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

(11) EP 0 469 997 B1

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention de la délivrance du brevet:
11.09.1996 Bulletin 1996/37

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

(21) Numéro de dépôt: **91402153.0**

(22) Date de dépôt: **31.07.1991**

(54) **Procédé de codage et codeur de parole à analyse par prédiction linéaire**

Kodierungsverfahren und Sprachkodierer unter Anwendung von Analyse durch lineare Prädiktion

Coding method and speech coder using linear prediction analysis

(84) Etats contractants désignés:
BE CH DE DK ES GB IT LI NL SE

(30) Priorité: **02.08.1990 FR 9009905**

(43) Date de publication de la demande:
05.02.1992 Bulletin 1992/06

(73) Titulaire: **MATRA COMMUNICATION**
F-29101 Quimper (FR)

(72) Inventeurs:
- **Delprat, Marc**
  F-78150 Le Chesnais (FR)
- **Gruet, Christophe,**
  Résidence les Grandes Coudraies
  F-91190 Gif sur Yvette (FR)

(74) Mandataire: **Fort, Jacques**
CABINET PLASSERAUD
84, rue d'Amsterdam
75009 Paris (FR)

(56) Documents cités:
EP-A- 0 296 763

- **ICASSP'90, 1990 INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Albuquerque, New Mexico, 3-6 avril 1990, vol. 1, pages 469-472, IEEE, New York, US; I. TRANCOSO et al.: "Celp: a candidate for GSM half-rate coding"**
- **FREQUENZ, vol. 43, no. 9, septembre 1989, pages 242-252, Berlin, DE; J.-M. MÜLLER et al.: "Ein Beitrag zur Sprachcodierung für Bitraten unter 8 kbit/s"**
- **SIGNAL PROCESSING IV: THEORIES AND APPLICATIONS, (EUSIPCO-88, FOURTH EUROPEAN SIGNAL PROCESSING CONFERENCE, 5-8 septembre 1988, Grenoble, FR), vol. II, pages 871-874, North Holland Publishing Co., Amsterdam, NL; F. BOTTAU et al.: "On different vector predictive coding schemes and their application to low bit rates speech coding"**

EP 0 469 997 B1

Description

La présente invention concerne les procédés de codage et les codeurs de parole à analyse par synthèse, à prédiction linéaire, qui utilisent un modèle de production de la parole par passage d'un signal d'excitation, représentant la source vocale, à travers un filtre prédicteur à long terme de fonction de transfert $1/B(z)$ où $B(z) = 1-bz^{-T}$, où T est la période du fondamental de la parole et à travers un filtre prédicteur à court terme de fonction de transfert $1/A(z)$ représentant la contribution du conduit vocal et dont les caractéristiques spectrales varient lentement.

On connaît déjà de nombreux procédés et codeurs de ce type.

Un procédé qui s'est montré particulièrement satisfaisant consiste à générer le signal d'excitation à partir d'une séquence d'impulsions multiples (procédé dit MPLPC) ou d'un code choisi dans un dictionnaire (procédé dit CELP) et à utiliser un prédicteur à long terme qui prend en compte l'autocorrélation à long terme du signal de parole et qui est particulièrement efficace sur les sons voisés, car le signal est alors presque périodique. On peut ainsi obtenir un signal synthétique dont la forme d'onde se rapproche de celle de la parole à coder.

On trouvera une description de procédés de codage à prédiction linéaire CELP dans de nombreux documents auxquels on pourra se reporter, notamment dans "An efficient stochastically excited linear predictive coding algorithm for high quality low bit rates transmission of speech", W.B. KLEIJN et autres, Proceedings ICASSP, avril 1988, speech communication 7 (1988) 305-316, North Holland et dans "Synthesis filter optimization and coding applications to CELP", P. KABAL et autres, CH 2561 9880000-0147, 1988, IEEE. On connaît également (EP-A-0 347 307) un procédé de codage de parole, à prédiction linéaire et excitation par séquence extraite d'un dictionnaire et constituée par un vecteur d'excitation choisi dans le dictionnaire , permettant de coder des signaux de parole mis sous forme d'échantillons numérisés répartis en trames, procédé suivant lequel on représente chaque trame de signal de parole d'une part par des paramètres de prédiction, d'autre part par des séquences d'excitation contenues dans un dictionnaire et par des gains d'amplification de ces séquences, les séquences retenues étant déterminées par recherche du minimum de l'énergie d'un signal d'erreur obtenu par comparaison entre la trame de signal de parole d'origine et la trame synthètique reconstituée par synthèse à partir des paramètres. Un autre procédé de codage pertinent est décrit dans SIGNAL PROCESSING IV: THEORIES AND APPLICATIONS, (EUSIPCO-88, FOURTH EUROPEAN SIGNAL PROCESSING CONFERENCE, 5-8 septembre 1988, Grenoble, FR), vol. II, pages 871-874, North Holland Publishing Co., Amsterdam, NL; F. BOTTAU et al.: "On different vector predictive coding schemes and their application to low bit rates speech coding".

Les procédés de codage ci-dessus permettent une bonne reproduction de la parole avec un débit relativement faible, de l'ordre de 8 kbits par seconde. Les paramètres sont généralement transmis à intervalles correspondant chacun à une fenêtre pendant laquelle on conserve les mêmes caractéristiques de prédiction à court terme et à plusieurs trames. La durée des trames doit être suffisamment courte pour que les caractéristiques spectrales du signal de parole évoluent peu pendant leur durée. Le nombre d'échantillons par trame constitue un compromis. Il est souhaitable que la durée d'une trame soit inférieure à la période minimum du fondamental de la parole (désignée par pitch lag dans les articles en anglais). Mais une trame longue a l'avantage de se traduire par de moindres variations de l'énergie moyenne du signal d'une trame à l'autre. Dans la pratique, on choisit généralement une trame d'une quarantaine d'échantillons.

