

12)

DEMANDE DE BREVET D'INVENTION

A1

22) Date de dépôt : 13.03.98.

30) Priorité : 14.03.97 FI 00971108.

43) Date de mise à la disposition du public de la demande : 09.10.98 Bulletin 98/41.

56) Liste des documents cités dans le rapport de recherche préliminaire : *Ce dernier n'a pas été établi à la date de publication de la demande.*

60) Références à d'autres documents nationaux apparentés :

71) Demandeur(s) : NOKIA MOBILE PHONES LIMITED
— FI.

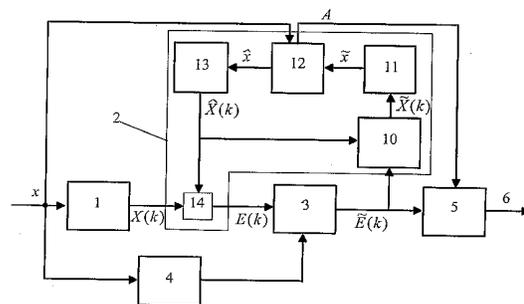
72) Inventeur(s) : YIN LIN.

73) Titulaire(s) :

74) Mandataire(s) : NOVAPAT.

54) METHODE DE CODAGE AUDIO ET DISPOSITIF.

57) Une méthode de codage d'un signal audio comporte la réception d'un signal audio x devant être codé, et la transformation du signal reçu du domaine de temps au domaine de fréquence. Un signal audio quantifié x est généré à partir du signal audio transformé x ; ainsi qu'un ensemble de coefficients de prévision à long terme A , qui peut être utilisé pour prévoir une trame de temps courante du signal audio reçu, directement à partir d'une ou de plusieurs trames de temps précédentes du signal audio quantifié x . Un signal audio prévu \hat{x} est généré en utilisant les coefficients de prévision A . Le signal audio prévu \hat{x} est ensuite transformé du domaine de temps vers le domaine des fréquences, et le signal du domaine de fréquence résultant est comparé avec celui du signal audio reçu x , pour générer un signal d'erreur $E(k)$ qui est ensuite quantifié pour générer un ensemble de signaux d'erreur quantifiés $\tilde{E}(k)$ qui sont combinés avec les coefficients de prévision A , pour générer un signal audio codé.



METHODE DE CODAGE AUDIO ET DISPOSITIF

La présente invention concerne une méthode et un dispositif de codage audio, ainsi qu'une méthode et un
5 dispositif de décodage audio.

Il est bien connu que la transmission de données sous forme digitale conduit à une augmentation des rapports signal/bruit et une augmentation des capacités d'information dans le canal de transmission. Il y a
10 cependant une tendance constante à améliorer encore la capacité de canal, en compressant les signaux numériques à des taux toujours plus élevés. En ce qui concerne les signaux audio, deux principes de compression de base sont habituellement appliqués. Le premier de ceux-ci consiste
15 à éliminer les redondances statistiques et déterministes dans le signal source, tandis que le second consiste à supprimer ou éliminer du signal source des éléments qui sont redondants dans la mesure où la perception humaine est concernée. Récemment, ce dernier principe est devenu
20 prédominant dans des applications audio de haute qualité, et implique habituellement la séparation d'un signal audio en ses composantes de fréquence (appelées souvent "sous-bandes"), chacune desquelles est analysée et quantifiée avec une précision de quantification
25 déterminée, pour retirer les données non pertinentes (à l'auditeur). Le standard de codage audio ISO (International Standards Organisation) MPEG (Moving Pictures Expert Group) ainsi que d'autres standards de codage audio utilisent et définissent ce principe.
30 Cependant, le standard MPEG (et d'autres standards) utilisent aussi une technique connue comme "prévision adaptative" pour réaliser une nouvelle réduction du débit de données.

Le fonctionnement d'un codeur selon le nouveau
35 standard MPEG-2 AAC est décrit en détail dans le document (provisoire) standard international ISO/IECC DIS 13818-7.

Ce nouveau standard MPEG-2 utilise une prévision linéaire de référence aux données antérieures ou "de retour" ("backward linear prediction"), avec 672 des 1024 composantes de fréquence. Il est envisagé que le nouveau standard MPEG-4 comporte des exigences similaires. Cependant, un tel nombre important de composantes de fréquence a pour conséquence une charge de calcul importante, du fait de la complexité de l'algorithme de prévision, et nécessite aussi la disponibilité d'une grande quantité de mémoire pour stocker les coefficients calculés et les coefficients intermédiaires. Il est bien connu que lorsque des dispositifs de prévision adaptative de retour de ce type sont utilisés dans le domaine de fréquence, il est difficile de réduire encore les charges de calcul et les exigences de mémoire. Cela est dû au fait que le nombre de dispositifs de prévision est si important dans le domaine des fréquences, que même un algorithme adaptatif très simple aura pour conséquence une grande complexité de calcul et des exigences de mémoire importantes. Bien qu'il soit connu d'éviter ce problème en utilisant des dispositifs de prévision adaptative avant qui sont mises à jour dans le codeur et transmises au décodeur, l'utilisation de dispositifs de prévision adaptative "avant" dans le domaine de fréquence a inévitablement pour conséquence une grande quantité d'informations "de côté" du fait du nombre de dispositifs de prévision importants.

Un objet de la présente invention est d'éviter ou au moins de limiter, les inconvénients des méthodes de prévision connues.

Cet objet ainsi que d'autres, sont réalisés par le codage d'un signal audio en utilisant des signaux d'erreur pour éliminer la redondance dans chaque sous-bande de la pluralité de sous-bandes de fréquence du signal audio, et, en addition, par la génération de

coefficients de prévision à long terme dans le domaine de temps, qui permet à une trame courante du signal audio, d'être prévue à partir d'une ou de plusieurs trames précédentes.

5

Selon un premier aspect de la présente invention, il est proposé une méthode de codage d'un signal audio, qui comporte les étapes de :

réception d'un signal audio x devant être codé;
10 génération d'un signal audio quantifié \tilde{x} à partir du signal audio reçu x ;

génération d'un ensemble de coefficients de prévision à long terme A qui peuvent être utilisés pour prévoir une trame de temps courante du signal audio reçu x , directement à partir d'au moins une trame de temps
15 précédente du signal audio quantifié \tilde{x} ;

utilisation des coefficients de prévision A pour générer un signal audio prévu \hat{x} ;

comparaison du signal audio reçu x avec le signal
20 audio prévu \hat{x} et génération d'un signal d'erreur $E(k)$ pour chaque sous-bande de la pluralité de sous-bandes de fréquence;

quantification des signaux d'erreur $\tilde{E}(k)$ pour générer un ensemble de signaux d'erreur quantifié $E(k)$;
25 et

combinaison du signal d'erreur quantifié $\tilde{E}(k)$ et des coefficients de prévision A , pour générer un signal audio codé.

