



⑫ **DEMANDE DE BREVET EUROPEEN**

⑲ Numéro de dépôt : **95401262.1**

⑤① Int. Cl.⁶ : **G10L 3/00**

⑳ Date de dépôt : **31.05.95**

③① Priorité : **03.06.94 FR 9406825**

⑦② Inventeur : **Scott, Sophie**
18, rue Lecourbe
F-75015 Paris (FR)
 Inventeur : **Navarro, William**
3, rue Roland Garros
F-78140 Velizy Villacoublay (FR)

④③ Date de publication de la demande :
06.12.95 Bulletin 95/49

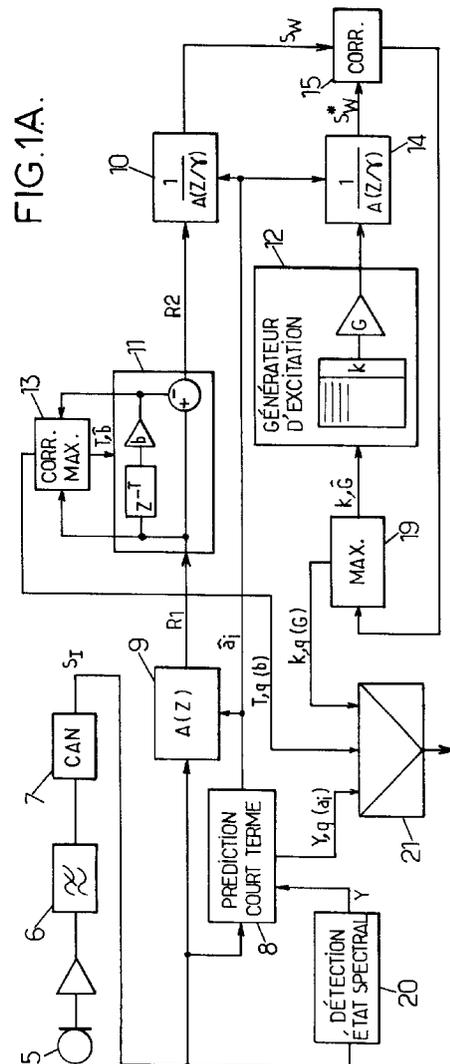
⑥④ Etats contractants désignés :
DE ES GB IT NL SE

⑦④ Mandataire : **Loisel, Bertrand et al**
Cabinet Plasseraud,
84, rue d'Amsterdam
F-75440 Paris Cédex 09 (FR)

⑦① Demandeur : **MATRA COMMUNICATION**
50, rue du Président Sadate,
BP 32 - Creac'h Gwenn
F-29101 Quimper Cédex (FR)

⑤④ **Procédé de codage de parole à prédiction linéaire.**

⑤⑦ Un signal de parole (S_i) numérisé en trames successives est soumis à une analyse par synthèse pour obtenir, pour chaque trame, des valeurs de quantification de paramètres de synthèse (a_i, b, T, k, G) permettant de reconstituer une estimation du signal de parole. L'analyse par synthèse comprenant une prédiction linéaire à court terme du signal de parole pour déterminer les valeurs de quantification des coefficients d'un filtre de synthèse à court terme. On détermine un état spectral (Y) du signal de parole parmi des premier et second états tels que le signal contienne proportionnellement moins d'énergie aux basses fréquences dans le premier état que dans le second état, et on applique l'un ou l'autre de deux modes de quantification pour obtenir les valeurs de quantification des coefficients du filtre de synthèse à court terme suivant l'état spectral déterminé (Y) du signal de parole.



La présente invention concerne un procédé de codage de parole à prédiction linéaire, dans lequel un signal de parole numérisé en trames successives est soumis à une analyse par synthèse pour obtenir, pour chaque trame, des valeurs de quantification de paramètres de synthèse permettant de reconstituer une estimation du signal de parole, l'analyse par synthèse comprenant une prédiction linéaire à court terme du signal de parole pour déterminer les coefficients d'un filtre de synthèse à court terme.

Les codeurs de parole à bas débit (typiquement 5 kbit/s pour une fréquence d'échantillonnage de 8 kHz) actuels donnent leur meilleure performance sur des signaux présentant un spectre "téléphonique", c'est-à-dire dans la bande 300-3400 Hz et avec une préaccentuation dans les fréquences élevées. Ces caractéristiques spectrales correspondent au gabarit IRS (Intermediate Reference System) défini par le CCITT dans la Recommandation P48. Ce gabarit a été défini pour les combinés téléphoniques, aussi bien en entrée (microphone) qu'en sortie (écouteurs).

Cependant, il arrive de plus en plus fréquemment que le signal d'entrée d'un codeur de parole présente un spectre plus "plat", par exemple lorsqu'une installation mains libres est utilisée, employant un microphone à réponse en fréquence linéaire. Les vocodeurs habituels sont conçus pour être indépendants de l'entrée avec laquelle ils fonctionnent, et ils ne sont d'ailleurs pas informés des caractéristiques de cette entrée. Si des microphones de caractéristiques différentes sont susceptibles d'être raccordés au vocodeur, ou plus généralement si le vocodeur est susceptible de recevoir des signaux acoustiques présentant des caractéristiques spectrales différentes, il y a alors des cas où le vocodeur est utilisé de façon sous-optimale.

Dans ce contexte, un but principal de la présente invention est d'améliorer les performances d'un vocodeur en les rendant moins dépendantes des caractéristiques spectrales du signal qui lui est destiné.

L'invention propose un procédé de codage de parole du type indiqué au début, dans lequel on détermine un état spectral du signal de parole parmi des premier et second états tels que le signal contienne proportionnellement moins d'énergie aux basses fréquences dans le premier état que dans le second état, et on applique l'un ou l'autre de deux modes de quantification pour obtenir les valeurs de quantification des coefficients du filtre de synthèse à court terme suivant l'état spectral déterminé du signal de parole.

Ainsi, la détection de l'état spectral permet d'adapter le codeur aux caractéristiques du signal d'entrée. Les performances du codeur peuvent être améliorées ou, à performances identiques, on peut réduire le nombre de bits nécessaires au codage.

De préférence, les coefficients du filtre de synthèse à court terme sont représentés par un ensemble de p paramètres fréquentiels de lignes spectrales ordonnés, dits "paramètres LSP", p étant l'ordre de la prédiction linéaire. La distribution de ces p paramètres LSP peut être analysée pour renseigner sur l'état spectral du signal et contribuer à la détection de cet état.

Les paramètres LSP peuvent faire l'objet d'une quantification scalaire ou vectorielle. Dans le cas d'une quantification scalaire, on quantifie le i -ième paramètre LSP en subdivisant un intervalle de variation inclus dans un intervalle de référence respectif en 2^{N_i} segments, N_i étant le nombre de bits de codage consacré à la quantification de ce paramètre. Une première possibilité est d'utiliser au moins pour les premiers paramètres LSP ordonnés, des intervalles de référence choisis chacun parmi deux intervalles distincts suivant l'état spectral déterminé du signal de parole. Une possibilité supplémentaire est de donner à certains au moins des nombres de bits de codage N_i l'une ou l'autre de deux valeurs distinctes suivant l'état spectral déterminé du signal de parole, pour effectuer une allocation dynamique de bits.

Dans le cas d'une quantification vectorielle directe, on subdivise l'ensemble des p paramètres LSP ordonnés en m groupes de paramètres consécutifs, et on peut quantifier au moins le premier groupe en sélectionnant dans une table de quantification un vecteur présentant une distance minimale avec les paramètres LSP dudit groupe, cette table étant choisie parmi deux tables de quantification distinctes suivant l'état spectral déterminé du signal de parole.