Si les procédés ci-dessus permettent de réduire très considérablement le débit requis pour une reproduction satisfaisante de la parole, ils exigent en contrepartie un volume de calcul en temps réel qui, dans les premières mises en oeuvre du procédé, était inacceptable. Le procédé décrit dans le document EP-A-0 347 307 permet de réduire notablement ce volume de calcul. La présente invention vise notamment à le réduire encore de façon significative, permettant ainsi une simplification des moyens mis en oeuvre, sans dégrader la restitution de la parole.

Dans ce but, l'invention telle que définie par les revendications propose notamment un procédé du type ci-dessus défini, caractérisé en ce qu'un signal d'erreur est obtenu en soustrayant, de la trame du signal de parole d'origine, deux termes, en ce que ladite trame de signal de parole est, avant soustraction, soumise à un filtrage d'analyse à court terme $\underline{r}$ prédéterminé et à un filtrage de synthèse pondéré perceptuel $\underline{H}$, le premier terme étant représentatif de la sortie du prédicteur à long terme avec un décalage temporel $\underline{T}$ égal à la période du fondamental, soumis au filtrage de synthèse avec un coefficient de proportionnalité tandis que le second terme est représentatif de chacune des séquences d'excitation à son tour, chaque séquence étant préalablement soumise à une amplification $\underline{G}$ d'un même filtrage de synthèse pondéré que la trame du signal de parole, et en ce que, au cours d'une même séquence et pour chaque trame, on détermine un optimum du décalage $\underline{T}$ par recherche d'un minimum d'erreur perceptuelle (par exemple d'erreur perceptuelle quadratique) et on en déduit le coefficient $\underline{k}$, puis on calcule simultanément les valeurs optimales de $\underline{b}$ et de $\underline{G}$ pour les valeurs retenues de $\underline{T}$ et de $\underline{k}$.

L'invention sera mieux comprise à la lecture de la description qui suit d'un mode particulier de réalisation donné à titre d'exemple non limitatif, correspondant au cas d'un codage à excitation vectorielle, et de la comparaison qui en est faite avec des procédés antérieurs.

La description se réfère aux dessins qui l'accompagnent dans lesquels :

- la figure 1 est un schéma de principe montrant les opérations qui fournissent la parole synthétique à partir des paramètres qui la représentent, dans les procédés de codage à analyse par prédiction linéaire ;
- la figure 2 est un schéma synoptique de principe montrant une constitution possible d'un codeur suivant l'invention.

En transmission et synthèse de parole, la parole synthétique est obtenue par la séquence d'opérations schématisées en figure 1. Un signal d'excitation, c_n pour la trame d'ordre n (signal constitué par un vecteur d'excitation dans le cas du procédé CELP), est soumis à un filtrage prédicteur à long terme 10 et un filtrage prédicteur à court terme 12, ayant respectivement des fonctions de transfert 1/B(z) et 1/A(z), en utilisant la notation classique en z. La sortie $\hat{s}_n$ représente une synthèse d'estimation du signal de parole, c'est-à-dire une parole synthétique.

Le filtre A(z) est généralement un filtre "tous zéros" de la forme :

$$A(z) = \Sigma a_i z^{-i} \text{ , } i = 0,...,M \text{ avec } a_0 = 1 \quad (1)$$

et le filtre B(z) est généralement à un seul coefficient, donc a une fonction de transfert de la forme :

$$B(z) = 1 - bz^{-T}. \quad (2)$$

Pour permettre de reconstituer, dans un décodeur, la parole synthétique il est nécessaire de transmettre :

- les paramètres du signal d'excitation c_n, c'est-à-dire l'identification du vecteur d'excitation et un coefficient d'amplification G ;
- les paramètres caractéristiques du prédicteur à long terme, qui prend en compte la corrélation à long terme du signal de parole, donc tient compte de la "mélodie" (ou "pitch" suivant la terminologie anglo-saxonne), constitués par le décalage temporel T et par le coefficient unique b ;
- les coefficients a_i de la fonction de transfert du filtre à court terme, qui assurent la synthèse des formants, liés aux résonances du conduit vocal.

Dans certaines réalisations, les coefficients du filtre à court terme sont mis à jour à une cadence plus faible que celle du vecteur d'excitation et du filtre à long terme c'est-à-dire à intervalles de plusieurs trames, correspondant à des "fenêtres" : l'invention est applicable aussi bien à ce cas qu'à celui où le rafraîchissement est effectué pour tous les paramètres à la cadence de répétition des trames.

Les paramètres ci-dessus définis sont déterminés dans le codeur pour chaque trame. Le signal d'excitation est déterminé de façon à minimiser l'écart entre le signal de parole original s_n et le signal synthétique $\hat{s}_n$, reconstitué dans le codeur lui-même à partir des paramètres, c'est-à-dire qu'on met en oeuvre une analyse par synthèse. Pour cela, on calcule, pour chaque signal d'excitation disponible, l'énergie du signal d'erreur. En règle générale, l'erreur prise en compte n'est pas obtenue par comparaison directe entre le signal original s_n et le signal synthétique $\hat{s}_n$, mais une erreur dite perceptuelle, tenant compte de ce que l'incidence du bruit est plus faible dans les zones du spectre de fréquence où le niveau du signal est élevé. On adopte en général un signal d'erreur (généralement un signal d'erreur quadratique) fourni par un filtre dit de pondération perceptuelle (frequency weighted en terminologie anglo-saxonne) de la forme :

$$W(z) = A(z) / A(z/\gamma) \quad (3)$$

où γ est un coefficient fixe, généralement d'environ 0,8.

Dans la pratique, la pondération peut être effectuée non pas sur le signal d'erreur, mais sur le signal de parole et le signal synthétique avant de former la différence entre ces signaux. Le signal d'origine est alors filtré par A(z), puis 1/A(z/γ) avant d'être comparé au signal d'excitation lui-même filtré par 1/B(z) puis 1/A(z/γ).