La présente invention permet la composition d'un
30 signal audio en utilisant un dispositif de prévision adaptative avant dans le domaine de temps. Pour chaque trame de temps d'un signal reçu, il est seulement nécessaire de générer et de transmettre un seul ensemble de coefficients de prévision adaptative avant pour la
35 transmission au décodeur. Cela contraste avec les techniques de prévision adaptatives avant connues, qui

nécessitent la génération d'un ensemble de coefficients de prévision pour chaque sous-bande de fréquence de chaque trame de temps. En comparaison aux gains de prévision obtenus par la présente invention, l'information de côté du dispositif de prévision à long terme est négligeable.

Certains modes de réalisation de la présente invention permettent une réduction dans la complexité de calcul et dans les exigences de mémoire. En particulier, en comparaison à l'utilisation de prévision adaptative de retour, il n'y a aucune exigence de recalcul des coefficients de prévision dans le décodeur. Certains modes de réalisation de l'invention sont aussi capables de répondre plus rapidement aux changements des signaux, que les dispositifs de prévision adaptative de retour conventionnel.

Dans un premier mode de réalisation de l'invention, le signal audio reçu x est transformé dans des trames x_m du domaine de temps au domaine de fréquence, pour fournir un ensemble de signaux de sous-bande de fréquence $X(k)$. Le signal audio prévu \hat{x} est transformé de façon similaire, du domaine de temps au domaine de fréquence, pour générer un ensemble de signaux de sous-bande de fréquence prévue $\hat{X}(k)$, et la comparaison entre le signal audio reçu \hat{x} et le signal audio prévu x , est effectuée dans le domaine de fréquence, en comparant les signaux de sous-bande respectifs les uns aux autres, pour générer les signaux d'erreur de sous-bande de fréquence $E(k)$. Le signal audio quantifié \tilde{x} est généré en additionnant le signal prévu et le signal d'erreur quantifié, soit dans le domaine de temps, soit dans le domaine de fréquence.

Dans un autre mode de réalisation de l'invention, la comparaison entre le signal audio reçu x et le signal audio prévu \hat{x} , est effectuée dans le domaine de temps pour générer un signal d'erreur e aussi dans le domaine

de temps. Ce signal d'erreur e est ensuite converti du domaine de temps au domaine de fréquence, pour générer ladite pluralité de signaux d'erreur de sous-bande de fréquence $E(k)$.

5 De préférence, la quantification des signaux d'erreur est effectuée selon un modèle psycho-acoustique.

Selon un second aspect de la présente invention, il est proposé une méthode de décodage d'un signal audio
10 codé, comportant les étapes de :

réception d'un signal audio codé comportant un signal d'erreur quantifié $\tilde{E}(k)$ pour chaque sous-bande de la pluralité de sous-bandes de fréquence du signal audio et, pour chaque trame de temps du signal audio, un
15 ensemble de coefficients de prévision A qui peuvent être utilisés pour prévoir une trame de temps courante x_m du signal audio reçu, directement à partir d'au moins une trame de temps précédente d'un signal audio quantifié reconstitué \tilde{x} ;

20 génération dudit signal audio quantifié reconstitué \tilde{x} , à partir des signaux d'erreur quantifié $\tilde{E}(k)$;

utilisation des coefficients de prévision A et du signal audio quantifié \tilde{x} , pour générer un signal audio
25 prévu \hat{x} ;

transformation du signal audio prévu \hat{x} , du domaine de temps au domaine de fréquence, pour générer un ensemble de signaux de sous-bande de fréquence prévu $\hat{X}(k)$ pour la combinaison avec les signaux d'erreur
30 quantifiés $\tilde{E}(k)$ afin de générer un ensemble de signaux de sous-bande de fréquence reconstitué $\tilde{X}(k)$; et

application d'une transformation de domaine de fréquence en domaine de temps, sur les signaux de sous-bande de fréquence reconstitués $\tilde{X}(k)$ afin de générer le
35 signal audio quantifié reconstitué \tilde{x} .

Des modes de réalisation selon ce second aspect de l'invention sont particulièrement bien adaptés lorsque seulement un sous-ensemble de tous les signaux d'erreur quantifiés possibles $\tilde{E}(k)$ sont reçus, quelques données
5 de sous-bande étant transmises directement par la transmission des signaux de sous-bande audio $X(k)$. Les signaux $\tilde{X}(k)$ et $X(k)$ sont combinés de façon appropriée avant la mise en œuvre de la transformation de fréquence en temps.

10

Selon un troisième aspect de la présente invention, il est proposé un dispositif pour le codage d'un signal audio, le dispositif comportant :

15 une entrée pour la réception d'un signal audio x devant être codé;

un moyen de quantification couplé à ladite entrée pour générer, à partir du signal audio reçu x , un signal audio quantifié \tilde{x} ;

20 un moyen de prévision couplé audit moyen de quantification, pour générer un ensemble de coefficients de prévision à long terme A , pour prévoir une trame de temps courante x_m du signal audio reçu x , directement à partir d'au moins une trame de temps précédente du signal audio quantifié \tilde{x} ;

25 un moyen de génération pour générer un signal audio prévu \hat{x} en utilisant les coefficients de prévision A , et pour comparer le signal audio reçu x avec le signal audio prévu \hat{x} , pour générer un signal d'erreur $E(k)$ pour chaque sous-bande de la pluralité de sous-bandes de
30 fréquence;

un moyen de quantification pour quantifier les signaux d'erreur $E(k)$, afin de générer un ensemble de signaux d'erreur quantifié $\tilde{E}(k)$; et

35 un moyen de combinaison pour combiner les signaux d'erreur quantifiée $\tilde{E}(k)$ avec les coefficients de prévision A , afin de générer un signal audio codé.

Dans un premier mode de réalisation, ledit moyen de génération comporte un premier moyen de transformation pour transformer le signal audio reçu x du domaine de temps au domaine de fréquence et un second moyen de transformation pour transformer le signal audio prévu \hat{x} du domaine de temps au domaine de fréquence, et un moyen de comparaison prévu pour y comparer les signaux du domaine de fréquence résultant dans le domaine de fréquence.

Dans un autre mode de réalisation de l'invention, le moyen de génération est prévu pour comparer le signal audio reçu x et le signal audio prévu \hat{x} dans le domaine de temps.

