



(12)

DEMANDE DE BREVET EUROPEEN

(43) Date de publication:
11.02.2004 Bulletin 2004/07

(51) Int Cl.7: G10L 19/14

(21) Numéro de dépôt: 03291749.4

(22) Date de dépôt: 15.07.2003

(84) Etats contractants désignés:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IT LI LU MC NL PT RO SE SI SK TR
Etats d'extension désignés:
AL LT LV MK

(30) Priorité: 17.07.2002 EP 02015919

(71) Demandeur: STMicroelectronics N.V.
1118 BH Schiphol Airport Amsterdam (NL)

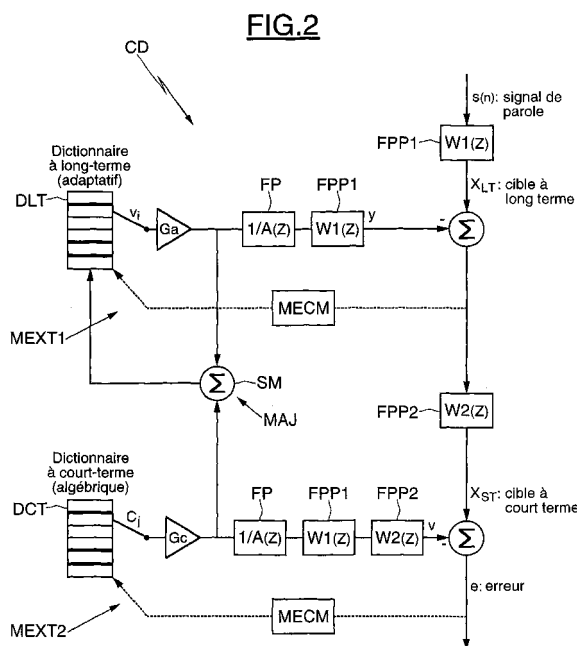
(72) Inventeurs:
• Ansorge, Michael
2068 Hauterive (CH)
• Biundo Lotito, Giuseppina
2000 Neuchatel (CH)
• Carnero, Benito
Santa Clara, CA 95050 (US)

(74) Mandataire: Dossmann, Gérard
Bureau D.A. Casalonga-Josse
Paul-Heyse-Strasse 33
80336 München (DE)

(54) Procédé et dispositif d'encodage de la parole à bande élargie capable de contrôler indépendamment les distorsions à court terme et à long terme

(57) On échantillonne la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons, et à chaque trame vocale on détermine des paramètres d'un modèle de prédiction linéaire à excitation par code, ces paramètres comportant un mot numérique d'excitation à long terme extrait d'un répertoire codé adaptatif, ainsi qu'un mot d'excitation à court terme extrait d'un répertoire codé fixe. L'extraction du mot d'excitation à long terme

s'effectue en utilisant un premier filtre de pondération perceptuelle comportant un premier filtre de pondération formantique (FPP1). On effectue l'extraction du mot d'excitation à court terme en utilisant le premier filtre de pondération perceptuelle (FPP1) cascadié à un deuxième filtre de pondération perceptuelle comportant un deuxième filtre de pondération formantique (FPP2). Le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.



Description

[0001] L'invention concerne l'encodage/décodage de la parole à bande élargie, notamment mais non limitativement pour la téléphonie mobile.

[0002] En bande élargie, la bande passante du signal de parole est comprise entre 50 et 7000 Hz.

[0003] Des séquences de parole successives échantillonnées à une fréquence d'échantillonnage prédéterminée, par exemple 16 kHz, sont traitées dans un dispositif de codage de type CELP, utilisant une prédiction linéaire à excitation par séquences codées (par exemple ACELP : « algebraic-code-excited linear-prediction »), bien connu de l'homme du métier, et décrit notamment dans la recommandation ITU-TG 729, version 3/96, intitulée « codage de la parole à 8 kbits/s par prédiction linéaire avec excitation par séquences codées à structure algébrique conjuguée ».

