(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 801 788 B1

(12) FASCICULE DE BREVET EUROPEEN

(45) Date de publication et mention de la délivrance du brevet:
09.06.1999 Bulletin 1999/23

(21) Numéro de dépôt: **96901008.1**

(22) Date de dépôt: **03.01.1996**

(51) Int Cl.[6]: **G10L 9/14**

(86) Numéro de dépôt international:
PCT/FR96/00004

(87) Numéro de publication internationale:
WO 96/21218 (11.07.1996 Gazette 1996/31)

(54) PROCEDE DE CODAGE DE PAROLE A ANALYSE PAR SYNTHESE

VERFAHREN ZUR SPRACHKODIERUNG MITTELS ANALYSE DURCH SYNTHESE

SPEECH CODING METHOD USING SYNTHESIS ANALYSIS

(84) Etats contractants désignés:
AT BE CH DE GB IT LI LU NL SE

(30) Priorité: **06.01.1995 FR 9500134**

(43) Date de publication de la demande:
22.10.1997 Bulletin 1997/43

(73) Titulaire: **MATRA NORTEL COMMUNICATIONS**
29000 Quimper (FR)

(72) Inventeurs:
- **NAVARRO, William**
  F-78140 Vélizy-Villacoublay (FR)
- **MAUC, Michel**
  F-91310 Leuville-sur-Orge (FR)

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

(56) Documents cités:
EP-A- 0 415 163
EP-A- 0 515 138
EP-A- 0 619 574
EP-A- 0 628 947

EP 0 801 788 B1

Printed by Jouve, 75001 PARIS (FR)

Description

[0001] La présente invention concerne le codage de la parole utilisant l'analyse par synthèse.

[0002] La demanderesse a notamment décrit de tels codeurs de parole qu'elle a développés dans ses demandes de brevet européen 0 195 487, 0 347 307 et 0 469 997.

[0003] Dans un codeur de parole à analyse par synthèse, on effectue une prédiction linéaire du signal de parole pour obtenir les coefficients d'un filtre de synthèse à court terme modélisant la fonction de transfert du conduit vocal. Ces coefficients sont transmis au décodeur, ainsi que des paramètres caractérisant une excitation à appliquer au filtre de synthèse à court terme. Dans la plupart des codeurs actuels, on recherche en outre les corrélations à plus long terme du signal de parole pour caractériser un filtre de synthèse à long terme rendant compte de la hauteur tonale de la parole. Lorsque le signal est voisé, l'excitation comporte en effet une composante prédictible pouvant être représentée par l'excitation passée, retardée de TP échantillons du signal de parole et affectée d'un gain g_P. Le filtre de synthèse à long terme, également reconstitué au décodeur, a alors une fonction de transfert de la forme 1/B(z) avec $B(z)=1-g_P.z^{-TP}$. La partie restante, non prédictible, de l'excitation est appelée excitation stochastique. Dans les codeurs dits CELP ("Code Excited Linear Prediction"), l'excitation stochastique est constituée par un vecteur recherché dans un dictionnaire prédéterminé. Dans les codeurs dits MPLPC ("Multi-Pulse Linear Prediction Coding"), l'excitation stochastique comporte un certain nombre d'impulsions dont les positions sont recherchées par le codeur. En général, les codeurs CELP sont préférés pour les bas débits de transmission, mais ils sont plus complexes à mettre en oeuvre que les codeurs MPLPC.

[0004] Pour déterminer le retard de prédiction à long terme, on utilise une analyse en boucle ouverte, une analyse en boucle fermée ou une combinaison des deux. L'analyse en boucle ouverte est peu exigeante en volume de calculs, mais sa précision est limitée. A l'inverse l'analyse en boucle fermée requiert beaucoup de calculs, mais elle est plus fiable car elle contribue directement à minimiser l'écart pondéré perceptuellement entre le signal de parole et le signal synthétique. Dans certains cas, une analyse en boucle ouverte est d'abord effectuée pour limiter l'intervalle dans lequel l'analyseur en boucle fermée recherchera le retard de prédiction. Cet intervalle de recherche doit néanmoins rester relativement large car on doit tenir compte de ce que le retard peut varier rapidement.

[0005] L'invention vise notamment à trouver un bon compromis entre la qualité de la modélisation de la partie à long terme de l'excication et la complexité de la recherche du retard correspondant dans un codeur de parole.

[0006] L'invention propose ainsi un procédé de codage à analyse par synthèse d'un signal de parole numérisé en trames successives divisées en nst sous-trames, comprenant les étapes suivantes : analyse par prédiction linéaire du signal de parole pour déterminer des paramètres d'un filtre de synthèse à court terme ; analyse en boucle ouverte du signal de parole pour détecter les trames voisées du signal et pour déterminer, pour chaque trame voisée, un degré de voisement du signal et un intervalle de recherche d'un retard de prédiction à long terme ; analyse prédictive en boucle fermée du signal de parole pour sélectionner, pour certaines au moins des sous-trames des trames voisées, un retard de prédiction à long terme contenu dans l'intervalle de recherche et constituant un paramètre d'un filtre de synthèse à long terme ; et détermination d'une excitation stochastique pour chaque sous-trame, de façon à minimiser un écart pondéré perceptuellement entre le signal de parole et l'excitation stochastique filtrée par les filtres de synthèse à long terme et à court terme. A l'étape d'analyse en boucle ouverte, on détermine l'intervalle de recherche relatif à chaque trame voisée de façon qu'il contienne un nombre de retards dépendant du degré de voisement de ladite trame.

[0007] Ainsi, le nombre de retards qui sont à tester en boucle fermée est adaptable au mode de voisement de la trame. En général, la largeur de l'intervalle de recherche sera plus faible pour les trames les plus voisées afin de tenir compte de leur plus grande stabilité harmonique. Pour ces trames très voisées, on peut gagner un ou plusieurs bits sur la quantification différentielle du retard dans l'intervalle de recherche, et réaffecter ce ou ces bits gagnés à des paramètres perceptuellement importants comme le gain de prédiction à long terme, ce qui améliore la qualité de restitution de la parole.

[0008] D'autres particularités et avantages de l'invention apparaîtront dans la description ci-après d'exemples de réalisation préférés, mais non limitatifs, en référence aux dessins annexés, dans lesquels :

- la figure 1 est un schéma synoptique d'une station de radiocommunication incorporant un codeur de parole mettant en oeuvre l'invention ;
- la figure 2 est un schéma synoptique d'une station de radiocommunication apte à recevoir un signal produit par celle de la figure 1 ;
- les figures 3 à 6 sont des organigrammes Illustrant un processus d'analyse LTP en boucle ouverte appliqué dans le codeur de parole de la figure 1 ;
- la figure 7 est un organigramme illustrant un processus de détermination de la réponse impulsionnelle du filtre de synthèse pondéré appliqué dans le codeur de parole de la figure 1 ;
- les figures 8 à 11 sont des organigrammes illustrant un processus de recherche de l'excitation stochastique appliqué dans le codeur de parole de la figure 1.

[0009] Un codeur de parole mettant en oeuvre l'invention est applicable dans divers types de systèmes de transmission et/ ou de stockage de parole faisant appel à une technique de compression numérique. Dans l'exemple de la figure 1, le codeur de parole 16 fait partie d'une station mobile de radiocommunication. Le signal de parole S est un signal numérique échantillonné à une fréquence typiquement égale à 8kHz. Le signal S est issu d'un convertisseur analogique-numérique 18 recevant le signal de sortie amplifié et filtré d'un microphone 20. Le convertisseur 18 met le signal de parole S sous forme de trames successives elles-mêmes subdivisées en *nst* sous-trames de *lst* échantillons. Une trame de 20 ms comporte typiquement *nst*=4 sous-trames de lst=40 échantillons de 16 bits à 8kHz. En amont du codeur 16, le signal de parole S peut également être soumis à des traitements classiques de mise en forme tels qu'un filtrage de Hamming. Le codeur de parole 16 délivre une séquence binaire de débit sensiblement plus faible que celui du signal de parole S, et adresse cette séquence à un codeur canal 22 dont la fonction est d'introduire des bits de redondance dans le signal afin de permettre une détection et/ou une correction d'éventuelles erreurs de transmission. Le signal de sortie du codeur canal 22 est ensuite modulé sur une fréquence porteuse par le modulateur 24, et le signal modulé est émis sur l'interface air.

[0010] Le codeur de paroie 16 est un codeur à analyse par synthèse. Le codeur 16 détermine d'une part des paramètres caractérisant un filtre de synthèse à court terme modélisant le conduit vocal du locuteur, et d'autre part une séquence d'excitation qui, appliquée au filtre de synthèse à court terme, fournit un signal synthétique constituant une estimation du signal de parole S selon un critère de pondération perceptuelie.

[0011] Le filtre de synthèse à court terme a une fonction de transfert de la forme 1/A(z), avec :

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

[0012] Les coefficients a_i sont déterminés par un module 26 d'analyse par prédiction linéaire à court terme du signal de parole S. Les a_i sont les coefficients de prédiction linéaire du signal de parole S. L'ordre q de la prédiction linéaire est typiquement de l'ordre de 10. Les méthodes applicables par le module 26 pour la prédiction linéaire à court terme sont bien connues dans le domaine du codage de la parole. Le module 26 met par exemple en oeuvre l'algorithme de Durbin-Levinson (voir J. Makhoul : "Linear Prediction : A tutorial review", Proc. IEEE, Vol.63, N°4, Avril 1975, p. 561-580). Les coefficients a_i obtenus sont fournis à un module 28 qui les convertit en paramètres de raies spectrales (LSP). La représentation des coefficients de prédiction a_i par des paramètres LSP est fréquemment utilisée dans des codeurs de parole à analyse par synthèse. Les paramètres LSP sont les q nombres $\cos(2\pi f_i)$ rangés en ordre décroissant, les q fréquences de raies spectrales (LSF) normalisées $f_i (1 \leq 1 \leq q)$ étant telles que les nombres complexes exp $(2\pi j f_i)$, avec i=1,3,...,q-1,q+1 et $f_{q+1}=0,5$, soient les racines du polynôme Q(z) défini par $Q(z)=A(z)+z^{-(q+1)}.A(z^{-1})$ et que les nombres complexes $\exp(2\pi j f_i)$, avec i=0,2,4,...,q et $f_0=0$, soient les racines du polynôme Q*(z) défini par $Q^*(z)=A(z)-z^{-(q+1)}.A(z^{-1})$.

[0013] Les paramètres LSP peuvent être obtenus par le module de conversion 28 par la méthode classique 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, pages 1419-1426). Ce sont des valeurs de quantification des paramètres LSF, obtenues par un module de quantification 30, qui sont transmises au décodeur pour que celui-ci retrouve les coefficients a_i du filtre de synthèse à court terme. Les coefficients a_i peuvent être retrouvés simplement, étant donné que:

$$Q(z) = (1+z^{-1}) \prod_{i=1,3,\ldots,q-1} (1-2\cos(2\pi f_i) z^{-1} + z^{-2})$$

$$Q^*(z) = (1-z^{-1}) \prod_{i=2,4,\ldots,q} (1-2\cos(2\pi f_i) z^{-1} + z^{-2})$$

et

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

[0014] Pour éviter des variations brusques dans la fonction de transfert du filtre de synthèse à court terme, les paramètres LSP font l'objet d'une interpolation avant qu'on en déduise les coefficients de prédiction a_i. Cette interpolation est effectuée sur les premières sous-trames de chaque trame du signal. Par exemple, si LSP_t et LSP_{t-1} désignent respectivement un paramètre LSP calculé pour la trame t et pour la trame précédente t-1, on prend: $LSP_t(0)=0,5.LSP_{t-1}+0,5.LSP_t$, $LSP_t(1)=0,25.LSP_{t-1}+0,75.LSP_t$ et $LSP_t(2)=...=LST_t(nst-1)=LSP_t$ pour les sous-trames 0,1,2,...,nst-1 de la trame t. Les coefficients a_i du filtre 1/A(z) sont alors déterminés, sous-trame par sous-trame à partir des paramètres LSP interpolés.

[0015] Les paramètres LSP non quantifiés sont fournis par le module 28 à un module 32 de calcul des coefficients d'un filtre de pondération perceptuelle 34. Le filtre de pondération perceptuelle 34 a de préférence une fraction de transfert de la forme $W(z)=A(z/\gamma_1)/A(z/\gamma_2)$ où γ_1 et γ_2 sont des coefficients tels que $\gamma_1>\gamma_2>0$ (par exemple $\gamma_1=0,9$ et $\gamma_2=0,6$). Les coefficients du filtre de pondération perceptuelle sont calculés par le module 32 pour chaque sous-trame après interpolation des paramètres LSP reçus du module 28.

[0016] Le filtre de pondération perceptuelle 34 reçoit le signal de parole S et délivre un signal SW pondéré perceptuellement qui est analysé par des modules 36, 38, 40 pour déterminer la séquence d'excitation. La séquence d'excitation du filtre à court terme se compose d'une excitation prédictible par un filtre de synthèse à long terme modélisant la hauteur tonale (pitch) de la parole, et d'une excitation stochastique non prédictible, ou séquence d'innovation.

[0017] Le module 36 effectue une prédiction à long terme (LTP) en boucle ouverte, c'est-à-dire qu'il ne contribue pas directement à la minimisation de l'erreur pondérée. Dans le cas représenté, le filtre de pondération 34 intervient en amont du module d'analyse en boucle ouverte, mais il pourrait en être autrement : le module 36 pourrait opérer directement sur le signal de parole S ou encore sur le signal S débarrassé de ses corrélations à court terme par un filtre de fonction de transfert A(z). En revanche, les modules 38 et 40 fonctionnent en boucle fermée, c'est-à-dire qu'ils contribuent directement à la minimisation de l'erreur pondérée perceptuellement.

[0018] Le filtre de synthèse à long terme a une fonction de transfert de la forme 1/B(z) avec $B(z)=1-g_P.z^{-TP}$ où g_P désigne un gain de prédiction à long terme et TP désigne un retard de prédiction à long terme. Le retard de prédiction à long terme peut typiquement prendre N=256 valeurs comprises entre rmin et rmax échantillons. Une résolution fractionnaire est prévue pour les plus petites valeurs de retard de façon à éviter les écarts trop perceptibles en termes de fréquence de voisement. On utilise par exemple une résolution 1/6 entre rmin=21 et 33+5/6, une résolution 1/3 entre 34 et 47+2/3, une résolution 1/2 entre 48 et 88+1/2, et une résolution entière entre 89 et rmax=142. Chaque retard possible est ainsi quantifié par un index entier compris entre 0 et N-1=255.

[0019] Le retard de prédiction à long terme est déterminé en deux étapes. Dans la première étape, le module 36 d'analyse LTP en boucle ouverte détecte les trames voisées du signal de parole et détermine, pour chaque trame voisée, un degré de voisement MV et un intervalle de recherche du retard de prédiction à long terme. Le degré de voisement MV d'une trame voisée peut prendre trois valeurs : 1 pour les trames faiblement voisées, 2 pour les trames modérément voisées, et 3 pour les trames très voisées. Dans les notations utilisées ci-après, on prend un degré de voisement MV=0 pour les trames non voisées. L'intervalle de recherche est défini par une valeur centrale représentée par son index de quantification ZP et par une largeur dans le domaine des index de quantification, dépendant du degré de voisement MV. Pour les trames faiblement ou modérément voisées (MV=1 ou 2) la largeur de l'intervalle de recherche est de N1 index, c'est-à-dire que l'index du retard de prédiction à long terme sera recherché entre ZP-16 et ZP+15 si N1=32. Pour les trames très voisées (MV=3), la largeur de l'intervalle de recherche est de N3 index, c'est-à-dire que l'index du retard de prédiction à long terme sera recherché entre ZP-8 et ZP+7 si N3=16.

[0020] Une fois que le degré de voisement MV d'une trame a été déterminé par le module 36, le module 30 opère la quantification des paramètres LSP qui ont auparavant été déterminés pour cette trame. Cette quantification est par exemple vectorielle, c'est-à-dire qu'elle consiste à sélectionner, dans une ou plusieurs tables de quantification prédéterminées, un jeu de paramètres quantifiés LSP_Q qui présente une distance minimale avec le jeu de paramètres LSP fourni par le module 28. De façon connue, les tables de quantification diffèrent suivant le degré de voisement MV fourni au module de quantification 30 par l'analyseur en boucle ouverte 36. Un ensemble de tables de quantification pour un degré de voisement MV est déterminé, lors d'essais préalables, de façon à être statistiquement représentatif de trames ayant ce degré MV. Ces ensembles sont stockés à la fois dans les codeurs et dans les décodeurs mettant en oeuvre l'invention. Le module 30 délivre le jeu de paramètres quantifiés LSP_Q ainsi que son index Q dans les tables des quantification applicables.

[0021] Le codeur de parole 16 comprend en outre un module 42 de calcul de la réponse impulsionnelle du filtre composé du filtre de synthèse à court terme et du filtre de pondération perceptueile. Ce filtre composé a pour fonction de transfert W(z)/A(z). Pour le calcul de sa réponse impulsionnelle h=(h(0),(h1),...h(lst-1)) sur la durée d'une sous-trame, le module 42 prend pour le filtre de pondération perceptuelle W(z) celui correspondant aux paramètres LSP interpolés mais non quantifiés, c'est-à-dire celui dont les coefficients ont été calculés par le module 32, et pour le filtre

de synthèse 1/A(z) celui correspondant aux paramètres LSP quantifiés et interpolés, c'est-à-dire celui qui sera effectivement reconstitué par le décodeur.

[0022] Dans la deuxième étape de la détermination du retard TP de prédiction à long terme, le module 38 d'analyse LTP en boucle fermée détermine le retard TP pour chaque sous-trame des trames voisées (MV=1, 2 ou 3). Ce retard TP est caractérisé par une valeur différentielle DP dans le domaine des index de quantification, codée sur 5 bits si MV=1 ou 2 (N1=32), et sur 4 bits si MV=3 (N3=16). L'index du retard TP vaut ZP+DP. De façon connue, l'analyse LTP en boucle fermée consiste à déterminer, dans l'intervalle de recherche des retards T de prédiction à long terme, le retard TP qui maximise, pour chaque sous-trame d'une trame voisée, la corrélation normalisée :

$$\frac{\left[\sum_{i=0}^{lst-1} x(i) \cdot y_T(i)\right]^2}{\sum_{i=0}^{lst-1} [y_T(i)]^2}$$

où x(i) désigne le signal de parole pondéré SW de la sous-trame auquel on a soustrait la mémoire du filtre de synthèse pondéré (c'est-à-dire la réponse à un signal nul, due à ses états initiaux, du filtre dont la réponse impulsionnelle h a été calculée par le module 42), et $y_T(i)$ désigne le produit de convolution :

$$y_T(i) = u(i-T) * h(i) = \sum_{j=0}^{i} u(j-T) \cdot h(i-j) \qquad (1)$$

u(j-T) désignant la composante prédictible de la séquence d'excitation retardée de T échantillons, estimée par la technique bien connue du répertoire adaptatif ("adaptive codebook"). Pour les retards T inférieurs à la longueur d'une sous-trame, les valeurs manquantes de u(j-T) peuvent être extrapolées à partir des valeurs antérieures. Les retards fractionnaires sont pris en compte en suréchantillonnant le signal u(j-T) dans le répertoire adaptatif. Un suréchantillonnage d'un facteur m est obtenu au moyen de filtres polyphasés interpolateurs.

