



(11) **EP 3 330 966 A1**

(12) **DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:
06.06.2018 Bulletin 2018/23

(51) Int Cl.:
G10L 21/038 (2013.01)

(21) Numéro de dépôt: **17206563.3**

(22) Date de dépôt: **04.02.2015**

(84) Etats contractants désignés:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Etats d'extension désignés:
BA ME

(71) Demandeur: **Koninklijke Philips N.V.**
5656 AE Eindhoven (NL)

(72) Inventeurs:
• **KANIEWSKA, Magdalena**
5656 AE Eindhoven (NL)
• **RAGOT, Stéphane**
5656 AE Eindhoven (NL)

(30) Priorité: **07.02.2014 FR 1450969**

(62) Numéro(s) de document de la (des) demande(s) initiale(s) en application de l'article 76 CBE:
15705687.0 / 3 103 116

(74) Mandataire: **de Haan, Poul Erik**
Philips International B.V.
Philips Intellectual Property & Standards
High Tech Campus 5
5656 AE Eindhoven (NL)

(54) **EXTENSION AMÉLIORÉE DE BANDE DE FRÉQUENCE DANS UN DÉCODEUR DE SIGNAUX AUDIOFRÉQUENCES**

(57) L'invention se rapport à un procédé d'extension de bande de fréquence d'un signal audiofréquence lors d'un processus de décodage ou d'amélioration comportant une étape d'obtention du signal décodé dans une première bande de fréquence dite bande basse, le procédé étant caractérisé en ce qu'il comporte les étapes suivantes :

- Extraction (E402) de composantes tonales et d'un signal d'ambiance à partir du signal issu du signal bande basse décodé ;

- Combinaison (E403) des composantes tonales et du signal d'ambiance par mixage adaptatif utilisant des facteurs de contrôle de niveau d'énergie pour obtenir un signal audio, dit combiné ;

- Extension (E401a) sur au moins une deuxième bande de fréquence supérieure à la première bande de fréquence du signal décodé bande basse avant l'étape d'extraction pour former un signal bande basse décodé étendu $U_{HB1}(k)$,

et selon lequel l'étape d'extraction (E402) des composantes tonales et du signal d'ambiance comporte les opérations suivantes :

(a) calcul de l'énergie tonale du signal de bande basse décodé étendu ;

(b) calcul de l'ambiance en valeur absolue correspondant au niveau moyen du spectre raie par raie et calcul de l'énergie des parties tonales dominantes dans le spectre hautes fréquences.

L'invention se rapporte également à un dispositif

d'extension de bande de fréquence mettant en oeuvre le procédé décrit et un décodeur comportant un tel dispositif.

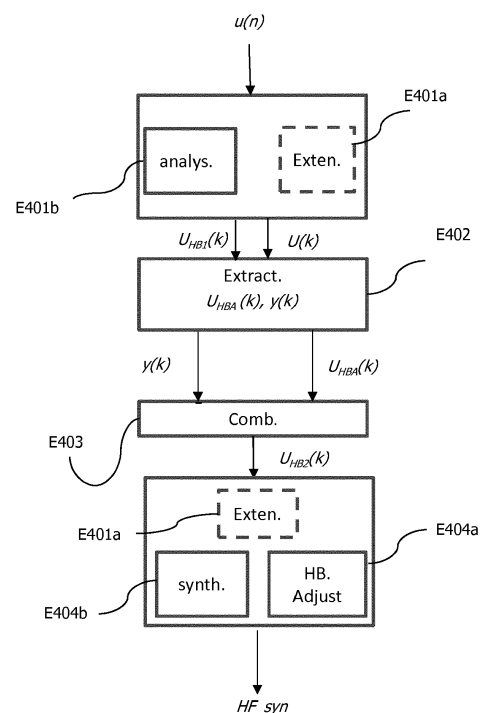


Fig.4

EP 3 330 966 A1

Description

[0001] La présente invention se rapporte au domaine du codage/décodage et du traitement de signaux audiofréquences (comme des signaux de parole, de musique ou autres) pour leur transmission ou leur stockage.

[0002] Plus particulièrement, l'invention concerne un procédé et un dispositif d'extension de bande de fréquence dans un décodeur ou un processeur réalisant une amélioration de signal audiofréquence.

[0003] De nombreuses techniques existent pour compresser (avec perte) un signal audiofréquence comme la parole ou la musique.

[0004] Les méthodes classiques de codage pour les applications conversationnelles sont en général classifiées en *codage de forme d'onde* (MIC pour "Modulation par Impulsion et codage", MICDA pour "Modulation par Impulsion et Codage Différentiel Adaptatif", codage par transformée...), *codage paramétrique* (LPC pour "Linear Predictive Coding" en anglais, codage sinusoïdal...) et codage hybride paramétrique avec une quantification des paramètres par "analyse par synthèse" dont le codage CELP (pour "Code Excited Linear Prediction" en anglais) est l'exemple le plus connu.

[0005] Pour les applications non conversationnelles, l'état de l'art en codage de signal audio (mono) est constitué par le codage perceptuel par transformée ou en sous-bandes, avec un codage paramétrique des hautes fréquences par réplication de bande (SBR pour Spectral Band Replication en anglais).

Une revue des méthodes classiques de codage de parole et audio se trouve dans les ouvrages W.B. Kleijn and K.K. Paliwal (Eds.), *Speech Coding and Synthesis*, Elsevier, 1995 ; M. Bosi, R.E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Springer 2002 ; J. Benesty, M.M. Sondhi, Y. Huang (Eds.), *Handbook of Speech Processing*, Springer 2008.

[0006] On s'intéresse ici plus particulièrement au codec (codeur et décodeur) normalisé 3GPP AMR-WB (pour "Adaptive Multi-Rate Wideband" en anglais) qui fonctionne à une fréquence d'entrée/sortie de 16 kHz et dans lequel le signal est divisé en deux sous-bandes, la bande basse (0-6.4 kHz) qui est échantillonnée à 12.8 kHz et codée par modèle CELP et la bande haute (6.4-7 kHz) qui est reconstruite de façon paramétrique par « *extension de bande* » (ou BWE pour "Bandwidth Extension" en anglais) avec ou sans information supplémentaire selon le mode de la trame courante. On peut noter ici que la limitation de la bande codée du codec AMR-WB à 7kHz est essentiellement liée au fait que la réponse en fréquence en émission des terminaux en bande élargie a été approximée au moment de la normalisation (ETSI/3GPP puis UIT-T) selon le masque fréquentiel défini dans la norme UIT-T P.341 et plus précisément en utilisant un filtre dit « P341 » défini dans la norme UIT-T G.191 qui coupe les fréquences au-dessus de 7 kHz (ce filtre respecte le masque défini dans P.341). Cependant, en théorie, il est bien connu qu'un signal échantillonné à 16 kHz peut avoir une bande audio définie de 0 à 8000 Hz ; le codec AMR-WB introduit donc une limitation de la bande haute en comparaison à la largeur de bande théorique de 8 kHz.

[0007] Le codec de parole 3GPP AMR-WB a été normalisé en 2001 principalement pour les applications de téléphonie en mode circuit (CS) sur GSM (2G) et UMTS (3G). Ce même codec a été aussi normalisé en 2003 à l'UIT-T en tant que recommandation G.722.2 "Wideband coding speech at around 16kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)".

[0008] Il comprend neuf débits, appelés modes, de 6.6 à 23.85 kbit/s, et comprend des mécanismes de transmission continue (DTX pour "Discontinuous Transmission") avec détection d'activité vocale (VAD pour "Voice Activity Detection") et génération de bruit de confort (CNG pour "Confort Noise Generation") à partir de trames de description de silence (SID pour "Silence Insertion Descriptor"), ainsi que des mécanismes de correction de trames perdues (FEC pour "Frame Erasure Concealment", parfois appelé PLC pour "Packet Loss Concealment").