[0004] On va maintenant rappeler brièvement, en se référant à la figure 1, les principales caractéristiques et fonctionnalités d'un tel codeur, l'homme du métier pouvant se référer à toutes fins utiles, pour plus de détails, à la recommandation G 729 précitée.

[0005] Le codeur de prédiction CD, du type CELP, est fondé sur le modèle de codage prédictif linéaire à excitation par code. Le codeur opère sur des super-trames vocales équivalentes par exemple à 20 ms de signal et comportant chacune 320 échantillons. L'extraction des paramètres de prédiction linéaire, c'est-à-dire les coefficients du filtre de prédiction linéaire également appelé filtre de synthèse à court terme $1/A(z)$, est effectuée pour chaque super-trame de parole. Par contre, chaque super-trame est subdivisée en trames de 5 ms comprenant 80 échantillons. Toutes les trames, le signal vocal est analysé pour en extraire les paramètres du modèle de prédiction CELP (c'est-à-dire notamment, un mot numérique d'excitation à long terme v_l extrait d'un répertoire codé adaptatif DLT, également dénommé « dictionnaire à long terme adaptatif », un gain à long terme associé G_a , un mot d'excitation à court terme c_s , extrait d'un répertoire codé fixe DCT, également dénommé « dictionnaire à court terme », et un gain à court terme associé G_c).

[0006] Ces paramètres sont ensuite codés et transmis.

[0007] A la réception, ces paramètres servent, dans un décodeur, à récupérer les paramètres d'excitation et du filtre prédictif. On reconstitue alors la parole en filtrant ce flux d'excitation dans un filtre de synthèse à court terme.

[0008] Alors que le dictionnaire adaptatif DLT contient des mots numériques représentatifs de délais tonaux représentatifs d'excitations passées, le dictionnaire à court terme DCT est fondé sur une structure fixe, par exemple de type stochastique, ou de type algébrique utilisant un modèle de permutation entrelacée d'impulsions de Dirac. Dans le cas d'une structure algébrique, le répertoire codé contient des excitations innovatrices,

également appelées excitations algébriques ou à court terme, et chaque vecteur contient un certain nombre d'impulsions non nulles, par exemple quatre, dont chacune peut avoir l'amplitude +1 ou -1 avec des positions prédéterminées.

[0009] Les moyens de traitement du codeur CD comportent fonctionnellement des premiers moyens d'extraction MEXT1 destinés à extraire le mot d'excitation à long terme, et des deuxièmes moyens d'extraction MEXT2 destinés à extraire le mot d'excitation à court terme. Fonctionnellement, ces moyens sont réalisés par exemple de façon logicielle au sein d'un processeur.

[0010] Ces moyens d'extraction comportent un filtre prédictif FP ayant une fonction de transfert égale à $1/A(z)$, ainsi qu'un filtre de pondération perceptuel FPP ayant une fonction de transfert $W(z)$. Le filtre de pondération perceptuelle est appliqué au signal pour modéliser la perception de l'oreille.

[0011] Par ailleurs, les moyens d'extraction comportent des moyens MECM destinés à effectuer une minimisation d'une erreur carrée moyenne.

[0012] Le filtre de synthèse FP de la prédiction linéaire modélise l'enveloppe spectrale du signal. L'analyse prédictive linéaire est effectuée toutes les super-trames, de façon à déterminer les coefficients de filtrage prédictif linéaire. Ceux-ci sont convertis en paires de lignes spectrales (LSP : « Line Spectrum Pairs ») et numérisés par quantification vectorielle prédictive en deux étapes.

[0013] Chaque super-trame de parole de 20 ms est divisée en quatre trames de 5 ms chacune contenant 80 échantillons. Les paramètres LSP quantifiés sont transmis au décodeur une fois par super-trame alors que les paramètres à long terme et à court terme sont transmis à chaque trame.

[0014] Les coefficients du filtre de prédiction linéaire, quantifiés et non quantifiés, sont utilisés pour la trame la plus récente d'une super-trame, tandis que les trois autres trames de la même super-trame utilisent une interpolation de ces coefficients. Le délai tonal en boucle ouverte est estimé par exemple toutes les deux trames sur la base du signal vocal pondéré perceptuellement. Puis, les opérations suivantes sont répétées à chaque trame :

[0015] Le signal cible à long terme X_{LT} est calculé en filtrant le signal de parole échantillonné $s(n)$ par le filtre de pondération perceptuelle FPP.