[0023] Le gain g_P de prédiction à long terme pourrait être déterminé par le module 38 pour chaque sous-trame, en appliquant la formule connue :

$$g_P = \frac{\sum_{i=0}^{lst-1} x(i) \cdot y_{TP}(i)}{\sum_{i=0}^{lst-1} [y_{TP}(i)]^2}$$

Toutefois, dans une version préférée de l'invention, le gain g_P est calculé par le module d'analyse stochastique 40.

[0024] L'excitation stochastique déterminée pour chaque sous-trame par le module 40 est de type multi-impulsionnelle. Une séquence d'innovation de lst échantillons comprend np impulsions de positions p(n) et d'amplitude g(n). Autrement dit, les impulsions ont une amplitude de 1 et sont associées à des gains respectifs g(n). Etant donné que le retard LTP n'est pas déterminé pour les sous-trames des trames non voisées, on peut prendre un nombre d'impulsions supérieur pour l'excitation stochastique relative à ces sous-trames, par exemple np=5 si MV=1, 2 ou 3 et np=6 si MV=0. Les positions et les gains calculés par le module 40 d'analyse stochastique sont quantifiés par un module 44.

[0025] Un module d'ordonnancement des bits 46 reçoit les différents paramètres qui seront utiles au décodeur, et constitue la séquence binaire transmise au codeur canal 22. Ces paramètres sont :

- l'index Q des paramètres LSP quantifiés pour chaque trame ;
- le degré MV de voisement de chaque trame ;
- l'index ZP du centre de l'intervalle de recherche des retards LTP pour chaque trame voisée ;

- l'index différentiel DP du retard LTP pour chaque sous-trame d'une trame voisée, et le gain associé g_P ;
- les positions p(n) et les gains g(n) des impulsions de l'excitation stochastique pour chaque sous-trame.

[0026] Certains de ces paramètres peuvent avoir une importance particulière dans la qualité de restitution de la parole ou une sensibilité particulière aux erreurs de transmission. On prévoit ainsi dans le codeur un module 48 qui reçoit les différents paramètres et qui ajoute à certains d'entre eux des bits de redondance permettant de détecter et/ou de corriger d'éventuelles erreurs de transmission. Par exemple, le degré de voisement MV codé sur deux bits étant un paramètre critique, on souhaite qu'il parvienne au décodeur avec aussi peu d'erreurs que possible. Pour cette raison, des bits de redondance sont ajoutés à ce paramètre par le module 48. On peut par exemple ajouter un bit de parité aux deux bits codant MV et répéter une fois les trois bits ainsi obtenus. Cet exemple de redondance permet de détecter toutes les erreurs simples ou doubles et de corriger toutes les erreurs simples et 75% des erreurs doubles.

[0027] L'allocation du débit binaire par trame de 20 ms est par exemple celle indiquée dans le tableau I.

[0028] Dans l'exemple considéré ici, le codeur canal 22 est celui utilisé dans le système paneuropéen de radiocommunication avec les mobiles (GSM). Ce codeur canal, décrit en détail dans la Recommandation GSM 05.03, a été mis au point pour un codeur de parole à 13 kbit/s de type RPE-LTP qui produit également 260 bits par trame de 20 ms. La sensibilité de chacun des 260 bits a été déterminée à partir de tests d'écoute. Les bits issus du codeur source ont été regroupés en trois catégories. La première de ces catégories IA regroupe 50 bits qui sont codés convolutionnellement sur la base d'un polynôme générateur donnant une redondance d'un demi avec une longueur de contrainte égale à 5. Trois bits de parité sont calculés et ajoutés aux 50 bits de la catégorie IA avant le codage convolutionnel. La seconde catégorie (IB) compte 132 bits qui sont protégés à un taux d'un demi par le même polynôme que la catégorie précédente. La troisième catégorie (II) contient 78 bits non protégés. Après application du code convolutionnel, les bits (456 par trame) sont soumis à un entrelacement. Le module d'ordonnancement 46 du nouveau codeur source mettant en oeuvre l'invention distribue les bits dans les trois catégories en fonction de l'importance subjective de ces bits.

TABLEAU I

paramètres quantifiés	MV=0	MV=1 ou 2	MV=3
LSP	34	34	34
MV + redondance	6	6	6
ZP	-	8	8
DP	-	20	16
g_{TP}	-	20	24
positions impulsions	80	72	72
gains impulsions	140	100	100
Total	260	260	260

[0029] Une station mobile de radiocommunication apte à recevoir le signal de parole traité par le codeur source 16 est représentée schématiquement sur la figure 2. Le signal radio reçu est d'abord traité par un démodulateur 50 puis par un décodeur canal 52 qui effectuent les opérations duales de celles du modulateur 24 et du codeur canal 22. Le décodeur canal 52 fournit au décodeur de parole 54 une séquence binaire qui, en l'absence d'erreurs de transmission ou lorsque les éventuelles erreurs ont été corrigées par le décodeur canal 52, correspond à la séquence binaire qu'a délivrée le module d'ordonnancement 46 au niveau du codeur 16. Le décodeur 54 comprend un module 56 qui reçoit cette séquence binaire et qui identifie les paramètres relatifs aux différentes trames et sous-trames. Le module 56 effectue en outre quelques contrôles sur les paramètres reçus. En particulier, le module 56 examine les bits de redondance introduits par le module 48 du codeur, pour détecter et/ou corriger les erreurs affectant les paramètres associés à ces bits de redondance.

[0030] Pour chaque trame de parole à synthétiser, un module 58 du décodeur reçoit le degré de voisement MV et l'index de Q de quantification des paramètres LSP. Le module 58 retrouve les paramètres LSP quantifiés dans les tables correspondant à la valeur de MV, et, après interpolation, les convertit en coefficients a_i pour le filtre de synthèse à court terme 60. Pour chaque sous-trame de parole à synthétiser, un générateur d'impulsions 62 reçoit les positions p(n) des np impulsions de l'excitation stochastique. Le générateur 62 délivre des impulsions d'amplitude unitaire qui sont chacune multipliées en 64 par le gain associé g(n). La sortie de l'amplificateur 64 est adressée au filtre de synthèse à long terme 66. Ce filtre 66 a une structure à répertoire adaptatif. Les échantillons u de sortie du filtre 66 sont mémorisés dans le répertoire adaptatif 68 de façon à être disponibles pour les sous-trames ultérieures. Le retard TP relatif à une

sous-trame, calculé à partir des index de quantification ZP et DP, est fourni au répertoire adaptatif 68 pour produire le signal u convenablement retardé. L'amplificateur 70 multiplie le signal ainsi retardé par le gain g_P de prédiction à long terme. Le filtre à long terme 66 comprend enfin un additionneur 72 qui ajoute les sorties des amplificateurs 64 et 70 pour fournir la séquence d'excitation u. Lorsque l'analyse LTP n'a pas été effectuée au codeur, par exemple si MV=0, un gain de prédiction g_P nul est imposé à l'amplificateur 70 pour les sous-trames correspondantes. La séquence d'excitation est adressée au filtre de synthèse à court terme 60, et le signal résultant peut encore, de façon connue, être soumis à un post-filtre 74 dont les coefficients dépendent des paramètres de synthèse reçus, pour former le signal de parole synthétique S'. Le signal de sortie S' du décodeur 54 est ensuite converti en analogique par le convertisseur 76 avant d'être amplifié pour commander un haut-parleur 78.

[0031] On va maintenant décrire, en référence au figures 3 à 6, le processus d'analyse LTP en boucle ouverte mis en oeuvre par le module 36 du codeur suivant un premier aspect de l'invention.

[0032] Dans une première étape 90, le module 36 calcule et mémorise, pour chaque sous-trame st=0,1,...,nst-1 de la trame courante, les autocorrélations $C_{st}(k)$ et les énergies retardées $G_{st}(k)$ du signal de parole pondéré SW pour les retards entiers k compris entre rmin et rmax :

$$C_{st}(k) = \sum_{i=st.lst}^{(st+1).lst-1} SW(i).SW(i-k)$$

$$G_{st}(k) = \sum_{i=st.lst}^{(st+1).lst-1} [SW(i-k)]^2$$

[0033] Les énergies par sous-trame $R0_{st}$ sont également calculées :

$$RO_{st} = \sum_{i=st.lst}^{(st+1).lst-1} [SW(i)]^2$$

[0034] A l'étape 90, le module 36 détermine en outre, pour chaque sous-trame st, le retard entier K_{st} qui maximise l'estimation en boucle ouverte $P_{st}(k)$ au gain de prédiction à long terme sur la sous-trame st, en excluant les retards k pour lesquels l'autocorrélation $C_{st}(k)$ est négative ou plus petite qu'une petite fraction ε de l'énergie $R0_{st}$ de la sous-trame. L'estimation $P_{st}(k)$ exprimée en décibels s'écrit :

$$P_{st}(k) = 20.log_{10}[RO_{st}/(RO_{st}-C_{st}^2(k)/G_{st}(k))]$$

Maximiser $P_{st}(k)$ revient donc à maximiser l'expression $X_{st}(k)=C_{st}^2(k)/G_{st}(k)$ comme indiqué sur la figure 6. Le retard entier K_{st} est le retard de base en résolution entière pour la sous-trame st. L'étape 90 est suivie par une comparaison 92 entre une première estimation en boucle ouverte du gain de prédiction global sur la trame courante et un seuil prédéterminé S0 typiquement compris entre 1 et 2 décibels (par exemple S0 =1,5 dB). La première estimation du gain de prédiction global est égale à :

$$20.\log_{10}\left[RO/\left[RO - \sum_{st=0}^{nst-1} X_{st}(K_{st})\right]\right]$$

où R0 est l'énergie totale de la trame ($R0 = R0_0 + R0_1 + ... + R0_{nst-1}$), et $X_{st}(K_{st})=C_{st}^2(K_{st})/G_{st}(K_{st})$ désigne le maximum déterminé à l'étape 90 relativement à la sous-trame st. Comme l'indique la figure 6, la comparaison 92 peut être effectuée sans avoir à calculer le logarithme.

[0035] Si la comparaison 92 montre une première estimation du gain de prédiction inférieure au seuil S0, on considère que le signal de parole contient trop peu de corrélations à long terme pour être voisé, et le degré de voisement MV de la trame courante est pris égal à 0 à l'étape 94, ce qui termine dans ce cas les opérations effectuées par le module 36 sur cette trame. Si au contraire le seuil S0 est dépassé à l'étape 92, la trame courante est détectée comme voisée et le degré MV sera égal à 1, 2 ou 3. Le module 36 calcule alors, pour chaque sous-trame st, une liste I_{st} contenant des retards candidats pour constituer le centre ZP de l'intervalle de recherche des retards de prédiction à long terme.

[0036] Les opérations effectuées par le module 36 pour chaque sous-trame st (st initialisé à 0 à l'étape 96) d'une trame voisée commencent par la détermination 98 d'un seuil de sélection SE_{st} en décibels égal à une fraction déterminée β de l'estimation $P_{st}(K_{st})$ du gain de prédiction en décibels sur la sous-trame, maximisée à l'étape 90 (β=0,75 typiquement). Pour chaque sous-trame st d'une trame voisée, le module 36 détermine le retard de base rbf en résolution entière pour la suite du traitement. Ce retard de base pourrait être pris égal à l'entier K_{st} obtenu à l'étape 90. Le fait de rechercher le retard de base en résolution fractionnaire autour de K_{st} permet toutefois de gagner en précision. L'étape 100 consiste ainsi, à rechercher, autour du retard entier K_{st} obtenu à l'étape 90, le retard fractionnaire qui maximise l'expression C_{st}^2/G_{st}. Cette recherche peut être effectuée à la résolution maximale des retards fractionnaires (1/6 dans l'exemple décrit ici) même si le retard entier K_{st} n'est pas dans le domaine où cette résolution maximale s'applique. On détermine par exemple le nombre Δ_{st} qui maximise $C_{st}^2(K_{st}+\delta/6)/G_{st}(K_{st}+\delta/6)$ pour $-6<\delta<+6$, puis le retard de base rbf en résolution maximale est pris égal à $K_{st}+\Delta_{st}/6$. Pour les valeurs fractionnaires T du retard, les autocorrélations $C_{st}(T)$ et les énergies retardées $G_{st}(T)$ sont obtenues par interpolation à partir des valeurs mémorisées à l'étape 90 pour les retards entiers. Bien entendu, le retard de base relatif à une sous-trame pourrait également être déterminé en résolution fractionnaire dès l'étape 90 et pris en compte dans la première estimation du gain de prédiction global sur la trame.

[0037] Une fois que le retard de base rbf a été déterminé pour une sous-trame, on procède à un examen 101 des sous-multiples de ce retard afin de retenir ceux pour lesquels le gain de prédiction est relativement important (figure 4), puis des multiples du plus petit sous-multiple retenu (figure 5). A l'étape 102, l'adresse j dans la liste I_{st} et l'index m du sous-multiple sont initialisés à 0 et 1, respectivement. Une comparaison 104 est effectuée entre le sous-multiple rbf/m et le retard minimal rmin. Le sous-multiple rbf/m est à examiner s'il est supérieur à rmin. On prend alors pour l'entier i la valeur de l'index du retard quantifié r_i le plus proche de rbf/m (étape 106), puis on compare, en 108, la valeur estimée du gain de prédiction $P_{st}(r_i)$ associée au retard quantifié r_i pour la sous-trame considérée au seuil de sélection SE_{st} calculé à l'étape 98 :

$$P_{st}(r_i) = 20.log_{10}[RO_{st}/[RO_{st} - C_{st}^2(r_i)/G_{st}(r_i)]]$$

avec, pour les retards fractionnaires une interpolation des valeurs C_{st} et G_{st} calculées à l'étape 90 pour les retards entiers. Si $P_{st}(r_i)<SE_{st}$, le retard r_i n'est pas pris en considération, et on passe directement à l'étape 110 d'incrémentation de l'index m avant d'effectuer de nouveau la comparaison 104 pour le sous-multiple suivant. Si le test 108 montre que $P_{st}(r_i) \geq SE_{st}$, le retard r_i est retenu et on exécute l'étape 112 avant d'incrémenter l'index m à l'étape 110. A l'étape 112, on mémorise l'index i à l'adresse j dans la liste I_{st}, on donne la valeur m à l'entier m0 destiné à être égal à l'index du plus petit sous-multiple retenu, puis on incrémente d'une unité l'adresse j.

[0038] L'examen des sous-multiples du retard de base est terminé lorsque la comparaison 104 montre rbf/m < rmin. On examine alors les retards multiples au plus petit rbf/m0 des sous-multiples précédemment retenus suivant le processus illustré sur la figure 5. Cet examen commence par une initialisation 114 de l'index n du multiple : n=2. Une comparaison 116 est effectuée entre le multiple n.rbf/m0 et le retard maximal rmax. Si n.rbf/m0 > rmax, on effectue le test 118 pour déterminer si l'index m0 du plus petit sous-multiple est un multiple entier de n. Dans l'affirmative, le retard n.rbf/m0 a déjà été examiné lors de l'examen des sous-multiples de rbf, et on passe directement à l'étape 120 d'incrémentation de l'index n avant d'effectuer de nouveau la comparaison 116 pour le multiple suivant. Si le test 118 montre que m0 n'est pas un multiple entier de n, le multiple n.rbf/m0 est à examiner. On prend alors pour l'entier i la valeur de l'index du retard quantifié r_i le plus proche de n.rbf/m0 (étape 122), puis on compare, en 124, la valeur estimée du gain de prédiction $P_{st}(r_i)$ au seuil de sélection SE_{st}. Si $P_{st}(r_i)<SE_{st}$, le retard r_i n'est pas pris en considération, et on passe directement à l'étape 120 d'incrémentation de l'index n. Si le test 124 montre que $P_{st}(r_i) \geq SE_{st}$, le retard r_i est retenu et on exécute l'étape 126 avant d'incrémenter l'index n à l'étape 120. A l'étape 126, on mémorise l'index i à l'adresse j dans la liste I_{st}, puis on incrémente d'une unité l'adresse j.

[0039] L'examen des multiples du plus petit sous-multiple est terminé lorsque la comparaison 116 montre que n.rbf/m0>rmax. A ce moment, la liste I_{st} contient j index de retards candidats. Si on souhaite limiter à jmax la longueur maximale de la liste I_{st} pour les étapes suivantes, on peut prendre la longueur j_{st} de cette liste égale à min(j,jmax) (étape 128) puis, à l'étape 130, ordonner la liste I_{st} dans l'ordre des gains $C_{st}^2(r_{Ist(j)})/G_{st}^2(r_{Ist(j)})$ décroissants pour $0 \leq j < j_{st}$ de façon à ne conserver que les j_{st} retards procurant les plus grandes valeurs de gain. La valeur de jmax est choisie en fonction du compromis visé entre l'efficacité de la recherche des retards LTP et la complexité de cette recherche.

Des valeurs typiques de jmax vont de 3 à 5.