Dans le cas d'une quantification vectorielle différentielle, on subdivise l'ensemble des p paramètres LSP ordonnés en m groupes de paramètres consécutifs, et, au moins pour le premier groupe, on peut effectuer une quantification différentielle par rapport à un vecteur moyenne choisi parmi deux vecteurs distincts suivant l'état spectral déterminé du signal de parole.

D'autres particularités et avantages de la présente invention apparaîtront dans la description ci-après d'exemples de réalisation préférés mais non limitatifs, en référence aux dessins annexés, dans lesquels :

- les figures 1A et 1B sont des schémas synoptiques respectivement d'un codeur de parole à analyse par synthèse pour la mise en oeuvre de l'invention et d'un décodeur associé ;
- la figure 2 est un schéma synoptique d'une unité de prédiction linéaire utilisable dans le codeur de la figure 1A ;
- la figure 3 est un diagramme illustrant les caractéristiques d'un signal acoustique de type IRS et d'un signal de type linéaire ;
- la figure 4 est un schéma d'un dispositif de détection de l'état spectral du signal, utilisable avec le codeur

de la figure 1A ; et

- la figure 5 montre des chronogrammes illustrant le mode de détection de l'état du signal par le dispositif de la figure 4.

Le codeur de parole illustré sur la figure 1A repose sur le principe de l'analyse par synthèse. Son organisation générale est classique sauf en ce qui concerne l'unité 8 de prédiction à court terme et l'unité 20 de détection de l'état spectral du signal.

Le codeur de parole traite le signal de sortie amplifié d'un microphone 5. Un filtre passe-bas 6 élimine les composantes fréquentielles de ce signal au-dessus de la limite supérieure (par exemple 4000 Hz) de la bande passante traitée par le codeur. Le signal est ensuite numérisé par le convertisseur analogique-numérique 7 qui délivre le signal d'entrée S_i sous la forme de trames successives de 10 à 30 ms constituées d'échantillons prélevés à un cadence de 8000 Hz par exemple.

L'analyse par synthèse repose sur une modélisation du conduit vocal du locuteur par un filtre purement récursif de fonction de transfert $H(z) = 1/A(z)$ où

15

$$A(z) = 1 - \sum_{i=1}^p a_i z^{-i}$$

20

Les coefficients a_i de ce filtre ($1 \leq i \leq p$) peuvent être obtenus par prédiction linéaire à court terme du signal d'entrée, le nombre p désignant l'ordre de la prédiction linéaire, qui est typiquement égal à 10 pour de la parole à bande étroite. L'unité 8 de prédiction à court terme détermine des estimations \hat{a}_i des coefficients a_i qui correspondent à une quantification de ces coefficients par des valeurs de quantification $q(a_i)$.

25

Chaque trame de signal d'entrée S_i est d'abord soumise au filtre inverse 9 de fonction de transfert $A(z)$, puis à un filtre 10 de fonction de transfert $1/A(z/\gamma)$ où γ désigne un facteur prédéfini, généralement compris entre 0,8 et 0,9. Le filtre combiné ainsi constitué, de fonction de transfert $W(z) = A(z)/A(z/\gamma)$, est un filtre de pondération perceptuelle de l'erreur résiduelle du codeur. Les coefficients utilisés dans les filtres 9 et 10 sont les estimations \hat{a}_i délivrées par l'unité 8 de prédiction à court terme.

30

La sortie R1 du filtre inverse 9 possède une périodicité à long terme, correspondant à la hauteur tonale (pitch) de la parole. Dans l'exemple considéré, on modélise le filtre correspondant par une fonction de transfert de la forme $1/B(z)$ avec $B(z) = 1 - bz^{-T}$. Le signal R1 est soumis à un filtre inverse 11 de fonction de transfert $B(z)$ dont la sortie R2 est fournie à l'entrée du filtre 10. La sortie S_w du filtre 10 correspond ainsi au signal d'entrée S_i débarrassé de sa corrélation à long terme par le filtre 11 de fonction de transfert $B(z)$, et pondéré perceptuellement par les filtres 9, 10 de fonction de transfert combinée $W(z)$.

35

Le filtre 11 comprend un soustracteur dont l'entrée positive reçoit le signal R1 et dont l'entrée négative reçoit une estimation à long terme obtenue en retardant le signal R1 de T échantillons et en l'amplifiant. Le signal R1 ainsi que l'estimation à long terme sont fournis à une unité 13 qui maximise la corrélation entre ces deux signaux pour déterminer le retard T et le gain b optimaux. L'unité 13 explore toutes les valeurs entières et/ou fractionnaires du retard T entre deux bornes pour sélectionner celui qui maximise la corrélation normalisée. Le gain b se déduit de la valeur de T , et est quantifié par discrétisation, ce qui conduit à une valeur de quantification $q(b)$; la valeur quantifiée \hat{b} correspondant à cette valeur de quantification $q(b)$ est celle fournie comme gain de l'amplificateur du filtre 11.

40

La synthèse de parole dans le codeur s'effectue dans une boucle fermée comprenant un générateur d'excitation 12, un filtre 14 ayant la même fonction de transfert que le filtre 10, un corrélateur 15, et une unité 19 de maximisation de la corrélation normalisée.

45

La nature du générateur d'excitation 12 permet de distinguer entre différents types de codeur à analyse par synthèse, suivant la forme de l'excitation. On distingue ainsi les procédés de codage à analyse par prédiction linéaire et excitation multi-impulsionnelle (MPLPC), dont un exemple est donné dans le document EP-A-0 195 487, et les procédés de codage à analyse par prédiction linéaire et excitation vectorielle (CELP), réputés avoir de bonnes performances lorsqu'un faible débit binaire est requis, dont un exemple est donné dans l'article de Schroeder et Atal "Code Excited Linear Prediction (CELP) : High Quality Speech At Very Low Bits Rates", Proc. ICASSP, Mars 1985, pp. 937-940. Ces différentes façons de modéliser l'excitation sont utilisables dans le cadre de la présente invention. La demanderesse a utilisé une excitation par séquences d'impulsions régulières, ou RPCELP, telle que décrite dans sa demande de brevet européen N° 0 347 307. S'agissant d'un codeur de type CELP, l'excitation est représentée par une adresse d'entrée k dans un dictionnaire de vecteurs d'excitation, et par un gain associé G .

55

Le vecteur d'excitation sélectionné et amplifié est soumis au filtre 14 de fonction de transfert $1/A(z/\gamma)$, dont

les coefficients \hat{a}_i ($1 \leq i \leq p$) sont fournis par l'unité 8 de prédiction à court terme. Le signal résultant S_{W^*} est fourni à une entrée du corrélateur 15, dont l'autre entrée reçoit le signal de sortie S_W du filtre 10. La sortie du corrélateur 15 est constituée par la corrélation normalisée qui est maximisée par l'unité 19, ce qui revient à minimiser l'erreur de codage. L'unité 19 sélectionne l'adresse k et le gain G du générateur d'excitation qui maximisent la corrélation issue du corrélateur 15. La maximisation consiste à déterminer l'adresse k optimale, le gain G se déduisant de k . L'unité 19 opère une quantification par discrétisation de la valeur numérique du gain G , ce qui conduit à une valeur de quantification $q(G)$. La valeur quantifiée \hat{G} correspondant à cette valeur de quantification $q(G)$ est celle qui est fournie comme gain de l'amplificateur du générateur d'excitation 12. La corrélation maximisée prend en compte la pondération perceptuelle par la fonction de transfert $W(z)=A(z)/A(z/\gamma)$, étant observé que cette fonction de transfert est appliquée au signal d'entrée S_i par les filtres 9 et 10, ainsi qu'au signal synthétisé à partir du vecteur d'excitation, puisque le signal S_{W^*} peut être considéré comme résultant du vecteur d'excitation amplifié auquel sont appliquées successivement les fonctions de transfert $H(z)=1/A(z)$ du filtre de synthèse à court terme, et $W(z)=A(z)/A(z/\gamma)$ du filtre de pondération perceptuelle.