Les valeurs optimales b_0 et T_0 du prédicteur à long terme peuvent être calculées avec le même critère d'erreur que les paramètres de l'excitation (constitués par exemple par l'indice k d'un vecteur d'excitation dans un dictionnaire et un coefficient d'amplification G). Cette méthode de prédiction à long terme est dite "en boucle fermée", car les paramètres sont calculés dans la boucle d'analyse du codeur. Ce processus d'analyse a l'inconvénient, même lorsqu'on met en oeuvre des procédés simplifiés tels que celui décrit dans l'article de KLEIJN mentionnée ci-dessus, d'une grande complexité.

Comme on l'a indiqué plus haut, le procédé selon l'invention vise à obtenir une restitution satisfaisante de la parole avec une charge de calcul réduite. Le premier résultat est obtenu par un calcul qu'on peut qualifier de simultané de tous les paramètres du fait qu'on ne fait pas d'hypothèse a priori sur le signal d'excitation pour calculer les paramètres de filtrage à long terme. Le second résultat est atteint en effectuant successivement

- le calcul de la valeur optimale T_0 du retard T, par une méthode de calcul rapide, mais en boucle fermée pour avoir une bonne restitution,
- le calcul simultané du coefficient unique b du prédicteur à long terme (qui constitue un gain) et des paramètres

de l'excitation, par un calcul d'optimisation utilisant des hypothèses simplificatrices que l'expérience a montré justifiées.

Il ne sera plus question du prédicteur à court terme, de fonction de transfert A(z) dont les coefficients peuvent être calculés par l'un quelconque des procédés habituels.

Avant de décrire l'invention proprement dite, il est nécessaire de rappeler des modes connus de calcul des valeurs optimales b_0 et T_0 des paramètres b et T en minimisant l'erreur E de prédiction du prédicteur à long terme sur le signal résiduel r provenant du prédicteur à court terme, qui constitue un codeur à prédiction linéaire, souvent désigné par l'abréviation LPC.

Un procédé en boucle ouverte utilise le fait que l'erreur de prédiction E_1 du prédicteur à long terme (PLT) peut s'écrire, pour la trame d'ordre n :

$$E_1 = \Sigma \, (r_n - b.r_{n-T})^2, \tag{4}$$

où l'indice n-T désigne r_n retardé de T.

Une première approche, dite "en boucle ouverte" utilise le fait que le minimum est atteint lorsque la dérivée partielle de E_1 par rapport à b est nulle. En conséquence, la recherche du minimum de E_1 revient à trouver le retard T_0 (retard de recopie du signal passé) pour laquelle on obtient un maximum de Q(T) défini par :

$$Q\,(T) = (r^t.r_T)^2 \,/\|r_T\|^2 \tag{5}$$

Le procédé en boucle ouverte est rapide mais dégrade la qualité, surtout dans le cas où les trames sont courtes c'est-à-dire où les paramètres sont renouvelés souvent.

Comme on l'a indiqué plus haut, il est préférable de calculer les paramètres du prédicteur à long terme en boucle fermée et pour cela de rechercher les valeurs b_0 et T_0 qui minimisent l'énergie E du signal d'erreur perceptuelle; jusqu'ici ce calcul était fait en supposant que le signal d'excitation est nul (ou a une énergie donnée). Cela revient à déterminer séquentiellement les paramètres du prédicteur à long terme et les paramètres d'excitation. Mais la procédure est très complexe.

Elle sera rapidement exposée afin de donner les notations qui sont utilisées plus loin pour décrire l'invention.

En supposant le signal d'excitation nul, c'est-à-dire dans le cas d'une détermination séquentielle, le signal en sortie du prédicteur à long terme s'écrit :

$$e_n = b.e_{n-T} \tag{6}$$

ou

$$\vec{e} = b.\overrightarrow{e_T} \tag{7}$$

en notation vectorielle.

Le signal d'erreur perceptuelle est alors obtenu en comparant le signal original filtré par A(z) (ce qui donne le signal résiduel r) puis par $1/A(z/\gamma)$, au signal e_T en sortie du PLT filtré par $1/A(z/\gamma)$. En supposant que le signal résiduel r prend en compte la mémoire des deux filtrages par $1/A(z/\gamma)$, ces filtrages peuvent s'exprimer simplement par un produit matriciel de la matrice H de la réponse impulsionnelle de $1/A(z/\gamma)$.

L'énergie du signal d'erreur est donc :

$$E = (H.r - b.H.e_T)^2 \tag{8}$$

Les paramètres optimaux sont alors déterminés en deux étapes :

1°) On recherche T_0 qui maximalise :

$$Q(T) = (\, r^t.H^t.H.e_T)^2 \,/\, \|H.e_T\|^2 \tag{9}$$

2°) On calcule le gain correspondant :

$$b_0 = (r^t.H^t.H.e_{T0}) \,/\, \|H.e_{T0}\|^2 \tag{10}$$

Cette procédure est très complexe.

Toutes les séquences e_T possibles doivent être filtrées par $1/A(z/\gamma)$ pour calculer $H.e_T$ qui intervient dans l'équation (8) et déterminer Q(T) pour chacune des valeurs possibles de T (une centaine en général) dans un intervalle prédéterminé dont les limites inférieure et supérieure sont généralement entre 25 et 40 et entre 100 et 150, respectivement, pour un signal de parole échantillonné à 8 kHz.

Le calcul du numérateur de Q(T) dans la formule (9) ne présente pas de difficulté et peut être effectué de façon bien connue : en effet ce numérateur N(T) s'écrit :

$$N(T) = (r^t.H^t.He_T)^2 = ([H^t.H.r]^t.e_T)^2 \quad (11)$$

Le facteur $y = H^t.H.r$ peut être calculé une seule fois par trame, par multiplication matricielle de la matrice $H^t.H$ et du vecteur r (signal résiduel du filtrage par A(z)). Le calcul du numérateur N(T) n'exige alors plus qu'un produit scalaire $y^t.e_T$ pour chacune des valeurs possible de T.