Selon un quatrième aspect de la présente invention, il est proposé un dispositif pour le décodage d'un signal audio codé x , dans lequel le signal audio codé comporte un signal d'erreur quantifié $\tilde{E}(k)$ pour chaque sous-bande de la pluralité de sous-bandes de fréquence du signal audio, et un ensemble de coefficients de prévision A pour chaque trame de temps du signal audio, et dans lequel les coefficients de prévision A peuvent être utilisés pour prévoir une trame de temps courante x_m du signal audio reçu, directement à partir d'au moins une trame de temps précédente d'un signal audio quantifié reconstitué \tilde{x} , le dispositif comportant :

une entrée pour recevoir le signal audio codé;
un moyen de génération pour générer ledit signal audio quantifié reconstitué \tilde{x} à partir des signaux d'erreur quantifié $\tilde{E}(k)$; et

un moyen de traitement de signal, pour générer un signal audio prévu \hat{x} à partir des coefficients de prévision A et dudit signal audio reconstitué \tilde{x} ;

dans lequel ledit moyen de génération comporte un premier moyen de transformation pour transformer le signal audio prévu \hat{x} du domaine du temps au domaine de

fréquence, pour générer un ensemble de signaux de sous-bande de fréquence prévus $\hat{X}(k)$, un moyen de combinaison pour combiner ledit ensemble de signaux de sous-bande de fréquence prévus $\hat{x}(k)$ avec les signaux d'erreur quantifiés $\tilde{E}(k)$, pour générer un ensemble de signaux de sous-bande de fréquence reconstitués $\tilde{X}(k)$, et un second moyen de transformation pour appliquer une transformation du domaine de fréquence au domaine du temps, aux signaux de sous-bande de fréquence reconstitués $\tilde{X}(k)$ pour générer le signal audio quantifié reconstitué \tilde{x} .

Pour une meilleure compréhension de la présente invention, et afin de montrer comment elle peut être mise en œuvre, référence sera maintenant faite aux figures accompagnantes au moyen d'exemples, dans lesquelles :

La Figure 1 représente de façon schématique, un codeur pour le codage d'un signal audio reçu;

La Figure 2 représente de façon schématique, un décodeur pour le décodage d'un signal audio codé par le codeur de la Figure 1;

La Figure 3 représente en détail le codeur de la Figure 1, comportant un outil de prévision du codeur;

La Figure 4 représente en détail le décodeur de la Figure 2, comportant un outil de prévision du décodeur; et

La Figure 5 représente en détail une modification du codeur de la Figure 1 et qui utilise un outil de prévision alternatif.

Sur la Figure 1 est représenté un bloc diagramme d'un codeur qui met en œuvre la fonction de codage définie en termes généraux dans le standard MPEG-2 AAC. L'entrée du codeur est un signal monophasé échantillonné x , dont les points d'échantillon sont groupés en trames de temps ou blocs de $2N$ points, c'est-à-dire :

$$\mathbf{x}_m = (x_m(0), x_m(1), \dots, x_m(2N - 1))^T \quad (1)$$

dans laquelle m est l'indice de bloc et T représente la transposition. Le groupement des points d'échantillon est effectué par un outil d'assemblage de filtres 1 qui met aussi en œuvre une transformée de cosinus discrète
 5 modifiée (MDCT) sur chaque trame individuelle du signal audio, pour générer un ensemble de coefficients de sous-bande de fréquence

$$\mathbf{X}_m = (X_m(0), X_m(1), \dots, X_m(N-1))^T \quad (2)$$

Les sous-bandes sont définies dans le standard MPEG.

10 La MDCT avant est définie par :

$$X_m(k) = \sum_{i=0}^{2N-1} f(i) x_m(i) \cos\left(\frac{\pi}{4N}(2i+1+N)(2k+1)\right), \quad (3)$$

15 $k = 0, \dots, N-1$

où $f(i)$ est la fenêtre d'analyse-synthèse, qui est une fenêtre symétrique telle que son effet ajouté-recouvert produit un gain unité dans le signal.

Les signaux de sous-bande de fréquence $X(k)$ sont
 20 à leur tour appliqués à l'outil de prévision 2 (décrit plus en détail ci-après) qui cherche à éliminer la redondance à long terme dans chacun des signaux de sous-bande. Le résultat est un ensemble de signaux d'erreur de sous-bande de fréquence

$$25 \quad \mathbf{E}_m(k) = (E_m(0), E_m(1), \dots, E_m(N-1))^T \quad (4)$$

qui sont représentatif des changements à long terme dans les sous-bandes respectives, et un ensemble de coefficient de prévision adaptative avant A pour chaque trame.

30 Les signaux d'erreur de sous-bande $E(k)$ sont appliqués à un quantificateur 3 qui quantifie chaque signal avec un nombre de bits déterminé par un modèle psycho-acoustique. Ce modèle est appliqué par un contrôleur 4. Tel que décrit, le modèle psycho-acoustique
 35 est utilisé pour modéliser le comportement de masquage du système auditif humain. Les signaux d'erreur quantifiés

$\tilde{E}(k)$ et les coefficients de prévision A , sont ensuite combinés dans un multiplexeur de flux de bits 5 pour être transmis via un canal de transmission 6.

La Figure 2 représente la disposition générale d'un décodeur pour le décodage d'un signal audio codé avec le codeur de la Figure 1. Un démultiplexeur de flux de bits 7 sépare d'abord les coefficients de prévision A des signaux d'erreur quantifiés $\tilde{E}(k)$, et sépare les signaux d'erreur dans les signaux de sous-bande séparés. Les coefficients de prévision A et les signaux de sous-bande d'erreur quantifié $\tilde{E}(k)$ sont fournis à un outil de prévision 8 qui inverse le procédé de prévision effectué dans le codeur, c'est-à-dire l'outil de prévision réinsère la redondance extraite dans le codeur, pour générer des signaux de sous-bande quantifiés reconstitués $\tilde{X}(k)$. Un outil d'assemblage de filtres 9 récupère ensuite le signal du domaine de temps \tilde{x} , par une transformation inverse sur la version reçue $\tilde{X}(k)$, décrite par :

$$\begin{aligned} \tilde{X}_m(i) &= \tilde{u}_{m-1}(i + N) + \tilde{u}_m(i) \\ i &= 0, \dots, N - 1 \end{aligned} \quad (5)$$

où $\tilde{u}_k(i)$, $i = 0, \dots, 2N - 1$, sont la transformation inverse de \tilde{X}

$$\begin{aligned} \tilde{u}_m(i) &= f(i) \sum_{k=0}^{N-1} X_m(k) \cos\left(\frac{\pi}{4N} (2i+1+N)(2k+1)\right), \\ i &= 0, \dots, N - 1 \end{aligned}$$

et qui approxime le signal audio original x .

La Figure 3 représente avec plus de détails, la méthode de prévision du codeur de la Figure 1. En utilisant les signaux d'erreur de sous-bande de fréquence quantifiés $E(k)$, un ensemble de signaux de sous-bande de fréquence quantifiés $\tilde{x}(k)$ sont générés par une unité de traitement de signal 10. Les signaux $\tilde{X}(k)$ sont appliqués à leur tour à un assemblage de filtres 11 qui applique

aux signaux une transformation cosinus discrète modifiée inverse (IMDCT), pour générer un signal du domaine de temps quantifié \tilde{x} . Le signal \tilde{x} est ensuite appliqué à un outil de prévision à long terme 12 qui reçoit aussi le signal d'entrée audio x . L'outil de prévision 12 utilise un dispositif de prévision à long terme (LT) pour retirer la redondance dans le signal audio présent dans une trame courante $m+1$, sur la base des données précédemment quantifiées. La fonction de transfert P de ce dispositif de prévision est :

$$P(Z) = \sum_{k=-m_1}^{m_2} b_k Z^{-(\alpha+k)} \quad (5)$$

où α représente un retard dans la gamme de 1 à 1024 échantillons, et b_k sont des coefficients de prévision. Pour $m_1 = m_2 = 0$, le dispositif de prévision est d'ordre 1, tandis que pour $m_1 = m_2 = 1$, le dispositif de prévision est d'ordre trois.