[0040] Une fois que les sous-multiples et les multiples ont été examinés et que la liste l_{st} a ainsi été obtenue (figure 3), le module d'analyse 36 calcule une quantité Ymax déterminant une seconde estimation en boucle ouverte du gain de prédiction à long terme sur l'ensemble de la trame, ainsi que des index ZP, ZP0 et ZP1 dans une phase 132 dont le déroulement est détaillé sur la figure 6. Cette phase 132 consiste à tester des intervalles de recherche de longueur N1 pour déterminer celui qui maximise une deuxième estimation du gain de prédiction global sur la trame. Les intervalles testés sont ceux dont les centres sont les retards candidats contenus dans la liste l_{st} calculée lors de la phase 101. La phase 132 commence par une étape 136 où l'adresse j dans la liste l_{st} est initialisée à 0. A l'étape 138, on vérifie si l'index $l_{st}(j)$ a déjà été rencontré en testant un intervalle précédent centré sur $l_{st'}(j')$ avec st'<st et $0 \leq j' < j_{st'}$, afin d'éviter de tester deux fois le même intervalle. Si le test 138 révèle que $l_{st}(j)$ figurait déjà dans une liste $l_{st'}$ avec st'<st, on incrémente directement l'adresse j à l'étape 140, puis on la compare à la longueur j_{st} de la liste l_{st}. Si la comparaison 142 montre que $j<j_{st}$, on revient à l'étape 138 pour la nouvelle valeur de l'adresse j. Lorsque la comparaison 142 montre que $j=j_{st}$, tous les intervalles relatifs à la liste l_{st} ont été testés, et la phase 132 est terminée. Lorsque le test 138 est négatif, on teste l'intervalle centré sur $l_{st}(j)$ en commençant par l'étape 148 où on détermine, pour chaque sous-trame st', l'index $i_{st'}$ du retard optimal qui maximise sur cet intervalle l'estimation en boucle ouverte $P_{st'}(r_i)$ du gain de prédiction à long terme, c'est-à-dire qui maximise la quantité $Y_{st'}(i)=C_{st'}^2(r_i)/G_{st'}(r_i)$ où r_i désigne le retard quantifié d'index i pour $l_{st}(j)-N1/2 \leq i < l_{st}(j)+N1/2$ et $0 \leq i < N$. Lors de la maximisation 148 relative à une sous-trame st', on écarte a priori les index i pour lesquels l'autocorrélation $C_{st'}(r_i)$ est négative, pour éviter de dégrader le codage. S'il se trouve que toutes les valeurs de i comprises dans l'intervalle testé [l(j)-N1/2, l(j)+N1/2[donnent lieu à des autocorrélations $C_{st'}(r_i)$ négatives, on sélectionne l'index $i_{st'}$ pour lequel cette autocorrélation est la plus petite en valeur absolue. Ensuite, en 150, la quantité Y déterminant la deuxième estimation du gain de prédiction global pour l'intervalle centré sur $l_{st}(j)$ est calculée selon :

$$Y = \sum_{st'=0}^{nst-1} Y_{st'}(i_{st'})$$

puis comparée à Ymax, où Ymax représente la valeur à maximiser. Cette valeur Ymax est par exemple initialisée à 0 en même temps que l'index st à l'étape 96. Si $Y \leq Ymax$, on passe directement à l'étape 140 d'incrémentation de l'index j. Si la comparaison 150 montre que Y > Ymax, on exécute l'étape 152 avant d'incrémenter l'adresse j à l'étape 140. A cette étape 152, l'index ZP est pris égal à $l_{st}(j)$ et les index ZP0 et ZP1 sont respectivement pris égaux au plus petit et au plus grand des index $i_{st'}$ déterminés à l'étape 148.

[0041] A la fin de la phase 132 relative à une sous-trame st, l'index st est incrémenté d'une unité (étape 154) puis comparé, à l'étape 156, au nombre nst de sous-trames par trame. Si st<nst, on revient à l'étape 98 pour effectuer les opérations relatives à la sous-trame suivante. Lorsque la comparaison 156 montre que st=nst, l'index ZP désigne le centre de l'intervalle de recherche qui sera fourni au module 38 d'analyse LTP en boucle fermée, et ZP0 et ZP1 sont des index dont l'écart est représentatif de la dispersion des retards optimaux par sous-trame dans l'intervalle centré sur ZP.

[0042] A l'étape 158, le module 36 détermine le degré de voisement MV, sur la base de la seconde estimation en boucle ouverte du gain exprimée en décibels : $Gp=20.\log_{10}(RO/RO\text{-}Y_{max})$. On fait appel à deux autres seuils S1 et S2. Si $Gp \leq S1$, le degré de voisement MV est pris égal à 1 pour la trame courante. Le seuil S1 est typiquement compris entre 3 et 5 dB; par exemple S1=4 dB. Si S1<Gp<S2, le degré de voisement MV est pris égal à 2 pour la trame courante. Le seuil S2 est typiquement compris entre 5 et 8 dB; par exemple S2=7 dB. Si Gp>S2, on examine la dispersion des retards optimaux pour les différentes sous-trames de la trame courante. Si ZP1-ZP<N3/2 et ZP-ZP0≤N3/2, un intervalle de longueur N3 centré sur ZP suffit à prendre en compte tous les retards optimaux et le degré de voisement est pris égal à 3 (si Gp>S2). Sinon, si ZP1-ZP≥N3/2 ou ZP-ZP0>N3/2, le degré de voisement est pris égal à 2 (si Gp>S2).

[0043] L'index ZP du centre de l'intervalle de recherche du retard de prédiction pour une trame voisée peut être compris entre 0 et N-1=255, et l'index différentiel DP déterminé pour le module 38 peut aller de -16 à +15 si MV=1 ou 2, et de -8 à +7 si MV=3 (cas N1=32, N3=16). L'index ZP+DP du retard TP finalement déterminé peut donc dans certains cas être plus petit que 0 ou plus grand que 255. Ceci permet à l'analyse LTP en boucle fermée de porter également sur quelques retards TP plus petits que rmin ou plus grands que rmax. On améliore ainsi la qualité subjective de la restitution des voix dites pathologiques et des signaux non vocaux (fréquences vocales DTMF ou fréquences de signalisation utilisées par le réseau téléphonique commuté). Une autre possibilité est de prendre pour l'intervalle de recherche les 32 premiers ou derniers index de quantification des retards si ZP<16 ou ZP>240 avec MV=1 ou 2, et les 16 premiers ou derniers index si ZP<8 ou ZP>248 avec MV=3.

[0044] Le fait de réduire l'intervalle de recherche des retards pour les trames très voisées (typiquement 16 valeurs

pour MV=3 au lieu de 32 pour MV=1 ou 2) permet de diminuer la complexité de l'analyse LTP en boucle fermée effectuée par le module 38 en réduisant le nombre de convolutions $y_T(i)$ à calculer suivant la formule (1). Un autre avantage est qu'un bit de codage de l'index différentiel DP est économisé. Le débit de sortie étant constant, ce bit peut être réalloué au codage d'autres paramètres. On peut en particulier allouer ce bit supplémentaire à la quantification au gain de prédiction à long terme g_P calculé par le module 40. En effet, une meilleure précision sur le gain g_P grâce à un bit de quantification supplémentaire est appréciable car ce paramètre est perceptuellement important pour les sous-trames très voisées (MV=3). Une autre possibilité est de prévoir un bit de parité pour le retard TP et/ou le gain g_P, permettant de détecter d'éventuelles erreurs affectant ces paramètres.

[0045] Il est possible d'apporter quelques modifications au processus d'analyse LTP en boucle ouverte décrit ci-dessus en référence aux figures 3 à 6.

[0046] Suivant une première variante de ce processus, les premières optimisations effectuées à l'étape 90 relativement aux différentes sous-trames sont rempiacées par une seule optimisation portant sur l'ensemble de la trame. Outre les paramètres $C_{st}(k)$ et $G_{st}(k)$ calculés pour chaque sous-trame st, on calcule également les autocorrélations C(k) et les énergies retardées G(k) pour l'ensemble de la trame :

$$C(k) = \sum_{st=0}^{nst-1} C_{st}(k)$$

$$G(k) = \sum_{st=0}^{nst-1} G_{st}(k)$$

[0047] On détermine alors le retard de base en résolution entière K qui maximise $X(k)=C^2(k)/G(k)$ pour $rmin \leq k \leq rmax$. La première estimation du gain comparée à S0 à l'étape 92 est alors $P(K)=20.\log_{10}[R0/[R0-X(K)]]$. On détermine ensuite, autour de K, un seul retard de base en résolution fractionnaire rbf et l'examen 101 des sous-multiples et des multiples est effectué une seule fois et produit une seule liste l au lieu de nst listes l_{st}. La phase 132 est ensuite effectuée une seule fois pour cette liste l, en ne distinguant les sous-trames qu'aux étapes 148, 150 et 152. Cette variante de réalisation a pour avantage de réduire la complexité de l'analyse en boucle ouverte.

[0048] Suivant une seconde variante du processus d'analyse LTP en boucle ouverte, le domaine [rmin, rmax] des retards possibles est subdivisé en nz sous-intervalles ayant par exemple la même longueur (nz=3 typiquement), et les premières optimisations effectuées à l'étape 90 relativement aux différentes sous-trames sont remplacées par nz optimisations dans les différents sous-intervalles portant chacune sur l'ensemble de la trame. On obtient ainsi nz retards de base $K_1', \ldots, K_{nz}'$ en résolution entière. La décision voisé/non voisé (étape 92) est prise sur la base de celui des retards de base K_i' qui procure la plus grande valeur pour la première estimation en boucle ouverte au gain de prédiction à long terme. Ensuite, si la trame est voisée, on détermine les retards de base en résolution fractionnaire par le même processus qu'à l'étape 100, mais en autorisant seulement les valeurs de retard quantifiées. L'examen 101 des sous-multiples et des multiples n'est pas effectué. Pour la phase 132 de calcul de la seconde estimation du gain de prédiction, on prend comme retards candidats les nz retards de base précédemment déterminés. Cette seconde variante permet de se dispenser de l'examen systématique des sous-multiples et des multiples qui sont en général pris en considération grâce à la subdivision du domaine des retards possibles.

[0049] Suivant une troisième variante du processus d'analyse LTP en boucle ouverte, la phase 132 est modifiée en ce que, aux étapes d'optimisation 148, on détermine d'une part l'index $i_{st'}$ qui maximise $C_{st}^2(r_i)/G_{st'}(r_i)$ pour $l_{st}(j)-N1/2 \leq i < l_{st}(j)+N1/2$ et $0 \leq i < N$, et d'autre part, au cours de la même boucle de maximisation, l'index $k_{st'}$ qui maximise cette même quantité sur un intervalle réduit $l_{st}(j)-N3/2 \leq i < l_{st}(j)+N3/2$ et $0 \leq i < N$. L'étape 152 est également modifiée : on ne mémorise plus les index ZP0 et ZP1, mais une quantité Ymax' définie de la même manière que Ymax mais en référence à l'intervalle de longueur réduite :

$$Y_{max}' = \sum_{st'=0}^{nst-1} Y_{st'}(k_{st'})$$

[0050] Dans cette troisième variante, la détermination 158 du mode de voisement conduit à sélectionner plus souvent le degré de voisement MV=3. On prend également en compte, en plus du gain Gp précédemment décrit, une troisième estimation en boucle ouverte du gain LTP, correspondant à Ymax': $Gp'=20.\log_{10}[R0/(R0-Ymax')]$. Le degré de voisement est MV=1 si Gp≤S1, MV=3 si Gp'>S2 et MV=2 si aucune de ces deux conditions n'est vérifiée. En augmentant ainsi la proportion de trames de degré MV=3, on réduit la complexité moyenne de l'analyse en boucle fermée et on améliore la robustesse aux erreurs de transmission.

[0051] Une quatrième variante du processus d'analyse LTP en boucle ouverte concerne surtout les trames faiblement voisées (MV=1). Ces trames correspondent souvent à un début ou à une fin d'une zone de voisement. Fréquemment, ces trames peuvent comporter de une à trois sous-trames pour lesquelles le coefficient de gain du filtre de synthèse à long terme est nul voire négatif. Il est proposé de ne pas effectuer l'analyse LTP en boucle fermée pour les sous-trames en question, afin de réduire la complexité moyenne du codage. Ceci peut être réalisé en mémorisant à l'étape 152 de la figure 6 nst pointeurs indiquant pour chaque sous-trame st' si l'autocorrélation $C_{st'}$ correspondant au retard d'index $i_{st'}$ est négative ou encore très petite. Une fois que tous les intervalles référencés dans les listes $I_{st'}$ les sous-trames pour lesquelles le gain de prédiction est négatif ou négligeable peuvent être identifiées en consultant les nst pointeurs. Le cas échéant le module 38 est désactivé pour les sous-trames correspondantes. Ceci n'affecte pas la qualité de l'analyse LTP puisque le gain de prédiction correspondant à ces sous-trames sera de toutes façons quasiment nul.

[0052] Un autre aspect de l'invention concerne le module 42 de calcul de la réponse impulsionnelle du filtre de synthèse pondéré. Le module 38 d'analyse LTP en boucle fermée a besoin de cette réponse impulsionnelle h sur la durée d'une sous-trame pour calculer les convolutions $y_T(i)$ selon la formule (1). Le module 40 d'analyse stochastique en a également besoin pour calculer des convolutions comme on le verra plus loin. Le fait d'avoir à calculer des convolutions avec une réponse h s'étendant sur la durée d'une sous-trame (lst=40 typiquement) implique une relative complexité du codage, qu'il serait souhaitable de réduire notamment pour augmenter l'autonomie de la station mobile. Dans certains cas il a été proposé de tronquer la réponse impulsionnelle à une longueur inférieure à la longueur d'une sous-trame (par exemple à 20 échantillons), mais ceci peut dégrader la qualité du codage. On propose selon l'invention de tronquer la réponse impulsionnelle h en tenant compte d'une part de la distribution énergétique de cette réponse et d'autre part du degré de voisement MV de la trame considérée, déterminé par le module 36 d'analyse LTP en boucle ouverte.

[0053] Les opérations effectuées par le module 42 sont par exemple conformes à l'organigramme de la figure 7. La réponse impulsionnelle est d'abord calculée à l'étape 160 sur une longueur pst supérieure à la longueur d'une sous-trame et suffisamment grande pour qu'on soit assuré de prendre en compte toute l'énergie de la réponse impulsionnelle (par exemple pst=60 pour nst=4 et lst=40 si la prédiction linéaire à court terme est d'ordre q=10). A l'étape 160, on calcule également les énergies tronquées de la réponse impulsionnelle :

$$Eh(i) = \sum_{k=0}^{i} [h(i)]^2$$

[0054] Les composantes h(i) de la réponse impulsionnelle et les énergies tronquées Eh(i) peuvent être obtenues en filtrant une impulsion unitaire au moyen d'un filtre de fonction de transfert W(z)/A(z) d'états initiaux nuls, ou encore par récurrence :

$$f(i) = \delta(i) + \sum_{k=1}^{q} a_k [\gamma_2^k . f(i-k) - \gamma_1^k . \delta(i-k)] \qquad (2)$$

$$h(i) = f(i) + \sum_{k=1}^{q} a_k . h(i-k) \qquad (3)$$

$$Eh(i) = Eh(i-1) + [h(i)]^2$$

pour 0<i<pst, avec f(i)=h(i)=0 pour i<0, δ(0)=f(0)=h(0)=Eh(0)=1, et δ(i)=0 pour i≠0. Dans l'expression (2), les coefficients

a_k sont ceux intervenant dans le filtre de pondération perceptuelle, c'est-à-dire les coefficients de prédiction linéaire interpolés mais non quantifiés, tandis que dans l'expression (3), les coefficients a_k sont ceux appliqués au filtre de synthèse, c'est-à-dire les coefficients de prédiction linéaire quantifiés et interpolés.

[0055] Ensuite le module 42 détermine la plus petite longueur $L\alpha$ telle que l'énergie $Eh(L\alpha-1)$ de la réponse impulsionnelle tronquée à $L\alpha$ échantillons soit au moins égale à une proportion α de son énergie totale $Eh(pst-1)$ estimée sur pst échantillons. Une valeur typique de α est 98%. Le nombre $L\alpha$ est initialisé à pst à l'étape 162 et décrémenté d'une unité en 166 tant que $Eh(L\alpha-2)>\alpha.Eh(pst-1)$ (test 164). La longueur $L\alpha$ cherchée est obtenue lorsque le test 164 montre que $Eh(L\alpha-2)\leq\alpha.Eh(pst-1)$.

[0056] Pour tenir compte du degré de voisement MV, un terme correcteur $\Delta(MV)$ est ajouté à la valeur de $L\alpha$ qui a été obtenue (étape 168). Ce terme correcteur est de préférence une fonction croissante du degré de voisement. On peut par exemple prendre $\Delta(0)=-5$, $\Delta(1)=0$, $\Delta(2)=+5$ et $\Delta(3)=+7$. De cette façon, la réponse impulsionnelle h sera déterminée de façon d'autant plus précise que le voisement de la parole est important. La longueur de troncature Lh de la réponse impulsionnelle est prise égale à $L\alpha$ si $L\alpha\leq nst$ et à nst sinon. Les échantillons restants de la réponse impulsionnelle ($h(i)=0$ avec $i\geq Lh$) peuvent être annulés.

[0057] Avec la troncature de la réponse impulsionnelle, le calcul (1) des convolutions $y_T(i)$ par le module 33 d'analyse LTP en boucle fermée est modifié de la façon suivante :

$$y_T(i) = \sum_{j=\max(0,\,i-Lh+1)}^{i} u(j-T).h(i-j) \qquad (1')$$

[0058] L'obtention de ces convolutions, qui représente une part importante des calculs effectués, nécessite donc sensiblement moins de multiplications, d'additions et d'adressages dans le répertoire adaptatif lorsque la réponse impulsionnelle est tronquée. La troncature dynamique de la réponse impulsionnelle faisant intervenir le degré de voisement MV permet d'obtenir une telle réduction de complexité sans affecter la qualité du codage. Les mêmes considérations s'appliquent pour les calculs de convolutions effectués par le module 40 d'analyse stochastique. Ces avantages sont particulièrement appréciables lorsque le filtre de pondération perceptuelle a une fonction de transfert de la forme $W(z)=A(z/\gamma_1)/A(z/\gamma_2)$ avec $0<\gamma_2<\gamma_1<1$ qui donne lieu à des réponses impulsionnelles généralement plus longues que celles de la forme $W(z)=A(z)/A(z/\gamma)$ plus communément employées dans les codeurs à analyse par synthèse.

[0059] Un troisième aspect de l'invention concerne le module 40 d'analyse stochastique servant à modéliser la partie non prédictible de l'excitation.

[0060] L'excitation stochastique considérée ici est de type multi-impulsionnelle. L'excitation stochastique relative à une sous-trame est représentée par np impulsions de positions p(n) et d'amplitudes, ou gains, g(n) ($1\leq n\leq np$). Le gain g_P de prédiction à long terme peut également être calculé au cours du même processus. De façon générale, on peut considérer que la séquence d'excitation relative à une sous-trame comporte nc contributions associées respectivement à nc gains. Les contributions sont des vecteurs lst échantillons qui, pondérés par les gains associés et sommés correspondent à la séquence d'excitation du filtre de synthèse à court terme. Une des contributions peut être prédictible, ou plusieurs dans le cas d'un filtre de synthèse à long terme à plusieurs prises ("multi-tap pitch synthesis filter"). Les autres contributions sont dans le cas présent np vecteurs ne comportant que des 0 sauf une impulsion d'amplitude 1. On a donc nc=np si MV=0, et nc=np+1 si MV=1, 2 ou 3.

[0061] L'analyse multi-impulsionnelle incluant le calcul du gain $g_P=g(0)$ consiste, de façon connue, à trouver pour chaque sous-trame des positions p(n) ($1\leq n\leq np$) et des gains g(n) ($0\leq n\leq np$) qui minimisent l'erreur quadratique pondérée perceptuellement E entre le signal de parole et le signal synthétisé, donnée par :

$$E = \left(X - \sum_{n=0}^{nc-1} g(n).F_{p(n)}\right)^2$$

les gains étant solution du système linéaire g.B=b.