Le vecteur d'excitation sélectionné dans le dictionnaire du générateur 12, le gain G associé, les paramètres b et T du filtre à long terme 13, et les coefficients a_i du filtre de prédiction à court terme, auxquels s'ajoute un bit d'état Y qui sera décrit plus loin, constituent les paramètres de synthèse dont les valeurs de quantification k , $q(G)$, $q(b)$, T , $q(a_i)$, Y sont émises vers le récepteur pour permettre de reconstituer une estimation du signal de parole S_i . Ces valeurs de quantification sont réunies sur un même canal par le multiplexeur 21 en vue de l'émission.

Le décodeur associé illustré sur la figure 1B comprend une unité 50 qui restitue les valeurs quantifiées k , \hat{G} , T , \hat{b} , \hat{a}_i sur la base des valeurs de quantification reçues. Un générateur d'excitation 52 identique au générateur 12 du codeur reçoit les valeurs quantifiées des paramètres k et G . La sortie $\hat{R}2$ du générateur 52 (qui est une estimation de $R2$) est soumise au filtre de prédiction à long terme 53 de fonction de transfert $1/B(z)$ dont les coefficients sont les valeurs quantifiées des paramètres T et b . La sortie $\hat{R}1$ du filtre 53 (qui est une estimation de $R1$) est soumise au filtre de prédiction à court terme 54 de fonction de transfert $1/A(z)$ dont les coefficients sont les valeurs quantifiées des paramètres a_i . Le signal résultant \hat{S} est l'estimation du signal d'entrée S_i du codeur.

La figure 2 montre une exemple de constitution de l'unité 8 de prédiction à court terme du codeur. Les coefficients de modélisation a_i sont calculés pour chaque trame, par exemple pour la méthode des autocorrélations. Le bloc 40 calcule les autocorrélations

$$R(j) = \sum_{n=1}^{L-j} S_I(n) \times S_I(n+j)$$

pour $0 \leq j \leq p$, n désignant l'indice d'un échantillon de la trame courante, et L le nombre d'échantillons par trame. De façon classique, ces autocorrélations permettent un calcul récursif des coefficients a_i optimaux au moyen de l'algorithme de Levinson-Durbin (voir J. Makhoul: "Linear Prediction: A Tutorial Review", Proc. IEEE, Vol. 63, N°4, Avril 1975 pp. 561-580), qui peut s'exprimer de la façon suivante :

$$E(0) = R(0)$$

Pour $i = 1$ à p faire :

45

50

55

$$\begin{aligned}
 & \left[\begin{aligned}
 k_i &= \left[R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right] / E(i-1) \\
 a_i^{(i)} &= k_i \\
 E(i) &= (1 - k_i^2) E(i-1) \\
 a_j^{(i)} &= a_j^{(i-1)} - k_i a_{i-j}^{(i-1)} \quad \text{pour } 1 \leq j \leq i-1
 \end{aligned} \right.
 \end{aligned}$$

20 La solution finale obtenue par le bloc 41 est donnée par : $a_i = a_i^{(p)}$ pour $1 \leq i \leq p$. Dans l'algorithme ci-dessus, la quantité $E(p)$ est l'erreur résiduelle de la prédiction linéaire, et les quantités k_i , comprises entre -1 et +1 sont appelées coefficients de réflexion.

En vue de la transmission des coefficients obtenus, on peut les représenter par différents paramètres à quantifier : les coefficients de prédiction eux-mêmes a_i , les coefficients de réflexion k_i , ou encore les rapports logarithmiques LAR donnés par :

$$LAR_i = \log_{10}[(1 + k_i)/(1 - k_i)]$$

Les paramètres de représentation ainsi obtenus sont quantifiés pour réduire le nombre de bits nécessaires à leur identification.

30 L'invention propose de déterminer l'état spectral du signal de parole parmi un premier état Y_A ($Y = 0$, type IRS) et un second état Y_B ($Y = 1$, type linéaire) tels que le signal contienne proportionnellement moins d'énergie aux basses fréquences dans l'état Y_A que dans l'état Y_B , et d'appliquer l'un ou l'autre de deux modes de quantification distincts pour obtenir les valeurs de quantification des coefficients du filtre de synthèse à court terme suivant l'état spectral déterminé.

35 Sur la figure 3, les deux lignes en traits pleins correspondent à l'encadrement du gabarit IRS, défini pour des microphones dans la Recommandation P48 du CCITT. On voit qu'un signal de microphone de type IRS présente une forte atténuation dans la partie basse du spectre (entre 0 et 300 Hz) et une relative accentuation dans les hautes fréquences. En comparaison, un signal de type linéaire, fourni par exemple par le microphone d'une installation mains libres présente un spectre plus plat, n'ayant notamment pas la forte atténuation aux basses fréquences (un exemple typique d'un tel signal de type linéaire est illustré par une ligne en tirets sur 40 le diagramme de la figure 3).

On tire parti de ces propriétés spectrales dans le dispositif de détection 20, représenté sur la figure 1A et détaillé sur la figure 4, qui délivre trame par trame le bit d'état Y .

Le dispositif de détection 20 comprend un filtre passe-haut 16 recevant le signal acoustique d'entrée S_i et délivrant un signal filtré S_i' . Le filtre 16 est typiquement un filtre numérique de type bi-quad ayant une coupe 45 pure abrupte à 400 Hz. Les énergies $E1$ et $E2$ contenues dans chaque trame du signal acoustique d'entrée S_i et du signal filtré S_i' sont calculées par deux unités 17, 18 effectuant chacune la somme des carrés des échantillons de chaque trame qu'elle reçoit.

L'énergie $E1$ de chaque trame du signal d'entrée S_i est adressée à l'entrée d'un comparateur à seuil 25 qui délivre un bit Z de valeur 0 lorsque l'énergie $E1$ est inférieure à un seuil d'énergie prédéterminé, et de valeur 1 lorsque l'énergie $E1$ est supérieure au seuil. Le seuil d'énergie est typiquement de l'ordre de -38 dB par rapport à l'énergie de saturation du signal. Le comparateur 25 sert à inhiber la détermination de l'état du signal lorsque celui-ci contient trop peu d'énergie pour être représentatif des caractéristiques de la source. Dans ce cas, l'état déterminé du signal reste inchangé.

Les énergies $E1$ et $E2$ sont adressées à un diviseur numérique 26 qui calcule le rapport $E2/E1$ pour chaque 55 trame. Ce rapport $E2/E1$ est adressé à un autre comparateur à seuil 27 qui délivre un bit X de valeur 0 lorsque le rapport $E2/E1$ est supérieur à un seuil prédéterminé, et de valeur 1 lorsque le rapport $E2/E1$ est inférieur au seuil. Ce seuil sur le rapport $E2/E1$ est typiquement de l'ordre de 0,93. Le bit X est représentatif d'une condition du signal sur chaque trame. La condition $X = 0$ correspond aux caractéristiques IRS du signal d'entrée

(état Y_A), et la condition $X = 1$ correspond aux caractéristiques linéaires (état Y_B). Pour éviter des changements d'état répétés et intempestifs à l'occasion des variations à court terme de l'excitation vocale, le bit d'état Y n'est pas pris directement égal au bit de condition X , mais il résulte d'un traitement des bits de condition successifs X par un circuit 29 de détermination d'état, qui permet de ne modifier l'état déterminé Y qu'après que plusieurs trames successives montrent une condition de signal X différente de celle correspondant à l'état précédemment déterminé.