Mais il subsiste le problème du calcul du dénominateur de Q(T) de la formule (9). On a déjà proposé de considérer que ce dénominateur est constant, ce qui revient à supposer que l'énergie du signal est constante ; cette hypothèse est inacceptable pour des trames courtes car elle conduit à une dégradation significative de la prédiction à long terme.

On pourrait penser qu'un compromis satisfaisant consiste à déterminer la valeur optimale T_0 du retard T en "boucle ouverte" et à déterminer ensuite la valeur optimale b_0 de b en "boucle fermée" conformément à l'équation (10), en partant de l'hypothèse que le prédicteur à long terme est essentiellement efficace pour les sons voisés et sur ces zones de dignal le retard du PLT est fortement lié à la période du fondamental, si bien que la méthode "boucle ouverte" conduira le plus souvent au même T_0 que la méthode "boucle fermée". En fait, cette procédure "mixte" donne des résultats intermédiaires entre les performances obtenues en boucle ouverte et en boucle fermée.

L'invention prend en considération cette constatation et écarte le calcul du T_0 en boucle ouverte au profit d'un calcul en boucle fermée. Elle propose donc :

(1) d'utiliser, pour le calcul de T_0, une fonction d'autocorrélation semblable à Q(T) de l'équation (9), normalisée pour tenir compte des variations de l'énergie de e_T avec T, et,

(2) pour simplifier les calculs, de faire l'hypothèse que $\|H.e_T\|^2$ dans l'équation (9) varie sensiblement de la même façon que $\|e_T\|^2$ c'est-à-dire que les variations des caractéristiques spectrales de la voix (répartition de l'énergie dans le spectre) sont négligeales d'une séquence e_T à la suivante.

Maximiser Q(T) pour déterminer T_0 revient alors à maximiser Q'(T) :

$$Q'(T) = N(T) / \|e_T\|^2 \quad (9bis)$$

Une fois T_0 déterminé, b_0 peut être calculé par la formule (10).

Ce mode de calcul élimine la nécessité de faire le calcul de filtrage de toutes les séquences e_T par 1/A (z/γ) et donne des performances pratiquement équivalentes à celle de la méthode rigoureuse.

Pour simplifier le calcul, on peut utiliser la méthode d'autocorrélation, déjà envisagée dans l'article de Kleijn mentionné plus haut. Dans ce cas la matrice $R = H^t.H$ est une matrice de Toeplitz symétrique dont la $i^{ième}$ diagonale contient le coefficient d'autocorrélation R_{i-1} de la réponse impulsionnelle h (de la matrice H) de 1/A(z/γ). Alors y = R.r peut être calculé efficacement sous forme du résultat d'une opération de filtrage particulière, tout comme pour la détermination de la séquence d'innovation dans la technique RPCELP décrite dans le docuent EP-A-0 347 307 déjà mentionné. En utilisant une pondération perceptuelle appropriée, R peut même devenir indépendant du temps sans perte de qualité subjective.

Le procédé suivant l'invention permet d'obtenir une qualité de parole synthétique meilleure que celle résultant :

- du calcul des paramètre T_0, b_0 du prédicteur à long terme, puis
- de la sélection de la séquence d'innovation et du gain G, c'est-à-dire des paramètres de l'excitation,

en optimisant conjointement tous les paramètres, T_0 étant seul calculé préalablement, et ce sans la charge excessive de calcul à laquelle on pourrait s'attendre. Le terme "conjointement" doit être compris comme indiquant qu'on ne fait pas d'hypothèse a priori sur la valeur de $b.e_T$ pour calculer G et K.

On supposera dans ce qui suit que la séquence d'innovation à multiplier par le gain G est un vecteur C_k choisi dans un dictionnaire et identifié par l'indice k. Mais le procédé est également applicable à une excitation multi-impulsionnelle car dans ce cas chaque impulsion peut être calculée à partir d'un dictionnaire dont toutes les séquences sont constituées de mono-impulsions dans toutes les positions possibles.

L'énergie E du signal d'erreur perceptuelle à minimiser est, puisque l'on tient compte de l'excitation, donnée par la formule (12) qui se substitue à (8) :

$$E = (H.r - b.H.e_T - G.H.C_k)^2 \quad (12)$$

En annulant les dérivées partielles de E par rapport à b et à G on obtient alors deux équations :

$$b = (R_1.E_2 - R_2.R_3) / (E_1.E_2 - R_3^2) \quad (13)$$

et

$$G = (R_2.E_1 - R_1.R_3) / (E_1.E_2 - R_3^2) \quad (14)$$

avec

$$\begin{cases} R_1 = r^t.R.e_T,\ R_2 = r^t.R.c_k\ ,\ R_3 = {e_T}^t.R.c_k \\ E_1 = {e_T}^t.R.e_T,\ E^2 = c_k.R.c_k \end{cases} \quad (15)$$

La valeur optimale T_0 du retard T et l'indice k de la séquence optimale sont alors déterminées en cherchant à maximiser C (T,k) :

$$C(T,k) = (R_2^2.E_1 + R_1^2 - 2R_1.R_2.R_3) / (E_1.E_2 - R_3^2) \quad (16)$$

Pour réduire la complexité du calcul sans dégrader les résultats, on détermine d'abord T_0, puis on cherche l'indice k qui maximise $C(T_0,k)$ ce qui revient à faire le calcul de (15) pour toutes les indices k possibles, et, simultanément, on calcule les valeurs à retenir pour b et G.

Le retard T_0 est avantageusement calculé en boucle fermée, par l'algorithme rapide de la formule (9bis).

