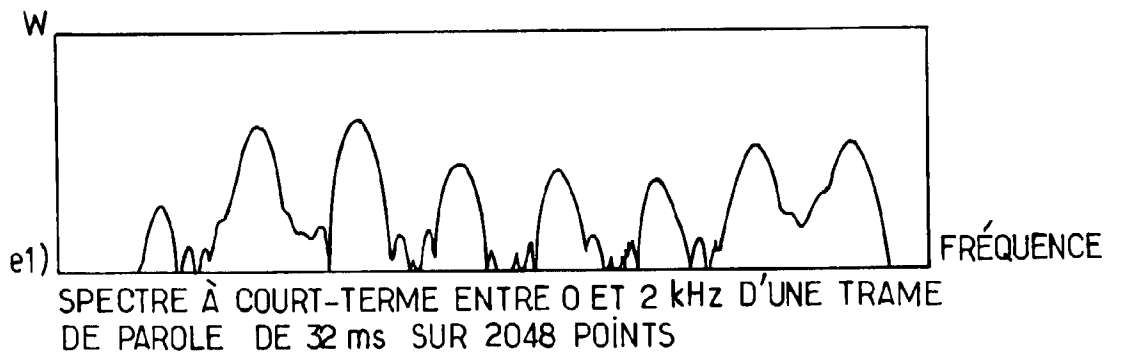


FIG. 2b.



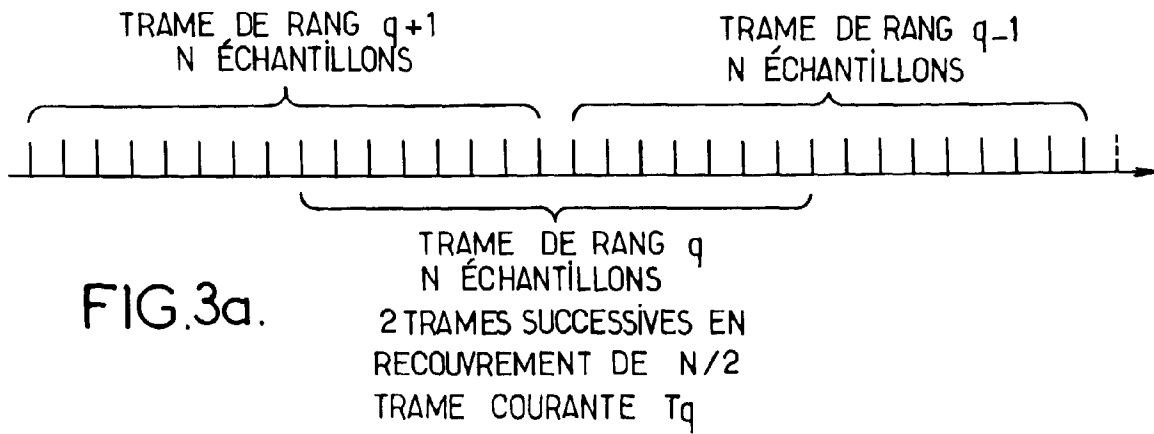


FIG. 3a.