[0009] On ne reprend pas ici les détails de l'algorithme de codage et de décodage AMR-WB, une description détaillée de ce codec se trouve dans les spécifications 3GPP (TS 26.190, 26.191, 26.192, 26.193, 26.194, 26.204) et UIT-T-G.722.2 (et les Annexes et Appendice correspondantes) ainsi que dans l'article de B. Bessette et al. intitulé « The adaptive multirate wideband speech codec (AMR-WB) », IEEE Transactions on Speech and Audio Processing, vol. 10, no. 8, 2002, pp. 620-636 et les codes sources des standards 3GPP et UIT-T associés.

[0010] Le principe de l'extension de bande dans le codec AMR-WB est assez rudimentaire. En effet, la bande haute (6.4-7 kHz) est générée en mettant en forme un bruit blanc par le biais d'une enveloppe temporelle (appliquée sous la forme de gains par sous-trame) et fréquentielle (par l'application d'un filtre de synthèse de prédiction linéaire ou LPC pour "Linear Predictive Coding"). Cette technique d'extension de bande est illustrée à la **figure 1**.

[0011] Un bruit blanc, $u_{HB1}(n)$, $n = 0, \dots, 79$, est généré à 16 kHz par sous-trame de 5 ms par générateur congruentiel linéaire (bloc 100). Ce bruit $u_{HB1}(n)$ est mis en forme dans le temps par application de gains par sous-trame ; cette opération est décomposée en deux étapes de traitement (blocs 102, 106 ou 109) :

- Un premier facteur est calculé (bloc 101) pour mettre le bruit blanc $u_{HB1}(n)$ (bloc 102) à un niveau semblable à celui de l'excitation, $u(n)$, $n = 0, \dots, 63$, décodée à 12.8 kHz dans la bande basse :

$$u_{HB2}(n) = u_{HB1}(n) \sqrt{\frac{\sum_{l=0}^{63} u(l)^2}{\sum_{l=0}^{79} u_{HB1}(l)^2}}$$

[0012] On peut noter ici que la normalisation des énergies se fait en comparant des blocs de taille différente (64 pour $u(n)$ et 80 pour $u_{HB1}(n)$), sans compensation des différences de fréquences d'échantillonnage (12.8 ou 16 kHz).

- L'excitation dans la bande haute est ensuite obtenue (bloc 106 ou 109) sous la forme :

$$u_{HB}(n) = \hat{g}_{HB} u_{HB2}(n)$$

où le gain \hat{g}_{HB} est obtenu différemment selon le débit. Si le débit de la trame actuelle est <23.85 kbit/s, le gain \hat{g}_{HB} est estimé « en aveugle » (c'est-à-dire sans information supplémentaire); dans ce cas, le bloc 103 filtre le signal décodé en bande basse par un filtre passe-haut ayant une fréquence de coupure à 400 Hz pour obtenir un signal $\hat{s}_{hp}(n), n=0, \dots, 63$ - ce filtre passe-haut élimine l'influence des très basses fréquences qui peuvent biaiser l'estimation faite dans le bloc 104 - puis on calcule le « tilt » (indicateur de pente spectrale) noté e_{tilt} du signal $\hat{s}_{hp}(n)$ par autocorrélation normalisée (bloc 104):

$$e_{tilt} = \frac{\sum_{n=1}^{63} \hat{s}_{hp}(n) \hat{s}_{hp}(n-1)}{\sum_{n=0}^{63} \hat{s}_{hp}(n)^2}$$

et enfin on calcule \hat{g}_{HB} sous la forme :

$$\hat{g}_{HB} = w_{SP} g_{SP} + (1 - w_{SP}) g_{BG}$$

où $g_{SP} = 1 - e_{tilt}$ est le gain appliqué dans les trames actives de parole (SP pour speech), $g_{BG} = 1.25 g_{SP}$ est le gain appliqué dans les trames inactives de parole associées à un bruit de fond (BG pour Background) et w_{SP} est une fonction de pondération qui dépend de la détection d'activité vocale (VAD). On comprend que l'estimation du tilt (e_{tilt}) permet d'adapter le niveau de la bande haute en fonction de la nature spectrale du signal ; cette estimation est particulièrement importante quand la pente spectrale du signal décodé CELP est telle que l'énergie moyenne décroît quand la fréquence augmente (cas d'un signal voisé où e_{tilt} est proche de 1, donc $g_{SP} = 1 - e_{tilt}$ est ainsi réduit). A noter aussi que le facteur \hat{g}_{HB} dans le décodage AMR-WB est borné pour prendre des valeurs dans l'intervalle [0.1, 1.0]. En fait, pour les signaux dont le spectre a plus d'énergie en hautes fréquences (e_{tilt} proche de -1, g_{SP} proche de 2), le gain \hat{g}_{HB} est d'habitude sous-estimé.

[0013] A 23.85 kbit/s, une information de correction est transmise par le codeur AMR-WB et décodée (blocs 107, 108) afin d'affiner le gain estimé par sous-trame (4 bits toutes les 5ms, soit 0.8 kbit/s).

[0014] L'excitation artificielle $u_{HB}(n)$ est ensuite filtrée (bloc 111) par un filtre de synthèse LPC de fonction de transfert $1/A_{HB}(z)$ et fonctionnant à la fréquence d'échantillonnage de 16 kHz. La réalisation de ce filtre dépend du débit de la trame courante:

- A 6.6 kbit/s, le filtre $1/A_{HB}(z)$ est obtenu en pondérant par un facteur $\gamma=0.9$ un filtre LPC d'ordre 20, $1/\hat{A}^{ext}(z)$ qui « extrapole » le filtre LPC d'ordre 16, $1/\hat{A}(z)$, décodé dans la bande basse (à 12.8 kHz) - les détails de l'extrapolation dans le domaine des paramètres ISF (pour "Imittance Spectral Frequency" en anglais) sont décrits dans la norme G.722.2 à la section 6.3.2.1; dans ce cas,

$$1/A_{HB}(z) = 1/\hat{A}^{ext}(z/\gamma)$$

EP 3 330 966 A1

- Aux débits > 6.6 kbit/s, le filtre $1/A_{HB}(z)$ est d'ordre 16 et correspond simplement à :

$$1 / A_{HB}(z) = 1 / \hat{A}(z / \gamma)$$

5

où $\gamma=0.6$. A noter que dans ce cas le filtre $1/\hat{A}(z/\gamma)$ est utilisé à 16 kHz, ce qui résulte en un étalement (par homothétie) de la réponse en fréquence de ce filtre de [0, 6.4 kHz] à [0, 8 kHz].

10

Le résultat, $s_{HB}(n)$, est enfin traité par un filtre passe-bande (bloc 112) de type FIR ("Finite Impulse Response"), pour ne garder que la bande 6 - 7 kHz ; à 23.85 kbit/s, un filtre passe-bas également de type FIR (bloc 113) se rajoute au traitement pour atténuer encore plus les fréquences supérieures à 7 kHz. La synthèse en hautes fréquences (HF) est finalement additionnée (bloc 130) à la synthèse en basses fréquences (BF) obtenue avec les blocs 120 à 123 et ré-échantillonnée à 16 kHz (bloc 123). Ainsi même si la bande haute s'étend en théorie de 6.4 à 7 kHz dans le codec AMR-WB, la synthèse HF est plutôt comprise dans la bande 6-7 kHz avant addition avec la synthèse BF.