[0062] Dans les notations ci-dessus :

- X désigne un vecteur-cible initial composé des lst échantillons du signal de parole pondéré SW sans mémoire : $X=(x(0),x(1),\ldots,x(lst-1))$, les x(i) ayant été calculés comme indiqué précédemment lors de l'analyse LTP en boucle fermée ;

- g désigne le vecteur ligne composé des np+1 gains : $g=(g(0)=g_P, g(1),...,g(np))$;
- les vecteurs-ligne $F_{p(n)}$ ($0\leq n\leq nc$) sont des contributions pondérées ayant pour composantes $i(0\leq i<lst)$ les produits de convolution entre la contribution n à la séquence d'excitation et la réponse impulsionnelle h du filtre de synthèse pondéré ;
- b désigne le vecteur ligne composé des nc produits scalaires entre le vecteur X et les vecteurs ligne $F_{p(n)}$;
- B désigne une matrice symétrique à nc lignes et nc colonnes dont le terme $B_{i,j}=F_{p(i)}.F_{p(j)}^{T}$ ($0\leq i,j<nc$) est égal au produit scalaire entre les vecteurs $F_{p(i)}$ et $F_{p(j)}$ précédemment définis ;
- $(.)^T$ désigne la transposition matricielle.

[0063] Pour les impulsions de l'excitation stochastique ($1\leq n\leq np=nc-1$) les vecteurs $F_{p(n)}$ sont simplement constitués par le vecteur de la réponse impulsionnelle h décalée de p(n) échantillons. Le fait de tronquer la réponse impulsionnelle comme décrit précédemment permet donc de réduire sensiblement le nombre d'opérations utiles au calcul des produits scalaires faisant intervenir ces vecteurs $F_{p(n)}$. Pour la contribution prédictible de l'excitation, le vecteur $F_{p(0)}=Y_{TP}$ a pour composantes $F_{p(0)}(i)$ ($0\leq i<lst$) les convolutions $y_{TP}(i)$ que le module 38 a calculées suivant la formule (1) ou (1') pour le retard de prédiction à long terme sélectionné TP. Si MV=0, la contribution n=0 est également de type impulsionnelle et la position p(0) est à calculer.

[0064] Minimiser l'erreur quadratique E définie ci-dessus revient à trouver l'ensemble des positions p(n) qui maximisent la corrélation normalisée $b.B^{-1}.b^T$ puis à calculer les gains selon $g=b.B^{-1}$.

[0065] Mais une recherche exhaustive des positions d'impulsion nécessiterait un volume de calculs excessif. Pour atténuer ce problème, l'approche multi-impulsionnelle applique en général une procédure sous-optimale consistant à calculer successivement les gains et/ou les positions d'impulsion pour chaque contribution. Pour chaque contribution n ($0\leq n<nc$), on détermine d'abord la position p(n) qui maximise la corrélation normalisée $(F_p.e_{n-1}^T)^2/(F_p.F_p^T)$, on recalcule les gains $g_n(0)$ à $g_n(n)$ selon $g_n=b_n.B_n^{-1}$, où $g_n=(g_n(0),...,g_n(n))$, $b_n=(b(0),...,b(n))$ et $B_n=\{B_{i,j}\}_{0\leq i,j\leq n}$, puis on calcule pour l'itération suivante le vecteur-cible en égal au vecteur-cible initial X auquel on retranche les contributions 0 à n du signal synthétique pondéré multipliées par leurs gains respectifs :

$$e_n = X - \sum_{i=0}^{n} g_n(i).F_{p(i)}$$

[0066] A l'issue de la dernière itération nc-1, les gains $g_{nc-1}(i)$ sont les gains sélectionnés et l'erreur quadratique minimisée E est égal à l'énergie au vecteur-cible e_{nc-1}.

[0067] La méthode ci-dessus donne des résultats satisfaisants, mais elle nécessite l'inversion d'une matrice B_n à chaque itération. Dans leur article "Amplitude Optimization and Pitch Prediction in Multipulse Coders" (IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol.37, N° 3, Mars 1989, pages 317-327), S. Singhal et B.S. Atal ont proposé de simplifier le problème de l'inversion des matrices B_n en utilisant la décomposition de Cholesky : $B_n=M_n.M_n^T$ où M_n est une matrice trianguiaire inférieure. Cette décomposition est possible parce que B_n est une matrice symétrique à valeurs propres positives. L'avantage de cette approche est que l'inversion d'une matrice triangulaire est relativement peu complexe, B_n^{-1} pouvant être obtenue par $B_n^{-1}=(M_n^{-1})^T.M_n^{-1}$.

[0068] La décomposition de Cholesky et l'inversion de la matrice M_n nécessitent toutefois d'effectuer des divisions et des calculs de racines carrées qui sont des opérations exigentes en termes de complexité de calcul. L'invention propose de simplifier considérablement la mise en oeuvre de l'optimisation en modifiant la décomposition des matrices B_n de la façon suivante :

$$B_n = L_n.R_n^T = L_n.(L_n.K_n^{-1})^T$$

où K_n est une matrice diagonale et L_n est une matrice triangulaire inférieure n'ayant que des 1 sur sa diagonale principale (soit $L_n=M_n.K_n^{1/2}$ avec les notations précédentes). Compte-tenu de la structure de la matrice B_n, les matrices $L_n=R_n.K_n$, R_n, K_n et L_n^{-1} sont construites chacune par simple adjonction d'une ligne aux matrices correspondantes de l'itération précédente :

$$B_n = \begin{pmatrix} & & B(n,0) \\ & B_{n-1} & \vdots \\ & & B(n,n-1) \\ B(n,0) \quad . \quad . \quad . & B(n,n-1) & B(n,n) \end{pmatrix}$$

$$L_n = \begin{pmatrix} & & 0 \\ & L_{n-1} & \vdots \\ & & 0 \\ L(n,0) \quad . \quad . \quad . & L(n,n-1) & 1 \end{pmatrix}$$

$$R_n = \begin{pmatrix} & & 0 \\ & R_{n-1} & \vdots \\ & & 0 \\ R(n,0) \quad . \quad . \quad . & R(n,n-1) & R(n,n) \end{pmatrix}$$

$$K_n = \begin{pmatrix} & & 0 \\ & K_{n-1} & \vdots \\ & & 0 \\ 0 \quad . \quad . \quad . & 0 & K(n) \end{pmatrix}$$

$$L_n^{-1} = \begin{pmatrix} & & 0 \\ & L_{n-1}^{-1} & \vdots \\ & & 0 \\ L^{-1}(n,0) \quad . \quad . \quad . & L^{-1}(n,n-1) & 1 \end{pmatrix}$$

[0069] Dans ces conditions, la décomposition de B_n, l'inversion de L_n, l'obtention de $B_n^{-1}=K_n.(L_n^{-1})^T.L_n^{-1}$ et le recalcul des gains ne nécessitent qu'une seule division par itération et aucun calcul de racine carrée.

[0070] L'analyse stochastique relative à une sous-trame d'une trame voisée (MV=1,2 ou 3) peut dès lors se dérouler comme indiqué sur les figures 8 à 11. Pour calculer le gain de prédiction à long terme, l'index de contribution n est initialisé à 0 à l'étape 180 et le vecteur $F_{p(0)}$ est pris égal à la contribution à long terme Y_{TP} fournie par le module 38. Si n>0, l'itération n commence par la détetmination 182 de la position p(n) de l'impulsion n qui maximise la quantité :

$$(F_p . e^T)^2/(F_p . F_p^T) = \frac{\left(\sum_{k=p}^{\min(Lh+p,\,lst)-1} h(k-p) . e(k)\right)^2}{\sum_{k=p}^{\min(Lh+p,\,lst)-1} h(k-p) . h(k-p)}$$

où e=(e(0),...,e(lst-1)) est un vecteur-cible calculé lors de l'itération précédente. Différentes contraintes peuvent être apportées au domaine de maximisation de la quantité ci-dessus inclus dans l'intervalle [0,lst[. L'invention utilise de préférence une recherche segmentaire dans laquelle la sous-trame d'excitation est subdivisée en ns segments de même longueur (par exemple ns=10 pour lst=40). Pour la première impulsion (n=1), la maximisation de $(F_p.e^T)^2/(F_p.F_p^T)$ est effectuée sur l'ensemble des positions possibles p dans la sous-trame. A l'itération n>1, la maximisation est effectuée à l'étape 182 sur l'ensemble des positions possibles à l'exclusion des segments dans lesquels ont été respectivement trouvées les positions p(1),...,p(n-1) des impulsions lors des itérations précédentes.

[0071] Dans le cas où la trame courante a été détectée comme non voisée, la contribution n=0 est également constituée par une impulsion de position p(0). L'étape 180 comprend alors seulement l'initialisation n=0, et elle est suivie par une étape de maximisation identique à l'étape 182 pour trouver p(0), avec $e=e_{-1}=X$ comme valeur initiale du vecteur-cible.

[0072] On remarque que lorsque la contribution n=0 est prédictible (MV=1, 2 ou 3), le module 38 d'analyse LTP en boucle fermée a effectué une opération de nature semblable à la maximisation 182, puisqu'il a déterminé la contribution à long terme, caractérisée par le retard TP, en maximisant la quantité $(Y_T.e^T)^2,(Y_T.Y_T^T)$ dans l'intervalle de recherche des retards T, avec $e=e_{-1}=X$ comme valeur initiale du vecteur-cible. On peut également, lorsque l'énergie de la contribution LTP est très faible, ignorer cette contribution dans le processus de recalcul des gains.

[0073] Après l'étape 180 ou 182, le module 40 procède au calcul 184 de la ligne n des matrices L, R et K intervenant dans la décomposition de la matrice B, ce qui permet de compléter les matrices L_n, R_n et K_n définies ci-dessus. La décomposition de la matrice B permet d'écrire :

$$B(n,j) = R(n,j) + \sum_{k=0}^{j-1} L(n,k) . R(j,k)$$

pour la composante située à la ligne n et à la colonne j. On peut donc écrire, pour j croissant de 0 à n-1 :

$$R(n,j) = B(n,j) - \sum_{k=0}^{j-1} L(n,k) . R(j,k)$$

$$L(n,j) = R(n,j).K(j)$$

et, pour j=n :

$$K(n) = 1/R(n,n) = 1/\left[B(n,n) - \sum_{k=0}^{n-1} L(n,k) \cdot R(n,k)\right]$$

$$L(n,n) = 1$$

[0074] Ces relations sont exploitées dans le calcul 184 détaillé sur la figure 9. L'index de colonne j est d'abord initialisé à 0, à l'étape 186. Pour l'index de colonne j, la variable tmp est d'abord initialisée à la valeur de la composante B(n,j), soit :

$$tmp = F_{p(n)} \cdot F_{p(j)}^{T}$$

$$= \sum_{k=\max(p(n),p(j))}^{\min(Lh+p(n),\,Lh+p(j),\,lst)-1} h(k-p(n)) \cdot h(k-p(j))$$

[0075] A l'étape 188, l'entier k est en outre initialisé à 0. On effectue alors une comparaison 190 entre les entiers k et j. Si k<j, on ajoute le terme L(n,k).R(j,k) à la variable tmp, puis on incrémente d'une unité l'entier k (étape 192) avant de réexécuter la comparaison 190. Quand la comparaison 190 montre que k=j, on effectue une comparaison 194 entre les entiers j et n. Si j<n, la composante R(n,j) est prise égale à tmp et la composante L(n,j) à tmp.K(j) à l'étape 196, puis l'index de colonne j est incrémenté d'une unité avant qu'on revienne à l'étape 188 pour calculer les composantes suivantes. Quand la comparaison 194 montre que j=n, la composante K(n) de la ligne n de la matrice K est calculée, ce qui termine le calcul 184 relatif à la ligne n. K(n) est pris égal à 1/tmp si tmp≠0 (étape 198) et à 0 sinon. On constate que le calcul 184 ne requiert qu'au plus une division 198, pour obtenir K(n). En outre, une éventuelle singularité de la matrice B_n n'entraîne pas d'instabilités puisqu'on évite les divisions par 0.

[0076] En référence à la figure 8, le calcul 184 des lignes n de L, R et K est suivi par l'inversion 200 de la matrice L_n constituée des lignes et des colonnes 0 à n de la matrice L. Le fait que L soit triangulaire avec des 1 sur sa diagonale principale en simplifie grandement l'inversion comme le montre la figure 10. On peut en effet écrire :

$$L^{-1}(n,j') = -L(n,j') - \sum_{k'=j'+1}^{n} L^{-1}(k',j') \cdot L(n,k') \qquad (4)$$

$$= -L(n,j') - \sum_{k'=j'+1}^{n} L(k',j') \cdot L^{-1}(n,k') \qquad (5)$$

pour $0 \le j' < n$ et $L^{-1}(n,n)=1$, c'est-à-dire que l'inversion peut être faite sans avoir à opérer une division. En outre, comme les composantes de la ligne n de L^{-1} suffisent à recalculer les gains, l'utilisation de la relation (5) permet de faire l'inversion sans avoir à mémoriser toute la matrice L^{-1}, mais seulement un vecteur Linv=(Linv(0),...,Linv(n-1)) avec $Linv(j')=L^{-1}(n,j')$. L'inversion 200 commence alors par une initialisation 202 de l'index de colonne j' à n-1. A l'étape 204, le terme Linv(j') est initialisé à -L(n,j') et l'entier k' à j'+1. On effectue ensuite une comparaison 206 entre les entiers k' et n. Si k'<n, on retranche le terme L(k',j').Linv(k') à Linv(j'), puis on incrémente d'une unité l'entier k' (étape 208) avant de réexécuter la comparaison 206. Quand la comparaison 206 montre que k'=n, on compare j' à 0 (test 210). Si j'>0, on décrémente l'entier j' d'une unité (étape 212) et on revient à l'étape 204 pour calculer la composante suivante. L'inversion 200 est terminée lorsque le test 210 montre que j'=0.

[0077] En référence à la figure 8 l'inversion 200 est suivie par le calcul 214 des gains réoptimisés et du vecteur-cible E pour l'itération suivante. Le calcul des gains réoptimisés est également très simplifié par la décomposition retenue

pour la matrice B. On peut en effet calculer le vecteur $g_n=(g_n(0),...,g_n(n))$ solution de $g_n.B_n=b_n$ selon :

$$g_n(n) = \left[b(n) + \sum_{i=0}^{n-1} b(i) \,.\, L^{-1}(n,i)\right] . K(n)$$

et $g_n(i')=g_{n-1}(i')+L^{-1}(n,i').g_n(n)$ pour $0\leq i'<n$. Le calcul 214 est détaillé sur la figure 11. On calcule d'abord la composante b(n) du vecteur b :

$$b(n) = F_{p(n)} \,.\, X^T = \sum_{k=p(n)}^{\min(Lh+p(n),\, lst)-1} h(k-p(n)) \,.\, x(k)$$

b(n) sert de valeur d'initialisation pour la variable tmq. A l'étape 216, on initialise également l'index i à 0. On effectue ensuite la comparaison 218 entre les entiers i et n. Si i<n, on ajoute le terme b(i). Linv(i) à la variable tmq et on incrémente i d'une unité (étape 220) avant de revenir à la comparaison 218. Quand la comparaison 218 montre que i=n, on calcule le gain relatif à la contribution n selon g(n)=tmq.K(n), et on initialise la boucle de calcul des autres gains et du vecteur-cible (étape 222) en prenant $e=X-g(n).F_{p(n)}$ et i'=0. Cette boucle comprend une comparaison 224 entre les entiers i' et n. Si i'<n, le gain g(i') est recalculé à l'étape 226 en ajoutant Linv(i').g(n) à sa valeur calculée lors de l'itération précédente n-1, puis on retranche au vecteur-cible e le vecteur $g(i').F_{p(i')}$. L'étape 226 comprend également l'incrémentation de l'index i' avant de revenir à la comparaison 224. Le calcul 214 des gains et du vecteur-cible est terminé lorsque la comparaison 224 montre que i'=n. On voit que les gains ont pu être mis à jour en ne faisant appel qu'à la ligne n de la matrice inverse L_n^{-1}.

[0078] Le calcul 214 est suivi par une incrémentation 228 de l'index n de la contribution, puis par une comparaison 230 entre l'index n et le nombre de contributions nc. Si n<nc, on revient à l'étape 182 pour l'itération suivante. L'optimisation des positions et des gains est terminée lorsque n=nc au test 230.

[0079] La recherche segmentaire des impulsions diminue sensiblement le nombre de positions d'impulsion à évaluer au cours des étapes 182 de la recherche de l'excitation stochastique. Elle permet en outre une quantification efficace des positions trouvées. Dans le cas typique où la sous-trame de lst=40 échantillons est divisée en ns=10 segments de ls=4 échantillons, l'ensemble des positions d'impulsion possibles peut prendre $ns!.ls^{np}/[np!(ns-np)!]=258\,048$ valeurs si np=5 (MV=1, 2 ou 3) ou 860 160 si np=6(MV=0), au lieu de $lst!/[np!(lst-np)!]=658\,008$ valeurs si np=5 ou 3 838 380 si np=6 dans le cas où on impose seulement que deux impulsions ne puissent pas avoir la même position. En d'autres termes, on peut quantifier les positions sur 18 bits au lieu de 20 bits si np=5, et sur 20 bits au lieu de 22 si np=6.

[0080] Le cas particulier où le nombre de segments par sous-trame est égal au nombre d'impulsions par excitation stochastique (ns=np) conduit à la plus grande simplicité de la recherche de l'excitation stochastique, ainsi qu'au plus faible débit binaire (si lst=40 et np=5, il y a $8^5=32768$ ensembles de positions possibles, quantifiables sur 15 bits seulement au lieu de 18 si ns=10). Mais en réduisant à ce point le nombre de séquences d'innovation possibles, on peut appauvrir la qualité du codage. Pour un nombre d'impulsions donné, le nombre des segments peut être optimisé selon un compromis visé entre la qualité du codage et sa simplicité de mise en oeuvre (ainsi que le débit requis).

[0081] Le cas où ns>np présente en outre l'avantage qu'on peut obtenir une bonne robustesse aux erreurs de transmission en ce qui concerne les positions des impulsions, grâce à une quantification séparée des numéros d'ordre des segments occupés et des positions relatives des impulsions dans chaque segment occupé. Pour une impulsion n, le numéro d'ordre s_n du segment et la position relative pr_n sont respectivement le quotient et le reste de la division euclidienne de p(n) par la longueur ls d'un segment : $p(n)=s_n.ls+pr_n$ ($0\leq s_n<ns$, $0\leq pr_n<ls$). Les positions relatives sont chacune quantifiées séparément sur 2 bits, si ls=4. En cas d'erreur de transmission affectant l'un de ces bits, l'impulsion correspondante ne sera que peu déplacée, et l'impact perceptuel de l'erreur sera limité. Les numéros d'ordre des segments occupés sont repérés par un mot binaire de ns=10 bits valant chacun 1 pour les segments occupés et 0 pour les segments dans lesquels l'excitation stochastique n'a pas d'impulsion. Les mots binaires possibles sont ceux ayant un poids de Hamming de np; ils sont au nombre de $ns!/[np!(ns-np)!]=252$ si np=5, ou 210 si np=6. Ce mot est quantifiable par un index de nb bits avec $2^{nb-1}<ns!/[np!(ns-np)!]\leq 2^{nb}$, soit nb=8 dans l'exemple considéré. Si, par exemple, l'analyse stochastique a fourni np=5 impulsions de positions 4, 12, 21, 34, 38, les positions relatives quantifiées scalairement sont 0,0,1,2,2 et le mot binaire représentant les segments occupés est 0101010011, ou 339 en traduction décimale.