Les paramètres α et b_k sont déterminés en minimisant l'erreur carrée moyenne après la prévision LT sur une période de $2N$ échantillons. Pour un dispositif de prévision d'ordre un, le résidu de prévision LT $r(i)$ est donné par :

$$r(i) = x(i) - b\tilde{x}(i - 2N + 1 - \alpha) \quad (6)$$

où x est le signal audio du domaine de temps, et \tilde{x} est le signal quantifié du domaine de temps. Le résidu carré moyen R est donné par :

$$R = \sum_{i=0}^{2N-1} r^2(i) = \sum_{i=0}^{2N-1} (x(i) - b\tilde{x}(i - 2N + 1 - \alpha))^2 \quad (7)$$

Considérant que $\partial R / \partial b = 0$,

$$b = \frac{\sum_{i=0}^{2N-1} x(i)\bar{x}(i-2N+1-\alpha)}{\sum_{i=0}^{2N-1} (\bar{x}(i-2N-\alpha))^2} \quad (8)$$

5

et en substituant b dans l'équation (7) :

$$R = \sum_{i=0}^{2N-1} x^2(i) - \frac{\left(\sum_{i=0}^{2N-1} x(i)\bar{x}(i-2N+1-\alpha)\right)^2}{\sum_{i=0}^{2N-1} (\bar{x}(i-2N+1-\alpha))^2} \quad (9)$$

10

La minimisation de R correspond à la minimisation du second terme du côté droit de l'équation (9). Ce terme est calculé pour toutes les valeurs possibles de α sur sa gamme spécifiée, et la valeur de α qui maximise ce terme, est choisie. L'énergie dans le dénominateur de l'équation (9), identifiée comme Ω , peut être facilement mise à jour à partir du retard ($\alpha - 1$) à α plutôt que de le recalculer à nouveau en utilisant :

20

$$\Omega_{\alpha} = \Omega_{\alpha+1} + \bar{x}^2(-\alpha) - \bar{x}^2(-\alpha + N) \quad (10)$$

Si un dispositif de prévision LT d'ordre un est utilisé, alors l'équation (8) est utilisée pour calculer le coefficient de prévision b_j . Pour un dispositif de prévision d'ordre, le retard de prévision LT α est d'abord déterminé en maximisant le second terme de l'équation (9), puis un ensemble de $j \times j$ équations est résolu pour calculer les coefficients de prévision j .

25

Les paramètres de prévision LT A sont le retard α et les coefficients de prévision b_j . Le retard est quantifié avec de 9 à 11 bits, dépendant de la gamme utilisée. Plus communément 10 bits sont utilisés avec 1024 valeurs possibles dans la gamme de 1 à 1024. Pour réduire le nombre de bits, les retards de prévision LT peuvent être codés en delta dans des trames égales avec 5 bits. Des expériences montrent qu'il est suffisant de

35

quantifier le gain avec de 3 à 6 bits. Du fait de la distribution non-uniforme du gain, une quantification non-uniforme doit être utilisée.

Dans la méthode décrite au-dessus, la stabilité du filtre de synthèse LT $1/P(z)$ n'est pas toujours garantie. Pour un dispositif de prévision d'ordre un, la condition de stabilité est $|b| \leq 1$. Par conséquent, la stabilisation peut être facilement mise en œuvre en ajustant $|b| = 1$, tant que $|b| > 1$. Pour un dispositif d'ordre trois, une autre procédure de stabilisation peut être utilisée telle que celle décrite dans "Stability and performance analysis of pitch filters in speech coders", de R.P. Ramachandran et P. Kabal, IEEE Trans. ASSP, volume 35, n° 7, pages 937 à 946, Juillet 1987. Cependant, l'instabilité du filtre de synthèse LT n'est pas si nuisible à la qualité du signal reconstitué. Le filtre instable persistera pendant quelques trames (augmentant l'énergie), mais par la suite, des périodes de stabilité sont rencontrées et la sortie ne continue donc pas à augmenter avec le temps.

Après que le coefficient du dispositif de prévision LT ait été déterminé, le signal prévu pour la $(m+1)$ ième trame peut être déterminé :

$$\hat{x}(i) = \sum_{j=-m_1}^{m_2} b_j \tilde{x}(i - 2N + 1 - j - \alpha), \quad (11)$$

$i = mN + 1, mN + 2, \dots, (m + 1)N$

Le signal de domaine de temps prévu x est ensuite appliqué à un assemblage de filtres 13 qui applique une MDCT au signal, pour générer des coefficients spectraux prévus $\hat{X}_{m+1}(k)$ pour la $(m+1)$ ième trame. Les coefficients spectraux prévus $\hat{X}(k)$ sont ensuite soustraits des coefficients spectraux $X(k)$ au niveau du soustracteur 14.

Afin de garantir que la prévision est utilisée seulement si elle résulte dans un gain de codage, un

contrôle de dispositif de prévision approprié est nécessaire et une petite quantité d'information de contrôle du dispositif de prévision doit être transmis au décodeur. Cette fonction est effectuée dans le soustracteur 14. Le schéma de commande du dispositif de prévision est le même que le schéma de commande du dispositif de prévision adaptative de retour qui a été utilisé dans le standard MPEG-2 Advanced Audio Coding (AAC). L'information de commande du dispositif de prévision pour chaque trame, qui est transmise comme information de côté, est déterminée en deux étapes. Premièrement, pour chaque bande de facteur de proportionnalité, il est déterminé si oui ou non la prévision mène à un gain de codage, et si oui, le bit utilisé par le dispositif de prévision pour ce facteur de proportionnalité est ajusté à zéro. Après que cela a été fait pour toutes les bandes de facteur de proportionnalité, il est déterminé si le gain de codage global par prévision dans cette trame, compense au moins la nécessité du bit additionnel pour l'information de côté du dispositif de prévision. Si oui, le bit présent de données du dispositif de prévision est ajusté à 1, et l'information de côté complète incluant celle nécessaire pour la mise à jour du dispositif de prévision, et transmise, et la valeur d'erreur de prévision est envoyée au quantificateur. Sinon, le bit présent des données du dispositif de prévision est ajusté à 0, et les bits utilisés par le dispositif de prévision sont tous remis à zéro et ne sont pas transmis. Dans ce cas, la valeur de composante spectrale est envoyée au quantificateur 3. Comme cela est décrit au-dessus, la commande du dispositif de prévision agit d'abord sur tous les dispositifs de prévision d'une bande de facteur de proportionnalité, et est ensuite suivie par une seconde étape sur toutes les bandes de facteur de proportionnalité.