Les algorithmes ci-dessus peuvent encore être simplifiés dans de nombreux cas. Par exemple on peut éviter la division dans le calcul de Q'(T) en utilisant des produits croisés pour comparer les valeurs successives. De plus, chaque terme d'énergie $\|e_T\|^2$ peut être calculé efficacement à partir de la valeur précédente $\|e_{T-1}\|^2$ en utilisant le fait que deux séquences successives e_T et e_{T-1} ne diffèrent que par deux échantillons (le premier et le dernier). Le calcul a alors une complexité très faible.

Une fois T_0 déterminé, R1 et E1 sont connus. Le calcul de R2 ne nécessite pas de filtrage supplémentaire car r^t. R a déjà été calculé pour déterminer T_0 et b_0 : il suffit de le mémoriser à l'issue de son calcul.

Le calcul de $R3 = {e_T}^t.R.C_k$ nécessite uniquement un seul filtrage, puisque la valeur T_0 de T est connue. E2 n'est calculé qu'une seule fois par trame, ou plus précisément chaque fois que H et la matrice R changent.

Si par ailleurs on utilise des séquences d'excitation c_k binaires régulières (procédé RPCELP), telles que celles décrites dans le document EP-A-0 347 307, le calcul de tous les produits scalaires de C_k des formules (15) est très simple. De plus E2 devient constant et il n'est plus nécessaire de le calculer.

Il est souvent justifié de faire une hypothèse supplémentaire conduisant à simplifier encore le calcul : si on néglige la corrélation entre les paramètres T et b du prédicteur à long terme et l'excitation, c'est-à-dire si on admet que T_0 et b_0 sont indépendants de G et k, R3 est nul. La formule (16) se simplifie et devient :

$$C(T,k) = R_2^2 / E_2 + R_1^2 / E1 \quad (16bis)$$

Il n'y a plus à proprement parler d'optimisation conjointe de T et de k ; mais l'utilisation du mode de calcul représenté par la formule (9bis) conduit à un allégement de la charge de calcul.

La figure 2 montre la constitution de principe d'un codeur permettant de mettre en oeuvre le procédé qui vient d'être décrit. Le signal de parole échantillonné s_n est appliqué à l'entrée du module d'analyse 10 qui calcule, pour chaque trame, les coefficients a_i de la fonction de filtrage A(z) de prédiction à court terme et les envoie au codeur 12. Le module 10 pouvant avoir une constitution classique, il n'est pas nécessaire de le décrire. Le module 10 est également prévu pour fournir le résidu r de la prédiction à court terme, constitué par s_n filtré par A(z) quantifié. Ce résidu est soumis à un filtrage par $R = H^t.H$ en (14) ce qui fournit en sortie le signal y = R.r.

Le calcul des coefficients du filtrage à long terme est assuré par un bloc 16 qui reçoit y et l'erreur de prédiction à long terme, mémorisée à partir de la trame précédente et donc retardée de T, fournie par une mémoire 18 constituant un dictionnaire évolutif.

Les paramètres y et e_T sont appliqués à un multiplieur 20 qui fournit une sortie, portée au carré en 22 pour former N(T) donnée par la formule (11). Un second circuit 24, identique à 22, forme $\|e_T\|^2$. Un diviseur 26 permet alors d'obtenir la grandeur Q′ à maximiser.

Le bloc 28 est constitué par un processeur permettant de calculer le retard T_0 pour lequel Q′ est maximum, par comparaison entre les valeurs de Q′ pour les différentes valeurs possibles de T. Le calcul peut s'effectuer soit par une logique câblée, soit par programmation. L'optimum T_0 obtenu, sans hypothèse en ce qui concerne la séquence d'excitation, est transmis à un processeur 30 de sélection de C_k parmi les séquences C fournies par un dictionnaire 32 et de calcul de b et G.

Le processeur 30 fournit les valeurs de k, b et G retenues, d'une part au codeur 12, pour transmission d'un jeu de

paramètres a_i, T_0, b, k, et G pour chaque trame, d'autre part au générateur 34 de e_T, qui calcule les séquences e_T, les fournit en sortie et les transmet à la mémoire 18. Les a_i peuvent cependant n'être renouvelés qu'à intervalles correspondant à des fenêtres de durée supérieure à la durée d'une trame et la transmission s'effectue alors à la cadence des fenêtres.

Pour des raisons de simplicité, le filtre de pondération perceptuelle n'a pas été représenté en détail sur la figure 2.

Revendications

1. Procédé de codage à analyse par synthèse, à prédiction linéaire, utilisant un modèle de production de la parole par passage d'un signal d'excitation C d'indice k parmi plusieurs signaux mémorisés, représentant la source vocale et soumis à amplification G, à travers un filtre prédicteur à long terme de fonction de transfert 1/B(z) où $B(z) = 1 - bz^{-T}$, où T est la période du fondamental de la parole, et à travers un filtre prédicteur à court terme de fonction de transfert $1/A(z) = \Sigma a_i z^i$ représentant la contribution du conduit vocal et dont les caractéristiques spectrales varient lentement, en représentant chaque trame par des valeurs des paramètres C_k, G, a_i, T et b, suivant lequel : on soumet la trame de signal de parole d'origine au filtrage d'analyse à court terme et à un filtrage de synthèse pondéré perceptuel H ; on génère un signal d'erreur en soustrayant, de la trame du signal de parole ainsi filtrée, un premier terme représentatif de la sortie du prédicteur à long terme avec le décalage temporel T, soumis au filtrage de synthèse, et un second terme représentatif de chacun des signaux d'excitation à son tour, chaque signal étant préalablement soumis à une amplification G et au même filtrage de synthèse pondéré que la trame du signal de parole ; et, au cours d'une même séquence et pour chaque trame, on détermine d'abord un optimum du décalage T par recherche d'un minimum d'erreur perceptuelle et on en déduit l'indice k, puis on calcule simultanément les valeurs optimales de b et de G pour les valeurs retenues de T et de k, la valeur optimale T_o du décalage T étant calculée en boucle fermée par recherche de la valeur pour laquelle :

$$(r^t.H^t.He_T)^2 / \|e_T\|^2$$

est maximum, l'exposant t désignant l'opération de transposition, H étant la matrice de la réponse impulsionnelle de $1/A(z/\gamma)$ où γ est un coefficient fixe, et e_T la sortie de la prédiction à long terme.

