



(12) **DEMANDE DE BREVET EUROPEEN**

(21) Numéro de dépôt : **94400525.5**

(51) Int. Cl.⁵ : **G10L 9/14**

(22) Date de dépôt : **10.03.94**

(30) Priorité : **12.03.93 FR 9302881**

(43) Date de publication de la demande :
21.09.94 Bulletin 94/38

(84) Etats contractants désignés :
DE GB

(71) Demandeur : **FRANCE TELECOM**
Etablissement autonome de droit public,
6, Place d'Alleray
F-75015 Paris (FR)

(72) Inventeur : **Massaloux, Dominique**
109, rue Saint Guirec,
Ploumanac'h
F-22700 Perros Guirec (FR)

(74) Mandataire : **Dubois-Chabert, Guy et al**
Société de Protection des Inventions
25, rue de Ponthieu
F-75008 Paris (FR)

(54) **Dispositif de codage et de décodage numérique de la parole, procédé d'exploration d'un dictionnaire pseudo-logarithmique de délais LTP, et procédé d'analyse LTP.**

(57) La présente invention concerne un dispositif de codage et décodage numériques de la parole comprenant, une prédiction à court terme (13), une prédiction à long terme (15) et une technique de codage de l'onde résiduelle utilisant une méthode d'analyse par synthèse (14). Le module d'analyse LTP utilise un dictionnaire de délais à structure pseudo-logarithmique dans lequel les délais sont rangés en ordre croissant; ce dictionnaire est constitué de segments, chacun d'une résolution donnée, les résolutions des segments successifs diminuant géométriquement dans un rapport k rationnel >1 , tandis que le nombre d'éléments de chaque segment reste constant. L'invention définit l'emploi des délais λ éléments de ce dictionnaire en étendant les techniques d'analyse LTP à haute résolution temporelle.

L'invention concerne également un procédé d'exploration rapide d'un tel dictionnaire pseudo-logarithmique de délais.

Elle concerne également un procédé de mise en place d'un critère de sélection du délai en boucle fermée avec filtrage perceptuel.

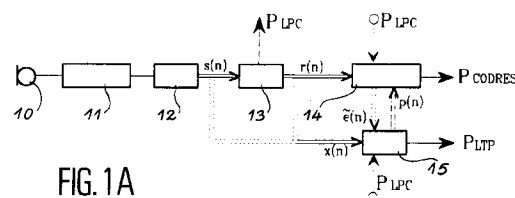


FIG. 1A

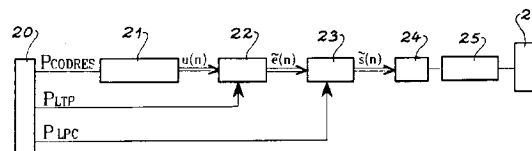


FIG. 1B

Domaine technique

La présente invention concerne un dispositif de codage et de décodage numériques de la parole, un procédé d'exploration d'un dictionnaire pseudo-logarithmique de délais LTP, et un procédé d'analyse LTP.

5

Etat de la technique antérieure

De manière connue, un dispositif de codage numérique de la parole consiste, après échantillonnage du signal analogique, à effectuer la compression des données binaires du signal de parole numérisé. Le dispositif de décodage effectue l'opération inverse et restitue un signal analogique différent du signal original, mais le plus voisin possible d'un point de vue perceptuel.

Un dispositif de codage-décodage numérique de la parole est caractérisé par le débit numérique des données à transmettre entre le codeur et le décodeur, la qualité du signal reconstitué au décodeur, et la complexité de la technique de compression mise en oeuvre.

Les codeurs prédictifs sont utilisés pour des débits assez faibles (de 4 à 16 kbits/s pour une fréquence d'échantillonnage de 8 kHz) et une bonne qualité de codage.

Ils intègrent des propriétés du signal de parole liées à sa production et d'autres liées à sa perception par un auditeur humain:

- Stationnarité locale du signal de parole: le signal de parole peut être prédit à partir de son passé récent (de 8 à 12 échantillons à 8 kHz) au moyen de paramètres évalués sur des fenêtres de 10 à 20 ms. Ces paramètres de prédiction à court terme, représentatifs de la fonction de transfert du conduit vocal, sont obtenus par des méthodes d'analyse "LPC" (Linear Prediction Coding).
- Périodicité des sons voisés (par exemple: les voyelles): cette corrélation à plus long terme est due à la vibration des cordes vocales. La cadence de vibration (fréquence fondamentale) varie de 60 à 400 Hz selon les locuteurs. Une analyse "LTP" (Long Term Prediction) permet d'évaluer les paramètres d'un prédicteur à long terme exploitant cette particularité.
- Masquage du bruit par le signal: dans les fréquences voisines d'un maximum d'énergie du signal, l'oreille est moins sensible au bruit de codage. Cette propriété est exploitée par l'introduction d'un "filtre perceptuel" au codage de l'onde résiduelle issue des prédicteurs à court et long terme, et éventuellement à l'analyse LTP. Ce filtre permet de redistribuer le bruit dans les zones fréquentielles où il est masqué par le signal.

De manière classique, un codeur prédictif est composé d'un module de prédiction à court terme, un module de prédiction à long terme, puis un module effectuant le codage de l'onde résiduelle à l'aide d'une méthode d'analyse par synthèse, ainsi que décrit dans l'article de P. Kroon et B.S. Atal intitulé "Predictive Coding of Speech Using Analysis by Synthesis Techniques" (Advances in Speech Signal Processing, Ed. Furui S., Sondhi M.M., pages 141-164, 1991).

Selon le type de codage de l'onde résiduelle, on distingue plusieurs familles de codeurs: codeurs APC, Multipulse-Excited, CELP, ..., comme décrit dans l'article de P. Kroon et B.S. Atal.

Ce type de dispositif de codage est largement utilisé, essentiellement dans des systèmes de transmission par voie terrestre ou satellite, ou dans des applications de stockage.

Différentes réalisations du module de prédiction à long terme ou module LTP, connues de l'homme de l'art, vont être passées en revue à présent.

La forme générale d'un prédicteur à long terme d'ordre p est:

45

$$P(Z) = 1 - \sum_{k=0}^{p-1} \beta(k) z^{-(\lambda+k)}$$

Le nombre p de coefficients de ce prédicteur varie généralement de 1 à 3. Si on considère le cas des prédicteurs du 1er ordre : $P(Z) = 1 - \beta z^{-\lambda}$.

A l'analyse, les paramètres β et λ sont déterminés par minimisation de l'énergie d'un signal d'erreur $e(n)$ sur un bloc de N échantillons du signal $x(n)$:

55

$$\text{Min} E(\lambda) = \sum_{n=0}^{N-1} e(n)^2 \quad \text{avec} \quad e(n) = x(n) - \beta x(n - \lambda) \quad (1)$$

$x(n)$ représente le signal d'entrée lui-même $s(n)$ ou le résidu LPC $r(n)$. Cette analyse dite "en boucle ouverte" est décrite dans l'article de B.S. Atal intitulé "Predictive Coding of Speech at Low Bit Rates" (IEEE Trans. Commun., COM-30, pages 600-614, avril 1982).

Ce type d'analyse peut être avantageusement remplacé par une analyse en boucle fermée, anticipant l'opération effectuée au décodeur pour produire le signal de synthèse $\tilde{s}(n)$.

A la synthèse on obtient:

$$\tilde{s}(z) = \frac{1}{A(z)} \cdot \frac{1}{P(z)} \cdot u(z)$$

(avec $u(n)$ = signal d'excitation)

Si $\tilde{e}(z) = \frac{1}{P(z)} \cdot u(z)$, alors $\tilde{e}(n) = u(n) + \beta \tilde{e}(n-\lambda)$ représente le signal résiduel reconstruit, ou excitation de synthèse du filtre LPC $1/A(z)$

La modélisation du résidu $r(n)$ par le signal $\tilde{e}(n)$ s'améliore lorsque l'on remplace le signal d'erreur $e(n)$ de l'équation (1) par:

$$e(n) = r(n) - \beta \tilde{e}(n - \lambda) \quad (2)$$

comme par exemple le codeur "RPELTP" décrit dans l'article de P. Vary, K. Hellwig, C. Galand, M. Rosso, JP. Petit, D. Massaloux intitulé "Speech Codec for the European Mobile Radio System" (Globecom, pages 1065-1069, 1986).

Le prédicteur à long terme décrit dans l'article de W.B. Kleijn, D.J. Krasinski et R.H. Ketchum intitulé "An Efficient Stochastically Excited Linear Predictive Coding Algorithm for High Quality Low Bit Rate Transmission of Speech" (Speech Commun., vol VII, pages 305-316, 1988) adopte une philosophie "CELP" pour une analyse LTP effectuée également en boucle fermée. A chaque période est associée une forme d'onde $u_\lambda = \tilde{e}(n-\lambda)$, $n = 0 \rightarrow N-1$ dans un dictionnaire CELP. Ce dictionnaire mis à jour à chaque analyse LTP, est appelé "dictionnaire adaptatif". L'analyse LTP est remplacée par la recherche du code optimal dans le dictionnaire adaptatif, résolue par les équations classiques du CELP, ce qui revient à remplacer $e(n)$ dans l'équation (1) et l'équation (2) par :

$$e(n) = h_g(n) * (t(n) - \beta u_\lambda(n)), n = 0 \rightarrow N-1$$

avec $h_g(n) = r.i.$ du filtre perceptuel $H_g(z) = \frac{A_{\gamma_1}(z)}{A_{\gamma_2}(z)A(z)}$ (lorsque l'on choisit $\gamma_1 = 1$, $\gamma_2 = g$, $H_g(z) = \frac{1}{A_g(z)}$)

Le signal $t(n)$ ("target") s'exprime à partir du résidu LPC $r(n)$ et du signal $\tilde{e}_p(n)$ obtenu en prolongeant l'excitation passée $\tilde{e}(n)$ par des échantillons nuls :

$$t(n) = r(n) - \tilde{e}_p(n) \quad n = 0 \rightarrow N-1 \quad \text{avec} \quad \tilde{e}_p(n) = \begin{cases} \tilde{e}(n) & n < 0 \\ 0 & 0 \leq n < N \end{cases}$$

On obtient alors pour $e(n)$ l'expression :

$$e(n) = h_g(n) * (r(n) - \tilde{e}_p(n) - \beta u_\lambda(n)) \quad (3)$$

essentiellement différente de l'équation (2) par l'introduction du filtre perceptuel et de sa mémoire.

D'autre part, les analyses en boucle fermée utilisent le signal $\tilde{e}(n)$ qui n'est connu, au début du bloc analysé, que pour $n < 0$, ce qui impose de restreindre l'analyse LTP aux valeurs $\lambda \geq N$. Cette restriction diminue l'efficacité d'un prédicteur à long terme sur les voix à fréquence fondamentale élevée (voix de femmes et d'enfants). On peut y remédier en extrapolant le signal $\tilde{e}(n)$ pour $n \geq 0$. Dans l'article de W.B. Kleijn, D.J. Krasinski et R.H. Ketchum cité précédemment, on utilise la périodicité supposée du signal pour chaque période candidate λ en remplaçant $\tilde{e}(n), n \geq 0$ par $\tilde{e}(n - \lambda)$ si $n < \lambda$ (ou $\tilde{e}(n - k\lambda)$ avec k = plus petit entier pour lequel $n < k\lambda$). Cependant, pour chaque période $\lambda < N$, il faut compléter $\tilde{e}$ avec $N - \lambda$ valeurs, ce qui accroît la complexité de l'analyse LTP.

Un certain nombre d'algorithmes rapides décrits dans l'article de W.B. Kleijn, D.J. Krasinski et R.H. Ketchum intitulé "Fast Methods for the CELP Speech Coding Algorithm", (IEEE Trans. on ASSP, vol. 38, No 8, pages 1330-1341, Août 1990) ont été conçus pour accélérer les calculs dans le prédicteur à long terme, principalement dans le cadre de l'analyse par dictionnaire adaptatif, fondamentalement plus complexe. Ces algo-

5 rithmes sont généralement perturbés par l'introduction d'éléments extrapolés de $\tilde{e}(n)$.

Un dernier point concerne la précision du prédicteur à long terme: pour un prédicteur d'ordre 1 avec délais λ entiers, la périodicité T recherchée est limitée aux multiples de la période d'échantillonnage T_e . Deux méthodes ont été proposées permettant d'améliorer la précision sur T :

- augmenter l'ordre du prédicteur, ce qui bien sûr augmente la complexité de l'analyse, mais augmente aussi le nombre des gains à coder;
- utiliser un prédicteur à "haute résolution temporelle", comme décrit dans l'article de P. Kroon et B.S. Atal intitulé "Pitch Predictors with High Temporal Resolution" (Proc. ICASSP, pages 661-664, avril 1990). Cette technique fait intervenir des délais fractionnaires du type $\lambda + \varphi/D$ avec $\lambda \in N$, $\varphi = 0, 1, \dots, D-1$, par interpolation du signal passé analysé. L'interpolation est effectuée par suréchantillonnage suivi d'un filtrage passe-bas. Cette opération peut être efficacement mise en place au moyen d'une structure polyphasée, telle que décrite dans l'article de R.E. Crochiere et L.R. Rabiner intitulé "Interpolation and Decimation of Digital Signals: A Tutorial Review" ("Proc. of the IEEE" vol. 69, No 3, mars 1981).