15

[0015] On peut identifier plusieurs inconvénients à la technique d'extension de bande du codec AMR-WB :

20

- Le signal dans la bande haute est un bruit blanc mis en forme (par gains temporels par sous-trame, par filtrage par $1/A_{HB}(z)$ et filtrage passe-bande), ce qui n'est pas un bon modèle général du signal dans la bande 6.4-7 kHz. Il existe par exemple des signaux de musique très harmoniques pour lesquels la bande 6.4-7 kHz contient des composantes sinusoïdales (ou tones) et aucun bruit (ou peu de bruit), pour ces signaux l'extension de bande du codec AMR-WB dégrade fortement la qualité.

25

- Le filtre passe-bas à 7 kHz (bloc 113) introduit un décalage de près de 1 ms entre les bandes basses et hautes, ce qui peut potentiellement dégrader la qualité de certains signaux en désynchronisant légèrement les deux bandes à 23.85 kbit/s - cette désynchronisation peut également poser problème lors d'une commutation de débit de 23.85 kbit/s à d'autres modes.

30

- L'estimation de gains par sous-trame (bloc 101, 103 à 105) n'est pas optimale. Pour partie, elle se base sur une égalisation de l'énergie « absolue » par sous-trame (bloc 101) entre des signaux à des fréquences différentes : l'excitation artificielle à 16 kHz (bruit blanc) et un signal à 12.8 kHz (excitation ACELP décodée). On peut noter en particulier que cette approche induit implicitement une atténuation de l'excitation bande haute (par un ratio $12.8/16=0.8$) ; en fait, on notera également qu'aucune désaccentuation (ou déemphasis) n'est effectuée sur la bande haute dans le codec AMR-WB, ce qui induit implicitement une amplification relative proche de 0.6 (qui correspond à la valeur de la réponse en fréquence de $1/(1-0.68z^{-1})$ à 6400 Hz). En fait, les facteurs de 1/0.8 et de 0.6 se compensent approximativement.

35

- Sur la parole, les tests de caractérisation du codec 3GPP AMR-WB documentés dans le rapport 3GPP TR 26.976 ont montré que le mode à 23.85 kbit/s a une qualité moins bonne qu'à 23.05 kbit/s, sa qualité est en fait similaire à celle du mode à 15.85 kbit/s. Ceci montre en particulier que le niveau du signal HF artificiel doit être contrôlé de façon très prudente, car la qualité est dégradée à 23.85 kbit/s alors que les 4 bits par trame sont sensés permettre de mieux approcher l'énergie des hautes fréquences originales.

40

- La limitation de la bande codée à 7 kHz résulte de l'application d'un modèle strict de la réponse en émission des terminaux acoustiques (filtre P.341 dans la norme UIT-T G.191). Or, pour une fréquence d'échantillonnage de 16 kHz, les fréquences dans la bande 7-8 kHz restent importantes, en particulier pour les signaux de musique, pour assurer un bon niveau de qualité.

45

[0016] L'algorithme de décodage AMR-WB a été amélioré en partie avec le développement du codec scalable UIT-T G.718 qui a été normalisé en 2008.

[0017] La norme UIT-T G.718 comprend un mode dit interopérable, pour lequel le codage coeur est compatible avec le codage G.722.2 (AMR-WB) à 12.65 kbit/s ; de plus, le décodeur G.718 a la particularité de pouvoir décoder un train binaire AMR-WB/G.722.2 à tous les débits possibles du codec AMR-WB (de 6.6 à 23.85 kbit/s).

50

[0018] Le décodeur interopérable G.718 en mode bas délai (« low delay » en anglais) (G.718-LD) est illustré à la **figure 2**. On liste ci-dessous les améliorations apportées à la fonctionnalité de décodage de train binaire AMR-WB dans le décodeur G.718, avec des références à la figure 1 lorsque nécessaire :

55

L'extension de bande (décrite par exemple dans la clause 7.13.1 de la recommandation G.718, bloc 206) est identique à celle du décodeur AMR-WB, sauf que le filtre passe-bande 6-7 kHz et le filtre de synthèse $1/A_{HB}(z)$ (blocs 111 et 112) sont en ordre inversé. De plus, à 23.85 kbit/s les 4 bits transmis par sous-frames par le codeur AMR-WB ne sont pas utilisés dans le décodeur G.718 interopérable ; la synthèse des hautes fréquences (HF) à 23.85 kbit/s est donc identique à 23.05 kbit/s ce qui évite le problème connu de qualité du décodage AMR-WB à 23.85 kbit/s. A fortiori, le filtre passe-bas à 7 kHz (bloc 113) n'est pas utilisé, et le décodage spécifique du mode à

23.85 kbit/s est omis (blocs 107 à 109).

Un post-traitement de la synthèse à 16 kHz (voir clause 7.14 de G.718) est mis en oeuvre dans G.718 par "noise gate" dans le bloc 208 (pour « améliorer » la qualité des silences par réduction du niveau), filtrage passe-haut (bloc 209), post-filtre de basses fréquences (dit « bass postfilter ») dans le bloc 210 atténuant le bruit inter-harmonique en basses fréquences et une conversion en entiers 16 bits avec contrôle de saturation (avec contrôle de gain ou AGC) dans le bloc 211.

[0019] Cependant l'extension de bande dans les codecs AMR-WB et/ou G.718 (mode interopérable) reste encore limitée sur plusieurs aspects.

En particulier, la synthèse de hautes fréquences par bruit blanc mis en forme (par une approche temporelle de type source-filtre LPC) est un modèle très limité du signal dans la bande des fréquences supérieures à 6.4 kHz.

Seule la bande 6.4-7 kHz est re-synthétisée de façon artificielle, alors qu'en pratique une bande plus large (jusqu'à 8 kHz) est en théorie possible à la fréquence d'échantillonnage de 16 kHz, ce qui peut potentiellement améliorer la qualité des signaux, s'ils ne sont pas prétraités par un filtre de type P.341 (50-7000 Hz) tel que définie dans la *Software Tool Library* (norme G.191) de l'UIT-T.

[0020] Il existe donc un besoin pour améliorer l'extension de bande dans un codec de type AMR-WB ou une version interopérable de ce codec ou plus généralement pour améliorer l'extension de bande d'un signal audio, notamment pour améliorer le contenu fréquentiel de l'extension de bande.

[0021] La présente invention vient améliorer la situation.

[0022] L'invention propose à cet effet, un procédé d'extension de bande de fréquence d'un signal audiofréquence lors d'un processus de décodage ou d'amélioration comportant une étape d'obtention du signal décodé dans une première bande de fréquence dite bande basse. Le procédé est tel qu'il comporte les étapes suivantes:

- extraction de composantes tonales et d'un signal d'ambiance à partir d'un signal issu du signal bande basse décodé;
- combinaison des composantes tonales et du signal d'ambiance par mixage adaptatif utilisant des facteurs de contrôle de niveau d'énergie pour obtenir un signal audio, dit signal combiné;
- extension sur au moins une deuxième bande de fréquence supérieure à la première bande de fréquence du signal décodé bande basse avant l'étape d'extraction ou du signal combiné après l'étape de combinaison.

[0023] On notera que par la suite l'« extension de bande » sera prise au sens large et inclura non seulement le cas de l'extension d'une sous-bande en hautes fréquences mais également le cas d'un remplacement de sous-bandes mises à zéro (de type « noise filling » en codage par transformée).

Ainsi, à la fois la prise en compte de composantes tonales et d'un signal d'ambiance extrait du signal issu du décodage de la bande basse permet d'effectuer l'extension de bande avec un modèle de signal adapté à la vraie nature du signal contrairement à l'utilisation d'un bruit artificiel. La qualité de l'extension de bande est ainsi améliorée et notamment pour certains types de signaux comme les signaux de musique.