Le fonctionnement du circuit 29 de détermination d'état est illustré sur la figure 5, où le chronogramme supérieur illustre un exemple d'évolution du bit X fourni par le comparateur 27. Le bit d'état Y (chronogramme inférieur) est initialisé à 0, car les caractéristiques IRS sont le plus fréquemment rencontrées. On calcule trame après trame une variable de comptage V initialement mise à 0. La variable V est incrémentée d'une unité chaque fois que la condition X du signal sur une trame diffère de celle correspondant à l'état déterminé Y ($X = 1$ et $Y = 0$, ou $X = 0$ et $Y = 1$). Dans le cas contraire ($X = Y = 0$ ou 1) la variable V est décrétementée de deux unités si elle est différente de 0 et de 1, décrétementée d'une unité si elle est égale à 1, et maintenue inchangée si elle est égale à 0. Dès que la variable V atteint un seuil prédéterminé (8 dans l'exemple considéré), on la remet à 0 et on change la valeur du bit Y , de sorte qu'on détermine que le signal a changé d'état. Ainsi, dans l'exemple représenté sur la figure 5, le signal est dans l'état Y_A jusqu'à la trame M, dans l'état Y_B entre les trames M et N (changement de la source de signal), puis de nouveau dans l'état Y_A à partir de la trame N. Bien entendu, d'autres modes d'incrémentement et de décrémentation et d'autres valeurs de seuil seraient utilisables.

Le mode de comptage ci-dessus peut par exemple être obtenu par le circuit 29 représenté sur la figure 4. Ce circuit comprend un compteur 32 sur quatre bits, dont le bit de poids fort correspond au bit d'état Y , et dont les trois bits de poids faible représentent la variable de comptage V . Les bits X et Y sont fournis à l'entrée d'une porte OU EXCLUSIF 33 dont la sortie est adressée à l'entrée d'incrémentement du compteur 32 par l'intermédiaire d'une porte ET 34 dont l'autre entrée reçoit le bit Z fourni par le comparateur à seuil 25. Ainsi, la variable V est incrémentée lorsque $X \neq Y$ et $Z = 1$. La sortie inversée de la porte 33 est fournie à une entrée de décrémentation du compteur 32 par l'intermédiaire d'une autre porte ET 35 dont les deux autres entrées reçoivent respectivement le bit Z fourni par le comparateur 25, et la sortie d'une porte OU à trois entrées 36 recevant les trois bits de poids faible du compteur 32. Le compteur 32 est agencé pour doubler les impulsions reçues sur son entrée de décrémentation lorsque son bit de poids le plus faible vaut 0 ou lorsque l'un au moins des deux bits suivants vaut 1, comme schématisé par la porte ou 37 sur la figure 4. Ainsi, le compteur 32 est décrétementé (d'une unité si $V = 1$ et de deux unités si $V > 1$) lorsque $X = Y$ et $Z = 1$ et $V \neq 0$. Lorsque l'énergie du signal d'entrée est insuffisante, on a $Z = 0$ et le circuit de détermination 29 n'est pas activé car les portes ET 34, 35 empêchent de modifier la valeur du compteur 32.

Le bit d'état Y ainsi déterminé est fourni à l'unité 8 de prédiction linéaire à court terme pour choisir le mode de quantification des coefficients du filtre de synthèse à court terme.

Dans l'exemple préféré illustré sur la figure 2, les paramètres utilisés pour représenter les coefficients a_i du filtre de synthèse à court terme sont les fréquences de lignes spectrales (LSF), ou paires de lignes spectrales (LSP). Ces paramètres sont connus comme ayant de bonnes propriétés statistiques et comme assurant aisément la stabilité du filtre synthétisé (voir N. Sugamura et F. Itakura : "Speech Analysis And Synthesis Method Developed At ECL In NTT : From LPC to LSP", Speech Communication, North Holland, Vol. 5, N° 2, 1986, pp. 199-215). Les paramètres LSP sont obtenus à partir des polynômes $Q(z)$ et $Q^*(z)$ définis ci-dessous :

$$Q(z) = A(z) + z^{-(p+1)} \times A(z^{-1})$$

$$Q^*(z) = A(z) - z^{-(p+1)} \times A(z^{-1})$$

On démontre que les racines complexes de ces deux polynômes sont sur le cercle unité et que, lorsqu'on parcourt le cercle unité, les racines de $Q(z)$ alternent avec celles de $Q^*(z)$. Les p racines autres que $z = +1$ et $z = -1$ s'écrivent e^{2njf_i} avec $j^2 = -1$, les p fréquences f_i étant définies comme les fréquences de lignes spectrales normalisées relativement à la fréquence d'échantillonnage. Les fréquences normalisées f_i sont comprises entre 0 et 0,5 et sont ordonnées de façon que chaque paire de fréquences consécutives comprenne une fréquence correspondant à une racine de $Q(z)$ et une fréquence correspondant à une racine de $Q^*(z)$. Dans cette modélisation, les lignes spectrales d'une paire encadrent un formant du signal de parole, et leur distance est inversement proportionnelle à l'amplitude de la résonance de ce formant.

Les paramètres LSP sont calculés par le bloc 42 à partir des coefficients de prédiction a_i obtenus par le bloc 41 au moyen des polynômes de Chebyshev (voir P. Kabal et R.P. Ramachandran : "The Computation Of Line Spectral Frequencies Using Chebyshev Polynomials", IEEE Trans. ASSP, Vol. 34, N°6, 1986 pp. 1419-1426). Ils peuvent également être obtenus directement à partir des autocorrélations du signal, par l'algorithme de Levinson éclaté (voir P. Delsarte et Y. Genin : "The Split Levinson Algorithm", IEEE Trans. ASSP, Vol. 34, N°3, 1986).

Le bloc 43 effectue la quantification des fréquences LSF, ou plus précisément des valeurs $\cos 2\pi f_i$, ci-après appelées paramètres LSP, comprises entre -1 et +1, ce qui simplifie les problèmes de dynamique. La

méthode de calcul des fréquences LSF permet de les obtenir dans l'ordre des fréquences croissantes, c'est-à-dire des cosinus décroissants.

Il existe pour ces paramètres LSP deux grandes familles de méthodes de quantification : la quantification scalaire où chaque paramètre est représenté séparément par la valeur quantifiée la plus proche ; et la quantification vectorielle, qui s'effectue sur un ou plusieurs groupes de paramètres, pour chacun desquels on cherche dans un dictionnaire multidimensionnel le vecteur le plus approchant.

Dans le cas d'une quantification vectorielle pour une analyse LPC d'ordre $p=10$, on effectue par exemple $m = 3$ quantifications vectorielles indépendantes, de dimensions respectives 3, 3 et 4, définissant les groupes LSP I(1,2,3), II(4,5,6) et III(7,8,9,10). Chaque groupe est quantifié en sélectionnant dans une table de quantification respective préenregistrée le vecteur présentant la distance euclidienne minimale avec les paramètres de ce groupe.