2. Procédé selon la revendication 1, caractérisé en ce que l'erreur minimisée est l'erreur perceptuelle quadratique.

3. Procédé selon la revendication 1, caractérisé en ce que le coefficient b_0 optimum de filtrage à long terme est calculé à partir de T_0 par la formule

$$b_0 = (r^t.H^t.H.e_{T0}) / \|H.e_{T0}\|^2$$

4. Procédé selon la revendication 1, 2 ou 3, caractérisé en ce que le signal d'excitation optimal est sélectionné par minimisation de $C(T_0,k)$:

$$C(T,k) = (R_2^2.E_1 + R_1^2 - 2R_1.R_2.R_3) / (E_1.E_2-R_3^2)$$

où :

$$R_1 = r^t.R.e_T,\ R_2 = r^t.R.c_k\ ,\ R_3 = e_T^t.R.c_k$$

$$E_1 = e_T^t.R.e_T,\ E_2 = c_k.R.c_k,\ R = H^t.H$$

5. Procédé selon la revendication 4, caractérisé en ce qu'on calcule $H^t.H.r$ par filtrage en assimilant $H^t.H$ à une matrice de Toeplitz symétrique.

6. Procédé selon la revendication 5, caractérisé en qu'on détermine le gain par la formule :

$$G = (R_2.E_1-R_1.R_3) / (E_1.E_2-R_3^2)$$

7. Procédé selon la revendication 4, caractérisé en ce qu'on néglige la corrélation enre T_0 et b_0, G et k pour annuler le terme R3 et sélectionner k par minimisation de

$$C(T_0,k) = R_2^2 /E_2 + R_1^2 /E_1E_2$$

8. Codeur mettant en oeuvre le procédé selon la revendication 1, caractérisé en ce qu'il comprend un module d'analyse (10) destiné à recevoir le signal de parole échantillonné et à calculer, pour chaque trame, les coefficients a_i de la fonction de filtrage A(z) de prédiction à court terme et le résidu r de la prédiction à court terme, un filtre recevant le résidu r et fournissant en sortie un signal $y = R.r_\lambda$ où $R = H^T.H$, et un bloc (16) de calcul des coefficients du filtrage à long terme qui reçoit y et l'erreur de prédiction à long terme, obtenue à partir de la trame précédente et stockée dans une mémoire (18), un ensemble multiplieur (20), un circuit de mise au carré (22) destiné à fournir le numérateur d'un terme à maximiser, un second ensemble formant $\|e_T\|^2$, et un diviseur (26) permettant d'obtenir le terme à maximiser, appliqué à un processeur (28) de calcul de la valeur optimum, par comparaison entre les valeurs du dit terme pour les différentes valeurs possibles de T.

Patentansprüche

1. Codierungsverfahren mit Analyse durch Synthese mit linearer Prediktion, bei welchem ein Modell zur Erzeugung der Sprache durch den Durchtritt eines Anregungssignals C eines Index k unter einer Mehrzahl von gespeicherten Signalen verwendet wird, welches die Sprachquelle bildet und einer Verstärkung G unterzogen wird, quer durch einen Prediktionsfilter mit einem langen Übertragungsfunktionsterm 1/B(z), wobei $B(z) = 1 - bz^{-T}$ ist, wobei T die Grundperiode der Sprache ist, quer durch einen Prediktionsfilter mit einem kurzen Übertragungsfunktionsterm $1/A(z) = \Sigma a_i z^i$, welcher die Verteilung der Sprachleitung repräsentiert und dessen spektrale Charakteristiken langsam variieren, wobei jedes Raster durch Werte von Parametern C_k, G, a_i, T und b dargestellt wird, gemäß welchem man das Raster des Signals der ursprünglichen Sprache einer Analysenfilterung mit einem kurzen Term und mit einer ausgeglichenen wahrnehmbaren Synthesefilterung H unterzieht, man ein Fehlersignal erzeugt, indem man von dem so gefilterten Raster des Sprachsignals einen ersten Term, der repräsentativ für den Ausgang des Prediktors mit langem Term mit einer zeitlichen Abweichung T ist und einer Synthesefilterung unterzogen wird, und einen zweiten Term zuliefert, der repräsentativ für jedes der Anregungssignale auf ihrem Umlauf ist, wobei jedes Signal vorhergehend einer Verstärkung G und derselben ausgeglichenen Synthesefilterung wie das Raster des Sprachsignals unterzogen wird, und man im Laufe einer selben Sequenz und für jedes Raster zuerst ein Optimum der Abweichung T durch Suchen eines wahrnehmbaren Fehlerminimums bestimmt und man daraus den Index k ableitet und man dann gleichzeitig die optimalen Werte von b und G für die erhaltenen Werte von T und k berechnet, wobei der optimale Wert T_0 der Abweichung T in einer geschlossenen Schleife durch Suche des Wertes bestimmt, für den

$$(r^t.H^t.He_T)^2 / |e_T|^2$$

maximal ist, wobei der Exponent t die Transponierungsoperation bezeichnet, H die impulsartige Antwortmatrix von $1/A(z/\gamma)$ ist, wobei γ ein fester Koeffizient ist und e_T der Ausgang des Prediktors mit langem Term ist.

2. Verfahren nach Anspruch 1,
dadurch gekennzeichnet,
daß der minimalisierter Fehler der wahrnehmbare quadratische Fehler ist.