Le problème de la réunion des techniques d'extrapolation du signal $\tilde{e}(n)$ et de prédiction à haute résolution temporelle est résolu par un processus récursif coûteux en complexité décrit dans la demande de brevet WO 91:03790 de I.A. Gerson, M.A. Jasiuk intitulée "Digital Speech Coder Having Improved Sub-Sample Resolution Long Term Predictor": pour chaque période fractionnaire $\lambda + \varphi/D$, les échantillons $\tilde{e}(n)$, $n \geq 0$ inconnus sont remplacés récursivement par des échantillons issus d'une interpolation du signal passé $\tilde{e}(n)$, $n < 0$.

L'invention a pour objet un dispositif numérique de codage et de décodage de la parole dans lequel on améliore le fonctionnement du module de prédiction à long terme tel que défini dans ces différents documents de l'art antérieur.

Exposé de l'invention

L'invention propose, à cet effet, un dispositif de codage et de décodage numériques de la parole comprenant, au codage: un module d'analyse LPC (prédiction à court terme), un module d'analyse LTP (prédiction à long terme), un module de codage de l'onde résiduelle utilisant une méthode d'analyse par synthèse; au décodage: un module de décodage de l'onde résiduelle, un module de synthèse LTP et un module de synthèse LPC; caractérisé en ce que le module d'analyse LTP utilise un dictionnaire de délais à structure pseudo logarithmique dans lequel les délais sont rangés en ordre croissant; ce dictionnaire étant constitué de Q segments adjacents, chacun d'une résolution donnée, les résolutions des segments successifs diminuant géométriquement dans un rapport k rationnel tel que $k > 1$ tandis que le nombre d'éléments L de chaque segment reste constant.

L'intérêt de ces précisions emboîtées est de maintenir à peu près constante la précision relative sur le délai, et de là l'erreur sur la périodicité du signal due à l'échantillonnage. L'invention permet de plus d'obtenir un codage simple et efficace du délai.

Les résolutions des délais dans les différents segments du dictionnaire pseudo-logarithmique sont des rationnels $R = p/q$, $p \in N$, $q \in N$ (N : ensemble des entiers naturels).

On a pour cela étendu les méthodes d'analyse à haute résolution temporelle évoquées (délais $\lambda = \lambda_1/R$ avec $\lambda_1 \in N$, $R \in N$) au cas des résolutions fractionnaires (délais $\lambda = \lambda_1 \times q/p$, $\lambda_1, q, p \in N$).

Avantageusement dans une première variante, le dictionnaire des délais est découpé en Q segments adjacents S_i ($i = 0 \rightarrow Q - 1$) comprenant chacun L délais. A chaque segment S_i correspond une résolution R_i , les résolutions des segments successifs décroissant dans un rapport k rationnel donné ($R_i = R_{i-1}/k$). Si l'on appelle γ_i le dernier délai du segment S_i , ce segment est formé des L délais $\lambda_j = \gamma_i - j/R_i$, $j = L-1 \rightarrow 0$ avec λ_j, R_i entiers. La condition d'adjacence entre segments est assurée par $\gamma_{i-1} = \gamma_i - L/R_i$, $i = 1 \rightarrow Q-1$. Si l'on introduit $\lambda_{\max}$ = dernier délai du dictionnaire et R_{Q-1} = résolution du dernier segment, on montre qu'un tel dictionnaire est entièrement défini par la donnée des valeurs $\{Q, L, k, \lambda_{\max}, R_{Q-1}\}$ et la condition $R_{Q-1} \cdot \lambda_{\max} \in N$.

Dans une seconde variante, le dictionnaire des délais est découpé en Q segments adjacents S_i ($i = 0 \rightarrow Q - 1$) comprenant chacun L délais. A chaque segment S_i correspond une résolution R_i , les résolutions des segments successifs décroissant dans un rapport k rationnel donné ($R_i = R_{i-1}/k$). Si l'on appelle β_i le premier délai du segment S_i , ce segment est formé des L délais: $\lambda_j = \beta_i + j/R_i$, $j = 0 \rightarrow L-1$ avec λ_j, R_i entiers. La condition d'adjacence entre segments est assurée par $\beta_i = \beta_{i-1} + L/R_{i-1}$, $i = 1 \rightarrow Q - 1$. Si l'on introduit β_{Q-1} = 1er délai du dernier segment et R_{Q-1} = résolution du dernier segment, on montre qu'un tel dictionnaire est entièrement

défini par la donnée des valeurs $\{Q, L, k, \beta_{Q-1}, R_{Q-1}\}$ et la condition $R_{Q-1} \cdot \beta_{Q-1} \in \mathbb{N}$.

Avantageusement, le dispositif permet un codage du délai LTP simple et peu coûteux en stockage, du type,

- selon la première variante:

$$\text{code}(\lambda_j) = L \cdot i + j,$$

5 avec

$$S_j = \{\lambda_j = \gamma_i - j/R_i, j = L - 1 \rightarrow 0\}$$

et

$$j = L - 1 - i$$

- selon la seconde variante:

10

$$\text{code}(\lambda_j) = L \cdot i + j$$

avec

$$S_i = \{\lambda_j = \beta_i + j/R_i, j = 0 \rightarrow L - 1\}$$

Avantageusement, une réalisation particulière de dictionnaire pseudo-logarithmique de délais tel que définis ci-dessus est le dictionnaire D, formé de délais fractionnaires, de résolution $R = p > 1$, ou entiers, que l'on peut décrire de la façon suivante: chaque segment $S_i, i = 0 \rightarrow 3$ de résolution $R_i = 2^{3-i}$ est formé des délais

15

$\lambda_{0-3}/R_i, i = 0 \rightarrow R_i - 1$, les délais λ_0 entiers formant un sous-ensemble S_i^0 de S_i comportant $n_i = 2^{i+3}$ éléments:

20

$$S_i^0 = \left\{ \lambda_0 = \mu_i + j/R_i, j = 0 \rightarrow n_i - 1, \text{ avec } \mu_i = \lambda_{\min}^0 + \sum_{j=0}^{i-1} n_j \right\} \text{ (en posant } n_{-1} = 0)$$

Avantageusement, une procédure sous-optimale efficace d'exploration d'un dictionnaire pseudo-logarithmique des délais tel que défini dans la première ou la seconde variante de l'invention, tirant parti de sa structure particulière, permet de réduire considérablement la complexité de la recherche du meilleur délai :

25

- dans une première passe, on sélectionne $K(i)$ maxima locaux du critère à maximiser parmi un ensemble restreint de $\alpha(i)$ délais de chaque segment S_i ;

- dans une seconde passe, on explore le dictionnaire de manière limitée au voisinage des valeurs sélectionnées dans la première passe.

30

Avantageusement la taille des segments L est un multiple de K^{L-1} , le choix pour $\alpha(0)$ de L/K^{L-1} ou d'un sous-multiple de L/K^{L-1} , introduisant un espacement régulier des délais explorés dans la première passe.

Avantageusement, on introduit une simplification supplémentaire à la recherche de la première passe en remplaçant la maximisation de $E'(\lambda) = N(\lambda)^2 / D(\lambda)$, où $N(\lambda)$ et $D(\lambda)$ représentent respectivement le numérateur et le dénominateur du gain optimal associé à chaque délai λ , par celle de $N(\lambda)$: on calcule donc les maxima locaux de l'intercorrélation $N(\lambda)$, pour tous les segments $i = 0 \rightarrow Q - 1$, dans la première passe.

35

L'invention propose également un procédé d'analyse LTP en boucle fermée avec filtrage perceptuel de performances équivalentes à l'analyse LTP par dictionnaire adaptatif et de complexité moindre, reposant sur l'expression suivante du signal d'erreur dont on minimise l'énergie:

40

$$e(n) = h_g(n) * (r(n) - \beta e(n - \lambda))$$

les points précédant le sous-bloc courant (tels que $n < 0$ si le sous-bloc courant commence à $n = 0$) étant les points $\tilde{e}(n - \lambda)$ (λ éventuellement fractionnaire, $\tilde{e}$ éventuellement extrapolé), et non $\tilde{e}(n)$, comme dans le cas du dictionnaire adaptatif.

45

La présente invention permet ainsi de définir une structure sur l'ensemble des délais explorés dans le module de prédiction à long terme, l'ensemble de délais ainsi structuré étant dénommé dans l'invention "dictionnaire pseudo-logarithmique de délais LTP". Il est connu que maintenir une grande précision sur les délais LTP, lorsque ces délais augmentent, est inutile d'un point de vue perceptuel. Le dictionnaire pseudo-logarithmique de l'invention exploite cette idée et permet de maintenir les performances de dictionnaires uniformes pour un débit moindre: à titre d'exemple, on a pu constater que les performances du dictionnaire D, composé de 256 éléments, étaient similaires à celles de l'ensemble des 960 délais obtenus en échantillonnant uniformément la même gamme de délais avec une précision de 1/8, ce qui représente un gain de plus de 20 % en débit.

50

La structure pseudo-logarithmique, outre qu'elle organise le concept précédemment énoncé, permet également d'établir une correspondance simple entre l'indice de chaque délai du dictionnaire pseudo-logarithmique et sa valeur, facilitant les opérations de codage et décodage du délai. Aucun stockage n'est nécessaire pour trouver les délais dans le dictionnaire.

55

Cette structure facilite de plus la conception d'un tel dictionnaire: un tel dictionnaire est en effet totalement défini par la donnée de quelques paramètres. Pour une application donnée, le choix de ces paramètres est gouverné par les contraintes de l'application. Il est alors simple de déterminer le ou les dictionnaire(s) pseudo-logarithmique(s) convenant à cette application.

La présente invention décrit en outre un procédé très peu complexe permettant la mise en place d'un module d'exploration d'un tel dictionnaire. Bien que sous optimale, une telle technique a montré des performances équivalentes à la recherche optimale. La réduction de complexité obtenue avec ce procédé est importante. Si l'on compare les temps de calcul dans un codeur de type CELP des deux techniques suivantes:

- technique de référence : analyse LTP par "codebook" adaptatif avec sélection du délai optimal par méthode d'autocorrélation comme définie dans l'article de messieurs Kleijn, Krasinski et Ketchum, intitulé "Fast Methods for the CELP Speech Coding Algorithm", cité plus haut;
- technique proposée dans l'invention: analyse LTP du type présenté selon la procédure sous optimale.

Bien que ne produisant pas les mêmes résultats, ces deux techniques ont été jugées d'une qualité subjective équivalente.

Sur un micro-ordinateur, le traitement du module LTP utilisant la technique proposée dans l'invention est trois fois plus rapide que celui du module utilisant une version optimisée de la technique de référence. Cette version optimisée exploite au maximum les méthodes permettant de diminuer la complexité de la technique de référence: si l'on compare les temps de calcul de la version non optimisée de la technique de référence à ceux de la technique proposée, on obtient un gain supérieur à 11.

Brève description des dessins

- Les figures 1A et 1B représentent le dispositif de codage et le dispositif de décodage de la parole selon l'invention; -la figure 2 représente une réalisation particulièrement intéressante du dispositif de codage représenté à la figure 1A ;
- la figure 3 illustre le fonctionnement d'un dictionnaire pseudo-logarithmique de délais ;
- la figure 4 illustre la procédure de calcul du signal $x(n-\lambda)$, λ rationnel intervenant dans le module LTP ;
- la figure 5 montre, sur une séquence de parole réelle, l'évolution du critère $E'(\lambda)$ lorsque parcourt le dictionnaire D ;
- la figure 6 détaille le dictionnaire D ;
- la figure 7 présente une procédure de codage et décodage des délais du dictionnaire D ;
- la figure 8 décrit les modules de calcul du signal $\tilde{e}_w(n-\lambda)$ intervenant dans la recherche du délai optimal de D ;
- les figures 9 à 12 montrent le fonctionnement de cette recherche du délai dans le cadre de la réalisation du module LTP.

