



(12) **DEMANDE DE BREVET EUROPEEN**

(21) Numéro de dépôt : **92403087.7**

(51) Int. Cl.⁵ : **G10L 3/00**

(22) Date de dépôt : **17.11.92**

(30) Priorité : **22.11.91 FR 9114411**

(43) Date de publication de la demande :
26.05.93 Bulletin 93/21

(84) Etats contractants désignés :
DE ES GB IT

(71) Demandeur : **THOMSON-CSF**
51, Esplanade du Général de Gaulle
F-92800 Puteaux (FR)

(72) Inventeur : **Laurent, Pierre-André,**
THOMSON-CSF
SCPI, 50, rue J.P. Timbaud
F-92402 Courbevoie Cedex (FR)

(74) Mandataire : **Lincot, Georges et al**
THOMSON-CSF, SCPI, B.P. 329, 50, rue
Jean-Pierre Timbaud
F-92402 Courbevoie Cédex (FR)

(54) **Procédé et dispositif de décision de voisement pour vocodeur à très faible débit.**

(57) Le procédé consiste à considérer l'évolution du signal de parole sur un nombre K déterminé de trames successives en affectant à chaque trame courante n un score à chacun des états possibles d'évolution du signal de parole en fonction des taux de corrélation d'échantillons répartis dans chaque trame (1, 2, 3), à déterminer (5) dans chaque trame courante l'état d'évolution du signal de parole qui a le score maximal pour remonter à partir de cet état à l'état le plus probable de la trame n-K en passant successivement par les états de score maximal de chaque trame précédente.

Application : Vocodeurs à faibles débits.

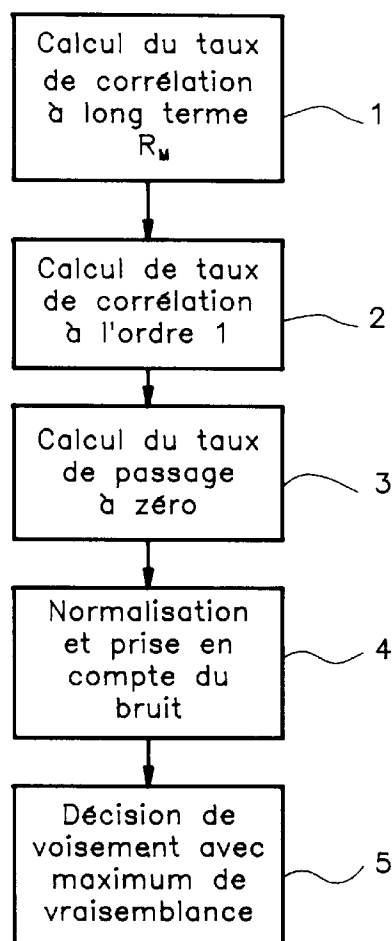


FIG 1

La présente invention concerne un procédé et un dispositif de décision de voisement pour vocodeur à très faible débit.

Dans les vocodeurs à très faible débit de 1200 bits/seconde et moins, le signal de parole est segmenté en trames de durées constantes de 10 à 30 millisecondes de façon à déterminer à l'intérieur de celles-ci la périodicité encore désignée par "Pitch" dans le langage anglo-saxon, du signal de parole. Ces trames sont positionnées arbitrairement dans le signal de parole et une seule valeur de voisement est fournie pour chaque trame. Comme les trames sont regroupées en blocs de 2, 6 ou 8 trames par exemple et pour que le débit alloué au voisement soit réduit, toutes les combinaisons de voisement possibles ne sont généralement pas autorisées. Il est généralement considéré en effet comme improbable qu'il puisse exister une ou deux trames voisées isolées au milieu d'un paquet de trames non voisées, ou l'inverse. Ceci conduit dans les réalisations habituelles à mettre en oeuvre un processus de décision de voisement qui fonctionne en deux passes, une première passe suivant laquelle une première décision de voisement locale est prise au niveau de chaque trame, celle-ci résultant de l'examen des valeurs de différents paramètres du signal de parole et une deuxième passe pour corriger les décisions prises lors de la première passe et supprimer dans les parties voisées les petits paquets de trames non voisées et vice versa. Naturellement le fonctionnement des dispositifs correspondants repose sur une grande part d'heuristique (basée sur des expérimentations, des apprentissages, etc...) qui est la plupart du temps satisfaisant dans des conditions peu sévères mais qui se dégrade très rapidement dès lors que le signal de parole est perturbé par exemple par du bruit.