[0082] Au niveau du décodeur, les mots binaires possibles sont stockés dans une table de quantification dans laquelle les adresses de lecture sont les index de quantification reçus. L'ordre dans cette table, déterminé une fois pour toutes, peut être optimisé de façon qu'une erreur de transmission affectant un bit de l'index (le cas d'erreur le plus fréquent, surtout lorsqu'un entrelacement est mis en oeuvre dans le codeur canal 22) ait, en moyenne, des conséquences minimales suivant un critère de voisinage. Le critère de voisinage est par exemple qu'un mot de ns bits ne puisse être remplacé que par des mots "voisins", éloignés d'une distance de Hamming au plus égale à un seuil np-2δ, de façon à conserver toutes les impulsions sauf δ d'entre elles à des positions valides en cas d'erreur de transmission de l'index portant sur un seul bit. D'autres critères seraient utilisables en substitution ou en complément, par exemple que deux mots soient considérés comme voisins si le remplacement de l'un par l'autre ne modifie pas l'ordre d'affectation des gains associés aux impulsions.

[0083] A des fins d'illustration, on peut considérer le cas simplifié où ns=4 et np=2, soit 6 mots binaires possibles quantifiables sur nb=3 bits. Dans ce cas, on peut vérifier que la table de quantification présentée au tableau II permet de conserver np-1=1 impulsion bien positionnée pour toute erreur affectant un bit de l'index transmis. Il y a 4 cas d'erreur (sur un total de 18), pour lesquels on reçoit un index de quantification qu'on sait être erroné (6 au lieu de 2 ou 4 ; 7 au lieu de 3 ou 5), mais le décodeur peut alors prendre des mesures limitant la distorsion, par exemple répéter la séquence d'innovation relative à la sous-trame précédente ou encore affecter des mots binaires acceptables aux index "impossibles" (par exemple 1001 ou 1010 pour l'index 6 et 1100 ou 0110 pour l'index 7 conduisent encore à np-1=1 impulsion bien positionnée en cas de réception de 6 ou 7 avec une erreur binaire).

TABLEAU II

index de quantification		mot d'occupation des segments	
décimal	binaire naturel	binaire naturel	décimal
0	000	0011	3
1	001	0101	5
2	010	1001	9
3	011	1100	12
4	100	1010	10
5	101	0110	6
(6)	(110)	(1001 ou 1010)	(9 ou 10)
(7)	(111)	(1100 ou 0110)	(12 ou 6)

[0084] Dans le cas général, l'ordre dans la table de quantification des mots peut être déterminé à partir de considérations arithmétiques ou, si cela est insuffisant, en simulant sur ordinateur les scénarios d'erreurs (de façon exhaustive ou par un échantillonnage statistique de type Monte-Carlo suivant le nombre de cas d'erreurs possibles).

[0085] Pour sécuriser la transmission de l'index de quantification des segments occupés, on peut en outre tirer parti des différentes catégories de protection offertes par le codeur canal 22, notamment si le critère de voisinage ne peut être vérifié de façon satisfaisante pour tous les casd'erreurs possibles affectant un bit de l'index. Le module d'ordonnancement 46 peut ainsi mettre dans la catégorie de protection minimale, ou dans la catégorie non protégée, un certain nombre nx des bits de l'index qui, s'ils sont affectés par une erreur de transmission, donnent lieu à un mot erroné mais vérifiant le critère de voisinage avec une probabilité jugée satisfaisante, et mettre dans une catégorie plus protégée les autres bits de l'index. Cette façon de procéder fait appel à un autre ordonnancement des mots dans la table de quantification. Cet ordonnancement peut également être optimisé au moyen de simulations si on souhaite maximiser le nombre nx des bits de l'index affectés à la catégorie la moins protégée.

[0086] Une possibilité est de commencer par constituer une liste de mots de ns bits par comptage en code de Gray de 0 à 2^{ns}-1, et d'obtenir la table de quantification ordonnée en supprimant de cette liste les mots n'ayant pas un poids de Hamming de np. La table ainsi obtenue est telle que deux mots consécutifs ont une distance de Hamming de np-2. Si les index dans cette table ont une représentation binaire en code de Gray, toute erreur sur le bit de poids le plus faible fait varier l'index de ±1 et entraîne donc le remplacement du mot d'occupation effectif par un mot voisin au sens du seuil np-2 sur la distance de Hamming, et une erreur sur le i-ième bit de poids le plus faible fait aussi varier l'index de ±1 avec une probabilité d'environ 2^{1-i}. En plaçant les nx bits de poids faible de l'index en code de Gray dans une catégorie non protégée, une éventuelle erreur de transmission affectant un de ces bits conduit au remplacement du mot d'occupation par un mot voisin avec une probabilité au moins égale à $(1+1/2+...+1/2^{nx-1})/nx$. Cette probabilité minimale décroît de 1 à $(2/nb)(1-1/2^{nb})$ pour nx croissant de 1 à nb. Les erreurs affectant les nb-nx bits de poids fort de l'index seront le plus souvent corrigées grâce à la protection que leur applique le codeur canal. La valeur de nx est dans ce cas choisie selon un compromis entre la robustesse aux erreurs petites valeurs) et un encombrement réduit

des catégories protégées (grandes valeurs).

[0087] Au niveau du codeur, les mots binaires possibles pour représenter l'occupation des segments sont rangés en ordre croissant dans une table de recherche. Une table d'indexage associe à chaque adresse le numéro d'ordre, dans la table de quantification stockée au décodeur, du mot binaire ayant cette adresse dans la table de recherche. Dans l'exemple simplifié évoqué ci-dessus, le contenu de la table de recherche et de la table d'indexage est donné dans le tableau III (en valeurs décimales).

TABLEAU III

Adresse	Table de recherche	Table d'indexage
0	3	0
1	5	1
2	6	5
3	9	2
4	10	4
5	12	3

[0088] La quantification du mot d'occupation des segments déduit des np positions fournies par le module d'analyse stochastique 40 est effectuée en deux étapes par le module de quantification 44. Une recherche dichotomique est d'abord effectuée dans la table de recherche pour déterminer l'adresse dans cette table du mot à quantifier. L'index de quantification est ensuite obtenu à l'adresse déterminée dans la table d'indexage puis fourni au module 46 d'ordonnancement des bits.

[0089] Le module 44 effectue en outre la quantification des gains calculés par le module 40. Le gain g_{TP} est par exemple quantifié dans l'intervalle [0;1,6], sur 5 bits si MV=1 ou 2 et sur 6 bits si MV=3 pour tenir compte de la plus grande importance perceptuelle de ce paramètre pour les trames très voisées. Pour le codage des gains associés aux impulsions de l'excitation stochastique, on quantifie sur 5 bits la plus grande valeur absolue Gs des gains g(1),...,g(np), en prenant par exemple 32 valeurs de quantification en progression géométrique dans l'intervalle [0;32767], et on quantifie chacun des gains relatifs g(1)/Gs,...,g(np)/Gs dans l'intervalle [-1;+1], sur 4 bits si MV=1, 2 ou 3, ou sur 5 bits si MV=0.

[0090] Les bits de quantification de Gs sont placés dans une catégorie protégée par le codeur canal 22, de même que les bits de poids fort des index de quantification des gains relatifs. Les bits de quantification des gains relatifs sont ordonnés de façon à permettre leur affectation aux impulsions associées appartenant aux segments localisés par le mot d'occupation. La recherche segmentaire selon l'invention permet en outre de protéger de manière efficace les positions relatives des impulsions associées aux plus grandes valeurs de gain.

[0091] Dans le cas où np=5 et ls=4, dix bits par sous-trame sont nécessaires pour quantifier les positions relatives des impulsions dans les segments. On considère le cas où 5 de ces 10 bits sont placés dans une catégorie peu ou pas protégée (II) et où les 5 autres sont placés dans une catégorie plus protégée (IB). La distribution la plus naturelle est de placer le bit de poids fort de chaque position relative dans la catégorie protégée IB, de sorte que les éventuelles erreurs de transmission affectent plutôt les bits de poids fort et ne provoquent donc qu'un décalage d'un échantillon pour l'impulsion correspondante. Il est toutefois judicieux, pour la quantification des positions relatives, de considérer les impulsions dans l'ordre décroissant des valeurs absolues des gains associés et de placer dans la catégorie IB les deux bits de quantification de chacune des deux premières positions relatives ainsi que le bit de poids fort de la troisième. De cette façon, les positions des impulsions sont protégées préférentiellement lorsqu'elles sont associées à des gains importants, ce qui améliore la qualicé moyenne particulièrement pour les sous-trames les plus voisées.

[0092] Pour reconstituer les contributions impulsionnelles de l'excitation, le décodeur 54 localise d'abord les segments au moyen du mot d'occupation reçu ; il attribue ensuite les gains associés ; puis il attribue les positions relatives aux impulsions sur la base de l'ordre d'importance des gains.

[0093] On comprendra que les différents aspects de l'invention décrits ci-dessus procurent chacun des améliorations propres, et qu'il est donc envisageable de les mettre en oeuvre indépendamment les uns des autres. Leur combinaison permet de réaliser un codeur de performances particulièrement intéressantes.

[0094] Dans l'exemple de réalisation décrit dans ce qui précède, le codeur de parole à 13 kbits/s requiert de l'ordre de 15 millions d'instructions par seconde (Mips) en virgule fixe. On le réalisera donc typiquement et programmant un processeur de signal numérique (DSP) du commerce, de même que le décodeur qui ne requiert que de l'ordre de 5 Mips.

Revendications

1. Procédé de codage à analyse par synthèse d'un signal de parole (S) numérisé en trames successives divisées en nst sous-trames, comprenant les étapes suivantes :

   - analyse par prédiction linéaire du signal de parole pour déterminer des paramètres d'un filtre (60) de synthèse à court terme ;
   - analyse en boucle ouverte du signal de parole pour détecter les trames voisées du signal et pour déterminer, pour chaque trame voisée, un degré de voisement du signal (MV) et un intervalle de recherche d'un retard de prédiction à long terme ;
   - analyse prédictive en boucle fermée du signal de parole pour sélectionner, pour certaines au moins des sous-trames des trames voisées, un retard de prédiction à long terme contenu dans l'intervalle de recherche et constituant un paramètre d'un filtre de synthèse à long terme (66); et
   - détermination d'une excitation stochastique pour chaque sous-trame, de façon à minimiser un écart pondéré perceptuellement entre le signal de parole et l'excitation stochastique filtrée par les filtres de synthèse à long terme et à court terme,

   caractérisé en ce qu'à l'étape d'analyse en boucle ouverte, on détermine l'intervalle de recherche relatif à chaque trame voisée de façon qu'il contienne un nombre de retards (N1, N3) dépendant du degré de voisement de ladite trame.

2. Procédé selon la revendication 1, caractérisé en ce que l'intervalle de recherche du retard de prédiction à long terme contient moins de retards pour les trames ayant le plus grand degré de voisement que pour les autres trames voisées.

3. Procédé selon la revendication 1 ou 2, caractérisé en ce que l'analyse en boucle ouverte relative à une trame comprend la détermination de nst retards de base (K_{st}) qui maximisent chacun une estimation en boucle ouverte du gain de prédiction à long terme sur une sous-trame respective de ladite trame, puis la comparaison entre un premier seuil prédéterminé (S0) et une première estimation en boucle ouverte du gain de prédiction à long terme sur la trame obtenue sur la base des nst retards de base relativement aux sous-trames correspondantes pour détecter si la trame est voisée, en ce que, si la trame est détectée comme voisée, l'analyse en boucle ouverte comprend en outre pour chaque sous-trame la détermination d'une liste (l_{st}) de retards candidats pour lesquels l'estimation en boucle ouverte au gain de prédiction sur la sous-trame est supérieure à une fraction déterminée (β) de l'estimation relative au retard de base pour la sous-trame, en ce qu'on sélectionne dans lesdites listes le retard candidat pour lequel une seconde estimation en boucle ouverte du gain de prédiction à long terme sur la trame est maximale, la seconde estimation en boucle ouverte sur la trame associée à un retard candidat étant obtenue sur la base de nst retards optimaux, compris dans un intervalle de N1 retards centré sur ledit retard candidat, qui, respectivement, maximisent sur ledit intervalle l'estimation en boucle ouverte du gain de prédiction sur les nst sous-trames, en ce que la détermination au degré de voisement de la trame comprend une comparaison entre la seconde estimation maximisée du gain de prédiction sur la trame et au moins un autre seuil prédéterminé (S1,S2), et en ce que l'intervalle de recherche déterminé à l'issue de l'analyse en boucle ouverte est centré sur ledit retard sélectionné.

4. Procédé selon la revendication 1 ou 2, caractérisé en ce que l'analyse en boucle ouverte relative à une trame comprend la détermination d'un retard de base (K) qui maximise une première estimation en boucle ouverte du gain de prédiction à long terme sur ladite trame, puis la comparaison entre un premier seuil prédéterminé (S0) et la première estimation maximisée du gain de prédiction à long terme sur la trame pour détecter si la trame est voisée, en ce que, si la trame est détectée comme voisée, l'analyse en boucle ouverte comprend en outre la détermination d'une liste (l) de retards candidats pour lesquels l'estimation en boucle ouverte du gain de prédiction sur la trame est supérieure à une fraction déterminée (β) de l'estimation relative au retard de base, en ce qu'on sélectionne dans ladite liste le retard candidat pour lequel une seconde estimation en boucle ouverte du gain de prédiction à long terme sur la trame est maximale, la seconde estimation en boucle ouverte sur la trame associée à un retard candidat étant obtenue sur la base de nst retards optimaux, compris dans un intervalle de N1 retards centré sur ledit retard candidat, qui, respectivement, maximisent sur ledit intervalle l'estimation en boucle ouverte du gain de prédiction sur les nst sous-trames, en ce que la détermination du degré de voisement de la trame comprend une comparaison entre la seconde estimation maximisée du gain de prédiction sur la trame et au moins un autre seuil prédéterminé (S1,S2), et en ce que l'intervalle de recherche déterminé à l'issue de l'analyse en boucle ouverte est centré sur ledit retard sélectionné.

5. Procédé selon la revendication 1 ou 2, caractérisé en ce que l'analyse en boucle ouverte relative à une trame comprend la détermination d'un nombre nz de retards de base (K_1' ,..., K_{nz}') qui maximisent chacun, sur un sous-intervalle respectif de valeurs de retard possibles, une première estimation en boucle ouverte du gain de prédiction à long terme sur ladite trame, puis la comparaison entre un premier seuil prédéterminé (S0) et la plus grande des nz premières estimations maximisées du gain de prédiction à long terme sur la trame pour détecter si la trame est voisée, en ce que, si la trame est détectée comme voisée, on sélectionne, parmi nz retards candidats obtenus à partir des nz retards de base, le retard candidat pour lequel une seconde estimation en boucle ouverte du gain de prédiction à long terme sur la trame est maximale, la seconde estimation en boucle ouverte sur la trame associée à un retard candidat étant obtenue sur la base de nst retards optimaux, compris dans un intervalle de N1 retards centré sur ledit retard candidat, qui, respectivement, maximisent sur ledit intervalle l'estimation en boucle ouverte au gain de prédiction sur les nst sous-trames, en ce que la détermination du degré de voisement de la trame comprend une comparaison entre la seconde estimation maximisée du gain de prédiction sur la trame et au moins un autre seuil prédéterminé (S1,S2), et en ce que l'intervalle de recherche déterminé à l'issue de l'analyse en boucle ouverte est centré sur ledit retard sélectionné.

6. Procédé selon l'une quelconque des revendications 3 à 5, caractérisé en ce que si la seconde estimation maximisée du gain de prédiction sur une trame voisée est supérieure à un des seuils (S2), on détermine si les nst retards optimaux sont compris dans un intervalle centré sur le retard sélectionné et contenant un nombre de retards N3 inférieur à N1 et, dans l'affirmative, on attribue à la trame un degré de voisement pour lequel l'intervalle de recherche du retard de prédiction à long terme contient N3 retards, l'intervalle de recherche contenant N1 retards pour au moins un autre degré de voisement.

7. Procédé selon l'une quelconque des revendications 3 à 5, caractérisé en ce que lors de la maximisation de la seconde estimation en boucle ouverte du gain de prédiction à long terme sur une trame voisée, on calcule en outre une troisième estimation en boucle ouverte du gain sur la trame sur la base de nst retards, compris dans un intervalle centré sur le retard sélectionné et contenant un nombre N3 de retards inférieur à N1, qui, respectivement, maximisent sur ledit intervalle de N3 retards l'estimation en boucle ouverte du gain de prédiction sur les nst sous-trames, et en ce qu'on attribue à la trame un degré de voisement pour lequel l'intervalle de recherche contient N3 retards si ladite troisième estimation dépasse un seuil prédéterminé (S2), l'intervalle de recherche contenant N1 retards pour au moins un autre degré de voisement.

8. Procédé selon la revendication 3 ou 4, caractérisé en ce que les retards candidats d'une liste sont choisis parmi les sous-multiples du retard de base associé à ladite liste et parmi les multiples du plus petit desdits sous-multiples pour lequel l'estimation en boucle ouverte au gain de prédiction est supérieure à ladite fraction déterminée de l'estimation relative au retard de base.

9. Procédé selon la revendication 8, caractérisé en ce que les retards de prédiction à long terme peuvent correspondre à des nombres entiers ou fractionnaires d'échantillons du signal de parole, en ce qu'on détermine les retards de base (rbf) en résolution fractionnaire pour chercher les sous-multiples et les multiples à inclure dans une liste des retards candidats, et en ce que les retards de base sont déterminés en résolution entière pour évaluer les premières estimations en boucle ouverte du gain de prédiction sur une trame.

10. Procédé selon l'une quelconque des revendications 3 à 9, caractérisé en ce qu'on n'effectue pas l'analyse prédictive en boucle fermée relativement à chaque sous-trame pour laquelle l'autocorrélation (C_{st}) du signal de parole associée au retard optimal pour ladite sous-trame est négative.