[0024] En effet, le signal décodé dans la bande basse comporte une partie correspondant à l'ambiance sonore qui peut être transposée en haute fréquence de telle sorte qu'un mixage des composantes harmoniques et de l'ambiance existante permet d'assurer une bande haute reconstruite cohérente.

[0025] On remarquera que même si l'invention est motivée par l'amélioration de la qualité de l'extension de bande dans le contexte du codage AMR-WB interopérable, les différents modes de réalisation s'appliquent au cas plus général de l'extension de bande d'un signal audio, en particulier dans un dispositif d'amélioration effectuant une analyse du signal audio pour extraire les paramètres nécessaires à l'extension de bande.

[0026] Les différents modes particuliers de réalisation mentionnés ci-après peuvent être ajoutés indépendamment ou en combinaison les uns avec les autres, aux étapes du procédé d'extension défini ci-dessus.

[0027] Dans un mode de réalisation, l'extension de bande est effectuée dans le domaine de l'excitation et le signal bande basse décodé est un signal d'excitation bande basse décodé.

[0028] L'avantage de ce mode de réalisation est qu'une transformation sans fenêtrage (ou de façon équivalente avec une fenêtre rectangulaire implicite de la longueur de la trame) est possible dans le domaine de l'excitation. Dans ce cas aucun artefact (effets de bloc) n'est alors audible.

[0029] Dans un premier mode de réalisation, l'extraction des composantes tonales et du signal d'ambiance s'effectue selon les étapes suivantes :

- détection des composantes tonales dominantes du signal bande basse décodé ou décodé et étendu, dans le domaine fréquentiel ;
- calcul d'un signal résiduel par extraction des composantes tonales dominantes pour obtenir le signal d'ambiance.

Ce mode de réalisation permet une détection précise des composantes tonales.

Dans un deuxième mode de réalisation, de faible complexité, l'extraction des composantes tonales et du signal d'ambiance s'effectue selon les étapes suivantes :

- obtention du signal d'ambiance par calcul d'une valeur moyenne du spectre du signal bande basse décodé ou étendu ;
- obtention des composantes tonales par soustraction du signal d'ambiance calculé au signal bande basse décodé ou décodé et étendu.

Dans un mode de réalisation de l'étape de combinaison, un facteur de contrôle du niveau d'énergie utilisé pour le mixage adaptatif, est calculé en fonction de l'énergie totale du signal bande basse décodé ou décodé et étendu et des composantes tonales.

[0030] L'application de ce facteur de contrôle permet à l'étape de combinaison de s'adapter aux caractéristiques du signal pour optimiser la proportion relative de signal d'ambiance dans le mélange. Le niveau d'énergie est ainsi contrôlé de façon à éviter les artefacts audibles.

[0031] Dans un mode de réalisation préféré, le signal bande basse décodé subit une étape de décomposition en sous-bandes par transformée ou par banc de filtres, les étapes d'extraction et de combinaison s'effectuant alors dans le domaine fréquentiel ou en sous-bandes.

[0032] La mise en oeuvre de l'extension de bande dans le domaine fréquentiel permet d'obtenir une finesse d'analyse en fréquence dont on ne dispose pas avec une approche temporelle, et permet aussi d'avoir une résolution fréquentielle suffisante pour détecter les composantes tonales.

[0033] Dans un mode de réalisation détaillé, le signal bande basse décodé et étendu est obtenu selon l'équation suivante:

$$U_{HB1}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ U(k) & k = 200, \dots, 239 \\ U(k + start_band - 240) & k = 240, \dots, 319 \end{cases}$$

avec k l'indice de l'échantillon, $U(k)$ le spectre du signal obtenu après une étape de transformée, $U_{HB1}(k)$ le spectre du signal étendu, et $start_band$ une variable prédéfinie. Ainsi, cette fonction comprend un ré-échantillonnage du signal en ajoutant des échantillons au spectre de ce signal. D'autres façons d'étendre le signal sont cependant possibles, par exemple par translation dans un traitement sous-bandes.

[0034] La présente invention vise également un dispositif d'extension de bande de fréquence d'un signal audiofréquence, le signal ayant été décodé dans une première bande de fréquence dite bande basse. Le dispositif est tel qu'il comporte:

- un module d'extraction de composantes tonales et d'un signal d'ambiance à partir d'un signal issu du signal bande basse décodé;
- un module de combinaison des composantes tonales et du signal d'ambiance par mixage adaptatif utilisant des facteurs de contrôle de niveau d'énergie pour obtenir un signal audio, dit signal combiné;
- un module d'extension sur au moins une deuxième bande de fréquence supérieure à la première bande de fréquence mis en oeuvre sur le signal décodé bande basse avant le module d'extraction ou sur le signal combiné après le module de combinaison.

[0035] Ce dispositif présente les mêmes avantages que le procédé décrit précédemment, qu'il met en oeuvre.

[0036] L'invention vise un décodeur comportant un dispositif tel que décrit.

[0037] Elle vise un programme informatique comportant des instructions de code pour la mise en oeuvre des étapes du procédé d'extension de bande tel que décrit, lorsque ces instructions sont exécutées par un processeur.

[0038] Enfin l'invention se rapporte à un support de stockage, lisible par un processeur, intégré ou non au dispositif d'extension de bande, éventuellement amovible, mémorisant un programme informatique mettant en oeuvre un procédé d'extension de bande tel que décrit précédemment.

[0039] D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description suivante, donnée uniquement à titre d'exemple non limitatif, et faite en référence aux dessins annexés, sur lesquels :

- la figure 1 illustre une partie d'un décodeur de type AMR-WB mettant en oeuvre des étapes d'extension de bande de fréquence de l'état de l'art et tel que décrit précédemment;
- la figure 2 illustre un décodeur de type interopérable G.718-LD à 16kHz selon l'état de l'art et tel que décrit précé-

demment;

- la figure 3 illustre un décodeur interopérable avec le codage AMR-WB et intégrant un dispositif d'extension de bande selon un mode de réalisation de l'invention;
- la figure 4 illustre sous forme d'organigramme, les étapes principales d'un procédé d'extension de bande selon un mode de réalisation de l'invention;
- la figure 5 illustre un mode de réalisation dans le domaine fréquentiel d'un dispositif d'extension de bande selon l'invention intégré dans un décodeur; et
- la figure 6 illustre une réalisation matérielle d'un dispositif d'extension de bande selon l'invention.

[0040] La figure 3 illustre un exemple de décodeur, compatible avec la norme AMR-WB/G.722.2 dans lequel on retrouve un post-traitement similaire à celui introduit dans G.718 et décrit en référence à la figure 2 et une extension de bande améliorée selon le procédé d'extension de l'invention, mis en oeuvre par le dispositif d'extension de bande illustré par le bloc 309.

[0041] Contrairement au décodage AMR-WB qui fonctionne avec une fréquence d'échantillonnage de sortie de 16 kHz et au décodage G.718 qui fonctionne à 8 ou 16 kHz, on considère ici un décodeur qui peut fonctionner avec un signal de sortie (synthèse) à la fréquence $f_s = 8, 16, 32$ ou 48 kHz. A noter qu'on suppose ici que le codage a été effectué selon l'algorithme AMR-WB avec une fréquence interne de 12.8 kHz pour le codage CELP en bande basse et à 23.85 kbit/s un codage de gain par sous-trame à la fréquence de 16 kHz, mais des variantes interopérables du codeur AMR-WB sont également possibles ; même si l'invention est décrite ici au niveau du décodage, on suppose ici que le codage peut aussi fonctionner avec un signal d'entrée à la fréquence $f_s = 8, 16, 32$ ou 48 kHz et des opérations de ré-échantillonnage adéquates, dépassant le cadre de l'invention, sont mises en oeuvre au codage en fonction de la valeur de f_s . On peut noter que quand $f_s=8$ kHz au décodeur, dans le cas d'un décodage compatible avec AMR-WB, il n'est pas nécessaire d'étendre la bande basse 0-6.4 kHz, car la bande audio reconstruite à la fréquence f_s est limitée à 0-4000 Hz.