Le but de l'invention est de pallier les inconvénients précités.

A cet effet, l'invention a pour objet un procédé de décision de voisement pour vocodeur à très faible débit suivant lequel le signal de parole est échantillonné et segmenté en trames de durée constante caractérisé en ce qu'il consiste à considérer l'évolution du signal de parole sur un nombre K déterminé de trames successives en affectant à chaque trame courante n un score à chacun des états possibles d'évolution du signal de parole en fonction des taux de corrélation d'échantillons répartis dans chaque trame, à déterminer dans chaque trame courante l'état d'évolution du signal de parole qui a le score maximal pour remonter à partir de cet état à l'état le plus probable de la trame n-K en passant successivement par les états de score maximal de chaque trame précédente.

D'autres caractéristiques et avantages de l'invention apparaîtront ci-après à l'aide de la description qui suit faite en regard des dessins annexes qui représentent :

- La figure 1 les différentes étapes du procédé selon l'invention mises sous la forme d'un organigramme.
- La figure 2 un diagramme d'états pour montrer le mécanisme des transitions mises en oeuvre dans le procédé selon l'invention permettant d'effectuer des décisions de voisement avec un maximum de vraisemblance.
- La figure 3 un exemple de fonctionnement du mécanisme d'état représenté à la figure 2 sur cinq trames successives.
- La figure 4 un graphe montrant comment une décision de voisement peut être prise en compte par un cheminement des états sur cinq trames successives.
- La figure 5 un mode de réalisation d'un dispositif pour la mise en oeuvre du procédé selon l'invention.

Le procédé selon l'invention qui est représenté schématiquement par les blocs 1 à 5 sur la figure 1 effectue une décision de voisement sur des plages de signal de parole de durée multiple du "pitch". Cette durée est calculée entre deux valeurs extrêmes prédéterminées, qui sont une valeur minimale pour prendre en compte suffisamment de signal et une valeur maximale pour à la fois limiter la charge de calcul et tenir compte de la vitesse de variation naturelle des caractéristiques de la parole. Le traitement commence par le calcul de trois paramètres qui sont un taux de corrélation à long terme noté R_M , un taux de corrélation à l'ordre 1 noté R_1 et un taux de passage à 0 noté T_{ppz} . Ces calculs sont figurés sur les blocs 1, 2 et 3 de la figure 1.

Le taux de corrélation à long terme R_M est calculé suivant l'expression

$$R_M = \text{MAX}_{m=M-\Delta \dots M+\Delta} \frac{\left(\sum_{n=1}^N S_n S_{n-m} \right)}{K_1 + \sum_{n=1}^N S_n^2 \sum_{n=1}^N S_{n-m}^2} \quad (1)$$

dans laquelle M représente une valeur de "pitch" en nombre d'échantillons, S_n et S_{n-m} sont des amplitudes d'échantillons de signal, N désigne le nombre d'échantillons analysés, K est une constante et Δ représente une fraction de M. Ce calcul permet de fournir la vraie valeur du "pitch" lorsque le taux de corrélation qui est

obtenu présente une valeur maximale. Pour un son parfaitement voisé, R_M est égal à 1, tandis qu'il est pratiquement nul pour un son aléatoire (non voisé). La constante K_1 évite les divisions par zéro et fournit une valeur très faible de R_M dans les silences de parole où la puissance est très faible.

Le taux de corrélation à l'ordre 1, R_1 est calculé suivant l'expression

$$R_1 = \frac{\sum_{n=1}^N S_n S_{n-1}}{\sqrt{K_2 + \sum_{n=1}^N S_n^2 \sum_{n=1}^N S_{n-1}^2}} \quad (2)$$

dans laquelle S_n et S_{n-1} représentent comme précédemment les amplitudes de N échantillons et K_2 est une constante.

Pour un son voisé, R_1 est très voisin de 1 et il se rapproche de -1 pour un son non voisé. La constante K_2 permet de donner à R_1 une valeur pratiquement nulle durant les silences.

Le taux de passage par zéro, T_{ppz} fournit le rapport entre le nombre de changements de signe de chaque échantillon S_n , et $N/2$ ce taux évolue entre une valeur presque nulle pour un son voisé et une valeur proche de 1 pour un son non voisé.