[0016] On soustrait ensuite du signal vocal pondéré la réponse à entrée nulle du filtre de synthèse pondéré FP, FPP, de façon à obtenir un nouveau signal cible long terme.

[0017] La réponse impulsionnelle du filtre de synthèse pondéré est calculé.

[0018] Une analyse tonale en boucle fermée utilisant une minimisation de l'erreur carrée moyenne, est ensuite effectuée afin de déterminer le mot d'excitation à long terme v_l et le gain associé G_a , au moyen du signal cible et de la réponse impulsionnelle, par recherche autour de la valeur du délai tonal en boucle ouverte.

[0019] Le signal cible long terme est ensuite mis à jour par soustraction de la contribution filtrée y du répertoire codé adaptatif DLT et ce nouveau signal cible court terme X_{ST} est utilisé lors de l'exploration du répertoire codé fixe DCT afin de déterminer le mot d'excitation court terme C_j et le gain G_c associé. Là encore, cette recherche en boucle fermée s'effectue par une minimisation de l'erreur carrée moyenne.

[0020] Finalement, le dictionnaire à long terme adaptatif DLT ainsi que les mémoires des filtres FP et FPP, sont mis à jour au moyen des mots d'excitation long terme et court terme ainsi déterminés.

[0021] La qualité d'un algorithme CELP dépend fortement de la richesse du dictionnaire d'excitation à court terme DCT, par exemple d'un dictionnaire d'excitation algébrique. Si l'efficacité d'un tel algorithme est incontestable pour les signaux à bande passante étroite (300-3400 Hz), des problèmes surviennent pour des signaux à bande élargie.

[0022] L'invention a pour but de contrôler indépendamment les distorsions à court terme et à long terme.

[0023] L'invention propose donc un procédé d'encodage de la parole à bande élargie, dans lequel on échantillonne la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons, et à chaque trame vocale, on détermine des paramètres d'un modèle de prédiction linéaire à excitation par code, ces paramètres comportant un mot numérique d'excitation à long terme extrait d'un répertoire codé adaptatif, ainsi qu'un mot d'excitation à court terme extrait d'un répertoire codé fixe associé.

[0024] Selon une caractéristique générale de l'invention, on effectue l'extraction du mot d'excitation à long terme en utilisant un premier filtre de pondération perceptuelle comportant un premier filtre de pondération formantique, on effectue l'extraction du mot d'excitation à court terme en utilisant le premier filtre de pondération perceptuelle cascadié à un deuxième filtre de pondération formantique. Le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.

[0025] Ainsi, selon l'invention, l'utilisation de deux filtres de pondération formantique différents permet de contrôler indépendamment les distorsions à court terme et à long terme. Le filtre de pondération à court terme est cascadié au filtre de pondération à long terme. En outre, le fait de lier le dénominateur du filtre de pondération à long terme au numérateur du filtre de pondération à court terme permet de contrôler séparément ces deux filtres et permet en outre une nette simplification lorsque ces deux filtres sont cascadiés.

[0026] L'invention a également pour objet un dispositif d'encodage de la parole à bande élargie, comportant

ner la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons,

- des moyens de traitement aptes à chaque trame vocale, à déterminer des paramètres d'un modèle de prédiction linéaire à excitation par code, ces moyens de traitement comportant des premiers moyens d'extraction aptes à extraire un mot numérique d'excitation à long terme d'un répertoire codé adaptatif, et des deuxièmes moyens d'extraction aptes à extraire un mot d'excitation à court terme d'un répertoire codé fixe.

[0027] Selon une caractéristique générale de l'invention, les premiers moyens d'extraction comprennent un premier filtre de pondération perceptuelle comportant un premier filtre de pondération formantique, par le fait que les deuxièmes moyens d'extraction comprennent le premier filtre de pondération perceptuelle et un deuxième filtre de pondération perceptuelle comportant un deuxième filtre de pondération formantique, et le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.