Exposé détaillé des modes de réalisation

La présente invention concerne un dispositif numérique de codage de la parole de type codeur prédictif utilisant une prédiction à court terme du signal permettant la modélisation des formants, une prédiction à long terme destinée à restituer la structure fine du spectre, puis un codage de l'onde résiduelle à l'aide d'une méthode d'analyse par synthèse; une description générale de ce type de codeurs étant réalisée dans l'article de messieurs Kroon et Atal cité précédemment. Les prédicteurs à court terme et à long terme sont calculés par des méthodes de prédiction linéaire connues sous le nom d'analyse LPC ("Linear Prediction Coding") et analyse LTP ("Long Term Prédiction")

Les figures 1A et 1B représentent un dispositif numérique de codage et un dispositif numérique de décodage de la parole faisant l'objet de la présente invention. Le dispositif de codage comprend successivement:

- un capteur 10;
- un filtre 11 ;
- un convertisseur analogique/numérique 12 ;
- un module LPC 13;
- un module de codage de résidu ou CODRES 14 ;
- un module LTP 15 recevant en entrée le signal d'entrée ou le signal de sortie du module LPC 13: $x(n)$ = $s(n)$ ou $r(n)$, et éventuellement le signal $\tilde{e}(n)$, résiduel reconstruit issu du module CODRES 14.

Ce dispositif de codage fonctionne de la manière suivante:

Le signal analogique, après conversion en numérique, est segmenté en trames de N_0 échantillons $s(n)$. Ces échantillons sont analysés dans le module LPC 13 par une méthode classique de prédiction linéaire. Ce

module 13 produit en sortie des paramètres PLPC transmis au décodeur et N_0 échantillons de signal résiduel $r(n)$.

Ensuite, le module LTP 15 accepte en entrée N échantillons d'un signal $x(n)$ qui peut provenir d'une sous-segmentation du signal $s(n)$ lui-même ou bien de $r(n)$. Lorsque le module LPT 15 fonctionne en boucle fermée, il doit recevoir également en entrée des échantillons de résiduel reconstruit (ou "excitation de synthèse") issus par bouclage du module de codage de résidu 14. Le module LTP peut éventuellement utiliser aussi les paramètres PLPC (dictionnaire adaptatif, filtre perceptuel). Ce module 15 produit les paramètres de sortie PLTP (gain quantifié β et indice i_d du délai) et élabore un signal de prédiction à long terme $p(n)$.

Ensuite, le module de codage de résidu 14 effectue le codage de l'excitation résiduelle. Les paramètres de codage de cette excitation sont transmis au décodeur. Lorsque nécessaire, ce module 14 comprend un décodeur local permettant le calcul de l'excitation de synthèse (ou résiduel reconstruit) $\tilde{e}(n)$.

La figure 1B présente le schéma du dispositif de décodage correspondant au dispositif de codage représenté sur la figure 1A. Ce dispositif de décodage comprend successivement:

- un module de démultiplexage 20;
- un module de décodage de résidu ou CODRES⁻¹ 21;
- un module de synthèse LTP (ou LTP⁻¹) 22;
- un module de synthèse LPC (ou LPC⁻¹) 23;
- un convertisseur numérique analogique 24;
- un filtre 25;
- un haut-parleur 26.

Le module de décodage de résidu 21 effectue le décodage des paramètres P_{CODRES} et calcule N échantillons d'un signal $u(n)$. Ce signal entre dans le module 22 conjointement avec les paramètres P_{LTP} qui y seront décodés. Après filtrage de $u(n)$ par $1/P(z)$, on obtient $\tilde{e}(n)$.

Ce signal entre alors dans le module 23 qui effectue le décodage des paramètres P_{LPC} et le filtrage de $\tilde{e}(n)$ par $1/A(z)$. Ce module 23 produit en sortie les N_0 échantillons du signal de synthèse $\tilde{s}(n)$, pour une trame, qui sont convertis en analogique.

Diverses réalisations du dispositif de l'invention étant possibles, on va considérer à présent une réalisation particulièrement intéressante à titre d'exemple.

Cette réalisation représentée à la figure 2, donnée à titre d'exemple, a les particularités suivantes:

- l'analyse LTP (module 13), qui sera décrite plus en détail ultérieurement, est une analyse en boucle fermée, utilisant les signaux $r(n)$ et $\tilde{e}(n)$ en entrée, avec un filtre perceptuel calculé à partir des paramètres P_{LPC} fournis par le module LPC;
- codage de l'excitation résiduelle: les signaux $r(n)$, $p(n)$ et $\tilde{e}(n)$ entrent dans un module 14 de type CELP qui utilise une procédure classique de recherche dans un dictionnaire CELP pour quantifier le signal résiduel tel que décrit dans l'article de B.S. Atal cité précédemment. Un tel dictionnaire est formé par exemple de NF formes d'ondes aléatoires de statistiques gaussiennes. Les paramètres P_{LPC} entrant dans le module CELP 14' permettent de calculer le filtre perceptuel $W(z) = A(z)/A_v(z)$, ($\gamma = 0.75$).

Après sélection de la meilleure forme d'onde du dictionnaire, le module 14' produit les paramètres PCELP (gain quantifié et indice i_c de la forme d'onde), et le signal résiduel reconstruit $\tilde{e}(n) = p(n) + \gamma u_{ic}(n)$.

Pour une fréquence d'échantillonnage de 8kHz, la présente réalisation du dispositif effectue un codage du signal de parole au débit de 8 kbit/s, avec les caractéristiques suivantes:

Trame LPC : 24 ms ($N = 192$)

Sous trames : 4 ms ($N_0 = 32$)

Débit LPC : 42 bits / trame (ordre 10)

Débit LTP : $\left. \begin{array}{l} i_d : 8 \text{ bits} \\ \beta : 3 \text{ bits} \end{array} \right\} 11 \times 6 \text{ bits / trame}$

Excitation : facteur d'échelle : 6 bits / trame

$\left. \begin{array}{l} \text{indice CELP } i_c : 10 \text{ bits} \\ \text{gain } \gamma : 3 \text{ bits} \\ (N_F = 1024) \end{array} \right\} 13 \times 6 \text{ bits / trame}$

La présente invention se situe au niveau du module LTP dont le fonctionnement va être décrit à présent.
Le module d'analyse LTP de l'invention repose sur l'exploration d'un dictionnaire de délais de type pseudo-logarithmique.

Un module d'analyse LTP d'ordre 1, quel que soit le type d'analyse, calcule le délai du prédicteur $P(z)$ qui minimise un certain critère d'erreur. La présente invention regroupe l'ensemble des délais explorés en un dictionnaire présentant une structure pseudo-logarithmique. Ces délais λ sont des nombres rationnels, rangés en ordre croissant dans le dictionnaire.

Le dictionnaire est découpé en Q segments adjacents $S_i (i = 0 \rightarrow Q-1)$ comprenant chacun L délais. A chaque segment S_i correspond une résolution R_i , et si on appelle γ_i le dernier délai du segment S_i , le segment S_i est formé de la façon suivante, comme représenté sur les figures 3A et 3B:

$$S_i = \{\lambda_j = \gamma_i - j / R_i, j = L-1 \rightarrow 0\} \quad (4)$$

Le délai γ_i peut éventuellement être fractionnaire mais les délais λ_j doivent vérifier $\lambda_j \cdot R_i$ entier $\forall j$, donc pour chaque segment S_i , il faut et il suffit que $\gamma_i \cdot R_i$ soit entier.

Les résolutions des segments successifs décroissent dans un rapport k rationnel donné:

$$R_i = R_{i-1} / k, i \rightarrow Q-1 \quad (5)$$

La condition d'adjacente entre segments (figure 3B) est assurée par:

$$\gamma_{i-1} = \gamma_i - L / R_i, i = 1 \rightarrow Q-1 \quad (6)$$

Si l'on appelle $\lambda_{\max}$ le dernier délai du dictionnaire ($\lambda_{\max} = \lambda_{Q-1}$), on montre que la condition $\gamma_i \cdot R_i \in \mathbb{N}$ est réalisée pour tout $i=0$ à $Q-1$ si et seulement si :

$$R_{Q-1} \cdot \lambda_{\max} \in \mathbb{N} \quad (7)$$

Le dictionnaire est alors totalement défini par la donnée des valeurs $\{Q=\text{nombre de segments}, L=\text{taille des segments}, k=\text{facteur de décroissance des résolutions}, \lambda_{\max}=\text{dernier délai du dictionnaire}, R_{Q-1}=\text{résolution du dernier segment telle que l'équation (7) soit vérifiée}\}$

On peut alors calculer $\lambda_{\min}$ (premier délai du dictionnaire) par la formule :

$$\lambda_{\min} = \lambda_{\max} - \left(L \frac{k^Q - 1}{k - 1} - 1 \right) / R_0,$$

et si on définit la longueur l_i des segments S_i comme $l_i = \gamma_i - \gamma_{i-1}$, on a alors (figure 3B):

$$l_i = k \cdot l_{i-1}, i = 1 \rightarrow Q-1 \quad (8)$$

La structure pseudo-logarithmique en base k du dictionnaire de délais apparaît dans les équations (5) et (8).

On peut construire un dictionnaire du même type en s'appuyant sur les premiers délais β_i de chaque segment:

$$S_i = \{\lambda_j = \beta_i + j / R_i, j = 0 \rightarrow L-1\}, \quad (4')$$

et en définissant la condition d'adjacence par (figure 3C) :

$$\beta_i = \beta_{i-1} + L / R_{i-1} \quad (6')$$

Il faut alors remplacer la donnée de $\lambda_{\max}$ par celle de β_{Q-1} = premier délai du dernier segment, et la condition (7) par:

$$R_{Q-1} \cdot \beta_{Q-1} \in \mathbb{N} \quad (7')$$

Bien que légèrement différent, ce dictionnaire est totalement équivalent à celui décrit en figure 3B.

Ces dictionnaires pseudo-logarithmiques de délais permettent un codage du délai simple et peu coûteux en stockage, du type :

$$\text{code}(\lambda_j) = L \cdot i + j',$$

$$\text{avec } (\lambda_j = \gamma_i + j / R_i) \in S_i \text{ (voir equation (4)) et } j' = L-1-j$$

pour un dictionnaire défini par les équations (4), (6) et (7).

Un codage du même type est réalisable pour un dictionnaire défini par les équations (4'), (6') et (7').

On va considérer ci-dessous un exemple donné de dictionnaire qui sera repris dans la suite de la description. Il constitue une réalisation particulièrement intéressante de la présente invention.

D=dictionnaire à 256 délais (8 bits) tel que:

$$\left\{ \begin{array}{l} Q = 4 \\ L = 64 \\ k = 2 \\ \lambda_{\max} = \lambda_{\min}^0 + 119 \text{ (avec } \lambda_{\min}^0 = \text{plus petit délai entier)} \\ R_3 = 1 \end{array} \right.$$

Tous les types d'analyse LTP utilisent un critère à minimiser qui fait intervenir un signal $x(n-\lambda)$ pour un certain délai λ et $n=0$ à $N-1$ (en boucle ouverte, $x(n)$ représente $s(n)$ ou $r(n)$, et en boucle fermée $\tilde{e}(n)$).

On va tout d'abord définir ce signal $x(n-\lambda)$ dans le cas particulier où le délai λ est un rationnel:

En effet lorsque λ appartient au dictionnaire défini précédemment, il est de la forme $\lambda = \lambda_1/R$ tel que $\lambda_1 \in \mathbb{N}$, R rationnel. R (résolution du segment qui contient λ) est un rationnel a priori quelconque, du type $R = p/q$, $p \in \mathbb{N}$ et $q \in \mathbb{N}$.

On définit $x(n-\lambda)$, $n = 0 \rightarrow N-1$ en étendant la technique décrite par P. Kroon au cas d'une résolution $R = p/q$ rationnelle. On passe du signal $x(n)$ au signal $y(n)$ de résolution multipliée par $x(p/q)$ à l'aide de méthodes classiques d'interpolation de signal comme décrit dans l'article de messieurs Crochière et Rabiner cité plus haut.

Comme montré à la figure 4, le signal $x(n)$ est d'abord suréchantillonné d'un facteur p dans un suréchantillonneur 30, produisant un signal $x'(n)$ qui entre dans un filtre $H(z)$ passe-bas 31 dont la fréquence de coupure est inférieure à $f_{\max}/\text{Max}(p,q)$ ($f_{\max} = f_{\text{échant}}/2$). Le signal $x''(n)$ issu de ce filtrage est ensuite sous-échantillonné d'un facteur q , dans un sous-échantillonneur 32 pour donner $y(n)$.