Le traitement continue, de la façon figurée par le bloc 4 sur la figure 1 par une normalisation des paramètres R_N , R_1 et T_{ppz} déterminés précédemment. Cette normalisation a pour effet de fixer les valeurs des paramètres R_N , R_1 et T_{ppz} entre les deux valeurs 0 et 1, la valeur 0 correspondant au son parfaitement non voisé et la valeur 1 correspondant au son parfaitement voisé. Les valeurs normalisées obtenues sont ensuite pondérées par des coefficients pondérateurs α_1 , α_2 et α_3 pour former une quantité Z définie par la relation

$$Z = \alpha_1 R_M + \alpha_2 \left(\frac{R_1 + 1}{2} \right) + \alpha_3 (-T_{ppz}) \quad (3)$$

En prenant soin de choisir les valeurs des coefficients pondérateurs telle que la relation $\alpha_1 + \alpha_2 + \alpha_3 = 1$ soit vérifiée, Z prend alors la valeur 1 pour les sons parfaitement voisés et la valeur 0 pour les sons parfaitement non voisés. Les paramètres α_1 , α_2 et α_3 peuvent par exemple être déterminés comme suit $\alpha_1 = 0,45$, $\alpha_2 = 0,35$ et $\alpha_3 = 0,2$ mais leurs valeurs optimales ne peuvent être réellement trouvées que par un ajustement expérimental des valeurs des coefficients α_1 en fonction notamment des filtres sur les échantillons de signal S_n .

L'adaptation à un niveau variable de bruit ambiant a lieu en constatant qu'ajouter du bruit aux signaux voisés a pour conséquence de diminuer les taux de corrélation R_M et R_1 et augmenter le taux de passage à zéro T_{ppz} , c'est-à-dire, de diminuer globalement la valeur maximale Z_{MAX} possible pour Z à un instant donné. Pour cela une mise à jour d'une estimée de Z_{MAX} est réalisée et la quantité relative

$$"z" = Z/Z_{MAX} \quad (4)$$

sert alors d'indicateur brut de voisement compte tenu du bruit.

La quantité Z_{MAX} peut être estimée à chaque nouvelle valeur de Z en mettant à jour Z_{MAX} par chaque nouvelle valeur de Z , si Z est supérieure à la valeur Z_{MAX} ou si la nouvelle valeur de Z est inférieure à Z_{MAX} , en calculant la quantité Z_{MAX} suivant la relation

$$Z_{MAX} = (1 - \varepsilon) Z_{MAX} \quad (5)$$

dont la décroissance est définie par une constante de temps qui est fonction de ε pour suivre l'évolution du bruit, cette constante de temps étant d'autant plus élevée que ε est petit.

La décision de voisement qui est figurée par le bloc 5 sur la figure 1 a lieu sur un critère de décision par maximum de vraisemblance, en considérant l'évolution possible du signal de parole sur plusieurs trames successives, en affectant dans chaque nouvelle trame un score à chacun des états possibles d'évolution dépendant de la valeur du nombre z obtenue de la relation (4).

Pour rendre compte d'un état de voisement sur une durée minimale, par exemple de trois trames consécutives, cela conduit à considérer une logique combinatoire à 6 états possibles de voisement, tels que définis dans le tableau de la figure 2, dans lequel "V" signifie "voisé" et "N" signifie "non voisé", et le graphe des transitions de la figure 3 représentant les transitions entre états possibles pour parvenir à la trame courante n à partir des états possibles de la trame précédente $n-1$. Il est à noter sur la figure 3 que chaque état d'une trame n est précédé d'un ou de deux états dans la trame $n-1$ précédente. Ainsi dans la trame n l'état NNV est précédé de l'état NNN, l'état NVV est précédé de l'état NNV, l'état VVV est précédé soit de l'état NVV soit de l'état VVV, l'état NNN est précédé soit de l'état NNN lui-même soit de l'état VNN, l'état VNN est précédé de l'état VVN et

l'état VVN est précédé de l'état VVV.

Naturellement, ce type de mécanisme peut très bien être adapté à des durées différentes d'où dépendront dans chaque cas le nombre d'états à considérer.

A chacun des états d'arrivée possibles à la trame n est affecté un score. Pour chacun d'eux, le score retenu est le plus grand possible parmi ceux obtenus en additionnant ou en soustrayant, suivant les signes indiqués sur les flèches représentant les transitions entre états sur la figure 3, la valeur de z calculée par la relation (4) précédente à laquelle il est éventuellement retranchée une valeur de seuil prédéterminée Z_0 comprise entre 0 et 1. Pour tenir compte du fait que certains états de départ deviennent interdits du fait que leur valeur de score devient infinie, les états d'arrivée correspondant sont également interdits.