Pour le groupe I, on définit deux tables de quantification $T_{I,1}$ et $T_{I,2}$ disjointes de tailles respectives 2^{n_1} et 2^{n_2} . Pour le groupe II, on définit deux tables de quantification $T_{II,1}$ et $T_{II,2}$ de tailles respectives 2^{p_1} et 2^{p_2} ayant une partie commune pour réduire l'espace mémoire nécessaire. Pour le groupe III, on définit une table de quantification unique T_{III} de taille 2^q . Les adresses AD_I , AD_{II} , AD_{III} des trois vecteurs issus de trois tables de quantification relatives aux trois groupes constituent les valeurs de quantification $q(a_i)$ des coefficients du filtre de synthèse à court terme, qui sont adressées au multiplexeur 21. Le bloc 43, qui opère la quantification des paramètres LSP, sélectionne les tables $T_{I,1}$ et $T_{II,1}$ pour rechercher les vecteurs de quantification des groupes I et II lorsque $Y = 0$ (signal de type IRS). En conséquence, on construit les échantillons des tables $T_{I,1}$ et $T_{II,1}$ de façon que leurs statistiques soient optimisées pour la quantification d'un signal de type de IRS. Lorsque $Y = 1$ (état linéaire), le bloc 43 sélectionne les tables $T_{I,2}$ et $T_{II,2}$, dont la statistique est établie pour être représentative d'un signal d'entrée de type linéaire. Pour le groupe III, la table T_{III} est utilisée dans tous les cas, car la partie haute du spectre est moins sensible aux différences entre les caractéristiques IRS et linéaire. Le bit d'état Y est également fourni au multiplexeur 21.

Une unité 45 calcule les estimations \hat{a}_i à partir des valeurs discrétisées des paramètres LSP données par les trois vecteurs retenus. Les paramètres LSP $\cos 2\pi f_i$ permettent aisément de déterminer les coefficients du filtre de synthèse à court terme étant donné que

$$Q(z) = (1+z^{-1}) \prod_{i=1,3,\dots,p-1} (1-2\cos(2\pi f_i) z^{-1}+z^{-2}) \quad ;$$

$$Q^*(z) = (1-z^{-1}) \prod_{i=2,4,\dots,p} (1-2\cos(2\pi f_i) z^{-1}+z^{-2}) \quad ;$$

et

$$A(z) = [Q(z) + Q^*(z)]/2$$

Les estimations \hat{a}_i ainsi obtenues sont fournies par l'unité 45 aux filtres à court terme 9, 10 et 14 du codeur. Dans le décodeur, le même calcul est effectué par l'unité de restitution 50, les vecteurs de cosinus quantifiés étant retrouvés à partir des adresses de quantification AD_I , AD_{II} et AD_{III} . Le décodeur contient les mêmes tables de quantification que le codeur, et leur sélection s'effectue en fonction du bit d'état Y reçu.

Outre l'optimisation des performances du codeur, l'utilisation de deux familles de tables de quantification sélectionnées suivant l'état spectral Y a l'avantage de procurer une meilleure efficacité en termes de nombre de bits de codage requis. En effet, le nombre total de bits utilisés, à performance égale, pour la quantification des paramètres LSP dans chacun des cas est inférieur au nombre de bits nécessaires lorsqu'une seule famille de tables est utilisée indépendamment d'une détection de l'état spectral. Dans le cas typique où $n_1 = 8$, $n_2 = 7$, $p_1 = 9$, $p_2 = 10$ et $q = 8$, le nombre de bits nécessaires au codage des paramètres LSP vaut $n_1 + p_1 + q + 1 = 26$ lorsque $Y = 0$, et $n_2 + p_2 + q + 1 = 26$ lorsque $Y = 1$ (ce qui assure le même débit global), tandis que, pour obtenir une statistique aussi riche sans faire appel à l'état Y , il faudrait au moins $n + p + q = 10 + 11 + 8 = 29$ bits d'adressage.

En variante, le bloc 43 peut être agencé pour effectuer une quantification vectorielle différentielle. Chaque groupe de paramètres I, II, III est alors quantifié différentiellement par rapport à un vecteur moyenne. Pour le groupe I, on définit deux vecteurs moyenne distincts $V_{I,1}$ et $V_{I,2}$ et une table de quantification des différences

TD_i. Pour le groupe II, on définit deux vecteurs moyenne distincts $V_{II,1}$ et $V_{II,2}$ et une table de quantification des différences TD_{II}. Pour le groupe III, on définit un unique vecteur moyenne V_{III} et une table de quantification des différences TD_{III}. Les vecteurs moyenne $V_{I,1}$ et $V_{I,2}$ sont établis pour être représentatifs d'une statistique de signaux de type IRS, tandis que les vecteurs moyenne $V_{II,1}$ et $V_{II,2}$ sont établis pour être représentatifs d'une statistique de signaux de type linéaire. Le bloc 43 opère la quantification différentielle des groupes I et II par rapport aux vecteurs $V_{I,1}$ et $V_{I,2}$ lorsque $Y = 0$ (état IRS) et par rapport aux vecteurs $V_{II,1}$ et $V_{II,2}$ lorsque $Y = 1$ (état linéaire). L'avantage de cette quantification différentielle est qu'elle permet de ne stocker, dans le codeur et dans le décodeur, qu'une seule table de quantification par groupe. Les valeurs de quantification $q(a_i)$ sont les adresses des trois vecteurs-différence optimaux dans les trois tables, auxquelles s'ajoute le bit Y déterminant quels sont les vecteurs moyenne à ajouter à ces vecteurs-différence pour restituer les paramètres LSP quantifiés.

Lorsqu'on procède à une quantification scalaire, chaque paramètre est représenté séparément par la valeur quantifiée la plus proche. On définit pour chaque paramètre LSP $\cos 2\pi f_i$ une borne supérieure M_i et une borne inférieure m_i , telles que, sur un grand nombre d'échantillons de parole, environ 90% des valeurs rencontrées de $\cos 2\pi f_i$ soient comprises entre m_i et M_i . L'intervalle de référence entre les deux bornes est divisé en 2^{N_i} segments égaux, où N_i est le nombre de bits de codage consacré à la quantification du paramètre $\cos 2\pi f_i$. Après avoir quantifié le premier paramètre LSP $\cos 2\pi f_1$, on utilise la propriété d'ordonnement des fréquences f_i pour remplacer dans certains cas la borne supérieure M_i par la valeur quantifiée du cosinus précédent $\hat{\cos} 2\pi f_{i-1}$. En d'autres termes, pour $1 < i \leq p$, on effectue la quantification de $\cos 2\pi f_i$ en subdivisant en 2^{N_i} segments égaux l'intervalle de variation $[m_i, \min\{M_i, \hat{\cos} 2\pi f_{i-1}\}]$. La quantification d'un paramètre LSP $\cos 2\pi f_i$ dans son intervalle de variation consiste à déterminer le nombre n_i de N_i bits tel que $\cos 2\pi f_i$ soit dans le n_i -ième segment de l'intervalle de référence (si $\cos 2\pi f_i < m_i$, on prend $n_i = 1$).

La détection de l'état spectral du signal permet de définir deux familles d'intervalles de référence $[m_{i,1}, M_{i,1}]$ et $[m_{i,2}, M_{i,2}]$ pour les r premiers paramètres ($1 \leq i \leq r \leq p$). La famille $[m_{i,1}, M_{i,1}]$ est établie statistiquement à partir d'échantillons de signaux de type IRS, et est sélectionnée pour procéder à la quantification lorsque $Y = 0$ (état IRS). La famille $[m_{i,2}, M_{i,2}]$ est établie statistiquement à partir d'échantillons de signaux de type linéaire et est sélectionnée pour procéder à la quantification lorsque $Y = 1$ (état linéaire). Ces deux familles sont mémorisées à la fois dans le codeur et dans le décodeur.