Il est bien évident que le but de la prévision LT est d'obtenir le gain de prévision complet le plus grand. Soit G_1 représentant le gain de prévision dans la 1-ième sous-bande de fréquence, le gain de prévision global dans une trame donnée peut être calculé comme suit :

$$G = \sum_{l=1 \& (G_l > 0)}^{N_s} G_l \quad (12)$$

Si le gain compense la nécessité du bit additionnel pour l'information de côté du dispositif de prévision, c'est-à-dire, $G > T$ (dB), l'information de côté complète est transmise, et les dispositifs de prévision qui produisent des gains positifs sont activés. Sinon, les dispositifs de prévision ne sont pas utilisés.

Les paramètres LP obtenus par la méthode utilisée au-dessus ne sont pas directement liés à la maximisation du gain. Cependant, en calculant le gain pour chaque bloc et pour chaque retard, dans la gamme sélectionnée (1 à 1024 dans cet exemple), et en sélectionnant ce retard qui produit le gain de prévision global le plus grand, le procédé de prévision est optimisé. Le retard sélectionné α et les coefficients correspondants b , sont transmis comme information de côté avec les signaux de sous-bande d'erreur quantifiés. Tandis que la complexité de calcul augmenté dans le codeur, il ne résulte pas d'augmentation de complexité au niveau du décodeur.

La Figure 4 représente en détail le décodeur de la Figure 2. Le signal audio codé est reçu à partir du canal de transmission 6 par le démultiplexeur de flux de bits 7 comme cela est décrit au-dessus. Le démultiplexeur de trains de bits 7 sépare les coefficients de prévision A et les signaux d'erreur quantifié $\tilde{E}(k)$, et fournit ceux-ci à l'outil de prévision 8. Cet outil comporte un combineur 24 qui combine les signaux d'erreur quantifiés $\tilde{E}(k)$ et un signal audio prévu du domaine de fréquence $\hat{X}(k)$, pour générer un signal audio reconstitué $\tilde{X}(k)$ aussi

dans le domaine de fréquence. L'assemblage de filtres 9 convertit le signal reconstitué $\tilde{X}(k)$ du domaine de fréquence au domaine de temps, pour générer un signal audio de domaine de temps reconstitué \tilde{x} . Ce signal est à son tour envoyé à un outil de prévision à long terme, qui reçoit aussi les coefficients de prévision A. L'outil de prévision à long terme 26 génère une trame de temps courante prévue à partir des trames de temps reconstituées précédentes, en utilisant les coefficients de prévision pour la trame courante. Un assemblage de filtres 25 transforme le signal prévu \tilde{x} .

On appréciera que les informations de commande de dispositif de prévision, transmises à partir du codeur peuvent être utilisées au niveau du décodeur pour commander l'opération de décodage. En particulier, les bits utilisés par le dispositif de prévision peuvent être utilisés dans le combinateur 24 pour déterminer si oui ou non la prévision a été utilisée dans une quelconque bande de fréquence donnée.

Sur la Figure 5, est représenté un autre mode de réalisation du codeur de signal audio de la Figure 1, dans lequel un signal audio x devant être codé est comparé avec le signal prévu \hat{x} dans le domaine de temps par un comparateur 15, pour générer un signal d'erreur e aussi dans le domaine de temps. Un outil d'assemblage de filtres 16 transforme alors le signal d'erreur du domaine de temps au domaine de fréquence, pour générer un ensemble de signaux d'erreur de sous-bande de fréquence $E(k)$. Ces signaux sont ensuite quantifiés par un quantificateur 17, pour générer un ensemble de signaux d'erreur quantifié $\tilde{E}(k)$.

Un second assemblage de filtres 18 est alors utilisé pour convertir les signaux d'erreur quantifiés $\tilde{E}(k)$ à nouveau dans le domaine de temps, résultant en un signal \tilde{e} . Ce signal d'erreur quantifié du domaine de temps \tilde{e} est ensuite combiné au niveau d'une unité de

traitement de signal 19, avec le signal audio du domaine
de temps prévu \hat{x} , pour générer un signal audio quantifié
 \tilde{x} . Un outil de prévision 20 met en œuvre la même fonction
que l'outil 12 du codeur de la Figure 3, générant le
5 signal audio prévu \hat{x} et les coefficients de prévision A.
Les coefficients de prévision et les signaux d'erreur
quantifiés sont combinés au niveau d'un multiplexeur de
train de bits 21, pour être transmis sur le canal de
transmission 22. Comme cela est décrit au-dessus, les
10 signaux d'erreur sont quantifiés selon un modèle psycho-
acoustique par un contrôleur 23.

Les algorithmes de codage audio décrits au-dessus
permettent la compression des signaux audio à des taux de
bits bas. La technique est basée sur la prévision à long
15 terme (LT). Comparées aux techniques de prévision
adaptatives de retour connues, les techniques décrites
ici délivrent des gains de prévision plus grands pour des
signaux de musique d'instrument seuls, et des signaux de
discours, tandis qu'ils nécessitent seulement une faible
20 complexité de calcul.

REVENDICATIONS

1 - Une méthode de codage d'un signal audio, la méthode comportant les étapes de :

5 réception d'un signal audio x devant être codé;
génération d'un signal audio quantifié \tilde{x} à partir du signal audio reçu x ;

génération d'un ensemble de coefficients de prévision à long terme A , qui peut être utilisé pour
10 prévoir une trame de temps courante du signal audio reçu, directement à partir d'au moins une trame de temps précédente du signal audio quantifié \tilde{x} ;

utilisation des coefficients de prévision A pour générer un signal audio prévu \hat{x} ;

15 comparaison du signal audio reçu x avec le signal audio prévu \hat{x} , et génération d'un signal d'erreur $E(k)$ pour chaque sous-bande de la pluralité de sous-bandes de fréquence;

20 quantification des signaux d'erreur $E(k)$ pour générer un ensemble de signaux d'erreur quantifiés $\tilde{E}(k)$;
et

combinaison des signaux d'erreur quantifiés $\tilde{E}(k)$ et des coefficients de prévision A pour générer un signal audio codé.

25

2 - Méthode selon la revendication 1 et comportant une transformation du signal audio reçu x dans des trames x_m du domaine de temps vers le domaine de fréquence, pour fournir un ensemble de signaux de sous-bande de fréquence $X(k)$, et une transformation du signal audio prévu \hat{x} du domaine de temps au domaine de fréquence, pour générer un ensemble de signaux de sous-bande de fréquence prévu $\hat{X}(k)$, dans lequel la comparaison entre le signal audio reçu x et le signal audio prévu \hat{x}
30 est effectuée dans le domaine de fréquence, en comparant des signaux de sous-bande respective les uns aux autres,
35

pour générer les signaux d'erreur de sous-bande de fréquence $E(k)$.