Par exemple, le score SA_2 de l'état d'arrivée 2 (VVV) sur la figure 3 est déterminé par l'une des relations :

$$- SA_2 = \text{MAX} ((SD_1 + Z - Z_0); SD_2 + (Z - Z_0))$$

si les états de départ SD_1 et SD_2 sont tous deux autorisés.

$$- SA_2 = SD_1 + (Z - Z_0) \text{ si l'état de départ 2 est interdit.}$$

$$- SA_2 = SD_2 + (Z - Z_0) \text{ si l'état de départ 1 est interdit}$$

où $SA_2 = -\text{Infini}$ si les états de départ 1 et 2 sont tous les deux interdits. Dans ce dernier cas l'état 2 est interdit comme état de départ à la trame suivante.

Naturellement à chaque nouvelle trame le procédé mémorise les scores d'arrivée qui servent ensuite de scores de départ pour la trame suivante et le tableau des états précédents est également mémorisé.

La décision finale a lieu sur toutes les trames avec un certain retard c'est-à-dire que la décision de voisement à l'instant d'arrivée de la trame n sera celle de la trame n-K de la façon représentée par l'organigramme de la figure 4. Le procédé consiste à rechercher parmi les six états d'arrivée celui dont le score est le plus élevé, puis il remonte K moins une fois de suite de l'état auquel il est parvenu à la trame n-i ($i = 0, 1, \dots, K-1$) à l'état précédent qui est mémorisé pour cette trame. Dans l'exemple de la figure 4 qui représenté une situation avec un retard de cinq trames, à la trame n l'état 1 est supposé avoir le meilleur score. Du fait que cet état a, suivant le diagramme des transitions de la figure 3, comme état précédent l'état 0 à la trame n-1, le procédé remonte à celui-ci et continue de façon similaire jusqu'à la trame n-5 suivant les traits gras continus de la figure 4. L'état auquel le procédé remonte ainsi est l'état VVN qui correspond à la trame n-5 non voisée de l'état 5. Cependant le procédé ne vaut que si la contrainte de trois trames successives de même état de voisement est respectée ce qui n'est pas toujours le cas si le procédé s'arrête au stade précédent. En effet étant donné que la trame n-5 correspond à l'état 5, la trame n-4 ne peut être dans ce cas que dans l'état 4, seul état auquel il est possible de parvenir d'après la figure 3 en partant de l'état 5, la remontée dans les états lors de la prochaine trame devant aboutir obligatoirement à l'état 5 dans la trame n-4. La suite du traitement consiste alors pour tous les états possibles à la trame n à éliminer les états (ici les états 0, 3, 4 et 5) à partir desquels il n'est pas possible de remonter à l'état 5, dans la trame n-4 en positionnant leurs scores à l'infini. Dans l'exemple de la figure 4 l'état 1 n'est bien entendu par éliminé, car c'est à partir de lui qu'a été amorcée la remontée dans les états et l'état 2 est conserve car le procédé peut remonter à partir de lui à l'état 5 dans la trame n-4 (traits gras tiretés).

Si dans un autre exemple, la remontée initiale aurait conduit à l'état 3 (trames n-5), sachant que l'état 3 ne peut être suivi que des états 0 ou 3 (trame n-4), tous les états dans la trame n qui ne remontent pas aux états 0 ou 3 dans la trame n-4 auraient été éliminés. Ce procédé d'élimination d'états a entre autres avantages celui de fournir en sortie un train de valeurs de voisement satisfaisant systématiquement la contrainte (absence d'îlot de voisement/non voisement isolés). Si côté réception, cette contrainte n'est pas vérifiée, cela signifie qu'il y a eu une erreur de transmission ce qui permet de prendre des mesures curatives par exemple, en positionnant à "voisé" l'indicateur de voisement dans la zone incriminée.