Patentansprüche

1. Synthese-Analyse-Verfahren zum Codieren eines digitalisierten Sprachsignals (S) in aufeinanderfolgende Raster, die in lst Unterraster aufgeteilt sind, welches die folgenden Schritte aufweist:

   - Analyse des Sprachsignals mittels linearer Prädiktion zum Bestimmen der Parameter eines Kurzzeitsynthesefilters (60);
   - rückkopplungslose Analyse des Sprachsignals zum Erfassen der stimmhaften Raster des Signals und zum Bestimmen für jedes stimmhafte Raster eines Stimmhaftigkeitsgrades (MV) des Signals und eines Suchintervalls einer Langzeitprädiktionsverzögerung;
   - prädiktive Analyse mit Rückkopplung des Sprachsignals zum Auswählen, für bestimmte mindestens der Un-

terraster der stimmhaften Raster, einer Langzeitprädiktionsverzögerung, welche in dem Suchintervall enthalten ist und einen Parameter eines Langzeitsynthesefilters (66) bildet; und

- Bestimmung einer stochastischen Anregung für jedes Unterraster, so daß ein Wahrnehmungswichtungsabstand zwischen dem Sprachsignal und der durch das Langzeit- und Kurzzeitsynthesefilter gefilterten stochastischen .Anregung minimiert wird,

dadurch gekennzeichnet, daß bei dem Schritt der rückkopplungslosen Analyse das Suchintervall relativ zu jedem stimmhaften Raster derart bestimmt wird, daß es eine Anzahl von Verzögerungen (N1, N3) enthält, welche von dem Stimmhaftigkeitsgrad des Rasters abhängt.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Suchintervall der Langzeitprädiktionsverzögerung für die Raster mit dem höheren Stimmhaftigkeitsgrad weniger Verzögerungen als für die weiteren stimmhaften Raster enthält.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die rückkopplungslose Analyse bezüglich eines Rasters die Bestimmung von nst Basisverzögerungen (K_{st}) aufweist, die jeweils eine rückkopplungslose Schätzung der Langzeitprädiktionsverstärkung an einem jeweiligen Unterraster des Rasters maximieren, daraufhin den Vergleich zwischen einem ersten vorgegebenen Schwellenwert (S0) und einer ersten rückkopplungslosen Schätzung der Langzeitprädiktionsverstärkung an dem Raster, welche auf der Grundlage der nst Basisverzögerungen relativ zu den entsprechenden Unterrastern erhalten wird, um zu erfassen, ob das Raster stimmhaft ist, dadurch, daß, wenn das Raster als stimmhaft erfaßt wird, die rückkopplungslose Analyse des weiteren für jedes Unterraster die Bestimmung einer Liste (l_{st}) von potentiellen Verzögerungen aufweist, bei denen die rückkopplungslose Schätzung der Prädiktionsverstärkung an dem Unterraster größer als ein bestimmter Bruchteil (β) der Schätzung bezüglich der Basisverzögerung für das Unterraster ist, dadurch, daß in diesen Listen die potentielle Verzögerung ausgewählt wird, bei der eine zweite rückkopplungslose Schätzung der Langzeitprädiktionsverstärkung an dem Raster maximal ist, wobei die zweite rückkopplungslose Schätzung an dem einer potentiellen Verzögerung zugeordneten Raster auf der Grundlage von nst optimalen Verzögerungen erhalten wird, welche in einem um die potentielle verzögerung zentrierten Intervall von N1 Verzögerungen enthalten sind, die jeweils an diesem Intervall die rückkopplungslose Schätzung der Prädiktionsverstärkung an den nst Unterrastern maximieren, dadurch, daß die Bestimmung des Stimmhaftigkeitsgrades des Rasters einen Vergleich zwischen der zweiten maximierten Schätzung der Prädiktionsverstärkung an dem Raster und mindestens einem weiteren vorgegebenen Schwellenwert (S1, S2) beinhaltet, sowie dadurch, daß das bei Beendigung der rückkopplungslosen Analyse bestimmte Suchintervall um die ausgewählte Verzögerung zentriert ist.

4. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die rückkopplungslose Analyse bezüglich eines Rasters die Bestimmung einer Basisverzögerung (K) aufweist, die eine erste rückkopplungslose Schätzung der Langzeitprädiktionsverstärkung an diesem Raster maximiert, daraufhin den Vergleich zwischen einem ersten vorgegebenen Schwellenwert (S0) und der ersten maximierten Schätzung der Langzeitprädiktionsverstärkung an dem Raster, um zu erfassen, ob das Raster stimmhaft ist, dadurch, daß, falls das Raster als stimmhaft erfaßt wird, die rückkopplungslose Analyse des weiteren die Bestimmung einer Liste (l) von potentiellen Verzögerungen aufweist, bei denen die rückkopplungslose Schätzung der Prädiktionsverstärkung an dem Raster größer als ein bestimmter Bruchteil (β) der Schätzung bezüglich der Basisverzögerung ist, dadurch, daß in dieser Liste die potentielle Verzögerung ausgewählt wird, bei der eine zweite rückkopplungslose Schätzung der Langzeitprädiktionsverstärkung an dem Raster maximal ist, wobei die zweite rückkopplungslose Schätzung an dem einer potentiellen Verzögerung zugeordneten Raster auf der Grundlage von ns optimalen Verzögerungen erhalten wird, die in einem um diese potentielle Verzögerung zentrierten Intervall von N1 Verzögerungen enthalten sind, die jeweils an dem Intervall die rückkopplungslose Schätzung der Prädiktionsverstärkung an den nst Unterrastern maximieren, dadurch, daß die Bestimmung des Stimmhaftigkeitsgrades des Rasters einen Vergleich zwischen der zweiten maximierten Schätzung der Prädiktionsverstärkung an dem Raster und mindestens einem weiteren vorgegebenen Schwellenwert (S1, S2) aufweist, sowie dadurch, daß das bei Beendigung der rückkopplungslosen Analyse bestimmte Suchintervall um die ausgewählte Verzögerung zentriert ist.

5. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die rückkopplungslose Analyse bezüglich eines Rasters die Bestimmung einer Anzahl nz von Basisverzögerungen ($K_1', \ldots, K_{nz}'$) aufweist, die jeweils an einem jeweiligen Unterintervall von möglichen Verzögerungswerten eine erste rückkopplungslose Schätzung der Langzeitprädiktionsverstärkung an diesem Raster maximieren, daraufhin den Vergleich zwischen einem ersten vorgegebenen Schwellenwert (S0) und der größten der nz ersten maximierten Schätzungen der Langzeitprädiktionsverstärkung an dem Raster, um zu erfassen, ob das Raster stimmhaft ist, dadurch, daß, falls das Raster als stimm-

haft erfaßt wird, unter nz potentiellen Verzögerungen, die ausgehend von den nz Basisverzögerungen erhalten wurden, die potentielle Verzögerung ausgewählt wird, bei der eine zweite rückkopplungslose Schätzung der Langzeitprädiktionsverstärkung an dem Raster maximal ist, wobei die zweite rückkopplungslose Schätzung an dem einer potentiellen Verzögerung zugeordneten Raster auf der Grundlage von nst optimalen Verzögerungen erhalten wird, welche in einem um diese potentielle Verzögerung zentrierten Intervall von N1 Verzögerungen enthalten sind, die jeweils an diesem Intervall die rückkopplungslose Schätzung der Prädiktionsverstärkung an den ns Unterrastern maximieren, dadurch, daß die Bestimmung des Stimmhaftigkeitsgrades des Rasters einen Vergleich zwischen der zweiten maximierten Schätzung der Prädiktionsverstärkung an dem Raster und mindestens einem weiteren vorgegebenen Schwellenwert (S1, S2) aufweist, sowie dadurch, daß das bei Beendigung der rückkopplungslosen Analyse bestimmte Suchintervall um diese ausgewählte Verzögerung zentriert ist.

6. Verfahren nach einem der Ansprüche 3 bis 5, dadurch gekennzeichnet, daß, falls die zweite maximierte Schätzung der Prädiktionsverstärkung an einem stimmhaften Raster größer als einer der Schwellenwerte (S2) ist, bestimmt wird, ob die nst optimalen Verzögerungen in einem Intervall enthalten sind, welches um die ausgewählte Verzögerung zentriert ist und eine Anzahl von Verzögerungen N3 geringer als N1 enthält und, wenn dies zutrifft, dem Raster ein Stimmhaftigkeitsgrad zugeordnet wird, bei dem das Suchintervall der Langzeitprädiktionsverzögerung N3 Verzögerungen enthält, wobei das Suchintervall N1 Verzögerungen für mindestens einen weiteren Stimmhaftigkeitsgrad enthält.

7. Verfahren nach einem der Ansprüche 3 bis 5, dadurch gekennzeichnet, daß bei der Maximierung der zweiten rückkopplungslosen Schätzung der Langzeitprädiktionsverstärkung an einem stimmhaften Raster des weiteren eine dritte rückkopplungslose Schätzung der Verstärkung an dem Raster auf der Grundlage von nst Verzögerungen berechnet wird, welche in in einem Intervall enthalten sind, das um die ausgewählte Verzögerung zentriert ist und eine Anzahl N3 von Verzögerungen geringer als N1 enthält, die jeweils an diesem Intervall von N3 Verzögerungen die rückkopplungslose Schätzung der Prädiktionsverstärkung an den nst Unterraster maximieren, sowie dadurch, daß dem Raster ein Stimmhaftigkeitsgrad zugeordnet wird, bei dem das Suchintervall N3 Verzögerungen enthält, falls die dritte Schätzung einen vorgegebenen Schwellenwert (S2) übersteigt, wobei das Suchintervall N1 Verzögerungen für mindestens einen weiteren Stimmhaftigkeitsgrad enthält.

8. Verfahren nach Anspruch 3 oder 4, dadurch gekennzeichnet, daß die potentiellen Verzögerungen einer Liste unter den Teilern der der Liste zugeordneten Basisverzögerung und unter den Vielfachen des kleinsten unter den Teiler ausgewählt werden, bei denen die rückkopplungslose Schätzung der Prädiktionsverstärkung größer als der bestimmte Bruchteil der Schätzung bezüglich der Basisverzögerung ist.

9. Verfahren nach Anspruch 8, dadurch gekennzeichnet, daß die Langzeitprädiktionsverzögerungen ganzen oder bruchartigen Zahlen von Abtastproben des Sprachsignals entsprechen können, dadurch, daß die Basisverzögerungen (rbf) in bruchartiger Auflösung bestimmt werden, um die in eine Liste der potentiellen Verzögerungen aufzunehmenden Teiler und Mehrfachen zu suchen, sowie dadurch, daß die Basisverzögerungen in ganzzahliger Auflösung bestimmt werden, um die ersten rückkopplungslosen Schätzungen der Prädiktionsverstärkung an einem Raster zu bewerten.

10. Verfahren nach einem der Ansprüche 3 bis 9, dadurch gekennzeichnet, daß die prädiktive Analyse mit Rückkopplung nicht durchgeführt wird bezüglich jedes Unterrasters, bei dem die Autokorrelation (C_{st}) des der optimalen Verzögerung für dieses Unterraster zugeordneten Sprachsignals negativ ist.

Claims

1. Analysis-by-synthesis speech coding method for a speech signal (S) digitised into successive frames which are divided into nst sub-frames of lst samples, comprising the steps of :

- linear prediction analysis of the speech signal in order to determine parameters of a short-term synthesis filter (60) ;
- open-loop analysis of the speech signal in order to detect the voiced frames of the signal and in order, for each voiced frame, to determine a degree of voicing of the signal (MV) and an interval for searching for a long-term prediction delay ;
- closed-loop predictive analysis of the speech signal in order, for at least some of the sub-frames of the voiced frames, to select a long-term prediction delay contained in the search interval and constituting a parameter of

a long-term synthesis filter (66) ; and

- determination of a stochastic excitation for each sub-frame, so as to minimise a perceptually weighted difference between the speech signal and the stochastic excitation filtered by the long-term and short-term synthesis filters,

characterised in that, in the open-loop analysis step, the search interval relating to each voiced frame is determined so that it contains a number (N1, N3) of delays which is dependent on the degree of voicing of said frame.

2. Method according to Claim 1, characterised in that the interval for searching for the long-term prediction delay contains fewer delays for those frames having the greatest degree of voicing than for the other voiced frames.

3. Method according to Claim 1 or 2, characterised in that the open-loop analysis relating to a frame comprises the determination of nst basic delays (K_{st}) which each maximise an open-loop estimate of the long-term prediction gain over a respective sub-frame of said frame, then the comparison between a first predetermined threshold (SO) and a first open-loop estimate of the long-term prediction gain over the frame obtained on the basis of the nst basic delays relating to the corresponding sub-frames in order to detect whether the frame is voiced, and in that, if the frame is detected as voiced, the open-loop analysis further comprises, for each sub-frame, the determination of a list (l_{st}) of candidate delays for which the open-loop estimate of the prediction gain over the sub-frame is higher than a defined fraction (β) of the estimate relating to the basic delay for the sub-frame, in that the candidate delay for which a second open-loop estimate of the long-term prediction gain over the frame is a maximum is selected from said lists, the second open-loop estimate over the frame associated with a candidate delay being obtained on the basis of nst optimal delays, lying in an interval of N1 delays which is centred on said candidate delay, which, respectively, over said interval, maximise the open-loop estimate of the prediction gain over the nst sub-frames, in that the determination of the degree of voicing of the frame comprises a comparison between the second maximised estimate of the prediction gain over the frame and at least one other predetermined threshold (S1, S2), and in that the search interval determined on completion of the open-loop analysis is centred on said selected delay.

4. Method according to Claim 1 or 2, characterised in that the open-loop analysis relating to a frame comprises the determination of a basic delay (K) which maximises a first open-loop estimate of the long-term prediction gain over said frame, then the comparison between a first predetermined threshold (S0) and the first maximised estimate of the long-term prediction gain over the frame in order to detect whether the frame is voiced, in that, if the frame is detected as voiced, the open-loop analysis further comprises the determination of a list (l) of candidate delays for which the open-loop estimate of the prediction gain over the frame is higher than a defined fraction (β) of the estimate relating to the basic delay, in that the candidate delay for which a second open-loop estimate of the long-term prediction gain over the frame is a maximum is selected from said list, the second open-loop estimate over the frame associated with a candidate delay being obtained on the basis of nst optimal delays, lying in an interval of N1 delays which is centred on said candidate delay, which, respectively, over said interval, maximise the open-loop estimate of the prediction gain over the nst sub-frames, in that the determining of the degree of voicing of the frame comprises a comparison between the second maximised estimate of the prediction gain over the frame and at least one other predetermined threshold (S1, S2), and in that the search interval determined on completion of the open-loop analysis is centred on said selected delay.

5. Method according to Claim 1 or 2, characterised in that the open-loop analysis relating to a frame comprises the determination of a number nz of basic delays (K_1', ..., K_{nz}',) which each, over a respective sub-interval of possible delay values, maximise a first open-loop estimate of the long-term prediction gain over said frame, then the comparison between a first predetermined threshold (S0) and the largest of the first nz maximised estimates of the long-term prediction gain over the frame in order to detect whether the frame is voiced, in that, if the frame is detected as voiced, the candidate delay for which a second open-loop estimate of the long-term prediction gain over the frame is a maximum is selected from among nz candidate delays obtained from the nz basic delays, the second open-loop estimate over the frame associated with a candidate delay being obtained on the basis of nst optimal delays, lying in an interval of N1 delays which is centred on said candidate delay, which, respectively, over said interval, maximise the open-loop estimate of the prediction gain over the nst sub-frames, in that the determining of the degree of voicing of the frame comprises a comparison between the second maximised estimate of the prediction gain over the frame and at least one other predetermined threshold (S1, S2), and in that the search interval determined on completion of the open-loop analysis is centred on said selected delay.

6. Method according to any one of Claims 3 to 5, characterised in that, if the second maximised estimate of the

prediction gain over a voiced frame is higher than one of the thresholds (S2), it is determined whether the nst optimal delays lie within an interval centred on the selected delay and containing a number N3 of delays which is less than N1 and, if so, the frame is assigned a degree of voicing for which the interval for searching for the long-term prediction delay contains N3 delays, the search interval containing N1 delays for at least one other degree of voicing.

7. Method according to any one of Claims 3 to 5, characterised in that, during the maximising of the second open-loop estimate of the long-term prediction gain over a voiced frame, a third open-loop estimate of the gain over the frame is also calculated on the basis of nst delays, lying within an interval centred on the selected delay and containing a number N3 of delays which is less than N1, which, respectively, over said interval of N3 delays, maximise the open-loop estimate of the prediction gain over the nst sub-frames, and in that the frame is assigned a degree of voicing for which the search interval contains N3 delays if said third estimate exceeds a predetermined threshold (S2), the search interval containing N1 delays for at least one other degree of voicing.

8. Method according to Claim 3 or 4, characterised in that the candidate delays of a list are chosen from among the sub-multiples of the basic delay which is associated with said list and from among the multiples of the smallest of said sub-multiples for which the open-loop estimate of the prediction gain is higher than said defined fraction of the estimate relating to the basic delay.

9. Method according to Claim 8, characterised in that the long-term prediction delays may correspond to integer or fractional numbers of samples of the speech signal, in that the basic delays (rbf) are determined in fractional resolution in order to search for the sub-multiples and the multiples to be included in a list of candidate delays, and in that the basic delays are determined in integer resolution in order to evaluate the first open-loop estimates of the prediction gain over a frame.

10. Method according to any one of Claims 3 to 9, characterised in that the closed-loop predictive analysis is not carried out in relation to each sub-frame for which the autocorrelation (C_{st}) of the speech signal associated with the optimal delay for said sub-frame is negative.

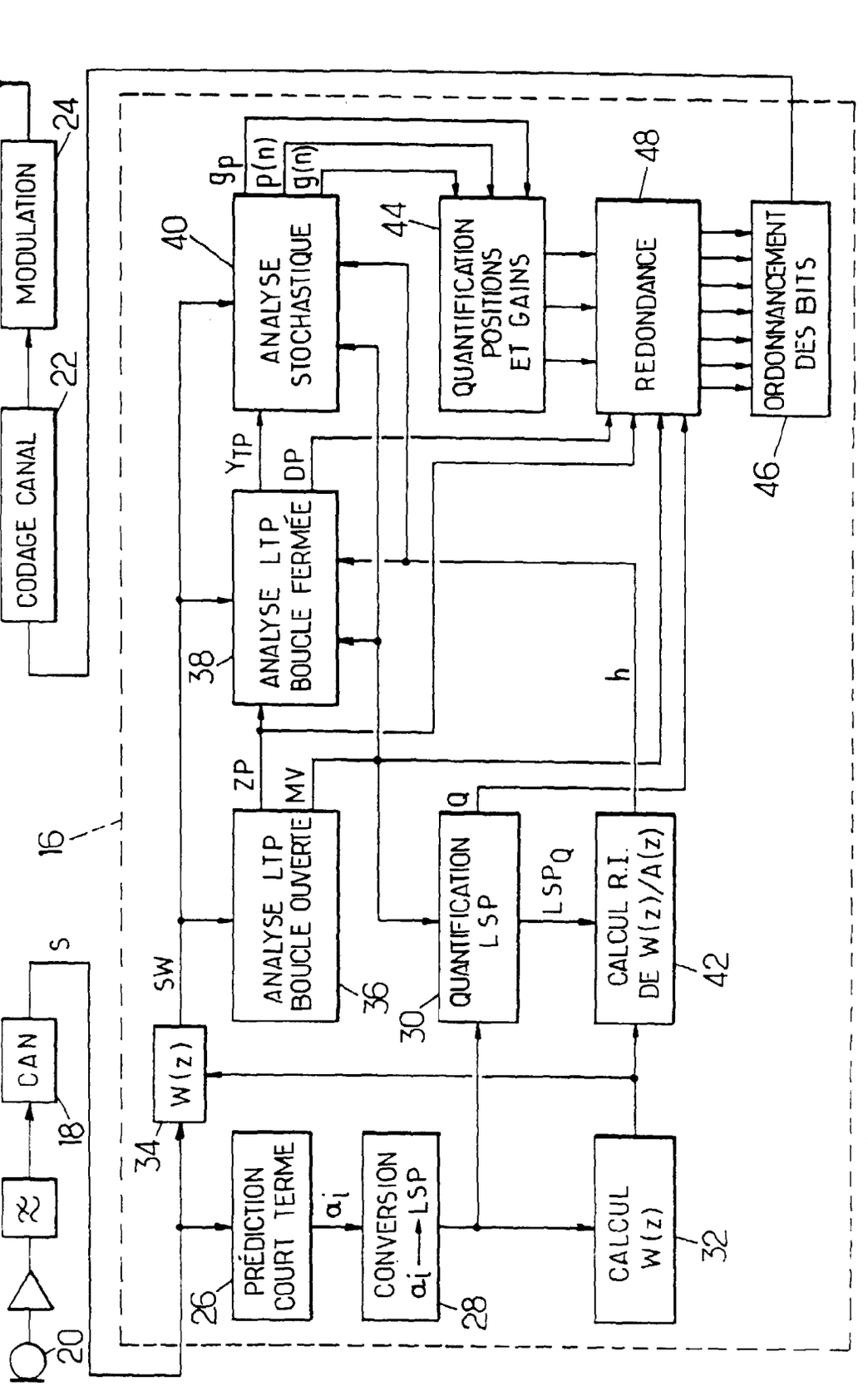
MODULATION
CODAGE CANAL
CAN
ANALYSE STOCHASTIQUE
QUANTIFICATION POSITIONS ET GAINS
REDONDANCE
ORDONNANCEMENT DES BITS
ANALYSE LTP BOUCLE FERMÉE
ANALYSE LTP BOUCLE OUVERTE
QUANTIFICATION LSP
CALCUL R.I. DE W(z)/A(z)
W(z)
PRÉDICTION COURT TERME
CONVERSION ai → LSP
CALCUL W(z)
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
s
SW
ZP
MV
YTP
DP
gp
p(n)
g(n)
q
LSPQ
h
ai

FIG. 1.