Une autre possibilité, qui peut compléter ou remplacer la précédente, consiste à définir pour certains des paramètres des nombres de bits de codage N_i différents suivant que le signal est de type IRS ou linéaire. Pour un même nombre total de bits de codage, on peut notamment prendre des nombres N_i plus faibles dans le cas IRS que dans le cas linéaire pour les premiers paramètres LSP (les cosinus les plus grands), étant donné que la dynamique des premiers paramètres LSP est réduite dans le cas IRS, la diminution des premiers N_i étant compensée par une augmentation des N_i relatifs aux derniers paramètres LSP, ce qui augmente la finesse de quantification de ces derniers paramètres. Ces différentes allocations des bits de codage sont mémorisées à la fois dans le codeur et dans le décodeur, les paramètres LSP pouvant ainsi être retrouvés en examinant le bit d'état Y.

En remplacement ou en complément du dispositif 20, on peut se servir des paramètres LSP calculés pour déterminer quel est l'état spectral Y du signal d'entrée. Ceci est illustré par le bloc 44 sur la figure 2. En effet, les lignes spectrales de chaque paire encadrent un formant du signal de parole, et leur distance est inversement proportionnelle à l'amplitude de la résonance. On voit qu'ainsi les paramètres LSP peuvent donner directement une idée assez précise de l'enveloppe spectrale du signal de parole. Dans le cas d'un signal de type IRS, l'amplitude des résonances situées dans la partie basse du spectre est plus faible que dans le cas linéaire. Ainsi, en analysant les écarts entre les premières fréquences LSF consécutives, on peut déterminer si le signal d'entrée est plutôt de type IRS (écarts grands) ou linéaire (écarts plus petits). Cette détermination peut s'effectuer pour chaque trame de signal pour obtenir le bit de condition X qu'on traite ensuite par un circuit de détermination d'état semblable au circuit 29 de la figure 4 pour obtenir le bit d'état Y utilisé par le bloc de quantification 43.

50

Revendications

1. Procédé de codage de parole à prédiction linéaire, dans lequel un signal de parole (S_i) numérisé en trames successives est soumis à une analyse par synthèse pour obtenir, pour chaque trame, des valeurs de quantification de paramètres de synthèse (a_i, b, T, k, G) permettant de reconstituer une estimation (\hat{S}) du signal de parole, et on émet lesdites valeurs de quantification, l'analyse par synthèse comprenant une prédiction linéaire à court terme du signal de parole pour déterminer les valeurs de quantification des coefficients d'un filtre de synthèse à court terme, caractérisé en ce qu'on détermine un état spectral (Y) du signal de

55

parole parmi des premier et second états (Y_A, Y_B) tels que le signal contienne proportionnellement moins d'énergie aux basses fréquences dans le premier état que dans le second état, et on applique l'un ou l'autre de deux modes de quantification pour obtenir les valeurs de quantification des coefficients du filtre de synthèse à court terme suivant l'état spectral déterminé (Y) du signal de parole.

5

2. Procédé selon la revendication 1, caractérisé en ce qu'on ne modifie pas l'état déterminé (Y) du signal de parole lorsqu'il a une énergie inférieure à un seuil prédéterminé.

10

3. Procédé selon la revendication 1 ou 2, caractérisé en ce qu'on détecte trame par trame si le signal est dans une première condition correspondant au premier état (Y_A) ou dans une seconde condition correspondant au second état (Y_B), et on détermine l'état (Y) du signal sur la base des conditions trame par trame (X), en ne modifiant l'état déterminé qu'après que plusieurs trames successives montrent une condition de signal différente de celle correspondant à l'état précédemment déterminé.

15

4. Procédé selon la revendication 3, caractérisé en ce qu'on incrémente une variable de comptage (V) lorsque la condition (X) du signal sur une trame diffère de celle correspondant à l'état déterminé (Y) du signal, en ce qu'on décrémente cette variable de comptage (V) lorsque la condition du signal sur une trame est celle correspondant à l'état déterminé du signal sauf si cette variable vaut zéro, et en ce que, lorsque la variable de comptage (V) atteint un seuil prédéterminé, on la remet à zéro et on détermine que le signal a changé d'état.

20

5. Procédé selon la revendication 3 ou 4, caractérisé en ce qu'on soumet le signal de parole (S_1) à un filtrage passe-haut, on compare l'énergie (E2) du signal filtré passe-haut (S_1') à celle (E1) du signal non filtré pour déterminer trame par trame si le signal est dans la première condition, pour laquelle l'énergie du signal filtré passe-haut est supérieure à une fraction prédéterminée de l'énergie du signal non filtré, ou dans la seconde condition, pour laquelle l'énergie du signal filtré passe-haut est inférieure à la fraction prédéterminée de l'énergie du signal non filtré.

25

6. Procédé selon la revendication 3 ou 4, caractérisé en ce que les coefficients (a_i) du filtre de synthèse à court terme sont représentés par un ensemble de fréquences de lignes spectrales (f_i), et en ce qu'on analyse la distribution des fréquences de lignes spectrales dans chaque trame du signal de parole (S_1) pour détecter si le signal est dans la première ou la seconde condition.

30

7. Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que les coefficients (a_i) du filtre de synthèse à court terme sont représentés par un ensemble de p paramètres fréquentiels de lignes spectrales ordonnés ($\cos 2\pi f_i$), subdivisé en m groupes de paramètres fréquentiels consécutifs, p étant l'ordre de la prédiction linéaire à court terme et m étant un nombre entier supérieur ou égal à 1, et en ce qu'au moins le premier groupe est quantifié différentiellement par rapport à un vecteur moyenne choisi parmi deux vecteurs distincts ($V_{1,1}, V_{1,2}$) suivant l'état spectral déterminé (Y) du signal de parole.

35

40

8. Procédé selon la revendication 7, caractérisé en ce que le nombre m est égal à 3, et en ce que chacun des deux premiers groupes de paramètres fréquentiels consécutifs est quantifié différentiellement par rapport à un vecteur moyenne respectif choisi parmi deux vecteurs distincts respectifs suivant l'état spectral déterminé (Y) du signal de parole.

45

9. Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que les coefficients (a_i) du filtre de synthèse à court terme sont représentés par un ensemble de p paramètres fréquentiels de lignes spectrales ordonnés ($\cos 2\pi f_i$), subdivisé en m groupes de paramètres fréquentiels consécutifs, p étant l'ordre de la prédiction linéaire à court terme et m étant un nombre entier supérieur ou égal à 1, et en ce qu'au moins le premier groupe est quantifié en sélectionnant dans une table de quantification un vecteur présentant une distance minimale avec les paramètres fréquentiels dudit groupe, cette table de quantification étant choisie parmi deux tables distinctes ($T_{1,1}, T_{1,2}$) suivant l'état spectral déterminé (Y) du signal de parole.

50

10. Procédé selon la revendication 9, caractérisé en ce que le nombre m est égal à 3, et en ce que chacun des deux premiers groupes de paramètres fréquentiels consécutifs est quantifié en sélectionnant dans une table de quantification respective un vecteur présentant une distance minimale avec les paramètres fréquentiels dudit groupe, chacune des deux tables de quantification relatives aux deux premiers groupes

55

étant choisie parmi deux tables distinctes respectives suivant l'état spectral déterminé (Y) du signal de parole.