On a donc:

$$y(n) = x''(nq) \text{ avec } x''(n) = \sum_{i=-I}^{i=I} h(i)x'(n-i)$$

si

$$H(z) = \sum_{i=-I}^{i=I} h(i)z^{-i}.$$

On peut aussi exprimer $x''(n)$ par

$$x''(n) = \sum_{i=-I/p}^{i=I/p} h(jp + \varphi)x(k-j)$$

si

$$k = E(n/p), n \equiv \varphi[p].$$

(On considère la notation $E(x)$ =partie entière de x)

Pour un délai $\lambda = \lambda_1/R$ avec $\lambda_1 \in \mathbb{N}$, on définit $x(n-\lambda)$ par:

$$\begin{aligned} x(n-\lambda) &= y(nR - \lambda_1) \quad \forall n \in \{0, 1, \dots, N-1\} \\ &= y(np/q - \lambda_1) \quad \text{si } R = p/q \end{aligned}$$

Alors

$$x(n - \lambda) = x''(np - \lambda_1 q)$$

On voit qu'il est intéressant de calculer à partir de $(\lambda_1 q)$ les valeurs $\lambda_0 \in \mathbb{N}$ et $\varphi \in \{0, 1, \dots, p-1\}$ telles que $\lambda_1 q = \lambda_0 p - \varphi$:

5

$$\begin{cases} \lambda_0 = E \left[\frac{\lambda_1 q - 1}{p} \right] + 1 \\ \varphi = (p - 1) - \text{mod}(\lambda_1 q - 1, p) \end{cases}$$

[La notation $q = \text{mod}(p, n)$ signifie $q = \text{reste de } p \text{ modulo } n$]

Alors :

10

$$x(n - \lambda) = \sum_{j=-1/p}^{1/p} h(jp + \varphi) x(n - \lambda_0 - j), n = 0 \rightarrow N - 1 \quad (9)$$

15

En pratique, on choisit par exemple pour $H(z)$ un sinus cardinal fenêtré et échantillonné d'un facteur $\text{Max}(p, q)$. Les p filtres $\{h_\varphi(j), j = -lp \rightarrow lp\}$, $\varphi = 0 \rightarrow p - 1$ sont les filtres polyphasés construits à partir de $H(z)$.

Lorsque $p > q$, on a alors h_0 défini par $\{h_0(0) = 1, \text{ et } h_0(j) = 0 \text{ si } j \neq 0\}$ et donc pour les valeurs de λ entières on trouve pour $x(n - \lambda)$ le signal $x(n)$ décalé de λ points.

20

Pour $q = 1$, on retrouve l'expression donnée précédemment dans le cadre de l'analyse LTP à haute résolution.

On décrit ci-dessous le procédé de recherche du délai optimum dans le dictionnaire pseudo-logarithmique défini dans la présente invention.

25

Quel que soit le type d'analyse LTP, la recherche du délai optimal revient à minimiser un critère:

$$E(\lambda) = \sum_{n=0}^{N-1} e(n)^2$$

30

Si l'on définit de façon générale $e(n)$ comme : $e(n) = v(n) - \beta x(n - \lambda)$, $v(n)$ étant un signal connu indépendant de λ et $x(n - \lambda)$ défini pour chaque délai λ candidat, les expressions de ces deux signaux dépendant du type d'analyse utilisé, alors la minimisation de $E(\lambda)$ revient à maximiser:

35

$$E'(\lambda) = \sum_{n=0}^{N-1} v(n)^2 - E(\lambda)$$

40

La recherche du délai optimum nécessite le calcul, pour chaque délai λ , les deux quantités : 2

45

$$\begin{cases} N(\lambda) = \sum_{n=0}^{N-1} v(n) x(n - \lambda) \\ D(\lambda) = \sum_{n=0}^{N-1} x(n - \lambda)^2 \end{cases}$$

(n).

50

$N(\lambda)$ et $D(\lambda)$ représentent respectivement le numérateur et le dénominateur du gain optimum β associé à chaque délai λ . Ces deux quantités interviennent dans $E'(\lambda)$. Par exemple, lorsque β n'est pas quantifié en boucle, on a $E'(\lambda) = N(\lambda)^2 / D(\lambda)$.

Dans tous les cas, l'évaluation de $E'(\lambda)$ pour chaque délai λ est un processus nécessitant de nombreux calculs, en particulier lorsque l'on utilise des délais non entiers, et dans le cas d'analyses en boucle fermée, dès que l'on doit extrapoler le signal $\tilde{x}(n)$.

55

Diverses méthodes ont été proposées pour réduire la complexité de cette recherche:

- Analyse LTP à haute résolution : calcul des critères $E'(\lambda_0)$ tels que $\lambda_0 \in \mathbb{N}$ et interpolation des critères

comme décrit dans l'article de P. Kroon et B.S. Atal cité précédemment. Cette méthode est une méthode approchée et reste relativement complexe.

- Dictionnaire adaptatif: prolongement de la sommation dans $E'(\lambda)$ pour utiliser une méthode d'autocorrélation telle que définie dans l'article de A. Le Guyader, D. Massaloux and J.P. Petit intitulé "Robust and Fast Code Excited Linear Predictive Coding of Speech Signals" (Proc. ICASSP, pages 120-123, Mai 1989), "Backward Filtering" pour le calcul des numérateurs tel que défini dans l'article de I.M. Trancoso et B.S. Atal intitulé "Efficient Procedures for Finding the Optimum Innovation in Stochastic Coders" (Proc. ICASSP, pages 2375-2378, avril 1986), récurrence dans le calcul des dénominateurs, comme décrit dans l'article de W.B. Kleijn, D.J. Krasinski et R.H. Ketchum intitulé "An Efficient Stochastically Excited Linear Predictive Coding Algorithm for High Quality Low Bit Rate Transmission of Speech" cité précédemment.

Ces procédures sont cependant perturbées par l'introduction de signaux $\tilde{e}(n)$ extrapolés et se compliquent avec l'usage de délais fractionnaires.

Il est donc intéressant de simplifier encore cette procédure de recherche, et dans le cadre du dictionnaire de délais de l'invention, de s'appuyer pour cela sur sa structure particulière.

Si l'on étudie l'évolution du critère $E'(\lambda)$ pour λ variant dans un dictionnaire de délais de l'invention tel que défini précédemment, on constate que la courbe

$$E_n'(\lambda) \left(\text{avec } E_n'(\lambda) = E'(\lambda) / \sum_{n=0}^{N-1} v(n)^2 \right)$$

présente elle-même une structure pseudo-logarithmique et que ses maxima en sont relativement amortis: à titre d'exemple, la figure 5 montre l'évolution de $E_n'(\lambda)$ pour $\lambda \in$ dictionnaire D, sur une trame voisée d'un échantillon de parole.

Cette étude suggère le fractionnement de la recherche en deux passes:

- dans une première passe : dans chaque segment S_i , calcul du critère sur un nombre restreint $\alpha(i)$ de délais tel que $\forall i = 1 \rightarrow Q - 1, \alpha(i) = k\alpha(i-1)$. Sélection d'un certain nombre $K(i)$ de maxima locaux pour chaque segment ;
- dans une seconde passe : exploration limitée au voisinage des extrema locaux sélectionnés dans la première passe, pour chaque segment.

Bien entendu, la progression $\alpha(i) = k\alpha(i-1)$ est limitée par L : si à partir de i_L on a $\alpha(i) \geq L$ alors $\alpha(i) = L$ pour $i \geq i_L$ et la recherche sous-optimale en deux passes est remplacée par une recherche optimale (en une seule passe) pour les segments i_L à $Q-1$.

Un cas est plus particulièrement intéressant: lorsque L est un multiple de k^{i_L-1} . Alors le choix pour $\alpha(0)$ de L/K^{i_L-1} ou d'un sous-multiple de L/K^{i_L-1} introduit un espacement régulier des délais explorés dans la première passe. On montre que ces délais forment alors l'ensemble:

$$\left\{ \gamma_0 - \frac{L}{R_0} + j \times a, j = 1 \rightarrow \frac{k^{i_L} - 1}{k - 1} \alpha(0) \right\},$$

le pas a étant égal à $L/(R_0\alpha(0))$.

Dans le cas particulier du dictionnaire D introduit plus haut, cette technique d'exploration en deux passes s'introduit de la manière suivante:

Pour ce dictionnaire $L = 64, k^{Q-1} = 8, R_0 = 8$. Le choix $\alpha(0) = 8$ permet d'explorer dans la première passe un sous-ensemble D0 de D constitué de délais régulièrement espacés de D avec un pas $\alpha = 1$. On montre que $\gamma_0 = \gamma_{\min}^0 + 7$ et que D0 est en fait formé des 120 délais entiers consécutifs $\{\lambda_0 = \lambda_{\min}^0, j, j = 0 \rightarrow 119\}$ extraits du dictionnaire D.

Il est possible d'introduire une simplification supplémentaire à la recherche de la première passe. On remplace la maximisation de $E'(\lambda) = N(\lambda)^2/D(\lambda)$ par celle de $N(\lambda)$. La normalisation qu'apporte la division par $D(\lambda)$ s'avère généralement superflue dans cette première passe par essence plus grossière que la recherche complète. On s'intéresse donc aux maxima locaux de l'intercorrélation $N(\lambda)$, pour tous les segments $i = 0 \rightarrow Q - 1$, dans la première passe.

La deuxième passe utilise par contre le critère complet $E'(\lambda)$ et doit être effectuée aussi sur tous les segments: même pour les segments $i \geq i_L$ $\alpha(i) \geq L$, car il faut évaluer $E'(\lambda)$ sur les extrema locaux de $N(\lambda)$ sélectionnés dans la première passe.

L'analyse LTP par dictionnaire adaptatif, très performante, est aussi très complexe, du fait de la présence de la boucle fermée d'une part, et du filtre perceptuel d'autre part. Une variante de cette analyse, diminuant la complexité intrinsèque du processus sans en dégrader les performances subjectives est proposée ici: elle repose sur une modification de l'expression (3) du signal d'erreur dont on minimise l'énergie (critère $E(\lambda)$ à minimiser).

On peut en effet conserver l'usage d'un filtre perceptuel sans souscrire entièrement à la philosophie "CELP" du dictionnaire adaptatif, en prenant:

$$e(n) = h_g(n) * (r(n) - \beta \tilde{e}(n - \lambda)) \quad (10)$$

Dans cette expression, le signal $\tilde{e}(n - \lambda)$ (λ éventuellement fractionnaire, $\tilde{e}$ éventuellement extrapolé) est continu au niveau de la frontière du sous-bloc: les points précédant le sous-bloc courant ($n = 0 \rightarrow N - 1$) sont les points ($\tilde{e}(n - \lambda), n < 0$), et non ($\tilde{e}(n), n < 0$), comme dans le cas du dictionnaire adaptatif.

L'intérêt de cette variante réside dans la possibilité de "préfiltrer" $\tilde{e}(n)$: le filtre perceptuel variant à la fréquence trame LPC, plusieurs analyses LTP étant pratiquées dans une trame LPC, un même échantillon filtré $\tilde{e}_w(n) = h_g(n) * \tilde{e}_w(n)$ sert à plusieurs analyses LTP.

En ce qui concerne les délais fractionnaires, on utilise la commutabilité des filtres linéaires et on applique le filtre d'interpolation aux échantillons préfiltrés $\tilde{e}_w(n)$ (ceci n'est toutefois pas applicable aux échantillons utilisant un signal $\tilde{e}(n)$ extrapolé).

On va à présent décrire un exemple de réalisation particulièrement intéressant de la présente invention: le dictionnaire D cité précédemment est d'abord présenté en détail. L'exploration de ce dictionnaire est présentée avec la procédure accélérée décrite dans le cadre de l'analyse LTP définie ci-dessus.

Le module LTP ainsi conçu est intégré à titre d'exemple au dispositif de codage présenté plus haut.

Ce dictionnaire a été défini précédemment. Les délais y sont du type fractionnaire, de résolution $R = p > 1$, ou entiers. On peut décrire D de la façon suivante (figure 6): chaque segment $S_i, i = 0 \rightarrow 3$ de résolution $R_i = 2^{3-i}$ est formé des délais $\lambda_0 - \varphi / R_i, \varphi = 0 \rightarrow 1$, les délais λ_0 , entiers formant un sous-ensemble S_i^0 de S_i comportant $n_i = 2^{i+3}$ éléments:

$$S_i^0 = \left\{ \lambda_0 = \mu_i + j / R_i, j = 0 \rightarrow n_i \text{ avec } \mu_i = \lambda_{\min}^0 + \sum_{j=0}^{i-1} n_j \right\} \quad (\text{en posant } n_{-1} = 0).$$

Un seul filtre d'interpolation $H(z)$ est nécessaire pour l'ensemble du dictionnaire. On prend en pratique: $h(i) = w(i) \cdot \sin(i\pi/8) \cdot (8/i\pi)$, $i = -I \rightarrow I$, $w(i)$ étant une fonction de fenêtrage, et I étant un multiple de codage On définit les filtres:

$$h_\varphi(j) = h(-I + 8j + \varphi), j = 0 \rightarrow 2J - 1 \text{ et } \varphi = 1, 2, \dots, 7.$$

Les algorithmes de codage et décodage des délais de ce dictionnaire D sont présentés à la figure 7 et implantés de façon simple à l'aide de décalages et opérateurs logiques, en utilisant la table des quatre valeurs μ_i (premier délai entier dans chaque segment). Le code décrit ici perturbe l'ordre naturel des délais dans le dictionnaire sans que cela change quoi que ce soit à la description qui précède.

$$\lambda = \lambda_0 - \varphi / 8 \in D, \lambda_0 \in N, \varphi \in \{0, 1, \dots, 7\}$$

On pose :

$$\begin{aligned} \lambda'_0 &= \lambda_0 - \mu(\text{iseg}) \\ \varphi' &= \varphi / 2^{\text{iseg}} \end{aligned} \quad \text{avec } \text{iseg} \in \{0, 1, 2, 3\} = n^\circ \text{ segment}$$

On a alors:

$$\text{code } \lambda = [\text{iseg}(2 \text{ bits}), \lambda'_0(3 + \text{iseg bits}), \varphi'(3 - \text{iseg bits})] = 8 \text{ bits}$$

L'analyse LTP utilise le critère modifié calculé à partir de l'équation (10) et fait donc intervenir un signal $\tilde{e}_w(n - \lambda) = h_g(n) \tilde{e}(n - \lambda)$, $n = 0 \rightarrow N - 1$, λ , éventuellement fractionnaire.