3. Verfahren nach Anspruch 1,
dadurch gekennzeichnet,
daß der Koeffizient b_0 des Optimums der Filterung mit langem Term ausgehend von T_0 durch die Formel

$$b_0 = r^t.H^t.H.e_{T0}) / |H.e_{T0}|^2$$

berechnet wird.

4. Verfahren nach Anspruch 1, 2 oder 3,
dadurch gekennzeichnet,
daß das optimale Anregungssignal durch Minimalisierung von $C(T_0,k)$ ausgewählt wird:

$$C(T,k) = (R_2^2.E_1 + R_1^2 - 2R_1.R_2.R_3) / (E_1.E_2-R_3^2),$$

wobei:

$$R_1 = r^t.R.e_T,\ R_2 = r^t.R.c_k,\ R_3 = e_T^t.R.c_k$$

$$E_1 = e_T^t.R.e_T,\ E_2 = c_k,R.c_k,\ R = H^t.H\ .$$

5. Verfahren nach Anspruch 4,
dadurch gekennzeichnet,
daß man $H^t.H.r$ durch Filterung berechnet, indem man $H^t.H$ mit einer symmetrischen Toeplitz-Matrix vergleicht.

6. Verfahren nach Anspruch 5,
dadurch gekennzeichnet,
daß man die Verstärkung durch folgende Formel berechnet:

$$G = (R_2.E_1-R_1.R_3) / (E_1.E_2-R_3^2).$$

7. Verfahren nach Anspruch 4,
dadurch gekennzeichnet,
daß man die Korrelation zwischen T_0 und b_0, G und k vernachlässigt, um den Term R3 zu annullieren und k durch Minimalisierung von

$$C(T_0,k) = R_2^2 /E_2 + R_2^2 /E_1.E_2$$

auszuwählen.

8. Codierer zur Durchführung des Verfahrens nach Anspruch 1,
dadurch gekennzeichnet,
daß er ein Analysemodul (10) zum Empfangen des ausgewählten Sprachsignals und zum Berechnen, für jedes Raster, der Koeffizienten a_i der Filterfunktion A(z) zur Prediktion mit einem kurzen Term und des Rückstandes r der Prediktion mit kurzem Term, einen Filter, der den Rückstand r empfängt und am Ausgang ein Signal $y = R.r_\lambda$ liefert, wobei $R = H^T.H$ ist, und einen Block (16) zur Berechnung der Filterkoeffizienten mit langem Term aufweist, der dort empfängt und den Fehler der Prediktion mit langem Term, der ausgehend vom vorhergehenden Raster erhalten und in einem Speicher (18) gespeichert worden ist, sowie einen Gesamtmultiplizierer (20), eine Kreuzschaltung (22) zur Lieferung eines Zählers eines zu maximierenden Termes, eine zweite Gesamtheit zur Bildung von $|e_T|^2$ und einen Divisor (26), der es erlaubt, den zu maximierenden Term zu erhalten, welcher auf einen Rechenprozessor (28) für den Optimalwert angewendet wird durch Vergleich zwischen den Werten dieses Termes für die verschiedenen möglichen Werte von T.

Claims

1. A synthesis analysis type linear predictive coding process, using a production model of the speech by passing an excitation signal C, with index k, through several stored signals, representing the speech source and subjected to amplification G, via a long-term predictive filter with transfer function 1/B(z), where $B(z) = 1 - bz^{-T}$, and T is the period of the fundamental of the speech, and via a short-term predictive filter with transfer function $1/A(z) = \Sigma a_i z^i$ representing the contribution of the speech channel and whose spectral characteristics slowly change, by representing each frame by the values of parameters C_k, G, a_i, T and b, according to which: the frame of the original speech signal undergoes short-term analysis filtering and undergoes weighted perceptual synthesis filtering H; an error signal is generated by subtracting from the frame of the speech signal thus filtered a first representative term from the output of the long-term predictor with the time lag T, subjected to synthesis filtering, and a second representative term of each of the excitation signals in turn, each signal having previously undergone amplification G and the same weighted synthesis filtering as the frame of the speech signal; and, during an identical sequence and for each frame, an optimum of the time lag T is first determined by searching for a minimum perceptual error and deducting the index k, then simultaneously calculating the optimum values of b and G for the values of T and k retained, the optimum value T_o of the time lag T being calculated in a closed loop by searching for the value for which:

$$(r^t.H^t.He_T)^2 / |e_T|^2$$

is maximum, the exponent t designating the transposition operation, H being the matrix of the pulse response of $1/A(z/\gamma)$, where γ is a fixed coefficient and e_T the output of the long-term prediction.

2. Process according to Claim 1, characterised in that the minimised error is the quadratic perceptual error.

3. Process according to Claim 1, characterised in that the optimum coefficient b_0 of long-term filtering is calculated from T_0 by the formula:

$$b_0 = (r^t.H^t.H.e_{T0}) / \|H.e_{T0}\|^2$$

4. Process according to Claim 1, 2 or 3, characterised in that the optimum excitation signal is selected by minimisation of $C(T_0,k)$:

$$C(T, k) = (R2^2.E1 + R1^2 - 2R1.R2.R3) / E1.E2-R3^2)$$

where:

$$R1 = r^t.R.e_T,\ R2 = r^t.R.C_k,\ R3 = e_T^t.R.C_k$$

$$E1 = e_T^t.R.e_T,\ E2 = C_k.R.C_k,\ R = H^t.H$$

5. Process according to Claim 4, characterised in that $H^t.H.r$ is calculated by filtering and comparing $H^t.H$ to a symmetrical Toeplitz matrix.