[0028] L'invention a également pour objet un terminal d'un système de communication sans fil, par exemple un téléphone mobile cellulaire, incorporant un dispositif tel que défini ci-avant.

[0029] D'autres avantages et caractéristiques de l'invention apparaîtront à l'examen de la description détaillée de modes de réalisation et de mise en oeuvre, nullement limitatifs, et des dessins annexés, sur lesquels :

- la figure 1, déjà décrite, illustre schématiquement un dispositif d'encodage de la parole, selon l'art antérieur ;
- la figure 2 illustre schématiquement un mode de réalisation d'un dispositif d'encodage, selon l'invention ; et
- la figure 3 illustre schématiquement l'architecture interne d'un téléphone mobile cellulaire incorporant un dispositif de codage, selon l'invention.

[0030] Le filtre de pondération perceptuelle FPP exploite les propriétés de masquage de l'oreille humaine par rapport à l'enveloppe spectrale du signal de parole, dont la forme est fonction des résonances du conduit vocal. Ce filtre permet d'attribuer plus d'importance à l'erreur apparaissant dans les vallées spectrales par rapport aux pics formantiques.

[0031] Dans l'art antérieur illustré sur la figure 1, le même filtre de pondération perceptuelle FPP est utilisé pour la recherche à court terme et pour celle à long terme. La fonction de transfert $W(z)$ de ce filtre FPP est donnée par la formule (I) ci-dessous.

- des moyens d'échantillonnage aptes à échantillon-

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (I)$$

dans laquelle $1/A(z)$ est la fonction de transfert du filtre prédictif FP et γ_1 et γ_2 sont les coefficients de pondération perceptuelle, les deux coefficients étant positifs ou nuls et inférieurs ou égaux à 1 avec le coefficient γ_2 inférieur ou égal au coefficient γ_1 .

[0032] D'une façon générale, le filtre de pondération perceptuelle est constitué d'un filtre de pondération formantique et d'un filtre de pondération de la pente de l'enveloppe spectrale du signal (tilt).

[0033] Dans le cas présent, on supposera que le filtre de pondération perceptuelle est uniquement formé du filtre de pondération formantique dont la fonction de transfert est donnée par la formule (I) ci-dessus.

[0034] Or, la nature spectrale de la contribution à long terme est différente de celle de la contribution à court terme. Par conséquent, il est avantageux d'utiliser deux filtres de pondération formantique différents, permettant de contrôler indépendamment les distorsions à court terme et à long terme.

[0035] Un tel mode de réalisation selon l'invention est illustré sur la figure 2, dans laquelle, par rapport à la figure 1, le filtre unique FPP a été remplacé par un premier filtre de pondération formantique FPP1 pour la recherche à long terme, cascadié avec un deuxième filtre de pondération formantique FPP2 pour la recherche à court terme.

[0036] Puisque le filtre de pondération à court terme FPP2 est cascadié au filtre de pondération à long terme, les filtres apparaissant dans la boucle de recherche long terme doivent aussi apparaître dans la boucle de recherche court terme.

[0037] La fonction de transfert $W_1(z)$ du filtre de pondération formantique FPP1 est donnée par la formule (II) ci-dessous.

$$W_1(z) = \frac{A(z/\gamma_{11})}{A(z/\gamma_{12})} \quad (II)$$

tandis que la fonction de transfert $W_2(z)$ du filtre de pondération formantique FPP2 est donnée par la formule (III) ci-dessous.

$$W_2(z) = \frac{A(z/\gamma_{21})}{A(z/\gamma_{22})} \quad (III)$$

[0038] Par ailleurs, le coefficient γ_{12} est égal au coefficient γ_{21} . Ceci permet une nette simplification lorsqu'on cascade ces deux filtres. Ainsi, le filtre équivalent à la cascade de ces deux filtres a une fonction de transfert donnée par la formule (IV) ci-dessous.

$$\frac{A(z/\gamma_{11})}{A(z/\gamma_{22})} \quad (IV)$$

[0039] Par ailleurs, si l'on utilise la valeur 1 pour le coefficient γ_{11} , alors le filtre de synthèse FP (ayant la fonction de transfert $1/A(z)$) suivi du filtre de pondération à long terme FPP1 et du filtre de pondération FPP2 équivaut alors au filtre dont la fonction de transfert est donnée par la formule (V) ci-dessous.

$$\frac{1}{A(z/\gamma_{22})} \quad (V)$$

[0040] Ce qui réduit encore considérablement la complexité de l'algorithme d'extraction des excitations.