On connaît les signaux $\tilde{e}(n)$ et $\tilde{e}_w(n)$ pour $n < 0$.

Selon les valeurs de λ , le calcul de $\tilde{e}_w(n-\lambda)$ fait intervenir l'un des quatre procédés suivants :

Délai $\lambda = \lambda_0$, entier $\geq N$: module ETW0 40 (voir figure 8A) $\tilde{e}_w(n-\lambda_0)$ est connu.

Délai $\lambda = \lambda_0$, entier $< N$: module ETW1 41 (voir figure 8B)

5 si $n < \lambda_0$, $\tilde{e}_w(n-\lambda_0)$ est connu

si $\lambda_0 \leq n < N$: extrapolation de $\tilde{e}(n-\lambda_0)$: $\tilde{e}(n-k\lambda_0)$ avec k = plus petit entier avec $n < k\lambda_0$, puis filtrage par $H_g(z)$

Délai $\lambda = \lambda_0 - \varphi/8$ fractionnaire, $\lambda_0 \geq N+J$: module ETW2 42 (voir figure 8C)

10

$$\tilde{e}_w(n-\lambda) = \sum_{j=0}^{2J-1} h_{\varphi}(j) \tilde{e}_w(n-\lambda_0+J-j) \quad (11)$$

15 **Délai $\lambda = \lambda_0 - \varphi/8$ fractionnaire, $\lambda_0 < N+J$: module ETW3 43** (voir figure 8D)

si $n < \lambda_0 - J$: $\tilde{e}_w(n-\lambda)$ est calculé par l'équation (11)

si $\lambda_0 - J \leq n < N$: $\tilde{e}$ est complété récursivement par:

20

$$\tilde{e}(0) = \tilde{e}(-\lambda) = \sum_{j=0}^{2J-1} h_{\varphi}(j) \tilde{e}_w(-\lambda_0+J-j)$$

25 puis $\tilde{e}(n) = \tilde{e}(n-\lambda)$ pour $n=1 \rightarrow (N-1-\lambda_0+J)$ $\tilde{e}_w(n-\lambda)$ est ensuite obtenu filtrant $\tilde{e}(n-\lambda)$ par $H_g(z)$.

Dans ces modules **ETW0, ETW1, ETW2, ETW3** représentés aux figures 8A, 8B, 8C et 8D, on a:

$$H_g(z) = \sum h_g(i) z^{-i} \quad \text{filtre perceptuel}$$

$$H_{\varphi}(z) = \sum h_{\varphi}(i) z^{-i} \quad \text{filtre polyphasé}$$

La recherche s'effectue en deux passes suivant le principe décrit plus haut.

30 Ainsi que mentionné précédemment, le dictionnaire D a l'avantage de permettre (en choisissant $\alpha(0) = 8$) la coïncidence entre l'ensemble des délais explorés dans la première et l'ensemble des délais entiers de D (c'est-à-dire

35

$$\bigcup_{i=0}^3 S_i^t$$

40 dans la description précédente). La première passe, effectuée uniquement sur les numérateurs $N(\lambda_0)$, est très rapide car elle ne fait intervenir aucune opération d'interpolation.

Le choix de $\lambda_{\min}^0 = N-8$ est particulièrement intéressant car il restreint au premier segment de D la nécessité d'extrapoler $\tilde{e}(n)$ dans la première passe.

Le module LTP donné ici à titre d'exemple s'intègre au dispositif présenté plus haut comme réalisation particulièrement intéressante de la présente invention.

45 On prend: $\lambda_{\min}^0 = N-8=24$ et $J=2$: $H(z)$ est un FIR (filtre à réponse impulsionnelle finie) de longueur 33.

Le nombre $K(i)$ de maxima locaux que l'on retient dans chaque segment S_i lors de la première passe de la recherche du délai est indiqué dans la table ci-dessous. Ces valeurs résultent de l'observation sur un certain nombre d'échantillons de parole, du nombre de maxima de $N(\lambda_0)$ qu'il est nécessaire de retenir pour s'assurer de la présence du délai optimal dans leur voisinage.

50

55

i/S_i	$K(i)$
0	1
1	1
2	2
3	1

5

10

La procédure complète de recherche du délai dans D sur le présent exemple est décrite dans la figure 9.

Les signaux $resw(n)$, $\tilde{e}_w(n)$ et $\tilde{e}(n)$ entrent dans le module de recherche 45. On trouve en sortie de ce module 45 le délai Λ sélectionné et le critère $E'(\Lambda)$ associé. Sur cette figure 9 on a la notation suivante:

$\Lambda, E'(\Lambda)$: délai Λ cherché et critère associé

15

$[\Lambda, E'(\Lambda)]^*$: Λ et $E'(\Lambda)$ sont éventuellement mis à jour.

On rappelle $\lambda_{\min}^0 = N-8$

Les modules $P1Si, i=0$ à 3 référencés respectivement 46, 47, 48 et 49 effectuent la première passe de la recherche sur les segments Si . Leur fonctionnement détaillé est illustré par la figure 10. Ces modules produisent en sortie $K(i), i=0$ à 3 (1 ou 2) valeurs de délais entiers λ_i , sélectionnés et les valeurs d'intercorrélation $N(\lambda_i)$ associées.

20

La deuxième passe de la recherche est décrite par les modules $P2Si, i=0$ à 3 référencés respectivement 50, 51, 52 et 53. En entrée de ces modules, outre les signaux $resw(n)$, $\tilde{e}_w(n)$ et $\tilde{e}(n)$, on trouve les sorties des modules $P1Si$ correspondants. Chaque module $P2Si$ effectue la maximisation du critère $E'(A)$ et sort le délai A associé au critère maximum.

25

Les figures 12A, 12B, 12C et 12D montrent le fonctionnement des modules $P2Si$, qui utilisent les modules de sélection $SEL_{ji} = 0$ à 3 décrits respectivement par les figures 11A, 11B, 11C et 11D:

- $SEL0$ présente les calculs effectués pour un délai entier lorsqu'aucune extrapolation de $\tilde{e}_w(n)$ n'est nécessaire;
- $SEL1$ présente les calculs effectués pour un délai entier avec extrapolation de $\tilde{e}_w(n)$;
- $SEL2$ présente les calculs effectués pour un délai fractionnaire lorsqu'aucune extrapolation de $\tilde{e}_w(n)$ n'est nécessaire;
- $SEL3$ présente les calculs effectués pour un délai fractionnaire avec extrapolation de $\tilde{e}_w(n)$.

35

Les modules PS 55 calculent le produit scalaire

$$\sum_{n=0}^{N-1} resw(n) \tilde{e}_w(n - \lambda_0).$$

40

Les modules NORM 56 calculent l'énergie

$$\sum_{n=0}^{N-1} \tilde{e}_w(n - \lambda)^2.$$

45

Les modules COMP 57 calculent $E'(\lambda)$ et sélectionnent $\Lambda = \lambda$, si $e'(\lambda) > E'(\Lambda)$

La valeur de délai Λ issue de la deuxième passe est le délai sélectionné par le module de recherche dans le dictionnaire D.

50

Revendications

55

1. Dispositif de codage et de décodage numériques de la parole comprenant, au codage: un module (13) d'analyse LPC (prédiction à court terme) un module (15) d'analyse LTP (prédiction à long terme), un module (14) de codage de l'onde résiduelle utilisant une méthode d'analyse par synthèse; au décodage: un module de décodage de l'onde résiduelle (21), un module de synthèse LTP (22) et un module de synthèse

LPC (23); caractérisé en ce que le module d'analyse LTP utilise un dictionnaire de délais à structure pseudo-logarithmique dans lequel les délais sont rangés en ordre croissant; ce dictionnaire étant constitué de Q segments adjacents, chacun d'une résolution donnée, les résolutions des segments successifs diminuant géométriquement dans un rapport k rationnel tel que $k > 1$, tandis que le nombre d'éléments L de chaque segment reste constant.

2. Dispositif selon la revendication 1, caractérisé en ce que le dictionnaire des délais est découpé en Q segments adjacents $S_i (i = 0 \rightarrow Q - 1)$ comprenant chacun L délais; en ce qu'une résolution R_i correspond à chaque segment S_i , les résolutions des segments successifs décroissant dans un rapport k rationnel donné ($R_i = R_{i-1}/k$); en ce qu'on appelle γ_i le dernier délai du segment S_i , ce segment étant formé des L délais $\lambda_j = \gamma_i - j/R_i, j = 0 \rightarrow L - 1$ avec λ_j, R_i entiers; en ce que la condition d'adjacence entre segments est assurée par $\gamma_{i-1} = \gamma_i - L / R_i, i = 1 \rightarrow Q - 1$; en ce que $\lambda_{\max}$ étant le dernier délai du dictionnaire et R_{Q-1} la résolution du dernier segment, un tel dictionnaire est entièrement défini par la donnée des valeurs $\{Q, L, k, \lambda_{\max}, R_{Q-1}\}$ et la condition $R_{Q-1} \cdot \lambda_{\max} \in \mathbb{N}$.

3. Dispositif selon la revendication 1, caractérisé en ce que le dictionnaire des délais est découpé en Q segments adjacents $S_i (i = 0 \rightarrow Q - 1)$ comprenant chacun L délais; en ce qu'une résolution R_i correspond à chaque segment S_i , les résolutions des segments successifs décroissant dans un rapport k rationnel donné ($R_i = R_{i-1}/k$); en ce qu'on appelle β_i le premier délai du segment S_i , ce segment étant formé des L délais $\lambda_j = \beta_i + j/R_i, j = 0 \rightarrow L - 1$ avec λ_j, R_i entiers; en ce que la condition d'adjacence entre segments est assurée par $\beta_i = \beta_{i-1} + L / R_{i-1}, i = 1 \rightarrow Q - 1$; en ce que β_{Q-1} étant le premier délai du dernier segment et R_{Q-1} la résolution du dernier segment, un tel dictionnaire est entièrement défini par la donnée des valeurs $\{Q, L, k, \beta_{Q-1}, R_{Q-1}\}$ et la condition $R_{Q-1} \cdot \beta_{Q-1} \in \mathbb{N}$.

4. Dispositif selon la revendication 2, caractérisé en ce qu'il utilise le codage suivant du délai LTP:

$$\text{code}(\lambda_j) = L \cdot i + j',$$

avec

$$S_i = \{\lambda_j = \gamma_i - j/R_i, j = 0 \rightarrow L - 1\}$$

et

$$j' = L - 1 - j$$

5. Dispositif selon la revendication 3, caractérisé en ce qu'il utilise le codage suivant du délai LTP:

$$\text{code}(\lambda_j) = L \cdot i + j$$

avec

$$S_i = \{\lambda_j = \beta_i + j/R_i, j = 0 \rightarrow L - 1\}$$

6. Dispositif selon la revendication 2, caractérisé en ce que dans le dictionnaire pseudo-logarithmique de délais LTP, chaque segment $S_i, i = 0 \rightarrow 3$ de résolution $R_i = 2^{3-i}$, est formé des délais $\lambda_{0-\varphi}/R_i, \varphi = 0 \rightarrow R_i - 1$, les délais λ_0 entiers formant un sous-ensemble S_i^0 de S_i comportant $n_i = 2^{i+3}$ éléments :

$$S_i^0 = \left\{ \lambda_0 = \mu_i + j / j = 0 \rightarrow n_i - 1, \text{ avec } \mu_i = \lambda_{\min}^0 + \sum_{j=0}^{i-1} n_j \right\} \text{ (en posant } n_{-1} = 0)$$

7. Procédé d'exploration d'un dictionnaire pseudo-logarithmique de délais LTP selon l'une quelconque des revendications précédentes, caractérisé en ce que dans une première passe, on sélectionne $K(i)$ maxima locaux du critère à maximiser parmi un ensemble restreint de $\alpha(i)$ délais de chaque segment S_i et dans une seconde passe, on explore le dictionnaire de manière limitée au voisinage des valeurs sélectionnées dans la première passe.

8. Procédé selon la revendication 7, caractérisé en ce que la taille des segments L étant un multiple de K^{L-1} , on choisit pour $\alpha(0)$ la valeur L/K^{L-1} ou d'un sous-multiple de L/K^{L-1} introduisant un espacement régulier des délais explorés dans la première passe.