3 - Méthode selon la revendication 1 et
5 comportant l'application de la comparaison entre le signal audio reçu x et le signal audio prévu \hat{x} dans le domaine de temps, pour générer un signal d'erreur e aussi dans le domaine de temps, et la conversion du signal d'erreur e du domaine de temps au domaine de fréquence,
10 pour générer ladite pluralité de signaux d'erreur de sous-bande de fréquence $E(k)$.

4 - Méthode de décodage d'un signal audio codé, la méthode comportant les étapes de :

15 réception d'un signal audio codé comportant un signal d'erreur quantifié $\tilde{E}(k)$ pour chaque sous-bande de la pluralité de sous-bandes de fréquence du signal audio, et, pour chaque trame de temps du signal audio, un ensemble de coefficients de prévision A qui peuvent être
20 utilisés pour prévoir une trame de temps courante x_m du signal audio reçu, directement à partir d'au moins une trame de temps précédente d'un signal audio quantifié reconstitué \tilde{x} ;

génération dudit signal audio quantifié
25 reconstitué \tilde{x} à partir des signaux d'erreur quantifié $\tilde{E}(k)$;

utilisation des coefficients de prévision A et du signal audio quantifié \tilde{x} pour générer un signal audio
prévu \hat{x} ;

30 transformation du signal audio prévu \hat{x} du domaine de temps au domaine de fréquence, pour générer un ensemble de signaux de sous-bande de fréquence prévus $\hat{X}(k)$, pour la combinaison avec les signaux d'erreur quantifiés $\tilde{E}(k)$, pour générer un ensemble de signaux de
35 sous-bande de fréquence reconstitués $\tilde{X}(k)$; et

application d'une transformation du domaine de fréquence au domaine de temps sur les signaux de sous-bande de fréquence reconstitués $\tilde{X}(k)$, pour générer le signal audio quantifié reconstitué \tilde{x} .

5

5 - Dispositif pour le codage d'un signal audio, le dispositif comportant :

une entrée pour recevoir un signal audio x devant être codé;

10 un moyen de traitement (2, 3 ; 15 à 19) couplé à ladite entrée, pour générer à partir du signal audio reçu x , un signal audio quantifié \tilde{x} ;

15 un moyen de prévision (12; 13) couplé audit moyen de traitement (3) pour générer un ensemble de coefficients de prévision à long terme A , pour prévoir une trame de temps courante x_m du signal audio reçu x , directement à partir d'au moins une trame de temps précédente du signal audio quantifié \tilde{x} ;

20 un moyen de génération (10 à 14; 20, 15) pour générer un signal audio prévu \hat{x} en utilisant les coefficients de prévision A , et pour comparer le signal audio reçu x avec le signal audio prévu \hat{x} , pour générer un signal d'erreur $E(k)$ pour chaque sous-bande de la pluralité de sous-bandes de fréquence;

25 un moyen de quantification (3; 17) pour quantifier les signaux d'erreur $E(k)$, pour générer un ensemble de signaux d'erreur quantifié $\tilde{E}(k)$; et

30 un moyen de combinaison (5; 21) pour combiner les signaux d'erreur quantifiés $\tilde{E}(k)$ avec les coefficients de prévision A , pour générer un signal audio codé.

6 - Dispositif selon la revendication 5, dans lequel ledit moyen de génération comporte un premier moyen de transformation (11), pour transformer le signal audio reçu x du domaine de temps au domaine de fréquence, et un second moyen de transformation (13), pour

35

transformer le signal audio prévu \hat{x} du domaine de temps au domaine de fréquence, et un moyen de comparaison (14), prévu pour comparer les signaux du domaine de fréquence résultant dans le domaine de fréquence.

5

7 - Dispositif selon la revendication 6, dans lequel le moyen de génération est prévu pour comparer le signal audio reçu x et le signal audio prévu \hat{x} dans le domaine de temps.

10

8 - Dispositif pour le décodage d'un signal audio codé x , dans lequel le signal audio codé comporte un signal d'erreur quantifié $\tilde{E}(k)$ pour chaque sous-bande d'une pluralité de sous-bandes de fréquence du signal audio, et un ensemble de coefficients de prévision A pour chaque trame de temps du signal audio, et dans lequel les coefficients de prévision A peuvent être utilisés pour prévoir une trame de temps courante x_m du signal audio reçu, directement à partir d'au moins une trame de temps précédente d'un signal audio quantifié reconstitué \tilde{X} , le dispositif comportant :

15

20

une entrée pour recevoir le signal audio codé;
un moyen de génération (24, 25, 9) pour générer ledit signal audio quantifié reconstitué \tilde{X} à partir des signaux d'erreur quantifié $\tilde{E}(k)$; et

25

un moyen de traitement de signal (26) pour générer un signal audio prévu \hat{x} à partir des coefficients de prévision A et dudit signal audio reconstitué \tilde{x} ;

30

dans lequel ledit moyen de génération comporte un premier moyen de transformation (25) pour transformer le signal audio prévu \hat{x} du domaine de temps au domaine de fréquence, pour générer un ensemble de signaux de sous-bande de fréquence prévus $\hat{X}(k)$,

35

un moyen de combinaison (24) pour combiner ledit ensemble de signaux de sous-bande de fréquence prévus $\hat{X}(k)$ avec les signaux d'erreur quantifiés $\tilde{E}(k)$, pour

généraliser un ensemble de signaux de sous-bande de fréquence reconstitués $\tilde{X}(k)$, et un second moyen de transformation (9) pour appliquer une transformation du domaine de fréquence au domaine de temps, sur les signaux de sous-bande de fréquence reconstitués $\tilde{X}(k)$, pour générer le signal audio quantifié reconstitué \tilde{x} .