[0042] A la figure 3, le décodage CELP (BF pour basses fréquences) fonctionne toujours à la fréquence interne de 12.8 kHz, comme dans AMR-WB et G.718, et l'extension de bande (HF pour hautes fréquences) faisant l'objet de l'invention fonctionne à la fréquence de 16 kHz, les synthèses BF et HF sont combinées (bloc 312) à la fréquence f_s après ré-échantillonnage adéquat (blocs 307 et 311). Dans des variantes de l'invention, la combinaison des bandes basse et haute pourra se faire à 16 kHz, après avoir ré-échantillonné la bande basse de 12.8 à 16 kHz, avant de ré-échantillonner le signal combiné à la fréquence f_s .

[0043] Le décodage selon la figure 3 dépend du mode (ou débit) AMR-WB associé à la trame courante reçue. A titre indicatif et sans que cela impacte le bloc 309, le décodage de la partie CELP en bande basse comporte les étapes suivantes:

- Démultiplexage des paramètres codés (bloc 300) en cas de trame correctement reçue ($bfi=0$ où bfi est le « *bad frame indicator* » valant 0 pour une trame reçue et 1 pour une trame perdue).
- Décodage des paramètres ISF avec interpolation et conversion en coefficients LPC (bloc 301) comme décrit dans la clause 6.1 de la norme G.722.2.
- Décodage de l'excitation CELP (bloc 302), avec une partie adaptative et fixe pour reconstruire l'excitation (exc ou $u'(n)$) dans chaque sous-trame de longueur 64 à 12.8 kHz:

$$u'(n) = \hat{g}_p v(n) + \hat{g}_c c(n) , n = 0, \dots, 63$$

en suivant les notations de la clause 7.1.2.1 de G.718 concernant le décodage CELP, où $v(n)$ et $c(n)$ sont respectivement les mots de code des dictionnaires adaptatif et fixe, et \hat{g}_p et \hat{g}_c sont les gains décodés associés. Cette excitation $u'(n)$ est utilisée dans le dictionnaire adaptatif de la sous-trame suivante ; elle est ensuite post-traitée et on distingue comme dans G.718 l'excitation $u'(n)$ (aussi notée exc) de sa version post-traitée modifiée $u(n)$ (aussi notée $exc2$) qui sert d'entrée au filtre de synthèse, $1/\hat{A}(z)$, dans le bloc 303. Dans des variantes qui peuvent être mises en oeuvre pour l'invention, les post-traitements appliqués à l'excitation peuvent être modifiés (par exemple, la dispersion de phase peut être améliorée) ou ces post-traitements peuvent être étendus (par exemple, une réduction du bruit inter-harmonique peut être mise en oeuvre), sans affecter la nature du procédé d'extension de bande selon l'invention.

- Filtrage de synthèse par $1/\hat{A}(z)$ (bloc 303) où le filtre LPC décodé $\hat{A}(z)$ est d'ordre 16
- Post-traitement bande étroite (bloc 304) selon la clause 7.3 de G.718 si $f_s=8$ kHz.
- Désaccentuation (bloc 305) par le filtre $1/(1-0.68z^{-1})$
- Post-traitement des basses fréquences (bloc 306) tel que décrit à la clause 7.14.1.1 de G.718. Ce traitement introduit un retard qui est pris en compte dans le décodage de la bande haute (>6.4 kHz).
- Ré-échantillonnage de la fréquence interne de 12.8 kHz à la fréquence de sortie f_s (bloc 307). Plusieurs réalisations

sont possibles. Sans perte de généralité, on considère ici à titre d'exemple que si $f_s=8$ ou 16 kHz, le ré-échantillonnage décrit dans la clause 7.6 de G.718 est repris ici, et si $f_s=32$ ou 48 kHz, des filtres à réponse impulsionnelle finie (FIR) supplémentaires sont utilisés.

- Calcul des paramètres du "noise gate" (bloc 308) qui est réalisé de façon préférentielle comme décrit dans la clause 7.14.3 de G.718.

Dans des variantes qui peuvent être mises en oeuvre pour l'invention, les post-traitements appliqués à l'excitation peuvent être modifiés (par exemple, la dispersion de phase peut être améliorée) ou ces post-traitements peuvent être étendus (par exemple, une réduction du bruit inter-harmonique peut être mise en oeuvre), sans affecter la nature de l'extension de bande. On ne décrit pas ici le cas du décodage de la bande basse lorsque la trame courante est perdue ($bfi=1$) qui est informatif dans la norme 3GPP AMR-WB ; en général, qu'il s'agisse du décodeur AMR-WB ou d'un décodeur général s'appuyant sur le modèle source-filtre, il s'agit typiquement d'estimer au mieux l'excitation LPC et les coefficients du filtre LPC de synthèse afin de reconstituer le signal perdu en gardant le modèle source-filtre. Lorsque $bfi=1$ on considère ici que l'extension de bande (bloc 309) peut fonctionner comme dans le cas $bfi=0$ et un débit <23.85 kbit/s ; ainsi, la description de l'invention supposera par la suite et sans perte de généralité que $bfi=0$.

On peut noter que l'utilisation des blocs 306, 308, 314 est optionnelle.

On notera également que le décodage de la bande basse décrit ci-dessus suppose une trame courante dite « active » avec un débit entre 6.6 et 23.85 kbit/s. En fait, quand le mode DTX (transmission continue en français) est activé, certaines trames peuvent être codées comme « inactives » et dans ce cas on peut soit transmettre un descripteur de silence (sur 35 bits) soit ne rien transmettre. En particulier, on rappelle que la trame SID du codeur AMR-WB décrit plusieurs paramètres : paramètres ISF moyennés sur 8 trames, énergie moyenne sur 8 trames, "flag de dithering" pour la reconstruction de bruit non stationnaire. Dans tous les cas, au décodeur, on retrouve le même modèle de décodage que pour une trame active, avec une reconstruction de l'excitation et d'un filtre LPC pour la trame courante, ce qui permet d'appliquer l'invention même sur des trames inactives. Le même constat s'applique pour le décodage de « trames perdues » (ou FEC, PLC) dans lequel le modèle LPC est appliqué.

[0044] Cet exemple de décodeur fonctionne dans le domaine de l'excitation et comporte donc une étape de décodage du signal d'excitation bande basse. Le dispositif d'extension de bande et le procédé d'extension de bande au sens de l'invention fonctionne également dans un domaine différent du domaine de l'excitation et notamment avec un signal direct décodé en bande basse ou un signal pondéré par un filtre perceptuel.

[0045] Contrairement au décodage AMR-WB ou G.718, le décodeur décrit permet d'étendre la bande basse décodée (50-6400 Hz en tenant en compte du filtrage passe-haut à 50 Hz au décodeur, 0-6400 Hz dans le cas général) à une bande étendue dont la largeur varie, allant approximativement de 50-6900 Hz à 50-7700 Hz en fonction du mode mis en oeuvre dans la trame courante. On peut ainsi parler d'une première bande de fréquence de 0 à 6400Hz et d'une deuxième bande de fréquence de 6400 à 8000Hz. En réalité, dans le mode de réalisation privilégié, l'excitation pour les hautes fréquences est générée dans le domaine fréquentiel dans une bande de 5000 à 8000 Hz, pour permettre un filtrage passe-bande de largeur 6000 à 6900 ou 7700 Hz dont la pente n'est pas trop raide dans la bande supérieure rejetée.