[0041] A titre indicatif, on peut par exemple utiliser pour les coefficients γ_{11} , $\gamma_{21} = \gamma_{12}$ et γ_{22} , les valeurs respectives 1 ; 0,1 et 0,9.

[0042] L'invention s'applique avantageusement à la téléphonie mobile, et en particulier à tous terminaux distants appartenant à un système de communication sans fil.

[0043] Un tel terminal, par exemple un téléphone mobile TP, tel que celui illustré sur la figure 3, comporte de façon classique une antenne reliée par l'intermédiaire d'un duplexeur DUP à une chaîne de réception CHR et à une chaîne de transmission CHT. Un processeur en bande de base BB est relié respectivement à la chaîne de réception CHR et à la chaîne de transmission CHT par l'intermédiaire de convertisseurs analogiques numériques CAN et numériques analogiques CNA.

[0044] Classiquement, le processeur BB effectue des traitements en bande de base, et notamment un décodage de canal DCN, suivi d'un décodage de source DCS.

[0045] Pour l'émission, le processeur effectue un codage de source CCS suivi d'un codage de canal CCN.

[0046] Lorsque le téléphone mobile incorpore un codeur selon l'invention, celui-ci est incorporé au sein des moyens de codage de source CCS, tandis que le décodeur est incorporé au sein des moyens de décodage de source DCS.

Revendications

1. Procédé d'encodage de la parole à bande élargie, dans lequel on échantillonne la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons, et à chaque trame vocale on détermine des paramètres d'un modèle de prédiction linéaire à excitation par code, ces paramètres comportant un mot numérique d'excitation à long terme extrait d'un répertoire codé adaptatif, ainsi qu'un mot d'excitation à court terme extrait d'un répertoire codé fixe,

caractérisé par le fait qu'on effectue l'extraction du mot d'excitation à long terme en utilisant un premier filtre de pondération perceptuelle comportant un premier filtre de pondération formantique (FPP1), **par le fait qu'on** effectue l'extraction du mot d'excitation à court terme en utilisant le premier filtre de pondération perceptuelle (FPP1) cascadié à un deuxième filtre de pondération perceptuelle comportant un deuxième filtre de pondération formantique (FPP2), et **par le fait que** le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.

2. Dispositif d'encodage de la parole à bande élargie, comportant des moyens d'échantillonnage aptes à échantillonner la parole de façon à obtenir des trames vocales successives comportant chacune un nombre prédéterminé d'échantillons, et des moyens de traitement aptes à chaque trame vocale, à déterminer des paramètres d'un modèle de prédiction linéaire à excitation par code, ces moyens de traitement comportant des premiers moyens d'extraction aptes à extraire un mot numérique d'excitation à long terme d'un répertoire codé adaptatif, ainsi que des deuxièmes moyens d'extraction aptes à extraire un mot d'excitation à court terme d'un répertoire codé fixe, **caractérisé par le fait que** les premiers moyens d'extraction (MEXT1) comprennent un premier filtre de pondération perceptuelle comportant un premier filtre de pondération formantique (FPP1), **par le fait que** les deuxièmes moyens d'extraction (MEXT2) comprennent le premier filtre de pondération perceptuelle cascadié à un deuxième filtre de pondération perceptuelle comportant un deuxième filtre de pondération formantique (FPP2), et **par le fait que** le dénominateur de la fonction de transfert du premier filtre de pondération formantique est égal au numérateur du deuxième filtre de pondération formantique.
3. Terminal d'un système de communication sans fil, **caractérisé par le fait qu'il** incorpore un dispositif selon la revendication 2.
4. Terminal selon la revendication 3, **caractérisé par le fait qu'il** forme un téléphone mobile cellulaire.