- 5 **11.** Procédé selon la revendication 10, caractérisé en ce que les deux tables de quantification distinctes ($T_{i,1}$, $T_{i,2}$) relatives au premier groupe sont disjointes, et en ce que les deux tables de quantification distinctes ($T_{ii,1}$, $T_{ii,2}$) relatives au second groupe présentent une partie commune.
- 10 **12.** Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que les coefficients (a_i) du filtre de synthèse à court terme sont représentés par un ensemble de p paramètres fréquentiels de lignes spectrales ordonnés ($\cos 2\pi f_i$), p étant l'ordre de la prédiction linéaire à court terme, en ce qu'on quantifie chacun de ces p paramètres en subdivisant un intervalle de variation ($[m_i, \min\{M_i, \hat{\cos 2\pi f_{i-1}}\}]$) inclus dans un intervalle de référence respectif ($[m_i, M_i]$) en 2^{N_i} segments, N_i étant le nombre de bits de codage consacré à la quantification de ce paramètre, et en ce qu'on utilise, au moins pour les premiers paramètres ordonnés, des intervalles de référence choisis chacun parmi deux intervalles distincts ($[m_{i,1}, M_{i,1}]$, $[m_{i,2}, M_{i,2}]$) suivant l'état spectral déterminé (Y) du signal de parole.
- 15 **13.** Procédé selon l'une quelconque des revendications 1 à 6 ou la revendication 12, caractérisé en ce que les coefficients (a_i) du filtre de synthèse à court terme sont représentés par un ensemble de p paramètres fréquentiels de lignes spectrales ordonnés ($\cos 2\pi f_i$), p étant l'ordre de la prédiction linéaire à court terme, en ce qu'on quantifie chacun de ces p paramètres en subdivisant un intervalle de variation ($[m_i, \min\{M_i, \hat{\cos 2\pi f_{i-1}}\}]$) inclus dans un intervalle de référence respectif ($[m_i, M_i]$) en 2^{N_i} segments, N_i étant le nombre de bits de codage consacré à la quantification de ce paramètre, et en ce qu'on donne à certains au moins des nombres de bits de codage N_i l'une ou l'autre de deux valeurs distinctes suivant l'état spectral déterminé (Y) du signal de parole.
- 20
- 25
- 30
- 35
- 40
- 45
- 50
- 55

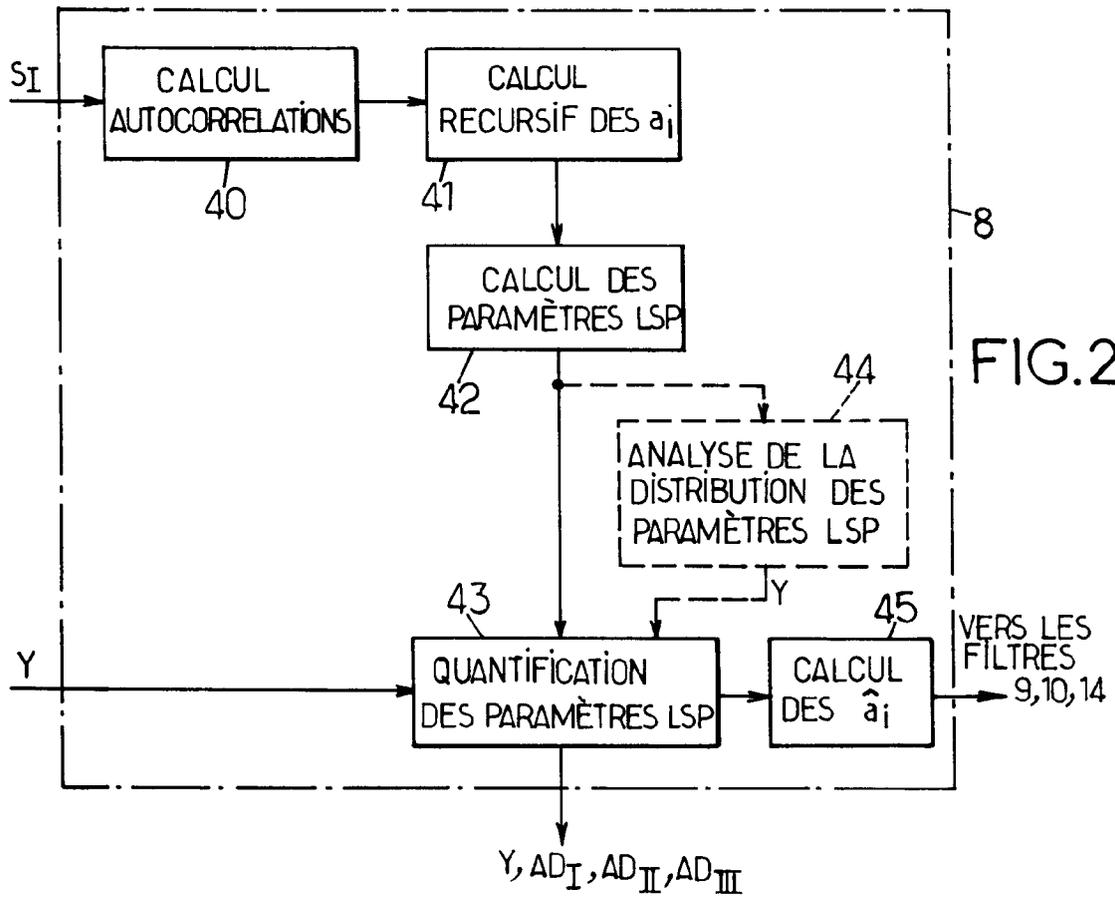


FIG. 2.

FIG. 3.

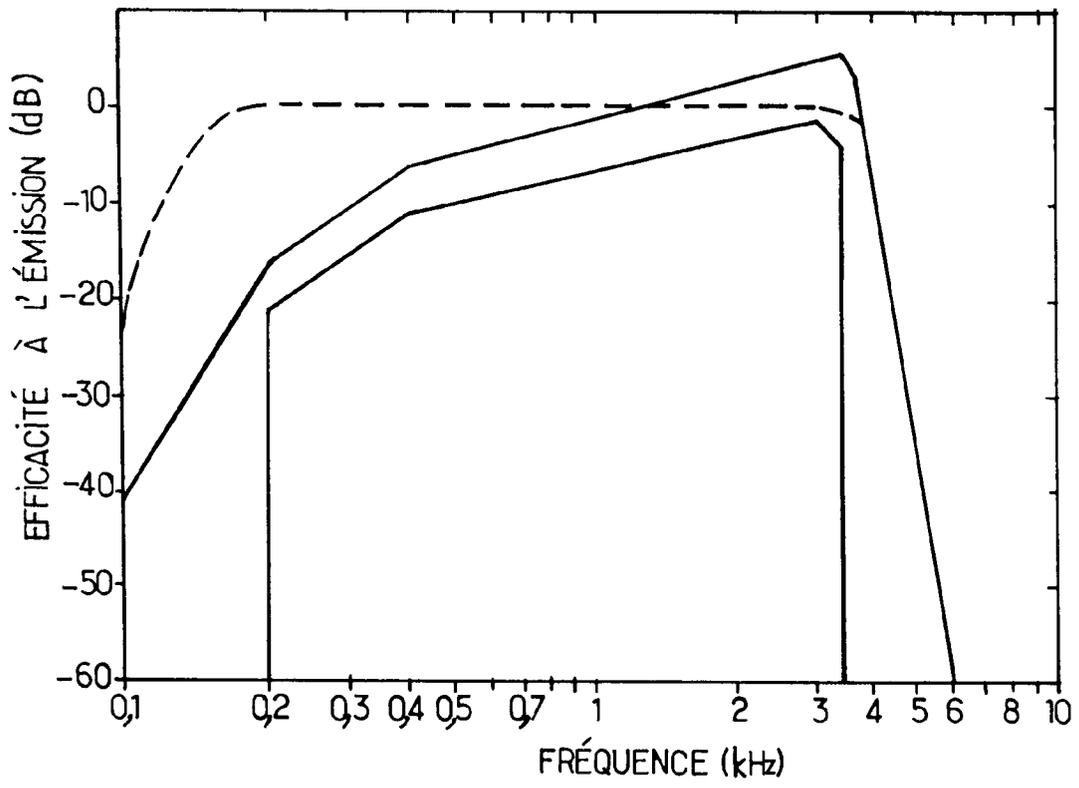


FIG. 4.