[0046] La partie synthèse bande haute est réalisée dans le bloc 309 représentant le dispositif d'extension de bande selon l'invention et qui est détaillé à la figure 5 dans un mode de réalisation.

[0047] Afin d'aligner les bandes basses et hautes décodées, un retard (bloc 310) est introduit pour synchroniser les sorties des blocs 306 et 309 et la bande haute synthétisée à 16 kHz est ré-échantillonnée de 16 kHz à la fréquence f_s (sortie de bloc 311). La valeur du retard T devra être adaptée pour les autres cas ($f_s=32, 48$ kHz) en fonction des traitements mis en oeuvre. On rappelle que quand $f_s=8$ kHz, il n'est pas nécessaire d'appliquer les blocs 309 à 311 car la bande du signal en sortie du décodeur est limitée à 0-4000 Hz.

[0048] A noter que le procédé d'extension de l'invention mis en oeuvre dans le bloc 309 selon le premier mode de réalisation n'introduit de façon préférentielle aucun retard supplémentaire par rapport à la bande basse reconstruite à 12.8 kHz ; cependant, dans des variantes de l'invention (par exemple en utilisant une transformation temps/fréquence avec recouvrement), un retard pourra être introduit. Ainsi, de façon générale la valeur de T dans le bloc 310 devra être ajustée en fonction de la mise en oeuvre spécifique. Par exemple dans le cas où le post-traitement des basses fréquences (bloc 306) n'est pas utilisé, le retard à introduire pour $f_s=16$ kHz pourra être fixé à $T=15$.

[0049] Les bandes basse et haute sont ensuite combinées (ajoutées) dans le bloc 312 et la synthèse obtenue est post-traitée par filtrage passe-haut à 50 Hz (de type IIR) d'ordre 2 dont les coefficients dépendent de la fréquence f_s (bloc 313) et post-traitement de sortie avec application optionnelle du "noise gate" de façon similaire à G.718 (bloc 314).

[0050] Le dispositif d'extension de bande selon l'invention, illustré par le bloc 309 selon le mode de réalisation du décodeur de la figure 5, met en oeuvre un procédé d'extension de bande (au sens large) décrit maintenant en référence à la figure 4.

[0051] Ce dispositif d'extension peut également être indépendant du décodeur et peut mettre en oeuvre le procédé décrit à la figure 4 pour effectuer une extension de bande d'un signal audio existant stocké ou transmis au dispositif, avec une analyse du signal audio pour en extraire par exemple une excitation et un filtre LPC.

Ce dispositif reçoit en entrée un signal décodé dans une première bande de fréquence dite bande basse $u(n)$ qui peut être dans le domaine de l'excitation ou dans celui du signal. Dans le mode de réalisation décrit ici, une étape de décomposition en sous-bandes (E401b) par transformée temps fréquence ou banc de filtres est appliquée au signal décodé bande basse pour obtenir le spectre du signal décodé bande basse $U(k)$ pour une mise en oeuvre dans le

5 domaine fréquentiel.
[0052] Une étape E401a d'extension du signal décodé bande basse dans une deuxième bande de fréquence supérieure à la première bande de fréquence, pour obtenir un signal décodé bande basse étendu $U_{HB1}(k)$, peut être effectuée sur ce signal décodé bande basse avant ou après l'étape d'analyse (décomposition en sous-bandes). Cette étape d'extension peut comporter à la fois une étape de ré-échantillonnage et une étape d'extension ou simplement une étape de translation ou transposition fréquentielle en fonction du signal obtenu en entrée. On notera que dans des variantes, l'étape E401a pourra être effectuée à la fin du traitement décrit à la figure 4,, c'est-à-dire sur le signal combiné, ce traitement étant alors principalement réalisé sur le signal bande basse avant extension, le résultat étant équivalent.

[0053] Cette étape est détaillée ultérieurement dans le mode de réalisation décrit en référence à la figure 5.

10 **[0054]** Une étape E402 d'extraction d'un signal d'ambiance ($U_{HBA}(k)$) et de composantes tonales ($y(k)$) est effectuée à partir du signal bande basse décodé ($U(k)$) ou décodé et étendu ($U_{HB1}(k)$). On définit ici l'ambiance comme le signal résiduel qui est obtenu en supprimant dans le signal existant les harmoniques (ou composantes tonales) principales (ou dominantes).

[0055] Dans la plupart des signaux en bande élargie (échantillonnés à 16 kHz), la bande haute (>6 kHz) contient une information d'ambiance qui est en général similaire à celle présente dans la bande basse.

20 **[0056]** L'étape d'extraction des composantes tonales et du signal d'ambiance comprend par exemple les étapes suivantes:

- détection des composantes tonales dominantes du signal bande basse décodé (ou décodé et étendu), dans le domaine fréquentiel; et
- 25 - calcul d'un signal résiduel par extraction des composantes tonales dominantes pour obtenir le signal d'ambiance.

Cette étape peut également être obtenue par:

- obtention du signal d'ambiance par calcul d'une moyenne du signal bande basse décodé (ou décodé et étendu); et
- 30 - obtention des composantes tonales par soustraction du signal d'ambiance calculé au signal bande basse décodé (ou décodé et étendu).

[0057] Les composantes tonales et le signal d'ambiance sont ensuite combinés de façon adaptative à l'aide de facteurs de contrôle de niveau d'énergie à l'étape E403 pour obtenir un signal dit combiné ($U_{HB2}(k)$). L'étape d'extension E401a peut alors être mise en oeuvre si elle n'a pas été déjà effectuée sur le signal bande basse décodé.

[0058] Ainsi, la combinaison de ces deux types de signaux permet d'obtenir un signal combiné avec des caractéristiques plus adaptées à certains types de signaux comme des signaux musicaux et plus riche en contenu fréquentiel et dans la bande de fréquence étendue correspondant à toute la bande de fréquence incluant la première et la deuxième bande de fréquence.

40 **[0059]** L'extension de bande selon le procédé améliore la qualité pour ce type de signaux par rapport à l'extension décrite dans la norme AMR-WB.

[0060] Le fait d'utiliser une combinaison de signal d'ambiance et de composantes tonales permet d'enrichir ce signal d'extension pour le rendre plus proche des caractéristiques du vrai signal et non pas d'un signal artificiel.

[0061] Cette étape de combinaison sera détaillée ultérieurement en référence à la figure 5.

45 **[0062]** Une étape de synthèse, qui correspond à l'analyse en 401b, est effectuée en E404b pour ramener le signal dans le domaine temporel.

[0063] De façon optionnelle, une étape d'ajustement de niveau d'énergie du signal bande haute peut être effectuée en E404a, avant et/ou après l'étape de synthèse, par application d'un gain et/ou par filtrage adéquat. Cette étape sera expliquée plus en détails dans le mode de réalisation décrit à la figure 5 pour les blocs 501 à 507.

50 **[0064]** Dans un exemple de réalisation, le dispositif d'extension de bande 500 est décrit maintenant en référence à la **figure 5** illustrant à la fois ce dispositif mais aussi des modules de traitement adaptés à la mise en oeuvre dans un décodeur de type interopérable avec un codage AMR-WB. Ce dispositif 500 met en oeuvre le procédé d'extension de bande décrit précédemment en référence à la figure 4.