Un dispositif pour la mise en oeuvre du procédé selon l'invention est représenté à la figure 5. Ce dispositif comporte représentés respectivement à l'intérieur de lignes fermées en pointillés un dispositif 6 de calcul du paramètre Z, un dispositif 7 de normalisation, un dispositif 8 de décision finale. Le dispositif de calcul du paramètre Z comporte un dispositif 9 de calcul d'autocorrélation à long terme de la valeur R_M , un dispositif 10 de corrélation à l'ordre 1 du coefficient R_1 et un dispositif 11 de calcul de taux de passage à zéro T_{ppz} . Les dispositifs de calcul 9, 10 et 11 reçoivent simultanément chacun sur une première entrée les échantillons de signal S_n et effectuent les calculs correspondant aux relations (1), (2) et (3) précédentes à l'intérieur d'une fenêtre d'analyse fourni par un dispositif 12 qui détermine une durée d'analyse pour chaque trame n en fonction de la valeur du "pitch" M. Les circuits multiplieurs 13, 14, 15 sont couplés à la sortie des dispositifs de calcul

9, 10 et 11 pour appliquer les coefficients de pondération α_1 , α_2 et α_3 aux taux de corrélation R_M et $\frac{R_1 + 1}{2}$ et au taux de passage à zéro- T_{ppz} . Les résultats des calculs effectués par les circuits multiplieurs 13, 14, 15 sont appliqués sur les entrées d'un circuit sommateur 16 fournissant sur sa sortie k valeur Z. La valeur Z de la relation (3) est appliquée à l'entrée du bloc de normalisation 7. Le bloc de normalisation 7 comporte un registre

17 de mémorisation du nombre Z maximum. La sortie du registre 17 est couplée à une première entrée d'un circuit comparateur 18 dont la deuxième entrée reçoit la valeur Z fournie par le dispositif de calcul 6. L'entrée du registre 17 est reliée à la sortie d'un circuit multiplexeur 19 commandé par la sortie du circuit comparateur 18 pour appliquer à l'entrée du registre 17 soit la valeur Z telle quelle, soit le résultat du calcul fourni par un circuit multiplieur 20 possédant une première entrée d'opérande reliée à la sortie du registre 17 et une deuxième entrée d'opérande sur laquelle est appliquée le coefficient $1-\varepsilon$ de la relation (5) décrite précédemment. Un circuit diviseur 21 est couplé par une première entrée d'opérande à la sortie du registre 17 et par une deuxième entrée d'opérande à la sortie du dispositif 6 pour calculer la quantité "z" égale à Z/Z_{MAX} . La quantité "z" est appliquée à l'entrée du dispositif de décision finale 8 qui comporte un dispositif de calcul 22 des scores, de remontées des états et d'éliminations des états interdits couplé à une mémoire 23 des scores de départ SDi, à une mémoire 24 des scores d'arrivée SAi et à une mémoire 25 des états précédents. Le dispositif de calcul des scores 22 reçoit d'autre part, sur une deuxième entrée la valeur de seuil Z_0 , sur une troisième entrée le numéro de trames n et fournit sur une sortie les décisions de voisement pour chacune des trames n-K.

Naturellement, la mise en oeuvre du procédé selon l'invention qui vient d'être décrite n'est pas unique et il va de soi que d'autres modes de réalisation mettant en oeuvre des microprocesseurs de traitement du signal notamment sont tout aussi envisageables, leur programmation étant à la portée de l'homme de métier.

Revendications

1. Procédé de décision de voisement pour vocodeur à très faible débit suivant lequel le signal de parole est échantillonné et segmenté en trames de durée constante caractérisé en ce qu'il consiste à considérer l'évolution du signal de parole sur un nombre K déterminé de trames successives en affectant à chaque trame courante n un score à chacun des états possibles d'évolution du signal de parole en fonction des taux de corrélation d'échantillons répartis dans chaque trame (1, 2, 3), à déterminer (5) dans chaque trame courante l'état d'évolution du signal de parole qui a le score maximal pour remonter à partir de cet état à l'état le plus probable de la trame n-K en passant successivement par les états de score maximal de chaque trame précédente.
2. Procédé selon la revendication 1 caractérisé en ce qu'il consiste à calculer (1) un taux de corrélation R_M à long terme pour un intervalle M égal à la valeur du "pitch" du signal de parole.
3. Procédé selon les revendications 1 et 2 caractérisé en ce qu'il consiste à calculer (2) un taux de corrélation à l'ordre 1 pour déterminer la nature voisé ou non du son correspondant aux échantillons corrélés.
4. Procédé selon l'une quelconque des revendications 1 à 3 caractérisé en ce qu'il consiste à déterminer (3) le taux de passages à zéro du signal de parole.
5. Procédé selon l'une quelconque des revendications 2 à 4 caractérisé en ce qu'il consiste (4) à effectuer une sommation pondérée du taux de corrélation à long terme, du taux de corrélation à l'ordre 1 et du taux de passage à zéro pour obtenir un incrément de score "z" et calculer à partir des scores possibles de chaque trame précédant une trame courante, les scores des états possibles d'arrivée.
6. Procédé selon la revendication 5 caractérisé en ce qu'il consiste à donner des valeurs de score infini aux états interdits.
7. Procédé selon l'une quelconque des revendications 5 et 6 caractérisé en ce que la valeur d'incrément est égale à la somme pondérée du taux de corrélation à long terme, du taux de corrélation à l'ordre 1 et du taux de passage à zéro diminuée d'une valeur de seuil prédéterminée Z_0 .
8. Dispositif pour la mise en oeuvre du procédé selon l'une quelconque des revendications 1 à 7 caractérisé en ce qu'il comprend un dispositif (6) de calcul de valeur d'incrément de score, couplé à un dispositif de décision finale (8) comportant un dispositif de calcul des scores, de remontée des états et d'élimination des états interdits par l'intermédiaire d'un dispositif de normalisation (7).
9. Dispositif selon la revendication 8 caractérisé en ce que le dispositif (6) de calcul de valeur d'incrément de score comporte un dispositif de calcul de corrélation à long terme (9), un dispositif de calcul de corrélation à l'ordre 1 (10) et un dispositif (11) de calcul de taux de passage à zéro, couplés à un circuit sommateur (16).