FIG.1
Art antérieur

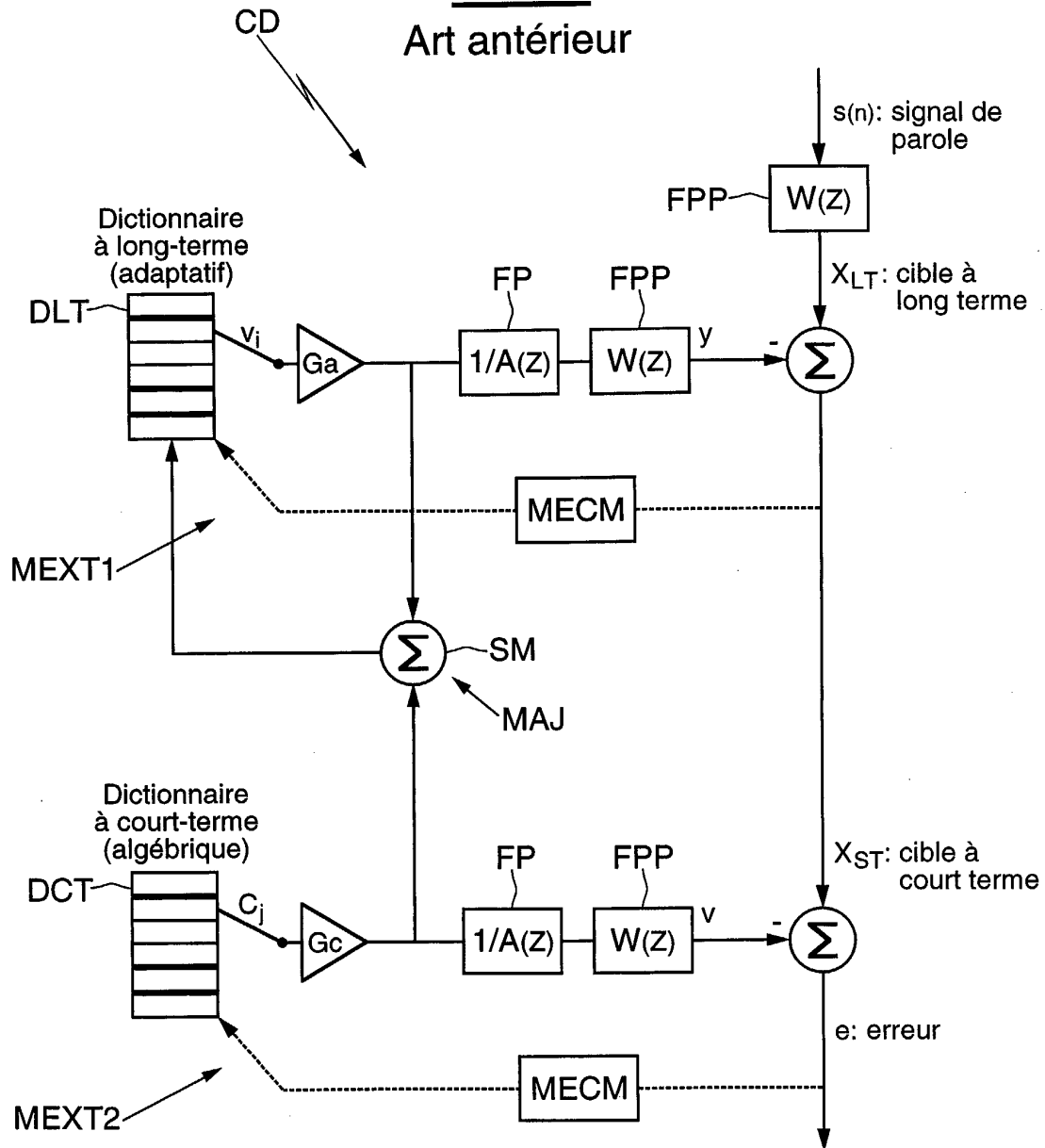