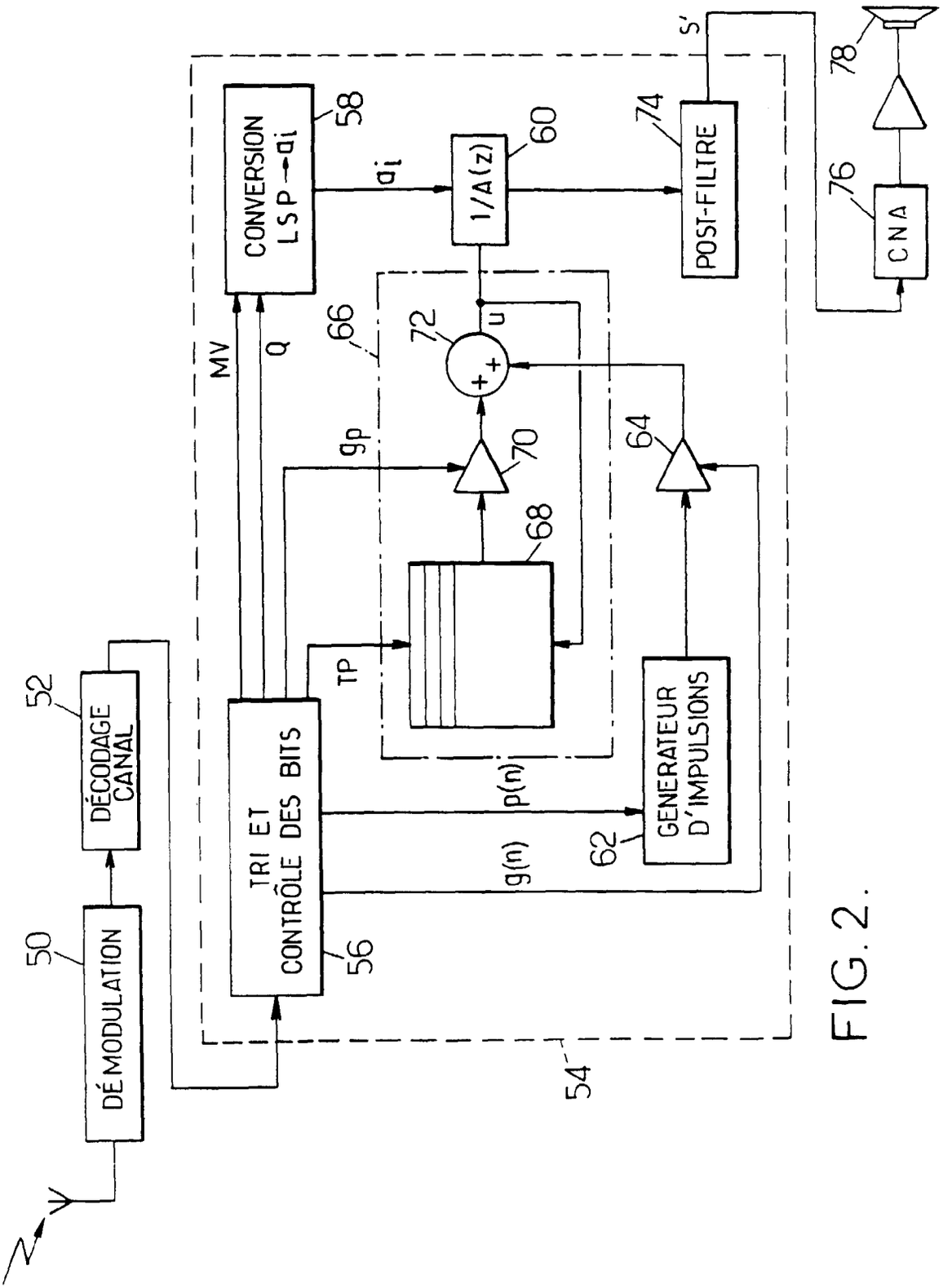
DÉMODULATION
50
DÉCODAGE CANAL
52
TRI ET CONTRÔLE DES BITS
56
CONVERSION LSP→ai
58
1/A(z)
60
GENERATEUR D'IMPULSIONS
62
64
66
68
70
72
+
+
u
POST-FILTRE
74
CNA
76
78
S'
54
MV
Q
ai
gp
TP
p(n)
g(n)

FIG. 2.

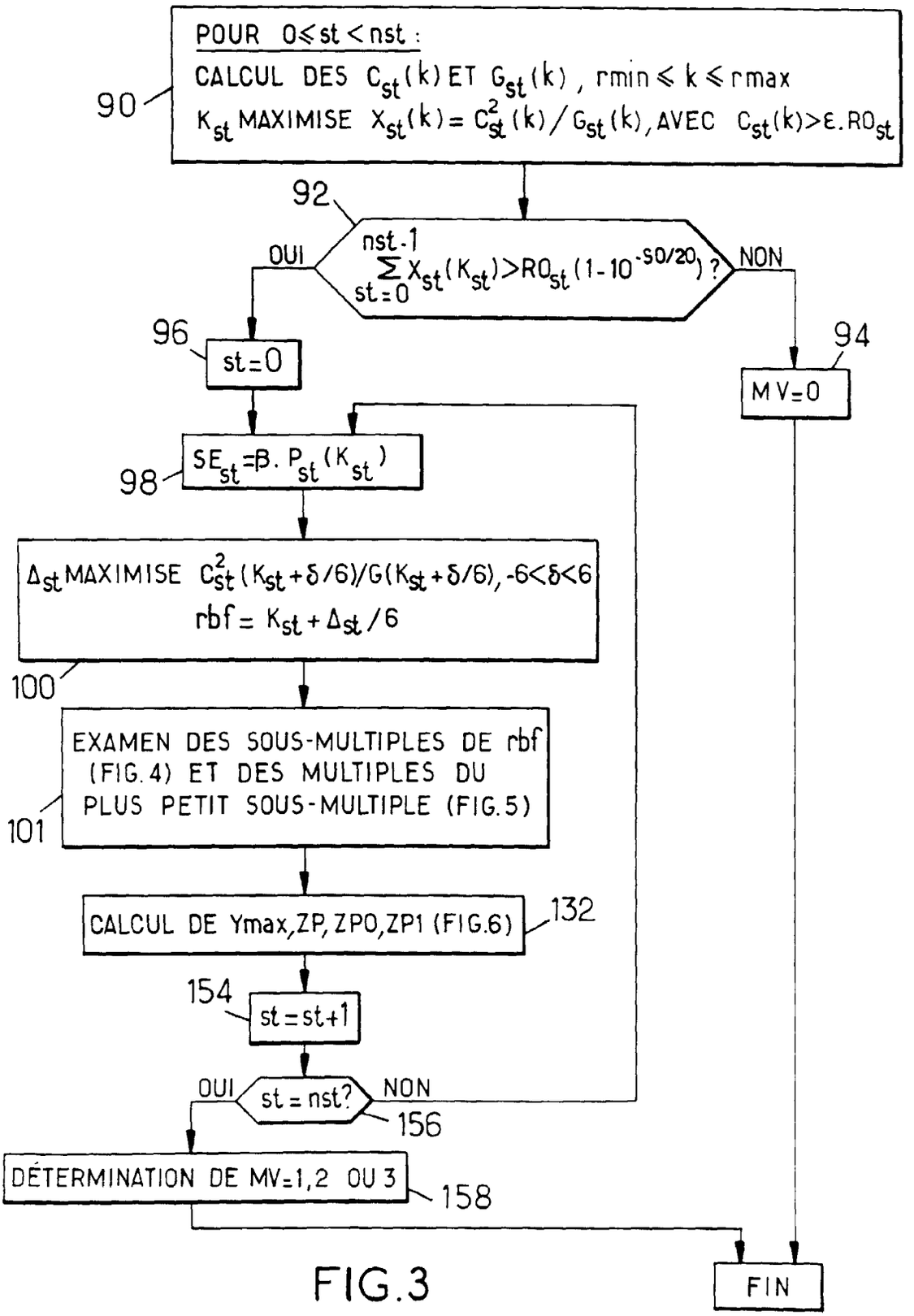
POUR 0≤st<nst :
CALCUL DES $C_{st}(k)$ ET $G_{st}(k)$, rmin≤k≤rmax
K_{st} MAXIMISE $X_{st}(k)=C^2_{st}(k)/G_{st}(k)$, AVEC $C_{st}(k)>\varepsilon.RO_{st}$
90
92
$\sum_{st=0}^{nst-1} X_{st}(K_{st})>RO_{st}(1-10^{-SO/20})$?
OUI
NON
96
st=0
94
MV=0
98
$SE_{st}=\beta.P_{st}(K_{st})$
Δ_{st} MAXIMISE $C^2_{st}(K_{st}+\delta/6)/G(K_{st}+\delta/6)$, $-6<\delta<6$
$rbf=K_{st}+\Delta_{st}/6$
100
EXAMEN DES SOUS-MULTIPLES DE rbf (FIG.4) ET DES MULTIPLES DU PLUS PETIT SOUS-MULTIPLE (FIG.5)
101
CALCUL DE Ymax,ZP,ZP0,ZP1 (FIG.6)
132
154
st=st+1
OUI
st=nst?
NON
156
DÉTERMINATION DE MV=1,2 OU 3
158
FIN

FIG.3

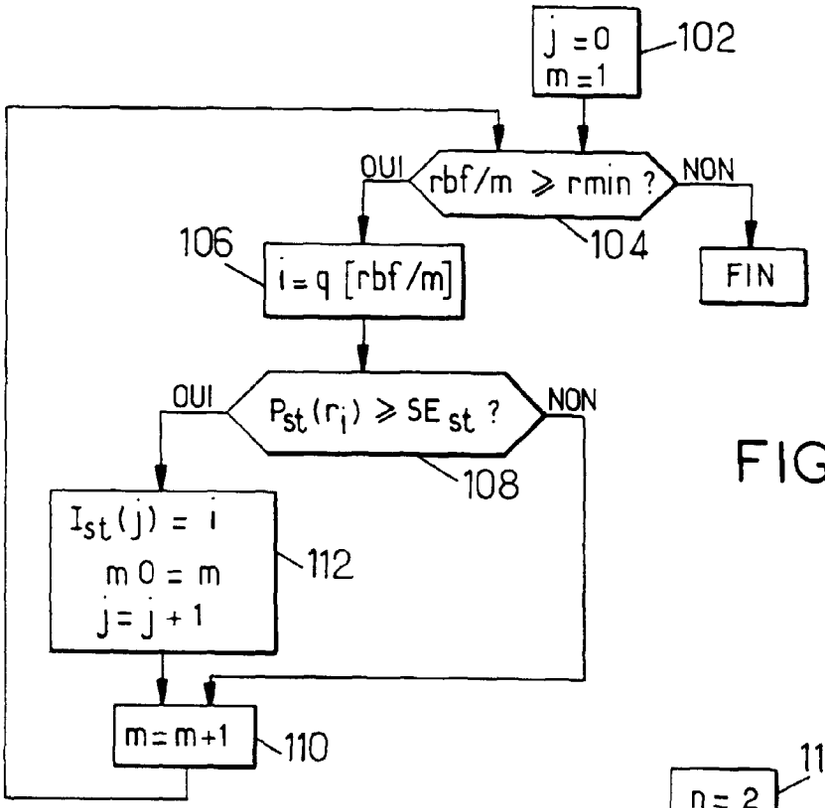
j = 0
m = 1
102
OUI
rbf/m ⩾ rmin ?
NON
104
106
i = q [rbf/m]
FIN
OUI
Pst(ri) ⩾ SE st ?
NON
108
Ist(j) = i
m 0 = m
j = j + 1
112
m = m+1
110

FIG. 4

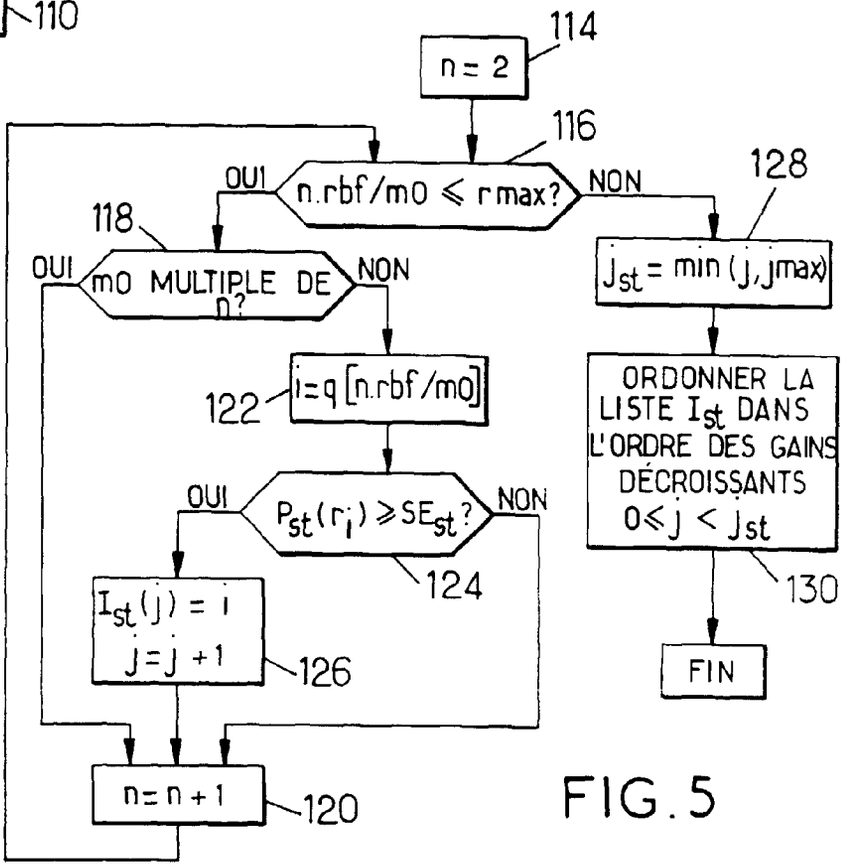
n = 2
114
116
OUI
n.rbf/m0 ⩽ rmax?
NON
128
118
OUI
m0 MULTIPLE DE n?
NON
Jst = min (j, jmax)
122
i=q [n.rbf/m0]
ORDONNER LA LISTE Ist DANS L'ORDRE DES GAINS DÉCROISSANTS
0⩽j < Jst
OUI
Pst(ri) ⩾SEst?
NON
124
130
Ist(j) = i
j = j + 1
126
FIN
n = n + 1
120

FIG. 5

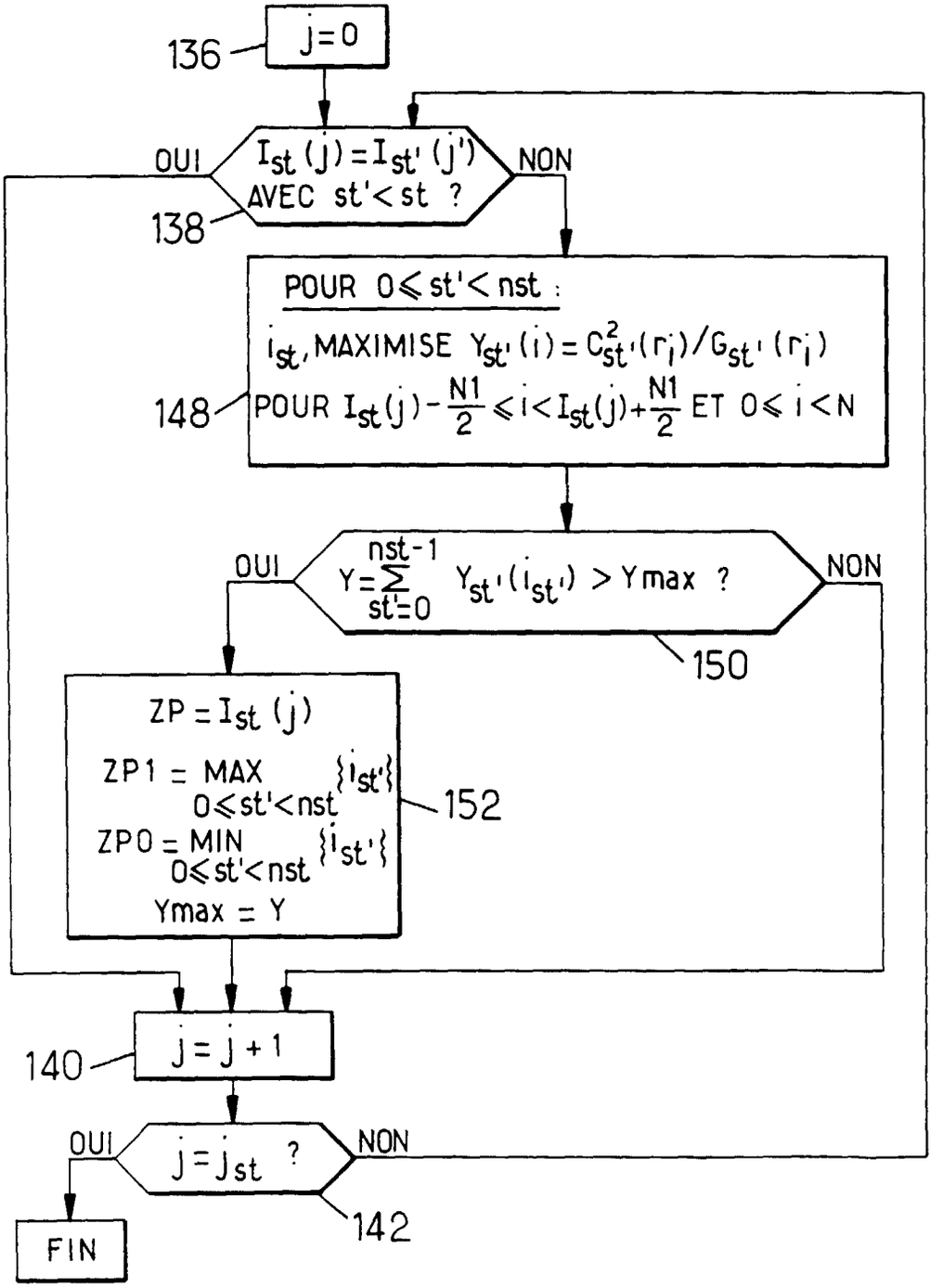
136
j=0
OUI
$I_{st}(j)=I_{st'}(j')$ AVEC st'<st ?
NON
138
POUR 0≤st'<nst :
$i_{st'}$, MAXIMISE $Y_{st'}(i)=C^2_{st'}(r_i)/G_{st'}(r_i)$
148
POUR $I_{st}(j)-\frac{N1}{2}\leq i<I_{st}(j)+\frac{N1}{2}$ ET 0≤i<N
OUI
$Y=\sum_{st'=0}^{nst-1} Y_{st'}(i_{st'})$ > Ymax ?
NON
150
ZP = $I_{st}(j)$
ZP1 = MAX $\{i_{st'}\}$ 0≤st'<nst
ZP0 = MIN $\{i_{st'}\}$ 0≤st'<nst
Ymax = Y
152
140
j = j + 1
OUI
j = j_{st} ?
NON
142
FIN