- 10.** Dispositif selon la revendication 9 caractérisé en ce que le dispositif (6) de calcul de valeur de score est formé par un processeur de traitement du signal.

5

10

15

20

25

30

35

40

45

50

55

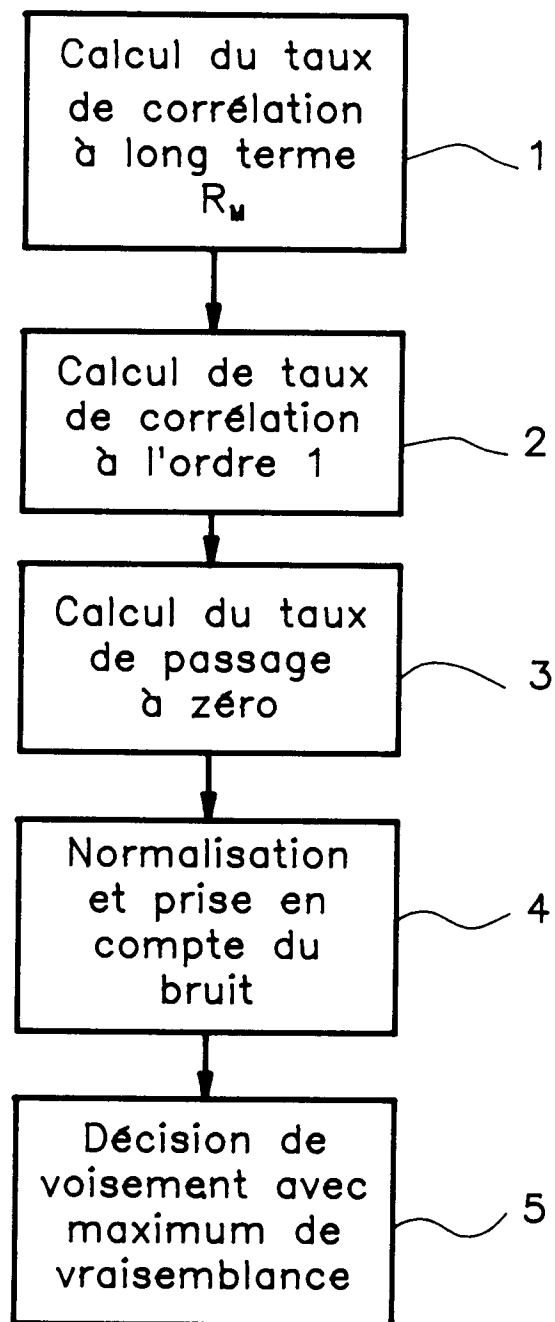


FIG 1

FIG 2

Etats :			
0	:	NNV	ou 001
1	:	NVV	ou 011
2	:	VVV	ou 111
3	:	NNN	ou 000
4	:	VNN	ou 100
5	:	VVN	ou 110