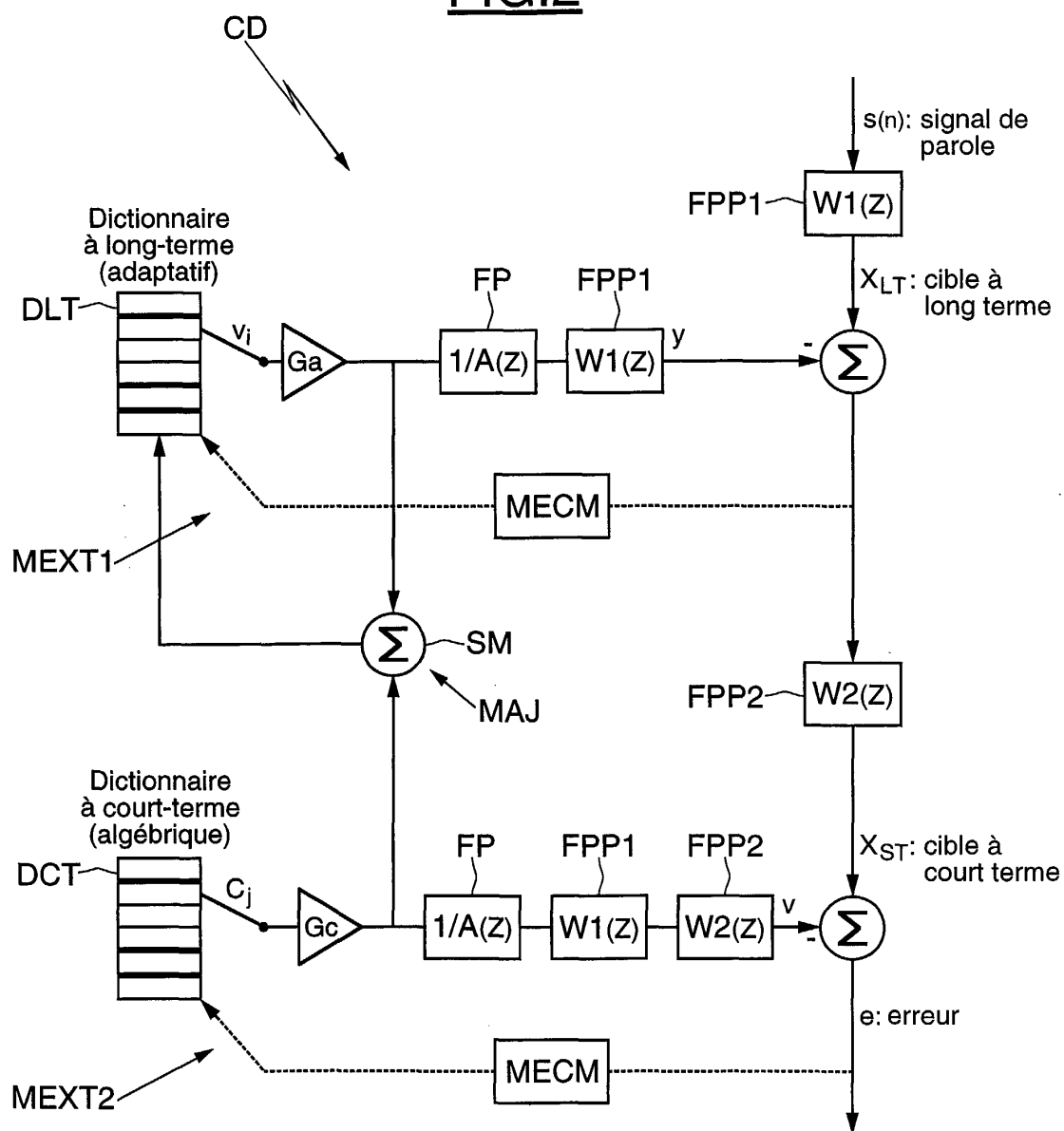
FIG.2

FIG.3