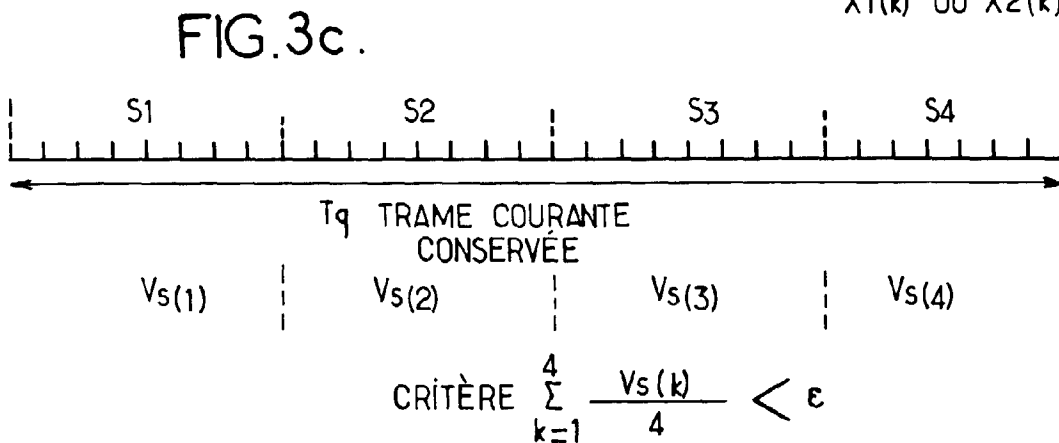
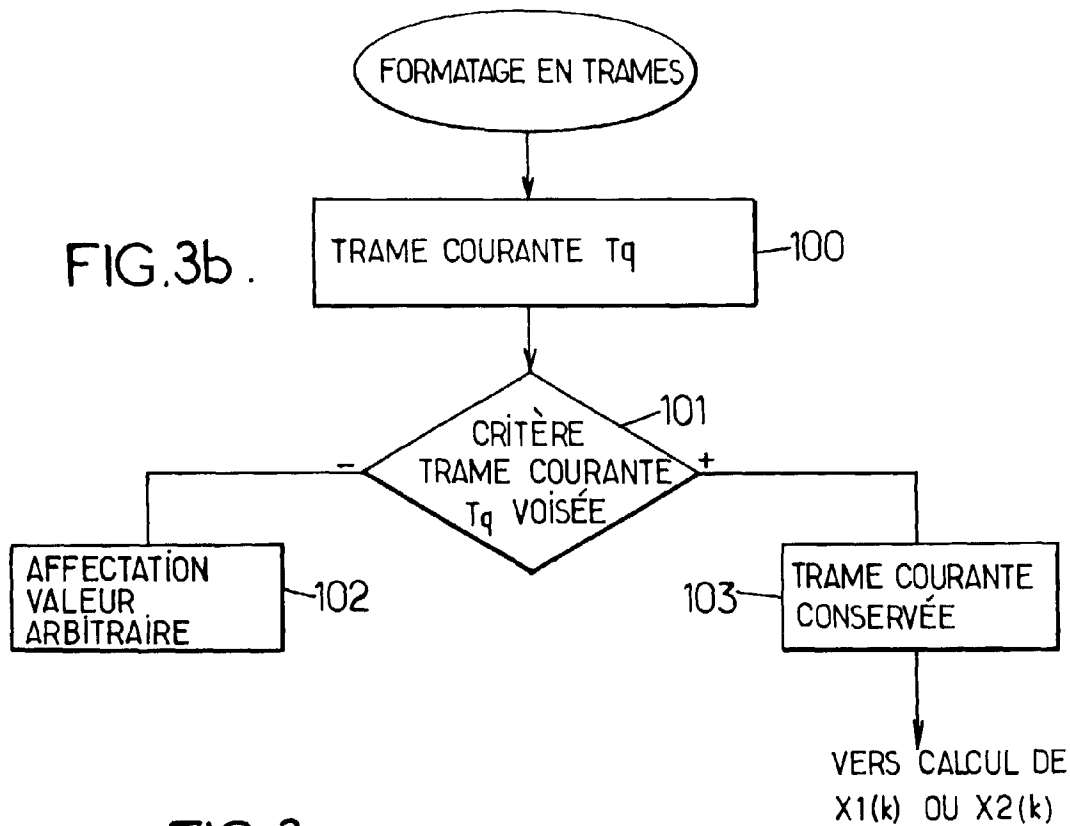


FIG. 3c.

FIG.3d.