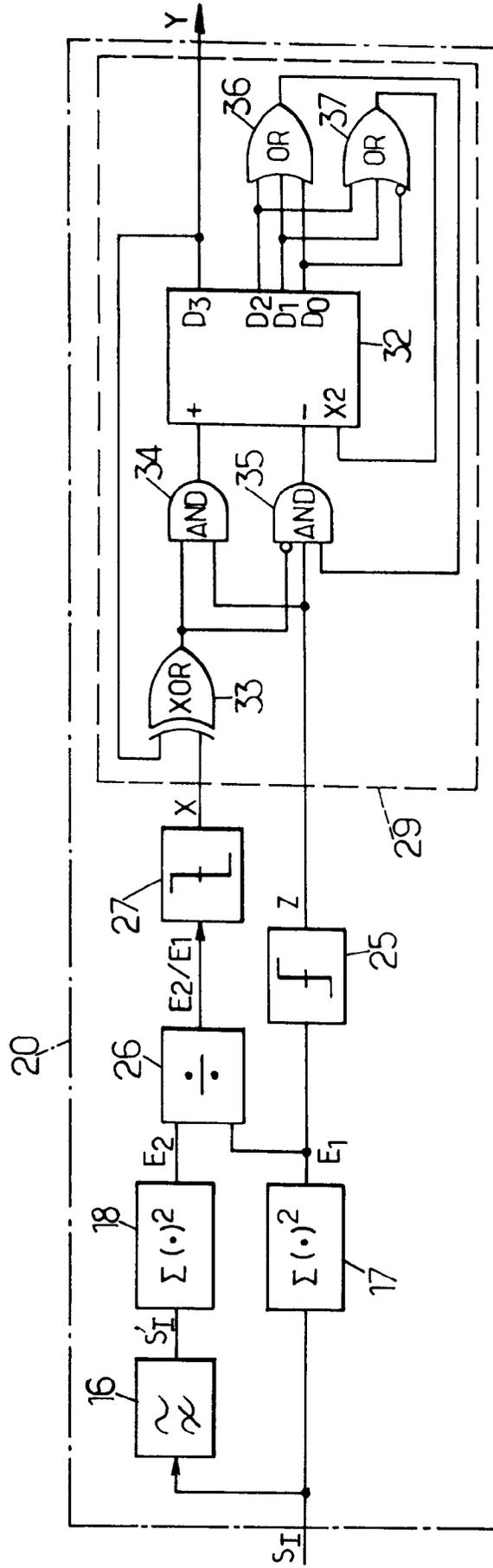
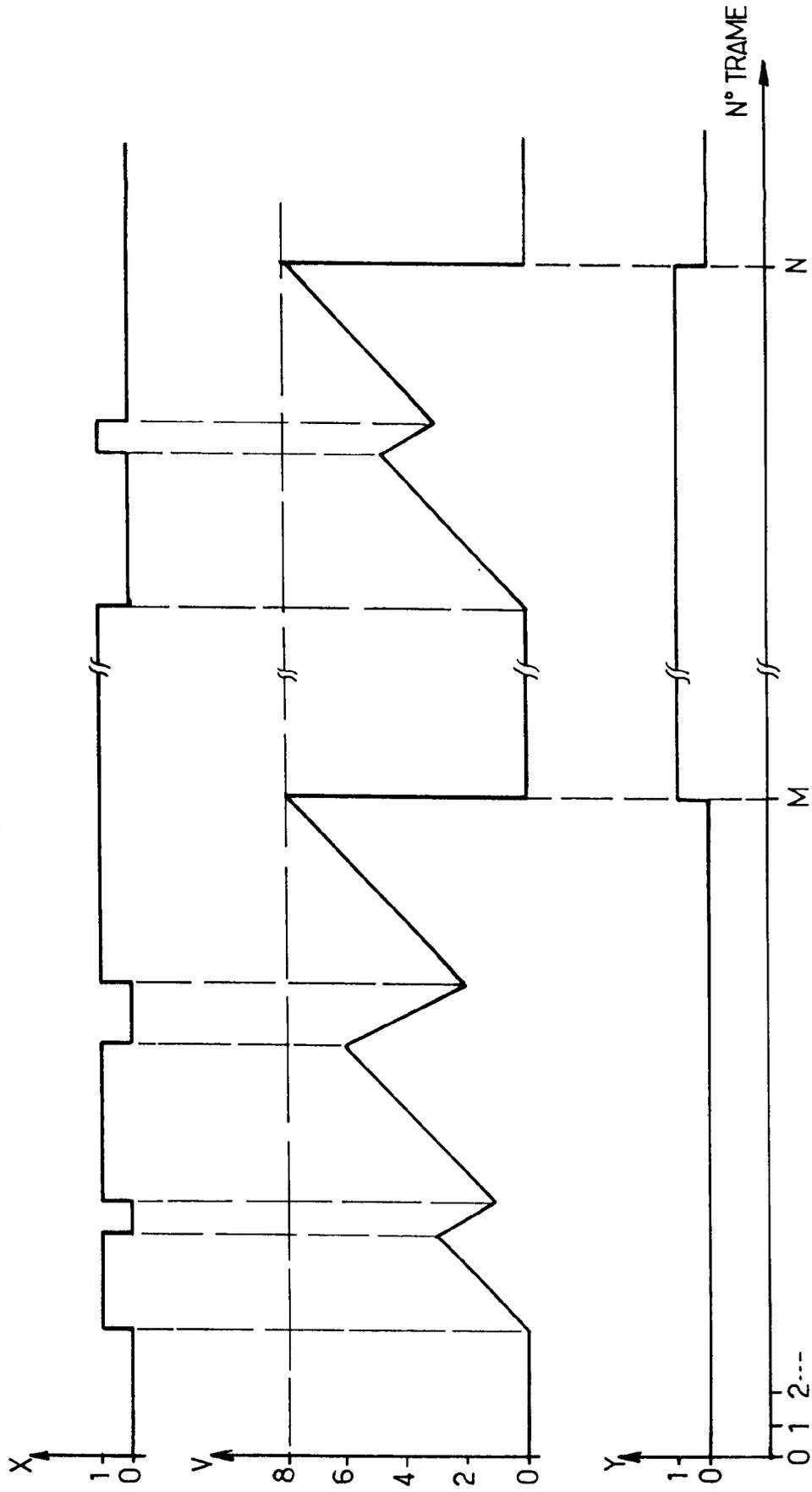


FIG.5.





Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande
EP 95 40 1262

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	INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 92, vol. 1, 14 Mai 1991 TORONTO, CA, pages 601-604, LIU 'A robust 440-bps speech coder against background noise' * alinéas 1., 4.1, 4.2, 5. * ---	1	G10L3/00
A	INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 93, vol. 2, 27 Avril 1993 MINNEAPOLIS, MN, US, pages 151-154, NISHIGUCHI ET AL. 'Vector quantized MBE with simplified V/UV division at 3.0 kbps' * page 153, alinéa 2.3 * ---	1	
D,A	INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 85, vol. 3, 26 Mars 1985 TAMPA, FL, US, pages 937-940, SCHROEDER ET AL. 'Code-excited linear prediction (CELP): high-quality speech at very low bit rates' * le document en entier * -----	1	
			DOMAINES TECHNIQUES RECHERCHES (Int.Cl.6)
			G10L
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche		Date d'achèvement de la recherche	Examineur
LA HAYE		29 Septembre 1995	Lange, J
CATEGORIE DES DOCUMENTS CITES		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	
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			

EPO FORM 1503 01.82 (P04C02)