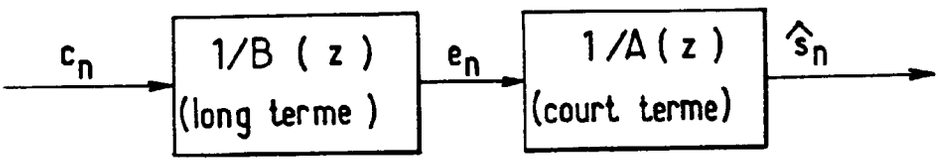
6. Process according to Claim 5, characterised in that the gain is determined by means of the formula:

$$G = (R2.E1-R1.R3) / (E1.E2-R3^2)$$

7. Process according to Claim 4, characterised in that the correlation between T_0 and b_0, G and k is ignored in order to cancel the term R3 and to select k by minimisation of $C(T_0,k) = R2^2/E2 + R1^2/E1E2$

8. Encoder implementing the process according to Claim 1, characterised in that it comprises an analysis module (10) intended to receive the sampled speech signal and to calculate for each frame the coefficients a_i of the short-term predictive filtering function A(z) and the remainder r of the short-term prediction, a filter receiving the remainder r and delivering at its output a signal $y = R.r_i$, where $R = H^T.H$, and a block (16) for calculating long-term filtering coefficients, which receives y and the long-term prediction error obtained from the preceding frame and stored in a memory (18), a multiplier unit (20), a squaring circuit (22) intended to supply the numerator of a term to be maximised, a second unit forming $|e_T|^2$, and a divider (26) enabling the term to be maximised to be obtained, applied to a processor (28) for calculating the optimum value by comparison between the values of said term for the various possible values of T.

FIG.1.


c_n
1/B (z)
(long terme)
e_n
1/A (z)
(court terme)
$\hat{s}_n$

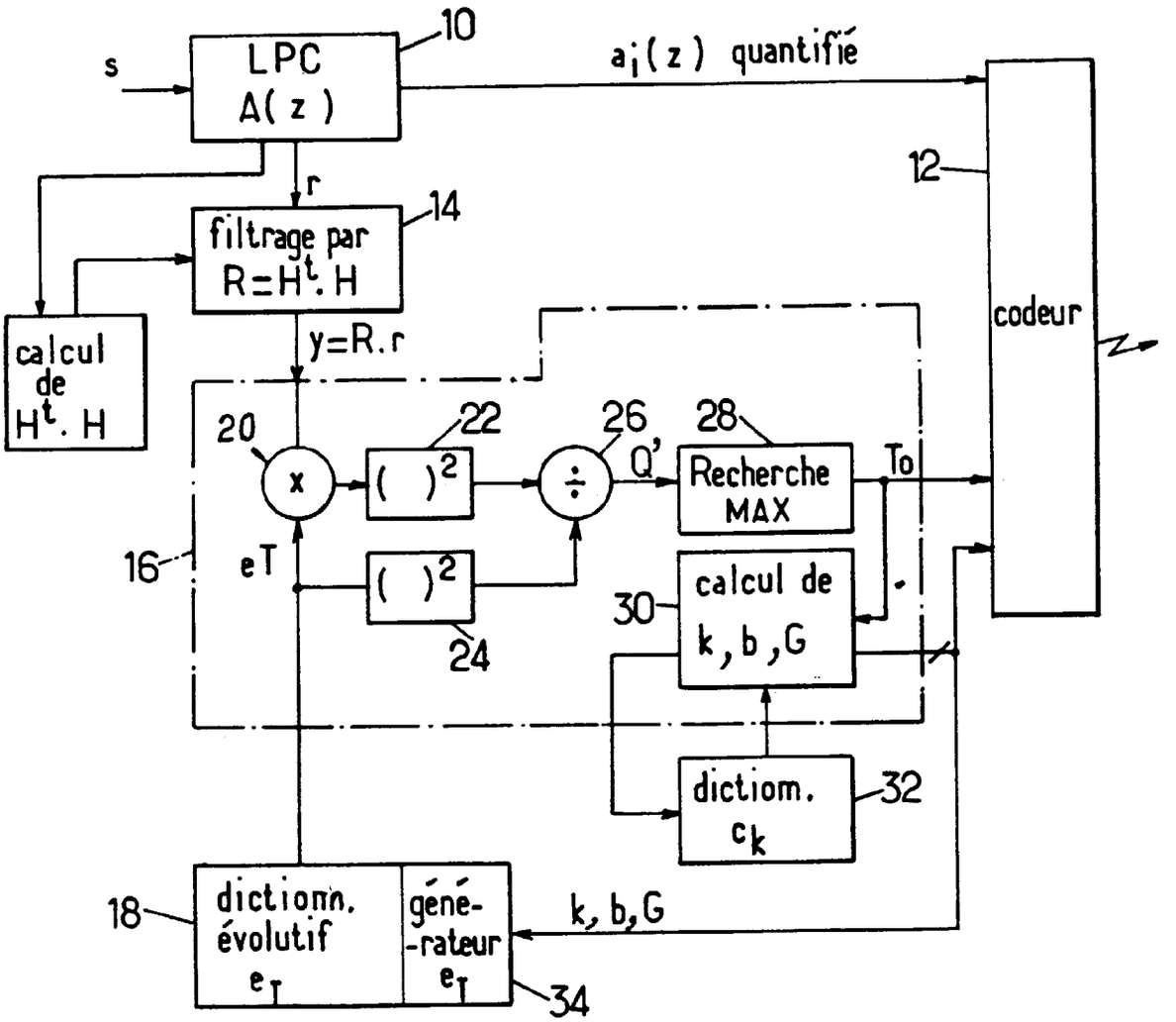

s
LPC
A(z)
10
a_i(z) quantifié
12
r
filtrage par
R=H^t. H
14
calcul
de
H^t. H
y=R. r
codeur
20
x
22
( )²
26
÷
Q'
28
Recherche
MAX
To
16
eT
( )²
24
30
calcul de
k, b, G
dictiom.
c_k
32
18
dictionn.
évolutif
e_T
géné-
-rateur
e_T
34
k, b, G

FIG.2.