FIG 3

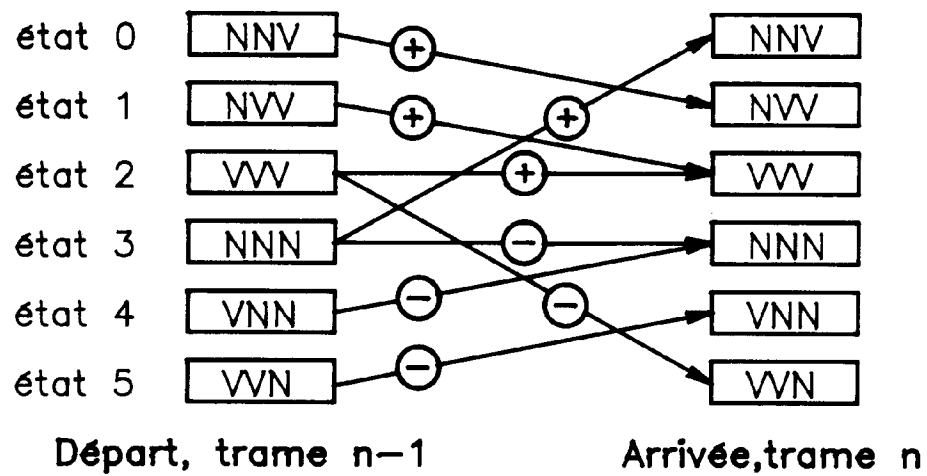
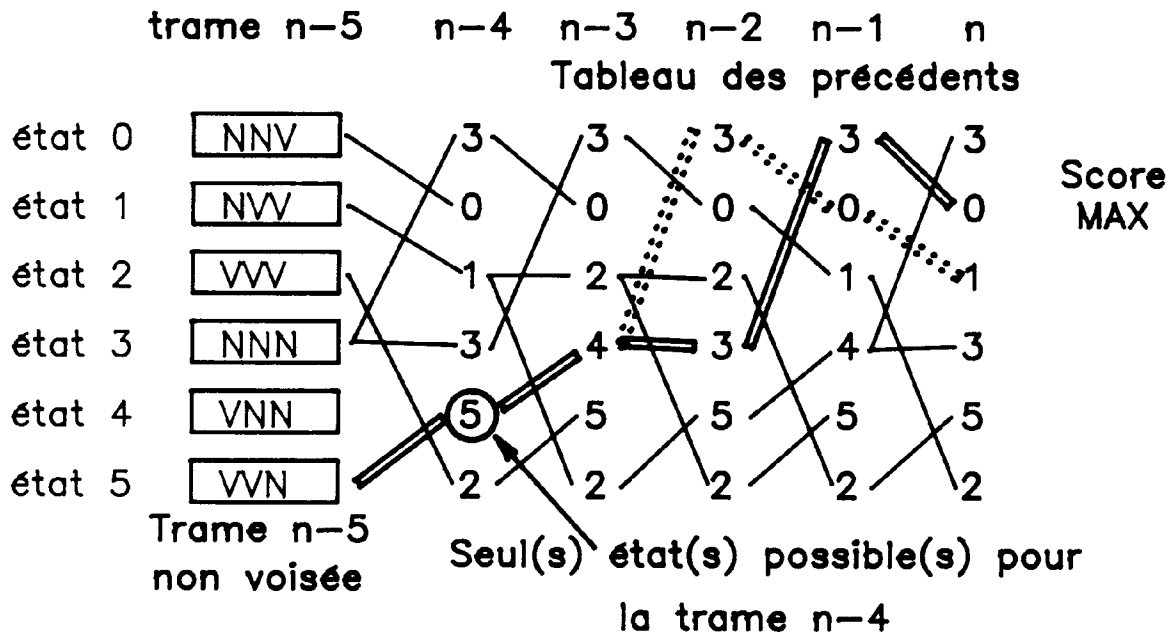