[0065] Ainsi, le bloc de traitement 510 reçoit un signal bande basse décodé ($u(n)$). Dans un mode de réalisation particulier, l'extension de bande utilise l'excitation décodée à 12.8 kHz (exc2 ou $u(n)$) en sortie du bloc 302 de la figure 3.

55 **[0066]** Ce signal est décomposé en sous-bandes de fréquence par le module de décomposition en sous-bandes 510 (qui met en oeuvre l'étape E401b de la figure 4) qui réalise en général une transformée ou applique un banc de filtres, pour obtenir une décomposition en sous-bandes $U(k)$ du signal $u(n)$.

[0067] Dans un mode de réalisation particulier, une transformée de type DCT-IV (pour "*Discrete Cosine Transform*" - Type IV en anglais) (bloc 510) est appliquée sur la trame courante de 20 ms (256 échantillons), sans fenêtrage, ce qui revient à transformer directement $u(n)$ avec $n = 0, \dots, 255$ selon la formule suivante :

5

$$U(k) = \sum_{n=0}^{N-1} u(n) \cos \left(\frac{\pi}{N} \left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right)$$

10 où $N = 256$ et $k = 0, \dots, 255$.

[0068] Une transformation sans fenêtrage (ou de façon équivalente avec une fenêtre rectangulaire implicite de la longueur de la trame) est possible lorsque le traitement est effectué dans le domaine de l'excitation, et non le domaine du signal. Dans ce cas aucun artefact (effets de bloc) n'est audible, ce qui constitue un avantage important de ce mode de réalisation de l'invention.

15 **[0069]** Dans ce mode de réalisation, la transformation DCT-IV est mise en oeuvre par FFT suivant l'algorithme dit « *Evolved DCT* (EDCT) » décrit dans l'article de D.M. Zhang, H.T. Li, A Low Complexity Transform - Evolved DCT, IEEE 14th International Conference on Computational Science and Engineering (CSE), Aug. 2011, pp. 144-149, et mis en oeuvre dans les normes UIT-T G.718 Annex B et G.729.1 Annex E.

20 **[0070]** Dans des variantes de l'invention et sans perte de généralité, la transformation DCT-IV pourra être remplacée par d'autres transformations temps-fréquences court-terme de même longueur et dans le domaine de l'excitation ou dans le domaine du signal, comme une FFT (pour "*Fast Fourier Transform*" en anglais) ou une DCT-II (*Discrete Cosine Transform* - Type II). De façon alternative, on pourra remplacer la DCT-IV sur la trame par une transformation avec recouvrement-addition et fenêtrage de longueur supérieure à la longueur de la trame courante, par exemple en utilisant une MDCT (pour "*Modified Discrete Cosine Transform*" en anglais). Dans ce cas le retard T dans le bloc 310 de la figure 3, devra être ajusté (réduit) de façon adéquate en fonction du retard additionnel dû à l'analyse/synthèse par cette transformée.

25 **[0071]** Dans un autre mode de réalisation, la décomposition en sous-bandes est effectuée par l'application d'un banc de filtres, par exemple de type PQMF (Pseudo-QMF) réels ou complexes. Pour certains bancs de filtres, on obtient, pour chaque sous-bande dans une trame donnée, non pas une valeur spectrale mais une série de valeurs temporelles associée à la sous-bande ; dans ce cas, le mode de réalisation privilégié dans l'invention peut être appliqué en réalisant par exemple une transformée de chaque sous-bande et en calculant le signal d'ambiance dans le domaine des valeurs absolues, les composantes tonales étant toujours obtenues par différence entre le signal (en valeur absolue) et le signal d'ambiance. Dans le cas d'un banc de filtre complexe, le module complexe des échantillons remplacera la valeur absolue.

30 **[0072]** Dans d'autres modes de réalisation, l'invention sera appliqué dans un système utilisant deux sous-bandes, la bande basse étant analysé par transformée ou par banc de filtres.

35 **[0073]** Dans le cas d'une DCT, le spectre DCT, $U(k)$, de 256 échantillons couvrant la bande 0-6400 Hz (à 12.8 kHz), est ensuite étendu (bloc 511) en un spectre de 320 échantillons couvrant la bande 0-8000 Hz (à 16 kHz) sous la forme suivante :

40

$$U_{HB1}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ U(k) & k = 200, \dots, 239 \\ U(k + start_band - 240) & k = 240, \dots, 319 \end{cases}$$

45

où on prend de façon préférentielle $start_band = 160$.

[0074] Le bloc 511 met en oeuvre l'étape E401a de la figure 4, c'est-à-dire l'extension du signal décodé bande basse. Cette étape peut également comporter un ré-échantillonnage de 12.8 à 16 kHz dans le domaine fréquentiel, en rajoutant $\frac{1}{4}$ d'échantillons ($k = 240, \dots, 319$) au spectre, le ratio entre 16 et 12.8 étant de 5/4.

50 **[0075]** Dans la bande de fréquence correspondant aux échantillons allant des indices 200 à 239, le spectre original est conservé, pour pouvoir y appliquer une réponse d'atténuation progressive du filtre passe-haut dans cette bande de fréquence et aussi pour ne pas introduire de défauts audibles lors de l'étape d'addition de la synthèse basse fréquence à la synthèse haute fréquence.

55 **[0076]** On notera que dans ce mode de réalisation, la génération du spectre étendu sur-échantillonné s'effectue dans une bande de fréquence allant de 5 à 8 kHz incluant donc une deuxième bande de fréquence (6.4-8kHz) supérieure à la première bande de fréquence (0-6.4 kHz).

[0077] Ainsi, l'extension du signal bande basse décodé s'effectue au moins sur la deuxième bande de fréquence mais aussi sur une partie de la première bande de fréquence.

[0078] Bien évidemment, les valeurs définissant ces bandes de fréquences peuvent être différentes selon le décodeur ou le dispositif de traitement dans lequel l'invention s'applique.

[0079] De plus, le bloc 511 réalise un filtrage passe-haut implicite dans la bande 0-5000 Hz puisque les 200 premiers échantillons de $U_{HB1}(k)$ sont mis à zéro ; comme expliqué plus tard, ce filtrage passe-haut peut également être complété par une partie d'atténuation progressive des valeurs spectrales d'indices $k = 200, \dots, 255$ dans la bande 5000-6400 Hz, cette atténuation progressive est mise en oeuvre dans le bloc 501 mais pourrait être réalisée séparément en dehors du bloc 501. De façon équivalente et dans des variantes de l'invention, la mise en oeuvre du filtrage passe-haut séparée en blocs de coefficients d'indice $k = 0, \dots, 199$ mis à zéro, de coefficients $k = 200, \dots, 255$ atténués, dans le domaine transformé, pourra donc être effectué en une seule étape.

[0080] Dans cet exemple de réalisation et selon la définition de $U_{HB1}(k)$, on remarque que la bande 5000-6000 Hz de $U_{HB1}(k)$ (qui correspond aux indices $k = 200, \dots, 239$) est copiée à partir de la bande 5000-6000 Hz de $U(k)$. Cette approche permet de conserver le spectre original dans cette bande et elle évite d'introduire des distorsions dans la bande 5000-6000 Hz lors de l'addition de la synthèse HF avec la synthèse BF - en particulier la phase du signal (implicitement représentée dans le domaine DCT-IV) dans cette bande est préservée.

[0081] La bande 6000-8000 Hz de $U_{HB1}(k)$ est ici définie en copiant la bande 4000-6000 Hz de $U(k)$ puisque la valeur de *start_band* est fixée préférentiellement à 160.