9. Procédé selon la revendication 7, caractérisé en ce que l'on remplace dans la première passe la maximisation de $E(\lambda) = N(\lambda)^2/D(\lambda)$, où $N(\lambda)$ et $D(\lambda)$ représentent respectivement le numérateur et le dénominateur du gain optimal associé à chaque délai λ , par celle de $N(\lambda)$.

10. Procédé d'exploration d'un dictionnaire pseudo-logarithmique de délais LTP selon la revendication 1, caractérisé en ce qu'on minimise l'énergie en considérant l'expression suivante du signal d'erreur :

$$e(n) = h_x(n) * (r(n) - \beta \tilde{e}(n - \lambda))$$

5

les points précédant le sous-bloc courant (tels que $n < 0$ si le sous-bloc courant commence à $n=0$) étant les points $\tilde{e}(n-\lambda)$ (λ éventuellement fractionnaire, $\tilde{e}$ éventuellement extrapolé).

10

11. Procédé d'analyse LTP en boucle fermée avec filtrage perceptuel, caractérisé en ce qu'on minimise l'énergie en considérant l'expression suivante du signal d'erreur:

$$e(n) = h_g(n) * (r(n) - \beta \tilde{e}(n - \lambda))$$

15

les points précédant le sous-bloc courant (tels $n < 0$ si le sous-bloc courant commence à $n=0$) étant les points $\tilde{e}(n-\lambda)$ (λ éventuellement fractionnaire, $\tilde{e}$ éventuellement extrapolé).

20

25

30

35

40

45

50

55

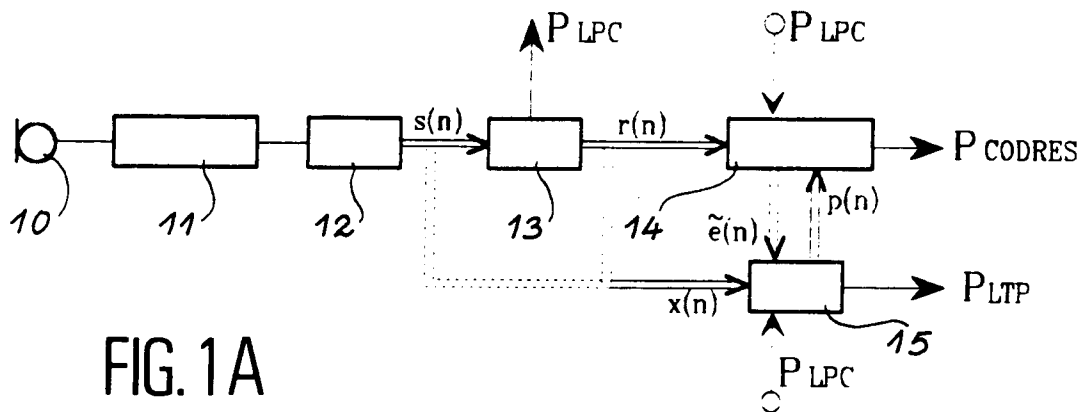


FIG. 1A

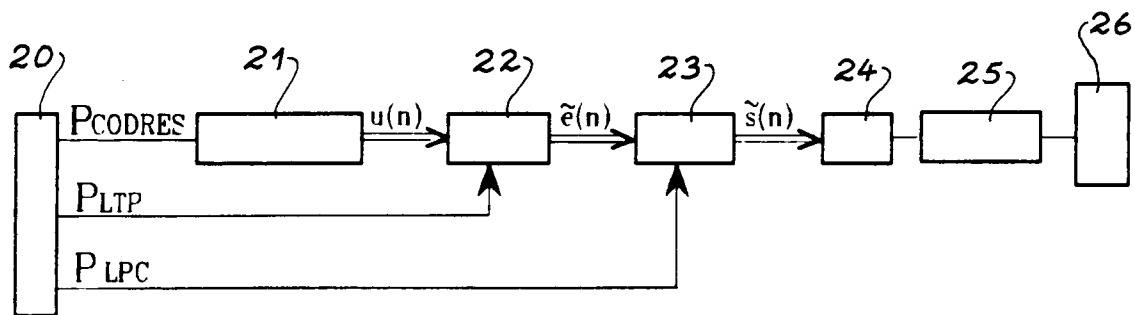


FIG. 1 B

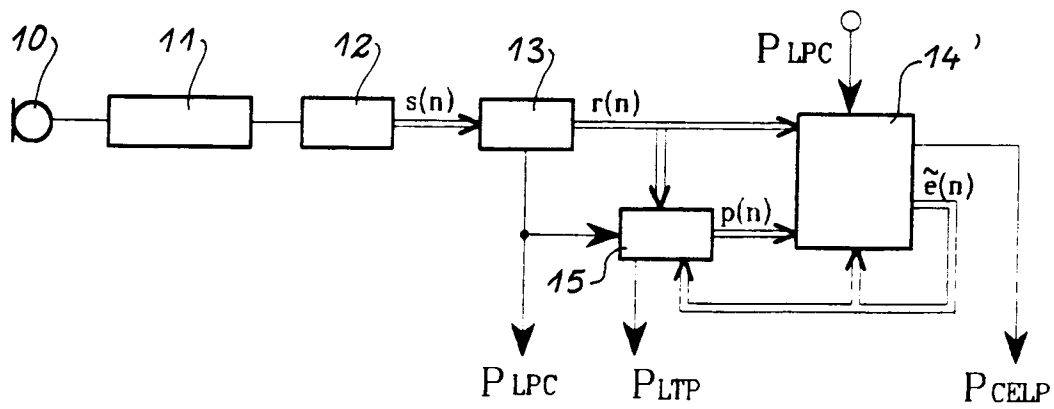


FIG. 2

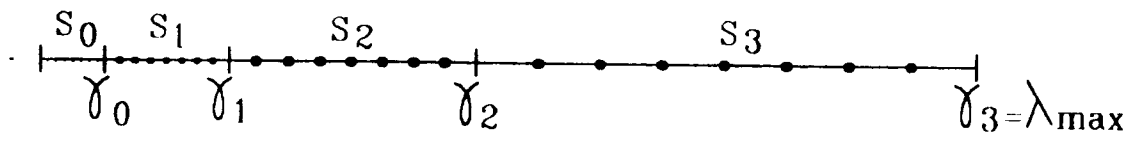


FIG. 3 A

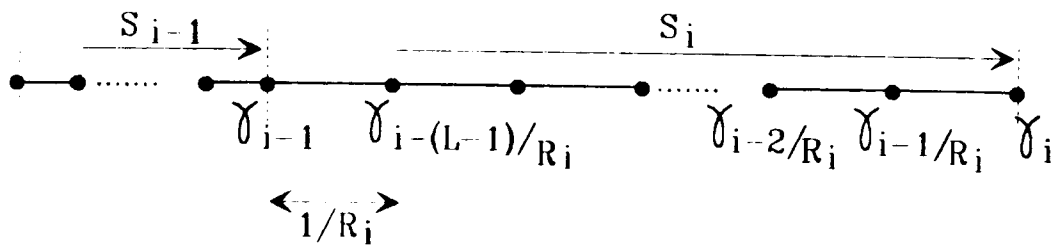


FIG. 3 B

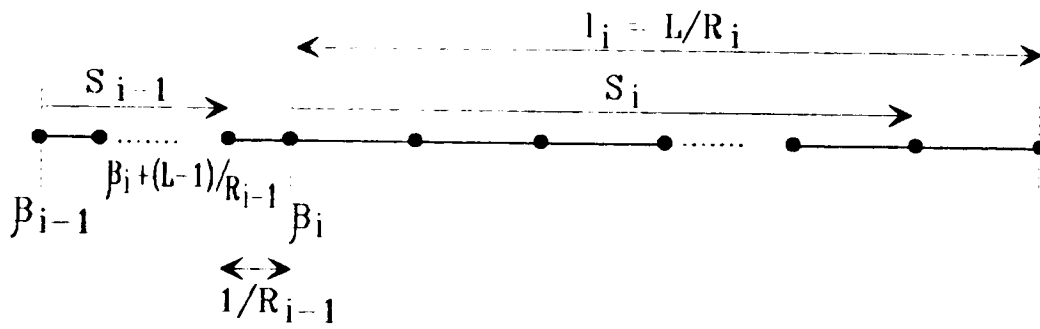


FIG. 3 C

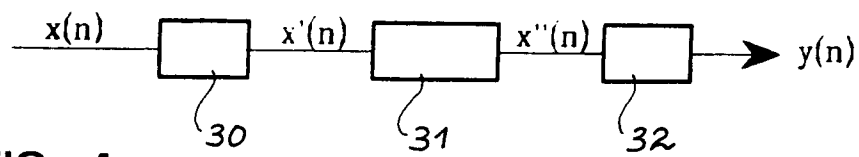


FIG. 4

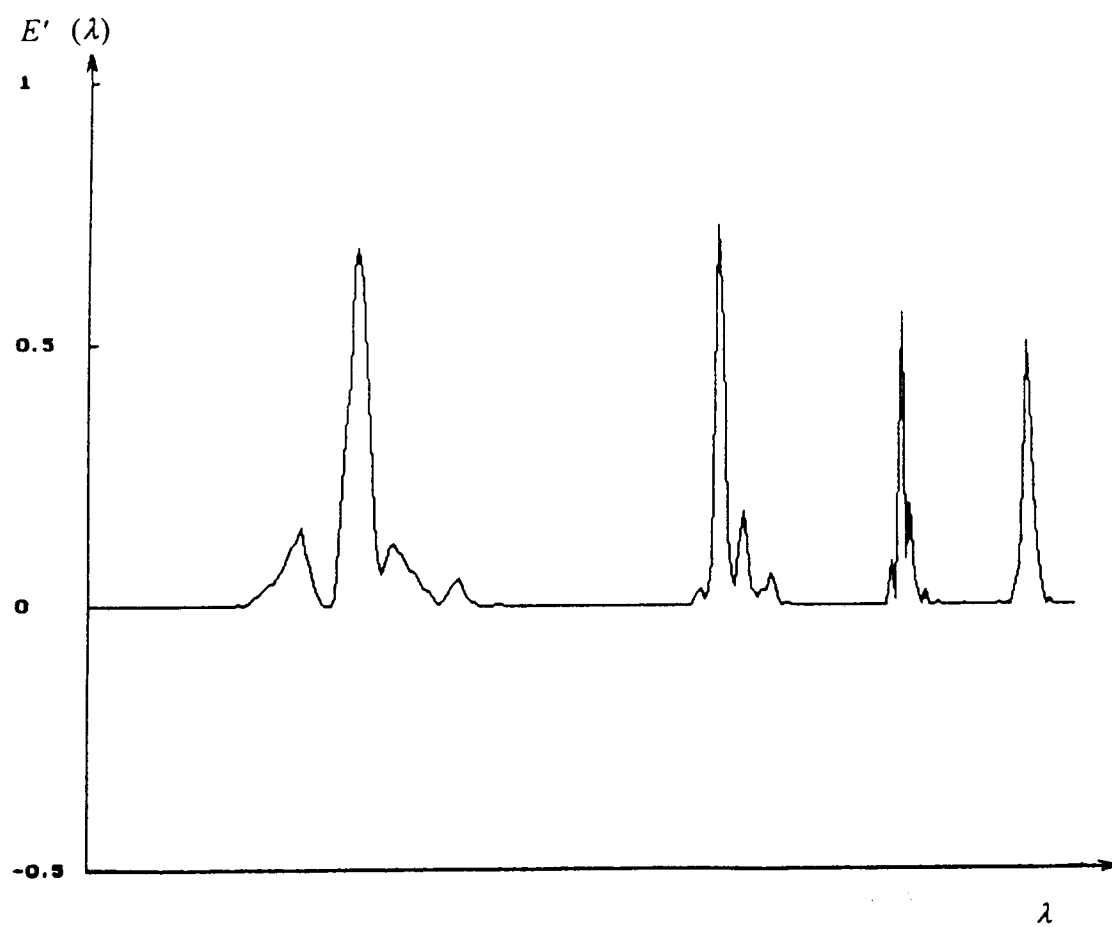


FIG. 5

i/S_i	μ_i	n_i	S_i^0	R_i	$\lambda/\lambda_0 \in S_i^0$
0	$\lambda_{\min}^0$	8	$\lambda_{\min}^0, \dots, \lambda_{\min}^0 + 7$	8	$\lambda_0^{-7/8}, \lambda_0^{-3/4}, \lambda_0^{-5/8}, \lambda_0^{-1/8}, \lambda_0$
1	$\lambda_{\min}^0 + 8$	16	$\lambda_{\min}^0 + 8, \dots, \lambda_{\min}^0 + 23$	4	$\lambda_0^{-3/4}, \lambda_0^{-1/2}, \lambda_0^{-1/4}, \lambda_0$
2	$\lambda_{\min}^0 + 24$	32	$\lambda_{\min}^0 + 24, \dots, \lambda_{\min}^0 + 56$	2	$\lambda_0^{-1/2}, \lambda_0$
3	$\lambda_{\min}^0 + 56$	64	$\lambda_{\min}^0 + 56, \dots, \lambda_{\min}^0 + 119$	1	λ_0