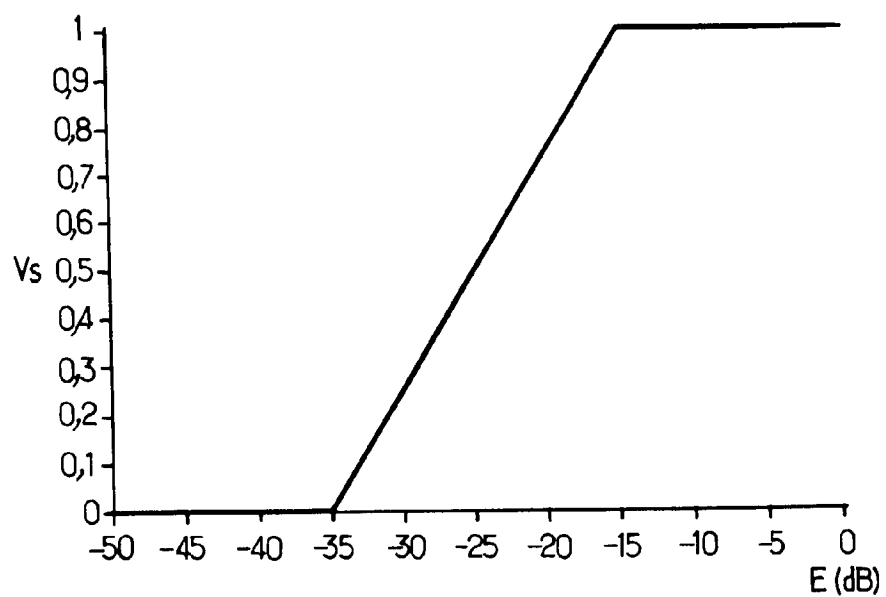
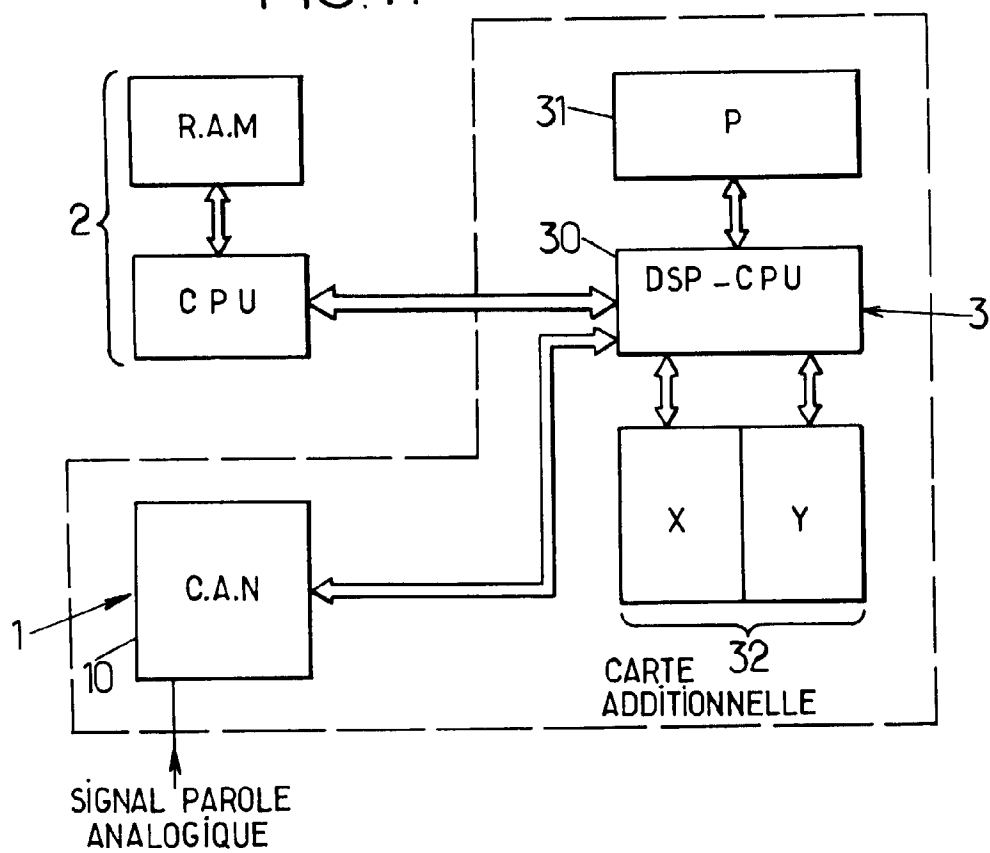


FIG.4.





Office européen  
des brevets

# RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande  
EP 97 40 1752

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (Int.Cl.6)
A	KEIKICHI HIROSE ET AL: "A SCHEME FOR PITCH EXTRACTION OF SPEECH USING AUTOCORRELATION FUNCTION WITH FRAME LENGTH PROPORTIONAL TO THE TIME LAG" SPEECH PROCESSING 1, SAN FRANCISCO, MAR. 23 - 26, 1992, vol. 1, 23 mars 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 149-152, XP000341105 * page 1-151, colonne de gauche, ligne 12 - ligne 14 * * figure 5 *	1	G10L3/00
A	HERMES D J: "Measurement of pitch by subharmonic summation" JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, JAN. 1988, USA, vol. 83, no. 1, janvier 1988, ISSN 0001-4966, pages 257-264, XP002027082 * abrégé; figure 1 * * page 258, colonne de droite, ligne 20 - ligne 24 *	1	DOMAINES TECHNIQUES RECHERCHES (Int.Cl.6) G10L
A	W. HESS: "Pitch Determination of Speech Signals: Algorithm and Methods" 1983, SPRINGER VERLAG, BERLING HEIDELBERG NY TOKYO XP002027083 * page 410 - page 415 *	1	
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche LA HAYE		Date d'achèvement de la recherche 1 octobre 1997	Examineur Van Doremalen, J
CATEGORIE DES DOCUMENTS CITES X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

EPO FORM 1503 03 82 (POAC02)