FIG.6

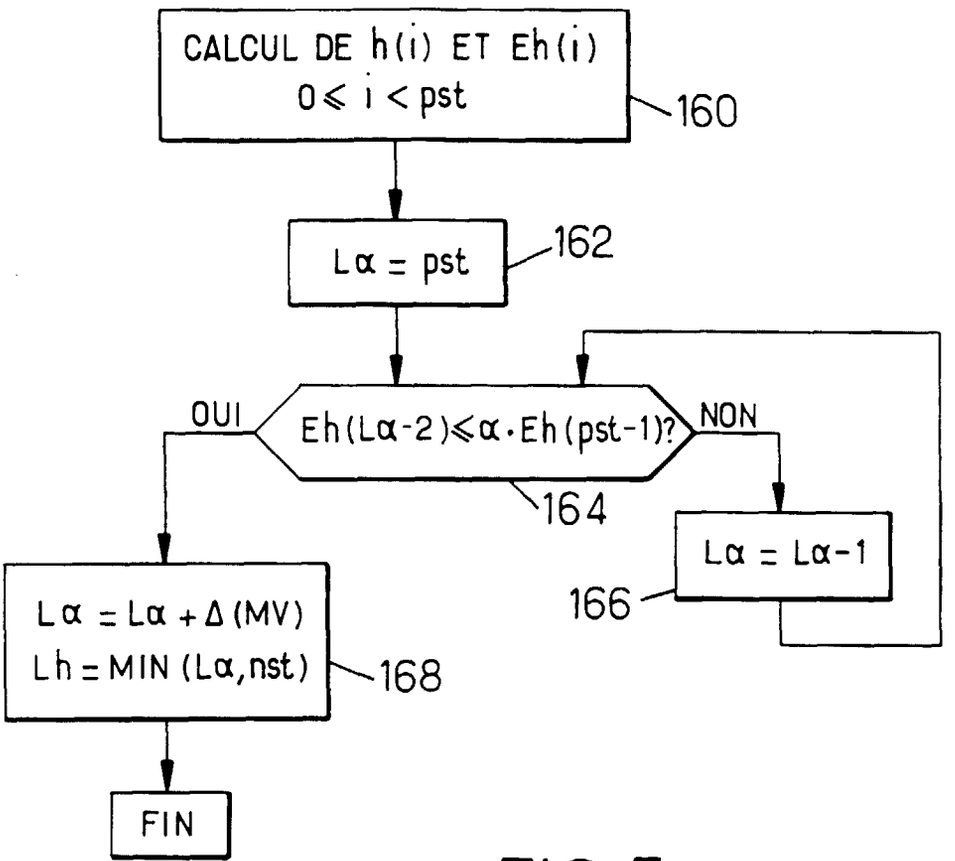
CALCUL DE h(i) ET Eh(i)
0 ⩽ i < pst
160
Lα = pst
162
OUI
Eh(Lα-2)⩽α·Eh(pst-1)?
NON
164
Lα = Lα-1
166
Lα = Lα + Δ(MV)
Lh = MIN (Lα, nst)
168
FIN

FIG.7

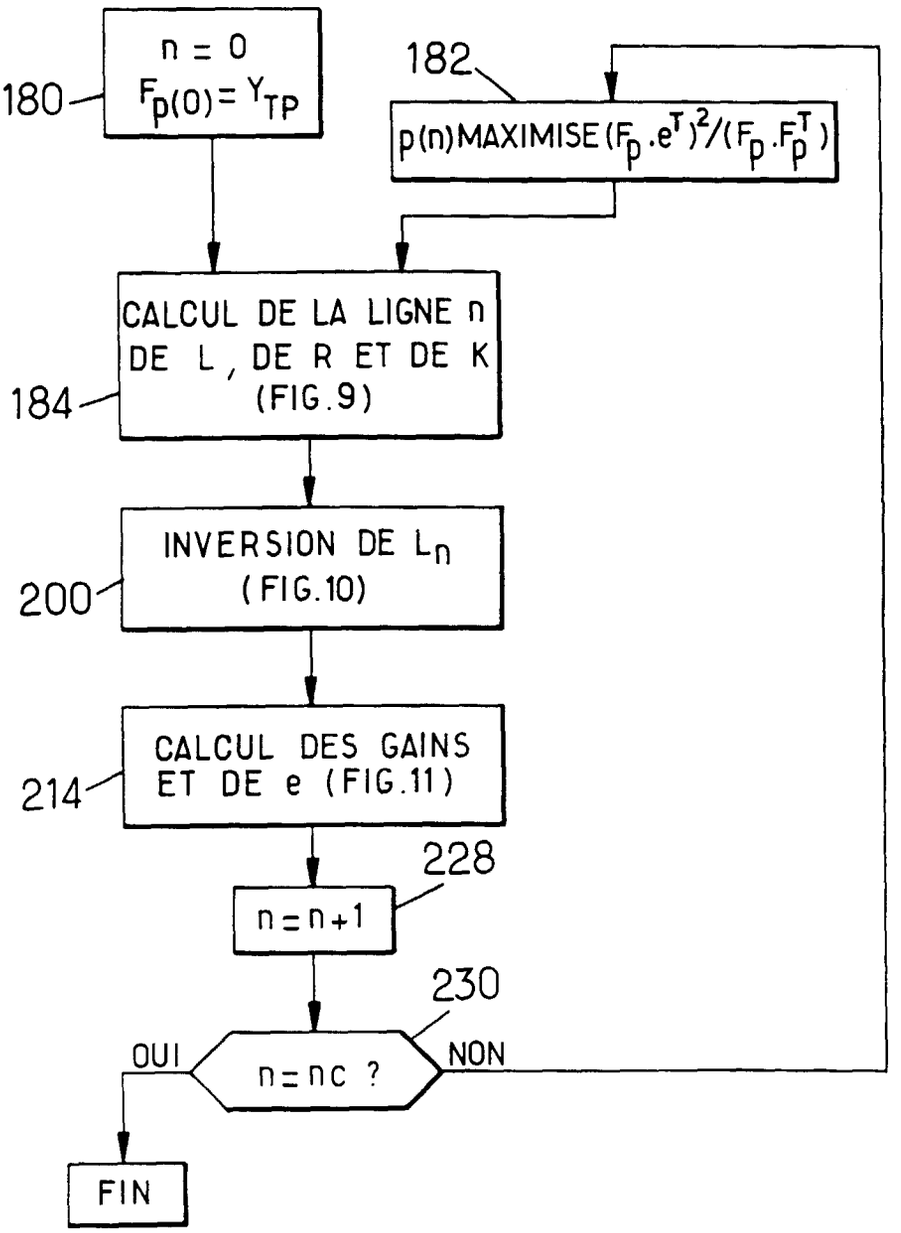
n = 0
$F_{p(0)} = Y_{TP}$
180
182
$p(n)$ MAXIMISE $(F_p . e^T)^2 / (F_p . F_p^T)$
CALCUL DE LA LIGNE n DE L, DE R ET DE K (FIG.9)
184
INVERSION DE L_n (FIG.10)
200
CALCUL DES GAINS ET DE e (FIG.11)
214
228
n = n+1
230
OUI
n = nc ?
NON
FIN

FIG.8

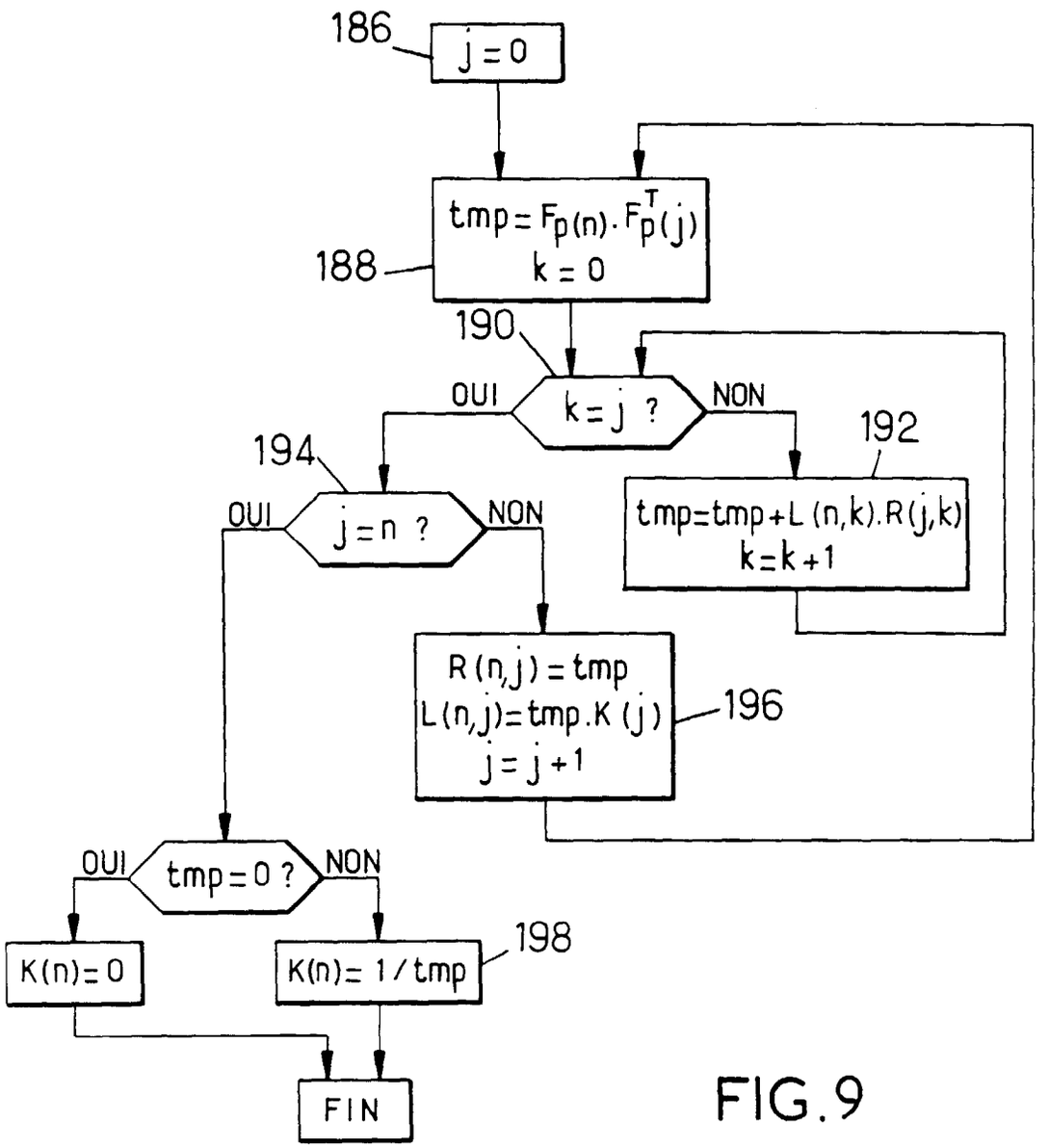
186
j = 0
tmp = Fp(n) . F^T p(j)
k = 0
188
190
OUI
k = j ?
NON
192
194
OUI
j = n ?
NON
tmp=tmp+L (n,k).R(j,k)
k=k+1
R (n,j) = tmp
L(n,j)=tmp.K (j)
j = j +1
196
OUI
tmp = 0 ?
NON
K(n)= 0
K(n)= 1 / tmp
198
FIN

FIG.9

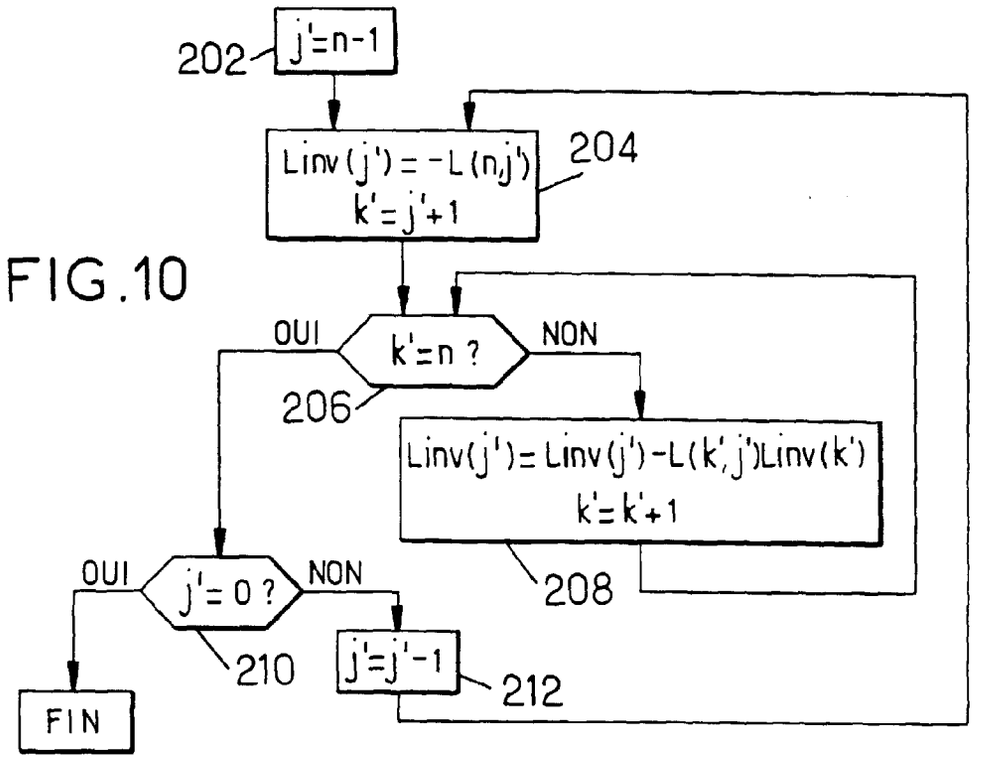
202
j'=n-1
Linv(j')=-L(n,j')
k'=j'+1
204
OUI
k'=n ?
NON
206
Linv(j')=Linv(j')-L(k',j')Linv(k')
k'=k'+1
208
OUI
j'=0 ?
NON
210
j'=j'-1
212
FIN

FIG.10

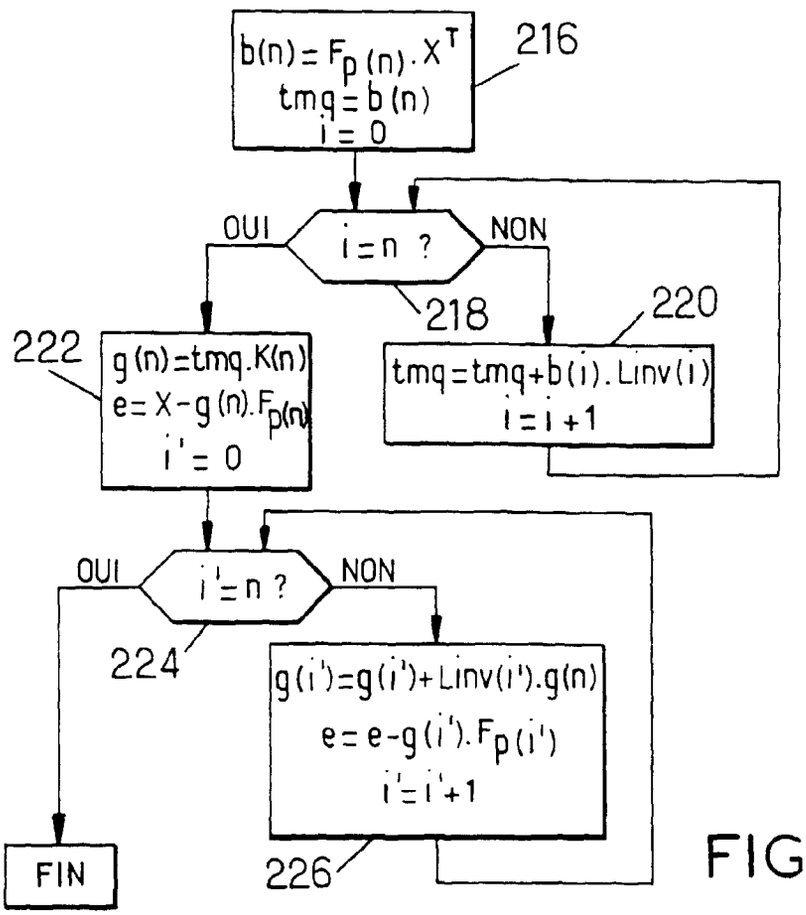
b(n)=Fp(n).X^T
tmq=b(n)
i=0
216
OUI
i=n ?
NON
218
220
222
g(n)=tmq.K(n)
e=X-g(n).Fp(n)
i'=0
tmq=tmq+b(i).Linv(i)
i=i+1
OUI
i'=n ?
NON
224
g(i')=g(i')+Linv(i').g(n)
e=e-g(i').Fp(i')
i'=i'+1
226
FIN

FIG.11