1/5

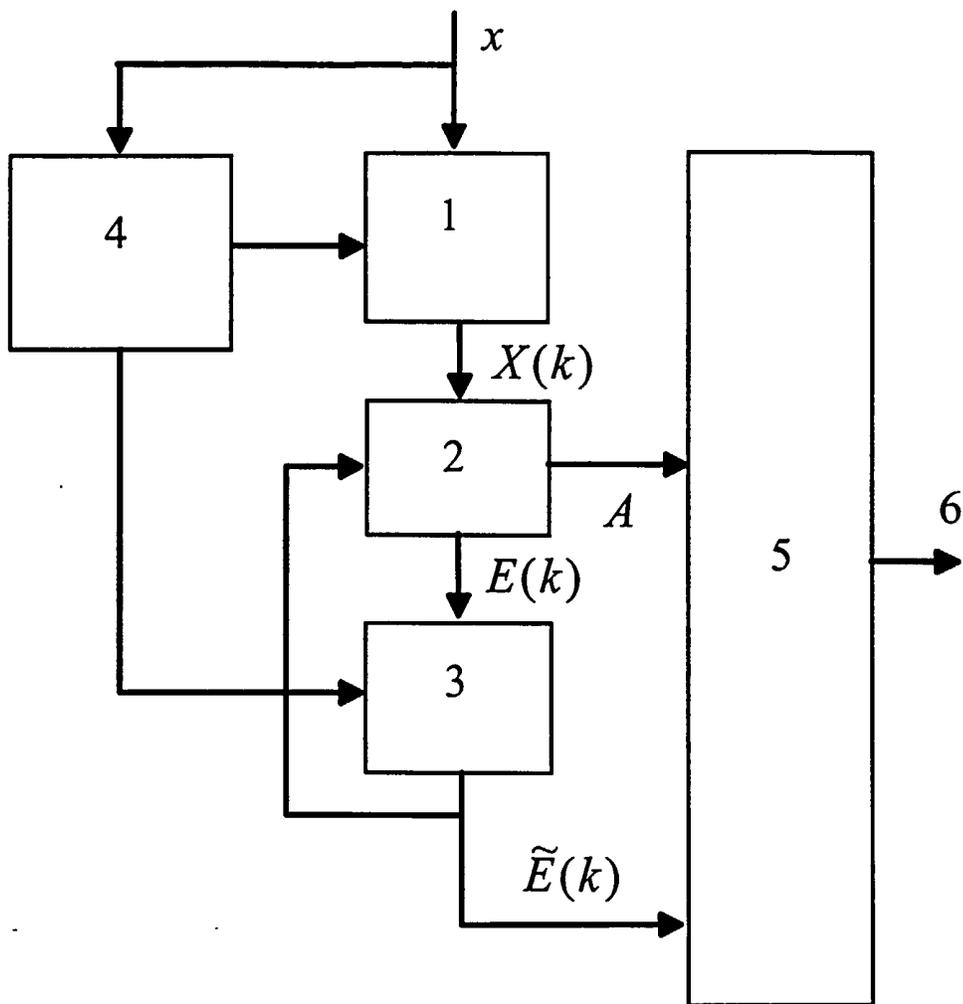


Figure 1

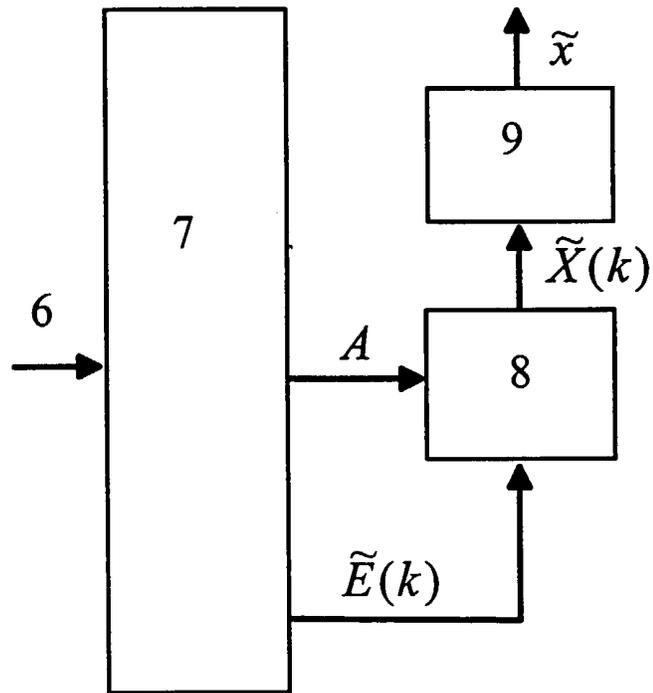


Figure 2

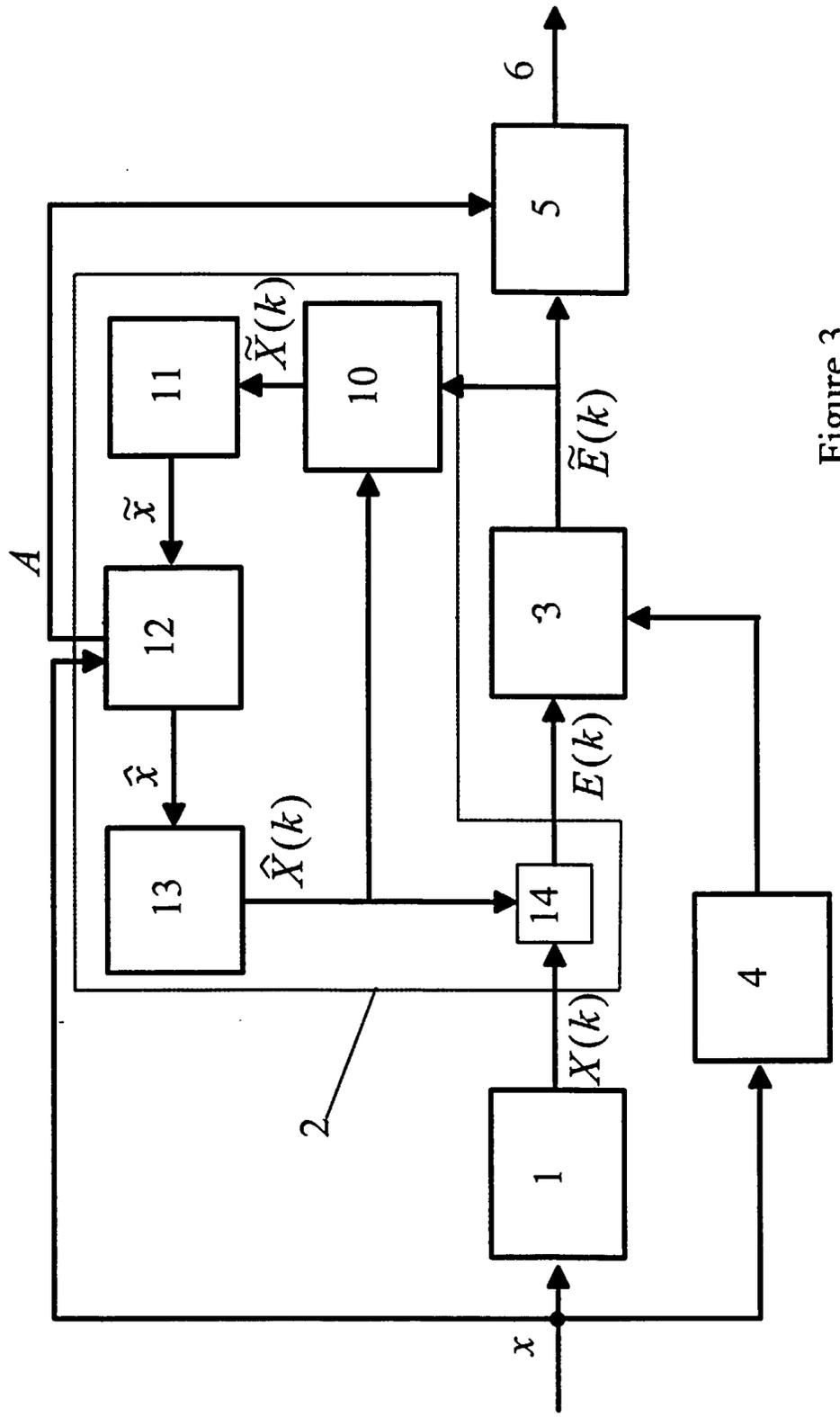