FIG. 6

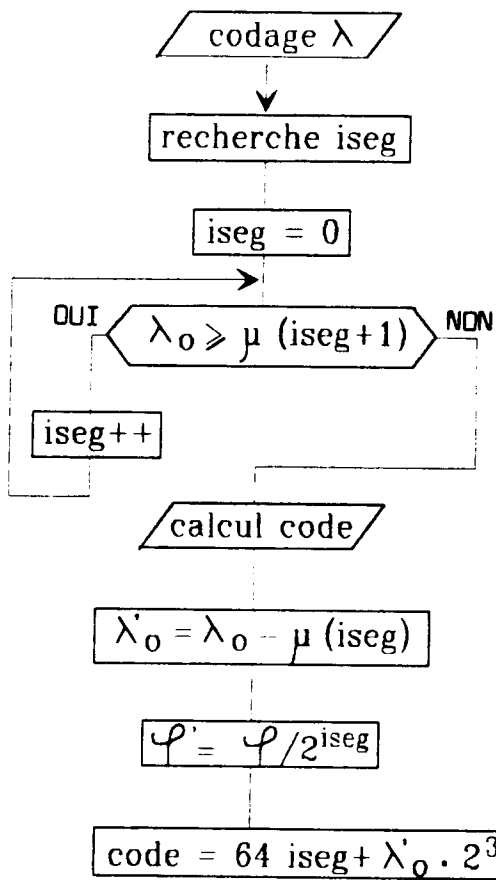
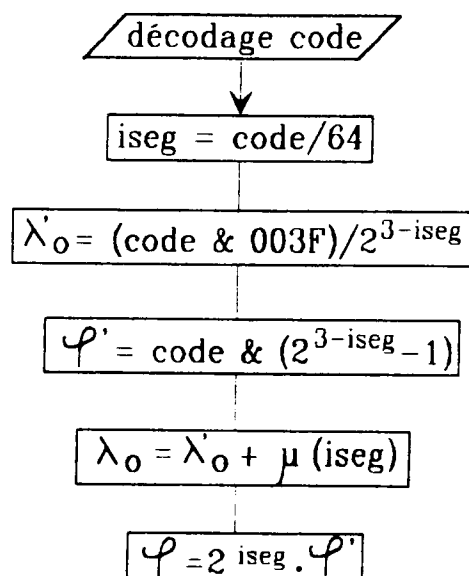