FIG 4



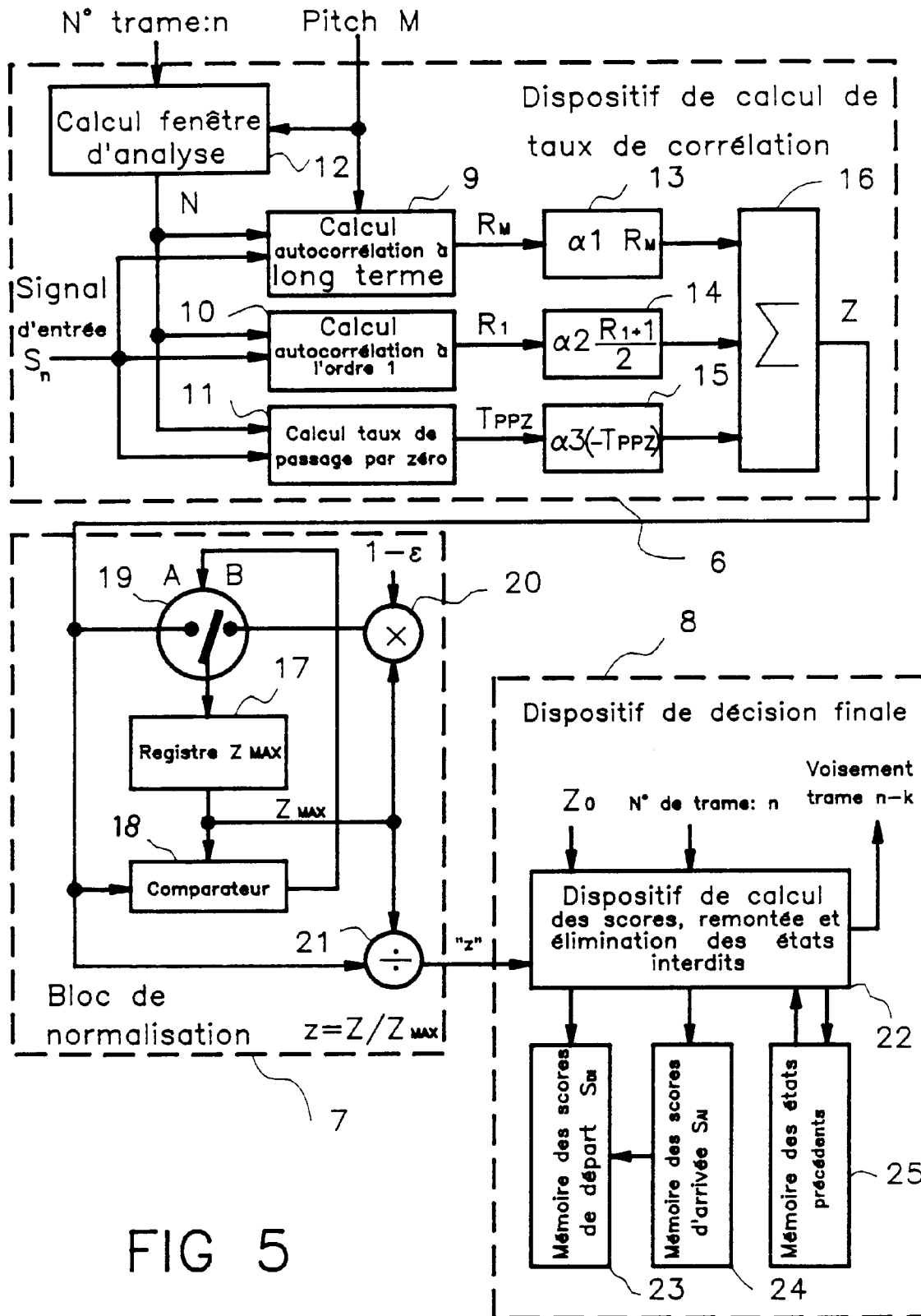


FIG 5



Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande

EP 92 40 3087

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.5)
A	US-A-4 589 131 (HORVATH) * revendications 4,11,13 * ---	1-4	G10L3/00
A	FR-A-2 316 682 (NTT) * revendications 1,3,5,7 * ---	2,3	
A	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. vol. 28, no. 4, Août 1980, NEW YORK US pages 398 - 407 CHONG K. ; HYEONG H. 'Voiced Unvoiced Silence discrimination of speech by delta modulation' * page 400, III D. ---	1,4 figure, 5**	
A	IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS. vol. 21, no. 3, Juin 1973, NEW YORK US pages 154 - 160 MARKEL '* Application of a digital inverse filter for automatic formant and Fo Analysis' * page 157, colonne de droite; figure 2 * ---	1,5	DOMAINES TECHNIQUES RECHERCHES (Int. Cl.5)
A	EUROSPEECH 89, EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY vol. 2, Septembre 1989, PARIS FRANCE pages 264 - 267 EL-SHEIKH T. ; EL-GHONEMY M. ; MANSOUR O. 'Toward a phoneme based word recognition system' * pages 264-265, II. -----	1,4	G10L
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche LA HAYE		Date d'achèvement de la recherche 25 FEVRIER 1993	Examineur FARASSOPOULOS A.
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 L503 03.82 (P0402)