Figure 3

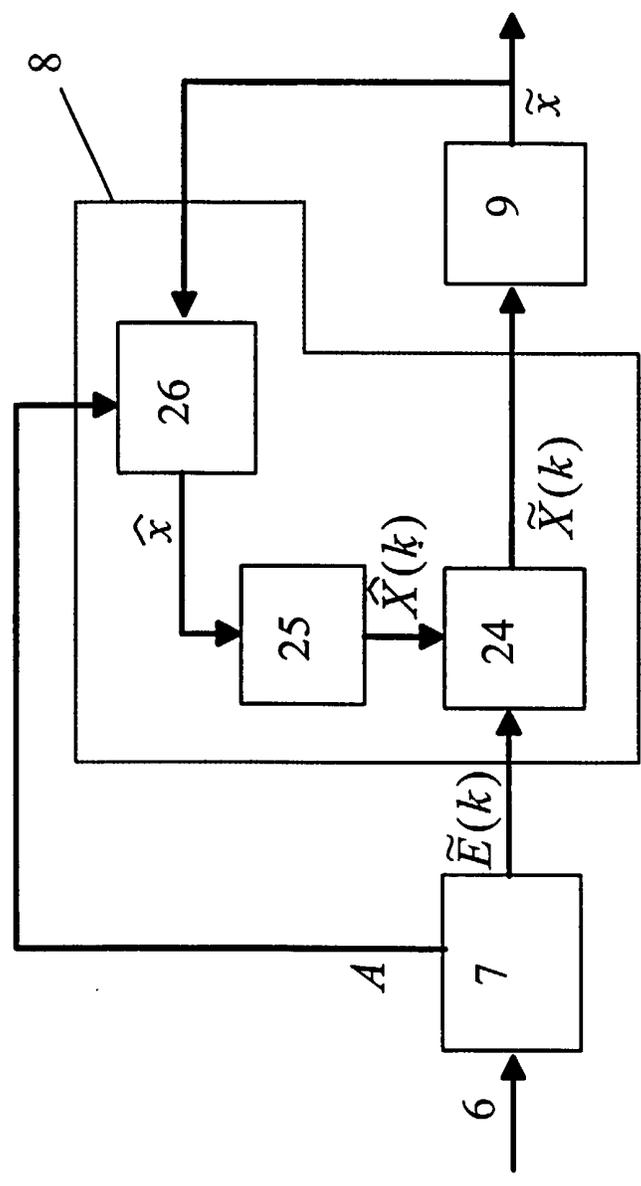


Figure 4

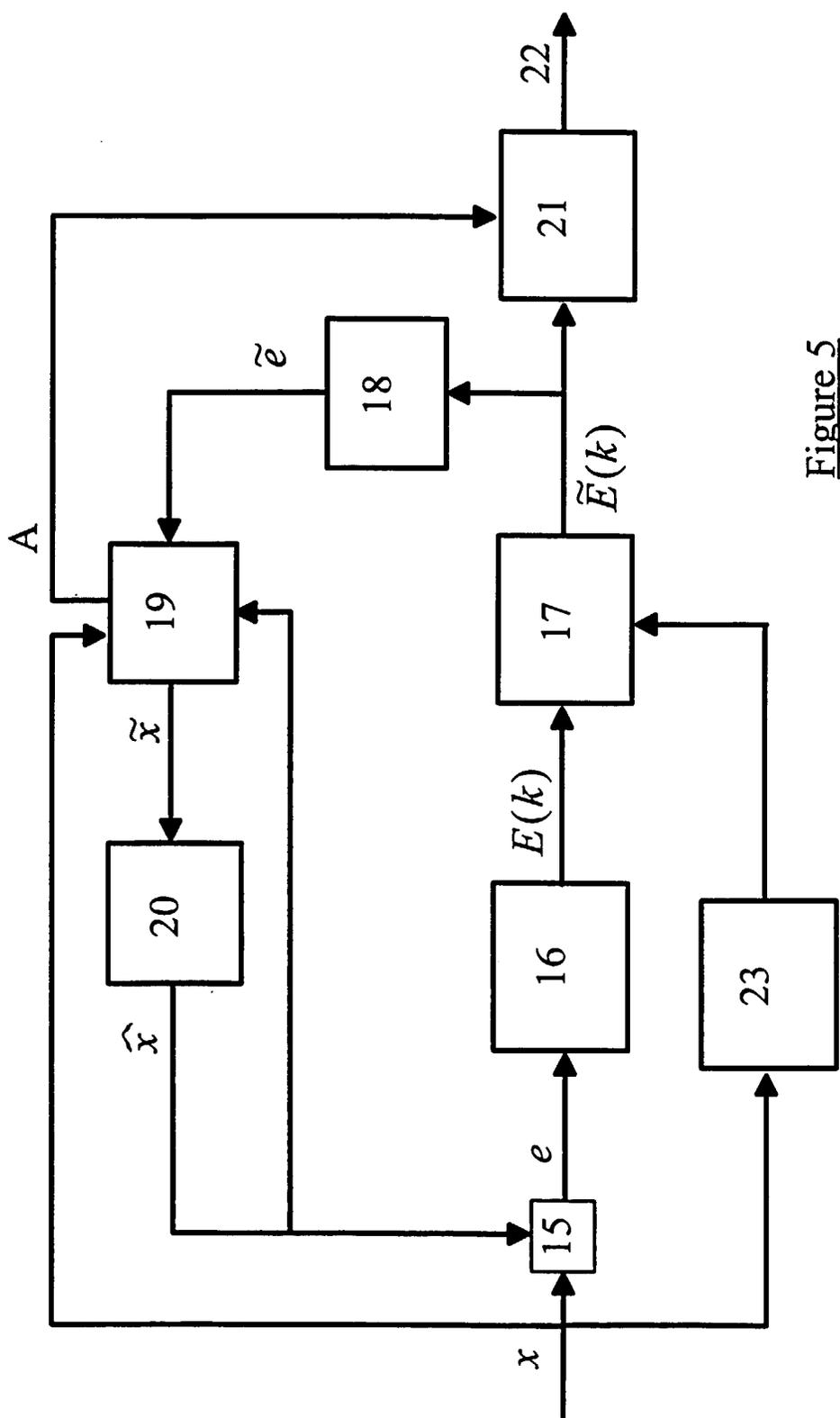


Figure 5