FIG. 7 A

FIG. 7 B



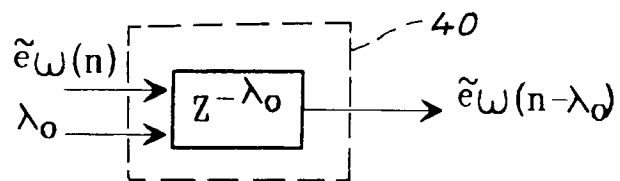


FIG. 8A

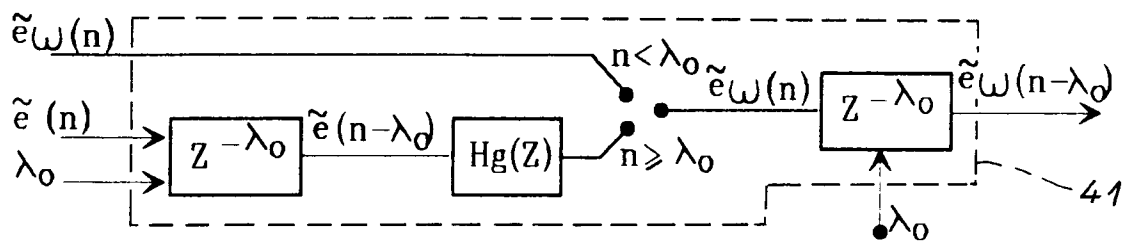


FIG. 8B

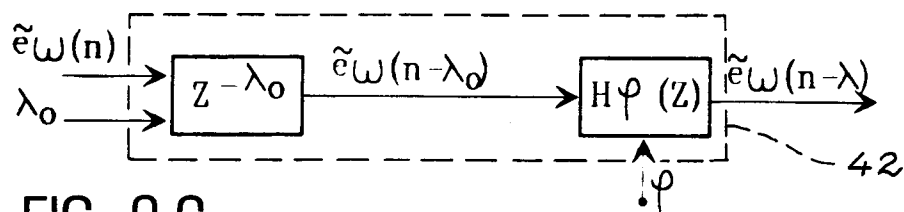


FIG. 8C

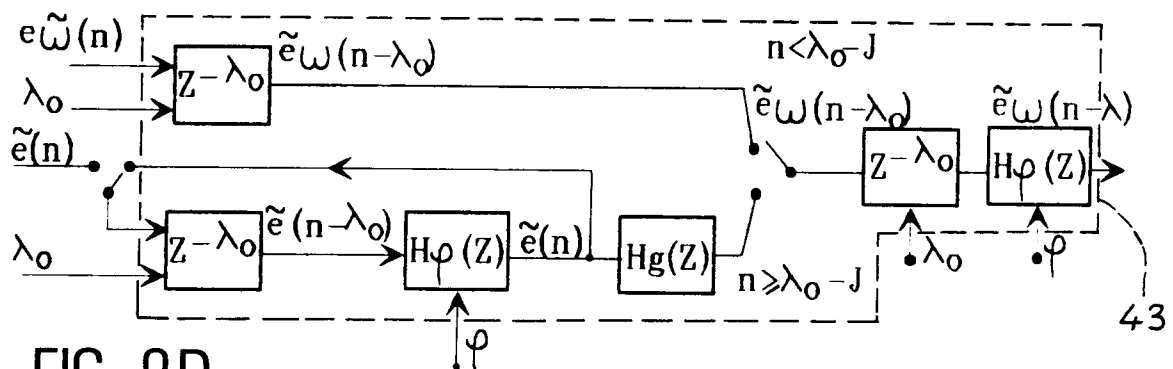


FIG. 8D

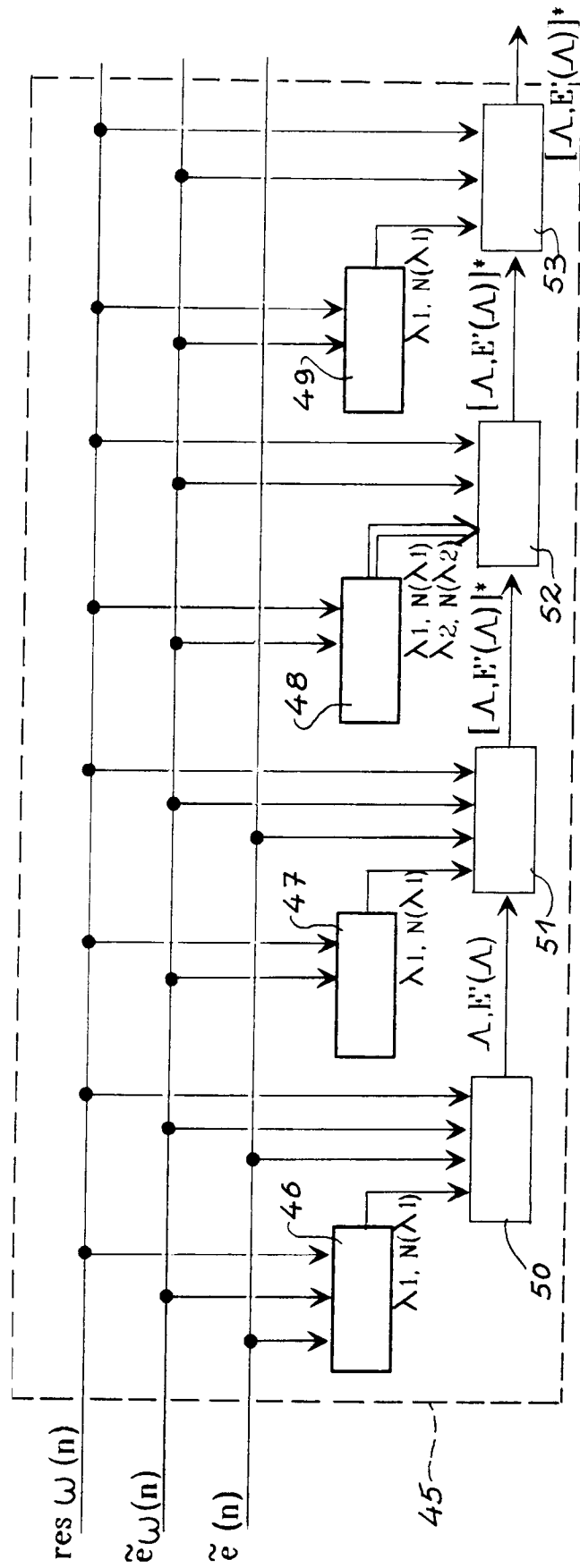


FIG. 9

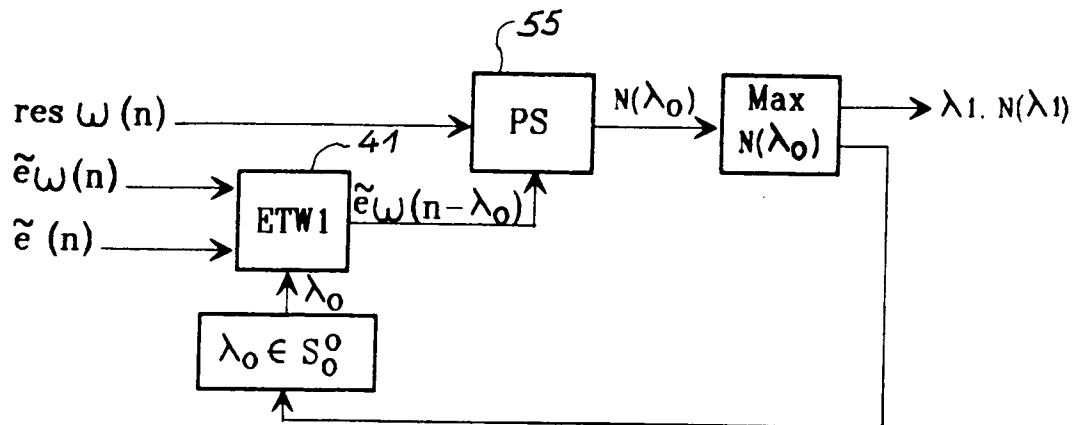


FIG. 10 A

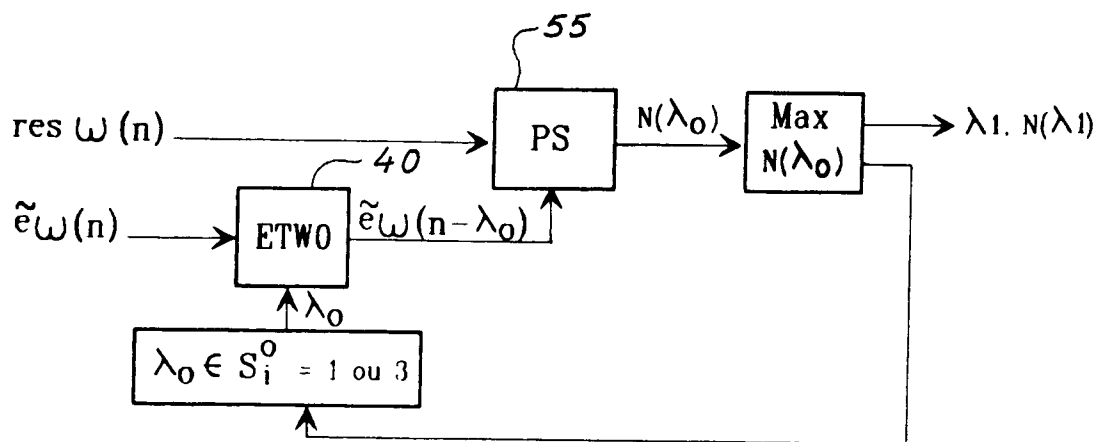


FIG. 10 B

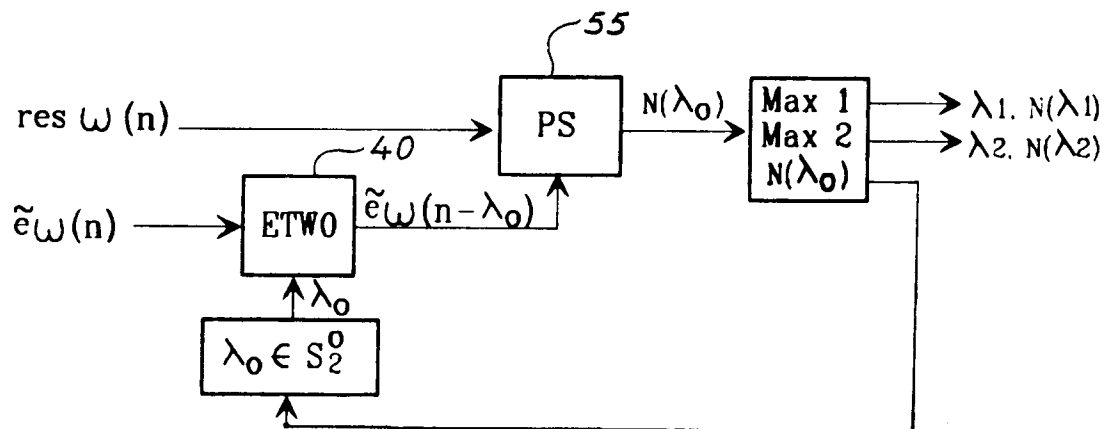


FIG. 10 C

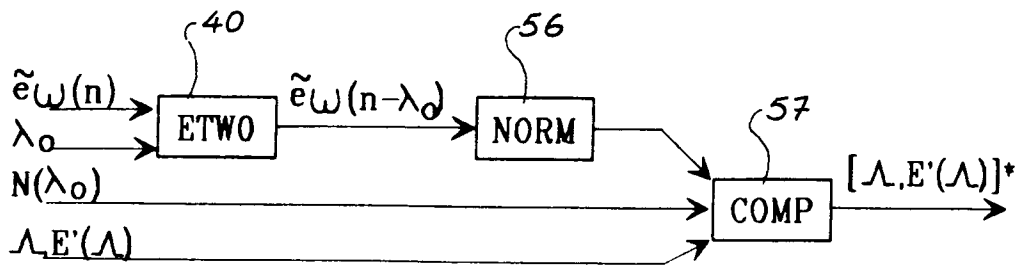


FIG. 11A

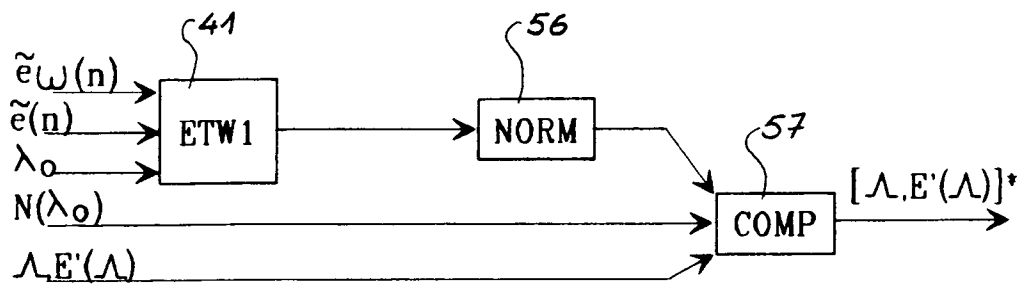


FIG. 11B

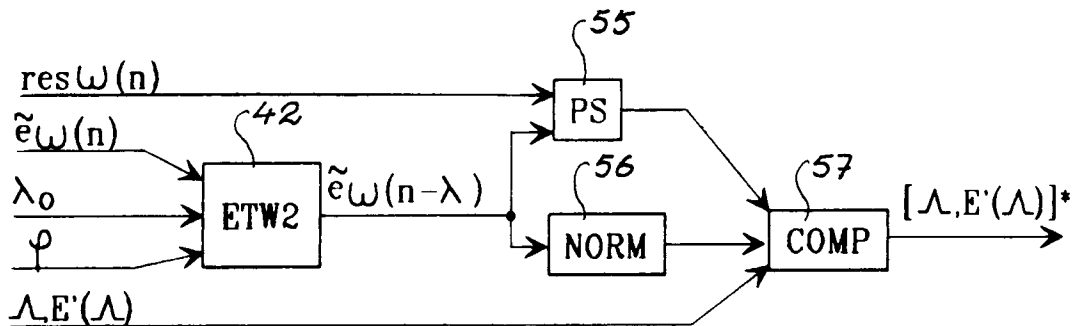


FIG. 11C

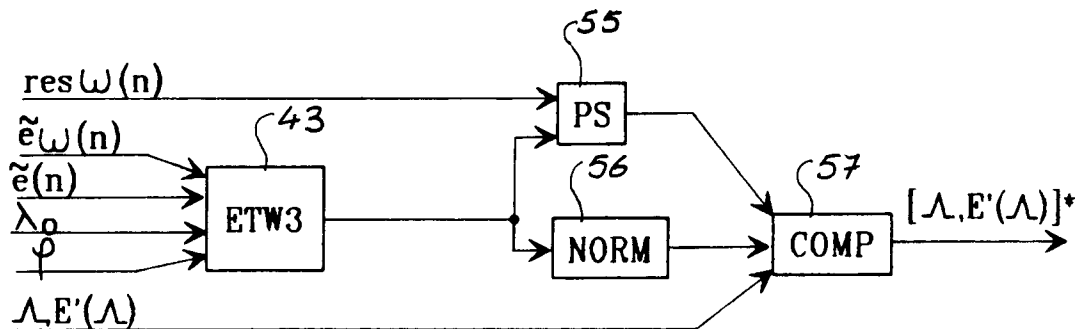


FIG. 11D

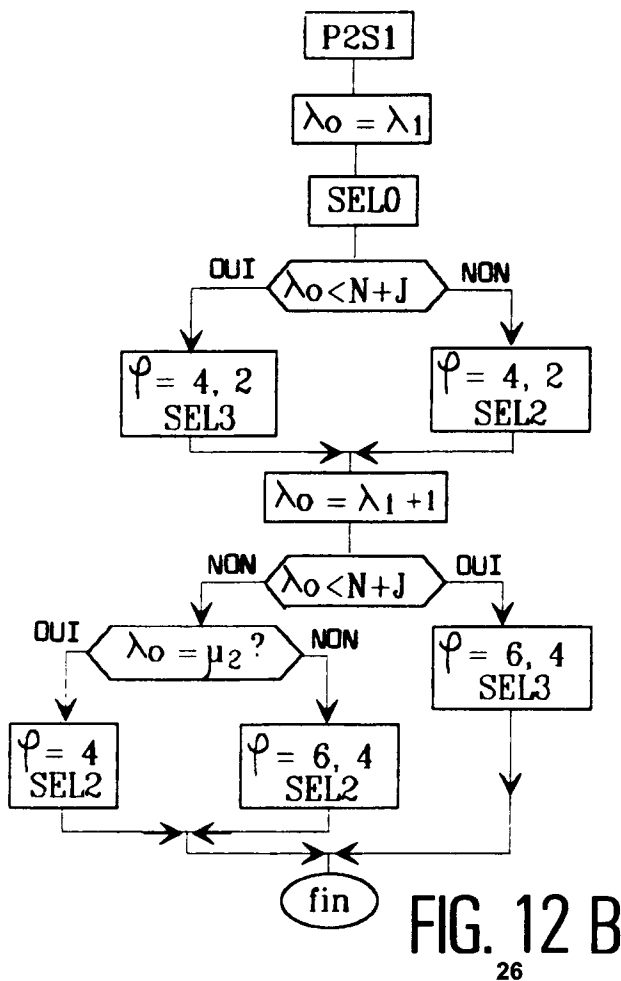
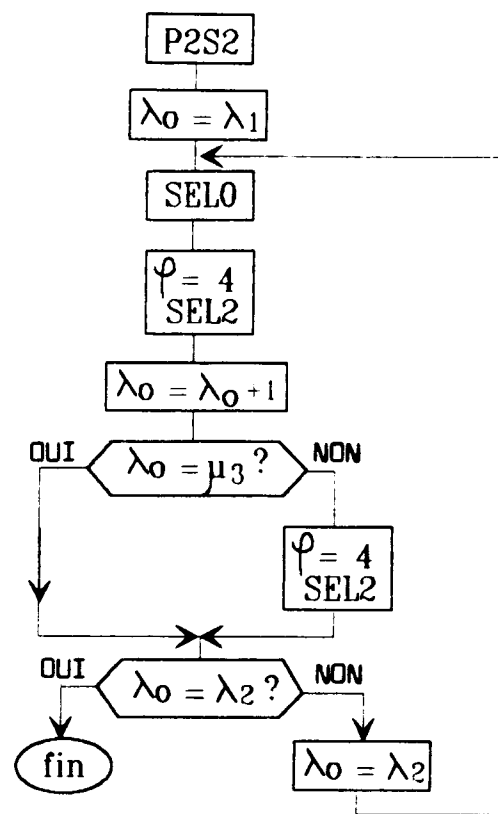
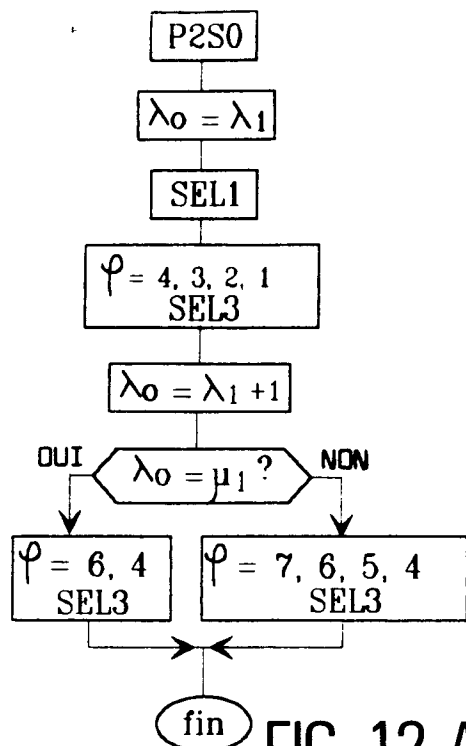
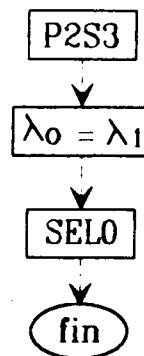


FIG. 12 C

FIG. 12 D





Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande
EP 94 40 0525

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	EP-A-0 523 979 (MOTOROLA) * page 9, ligne 40 - page 10, ligne 20 * ---	1	G10L9/14
A	INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 91, vol.1, 14 Mai 1991, TORONTO, ONTARIO, CA pages 609 - 612 KEMP ET AL. 'Multi-frame coding of LPC parameters at 600-800 bps' * page 610 * ---	1	
D,A	INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 90, vol.2, 3 Avril 1990, ALBUQUERQUE, NM, US pages 661 - 664 KROON ET AL. 'Pitch predictors with high temporal resolution' ---	1	
D,A	EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY 89, vol.2, 26 Septembre 1989, PARIS, FR pages 509 - 510 MARQUES ET AL. 'Pitch prediction with fractional delays in CELP coding' * page 509 - page 510 * ---	1	DOMAINES TECHNIQUES RECHERCHES (Int.Cl.5)
D,A	WO-A-91 03790 (MOTOROLA) * page 7, ligne 17 - ligne 34 * * page 12, ligne 7 - page 13, ligne 9 * * page 13, ligne 25 - page 15, ligne 2; revendications 1-7 * --- -/--	1	G10L
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche LA HAYE		Date d'achèvement de la recherche 15 Juillet 1994	Examineur Lange, 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 (P04.C02)



Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande
EP 94 40 0525

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)
D,A	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol.38, no.8, Août 1990, NEW YORK, US pages 1330 - 1342 KLEIJN ET AL. 'Fast methods for the CELP speech coding algorithm' * page 1332, colonne de gauche, ligne 2 - colonne de droite, ligne 3 *	1	
A	NACHRICHTENTECHNISCHE ZEITSCHRIFT, vol.33, no.3, Mars 1980, DE pages 150 - 155 PICARD ET AL. 'PCM code converter for two speech channels' * page 1 - page 2, colonne de gauche, ligne 33; figure 1 *	1	
A	EP-A-0 443 548 (NEC) * page 2, ligne 51 - page 4, ligne 18 *	11	
A	INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING 90, vol.2, 3 Avril 1990, ALBUQUERQUE, NM, US pages 677 - 680 OZAWA 'A hybrid speech coding based on multi-pulse and CELP at 3.2kb/2' * page 678 *	11	DOMAINES TECHNIQUES RECHERCHES (Int.Cl.5)
A	AEU ARCHIV FÜR ELEKTRONIK UND ÜBERTRAGUNGSTECHNIK, vol.43, no.5, Septembre 1989, STUTTGART, DE pages 307 - 312 REININGER ET AL. 'Prädiktive Sprachkodierung mit stochastischer Anregung' * page 308 - colonne de droite *	11	
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche LA HAYE		Date d'achèvement de la recherche 15 Juillet 1994	Examineur Lange, 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 I : cité pour d'autres raisons & : membre de la même famille, document correspondant	

EPO FORM 1503 03.82 (P04C02)