[0082] Dans une variante du mode de réalisation, la valeur de *start_band* pourra être rendue adaptative autour de la valeur de 160, sans modifier la nature de l'invention. Les détails de l'adaptation de la valeur *start_band* ne sont pas décrits ici car ils dépassent le cadre de l'invention sans en changer la portée.

[0083] Dans la plupart des signaux en bande élargie (échantillonnés à 16 kHz), la bande haute (>6 kHz) contient une information d'ambiance qui est naturellement similaire à celle présente dans la bande basse. On définit ici l'ambiance comme le signal résiduel qui est obtenu en supprimant dans le signal existant les harmoniques principales (ou dominantes). Le niveau d'harmonicité dans la bande 6000-8000 Hz est généralement corrélé à celui des bandes de fréquences inférieures.

[0084] Ce signal bande basse décodé et étendu est fourni en entrée du dispositif d'extension 500 et notamment en entrée du module 512. Ainsi le bloc 512 d'extraction de composantes tonales et d'un signal d'ambiance, met en oeuvre l'étape E402 de la figure 4 dans le domaine fréquentiel. Le signal d'ambiance, $U_{HBA}(k)$ pour $k = 240, \dots, 319$ (80 échantillons) est ainsi obtenu pour une deuxième bande de fréquence dite haute fréquence afin de le combiner ensuite de façon adaptative aux composantes tonales extraites $y(k)$, dans le bloc de combinaison 513.

[0085] Dans un mode de réalisation particulier, l'extraction des composantes tonales et du signal d'ambiance (dans la bande 6000-8000 Hz) est effectuée selon opérations suivantes:

- Calcul de l'énergie totale du signal bande basse décodé étendu $ener_{HB}$:

$$ener_{HB} = \sum_{k=240}^{319} U_{HB1}(k)^2 + \varepsilon$$

où $\varepsilon=0.1$ (cette valeur peut être différente, elle est fixée ici à titre d'exemple).

- Calcul de l'ambiance (en valeur absolue) qui correspond ici au niveau moyen du spectre $lev(i)$ (raie par raie) et calcul de l'énergie $ener_{tonal}$ des parties tonales dominantes (dans le spectre hautes fréquences)

Pour $i = 0 \dots L-1$, on obtient ce niveau moyen par l'équation suivante:

$$lev(i) = \frac{1}{fn(i) - fb(i) + 1} \sum_{j=fb(i)}^{fn(i)} |U_{HB1}(j + 240)|$$

Ceci correspond au niveau moyen (en valeur absolue) et représente donc une sorte d'enveloppe du spectre. Dans ce mode de réalisation, $L = 80$ et représente la longueur du spectre et l'indice i de 0 à $L-1$ correspond aux indices $j + 240$ de 240 à 319, soit le spectre de 6 à 8 kHz.

En général $fb(i)=i-7$ et $fn(i)=i+7$, cependant les 7 premiers et derniers indices ($i=0, \dots, 6$ et $i=L-7, \dots, L-1$) demandent un traitement spécial et sans perte de généralité on définit alors :

$$fb(i) = 0 \text{ et } fn(i) = i + 7 \text{ pour } i = 0, \dots, 6$$

$$fb(i) = i - 7 \text{ et } fn(i) = L - 1 \text{ pour } i = L - 7, \dots, L - 1$$

Dans des variantes de l'invention, la moyenne de $|U_{HB1}(j+240)|$, $j = fb(i), \dots, fn(i)$, pourra être remplacée par une valeur médiane sur le même ensemble de valeurs, soit $lev(i) = median_{j=fb(i), \dots, fn(i)}(|U_{HB1}(j+240)|)$. Cette variante a le défaut d'être plus complexe (en termes de nombre de calculs) qu'une moyenne glissante. Dans d'autres variantes une pondération non-uniforme pourra être appliquée aux termes moyennés, ou le filtrage médian pourra être remplacé par exemple par d'autres filtres non linéaires de type « *stack filters* ».

[0086] On calcule également le signal résiduel:

$$y(i) = |U_{HB1}(i + 240)| - lev(i), \quad i = 0, \dots, L - 1$$

qui correspond (approximativement) aux composantes tonales si la valeur $y(i)$ à une raie i donnée est positive ($y(i) > 0$). Ce calcul fait donc intervenir une détection implicite des composantes tonales. Les parties tonales sont donc implicitement détectées à l'aide du terme intermédiaire $y(i)$ représentant un seuil adaptatif. La condition de détection étant $y(i) > 0$. Dans des variantes de l'invention cette condition pourra être changée par exemple en définissant un seuil adaptatif fonction de l'enveloppe locale du signal ou sous la forme $y(i) > lev(i) + xdB$ où x a une valeur prédéfinie (par exemple $x=10$ dB).

L'énergie des parties tonales dominantes est définie par l'équation suivante:

$$ener_{tonal} = \sum_{i=0 \dots 7 | y(i) > 0} y(i)^2$$

[0087] D'autres méthodes d'extraction du signal d'ambiance peuvent bien sûr être envisagées. Par exemple, ce signal d'ambiance peut être extrait d'un signal basse fréquence ou éventuellement une autre bande de fréquence (ou plusieurs bandes de fréquence).

La détection des pics ou composantes tonales pourra être faite différemment.

L'extraction de ce signal d'ambiance pourrait également être fait sur l'excitation décodé mais pas étendu, c'est-à-dire avant l'étape d'extension ou de translation spectrale, c'est-à-dire par exemple sur une portion du signal basse fréquence plutôt que directement sur le signal haute fréquence.

Dans une variante de réalisation, l'extraction des composantes tonales et du signal d'ambiance s'effectue dans un ordre différent et selon les étapes suivantes :

- détection des composantes tonales dominantes du signal bande basse décodé (ou décodé et étendu), dans le domaine fréquentiel ;
- calcul d'un signal résiduel par extraction des composantes tonales dominantes pour obtenir le signal d'ambiance.

Cette variante peut par exemple être réalisée de la façon suivante : Un pic (ou composante tonale) est détecté à une raie d'indice i dans le spectre d'amplitude $|U_{HB1}(i+240)|$ si le critère suivant est vérifié:

$$|U_{HB1}(i + 240)| > |U_{HB1}(i + 240 - 1)| \text{ et } |U_{HB1}(i + 240)| > |U_{HB1}(i + 240 + 1)|,$$

pour $i=0, \dots, L-1$. Dès qu'un pic est détecté à la raie d'indice i on applique un modèle sinusoïdal afin d'estimer les paramètres d'amplitude, de fréquence et éventuellement de phase d'une composante tonale associé à ce pic. Les détails de cette estimation ne sont pas présentés ici mais l'estimation de la fréquence peut typiquement faire appel à une interpolation parabolique sur 3 points afin de localiser le maximum de la parabole approximant les 3 points d'amplitude $|U_{HB1}(i+240)|$ (ramené en dB), l'estimation d'amplitude étant obtenu par le biais de cette même interpolation. Le domaine par transformée utilisé ici (DCT-IV) ne permettant pas d'obtenir directement la phase, on pourra dans un mode de réalisation, négliger ce terme, mais dans des variantes on pourra appliquer une transformée en quadrature de type DST pour estimer un terme de phase. La valeur initiale de $y(i)$ est mise à zéro pour $i=0, \dots, L-1$. Les paramètres sinusoïdaux (fréquence, amplitude, et éventuellement phase) de chaque composante tonale étant estimés, on calcule alors le terme $y(i)$ comme la somme de prototypes (spectres) prédéfinis de sinusoïdes pures transformées dans le domaine DCT-IV (ou autre si une autre décomposition en sous-bandes est utilisée) selon les paramètres sinusoïdaux estimés. Enfin, on applique une valeur absolue aux termes $y(i)$ pour se ramener au domaine du spectre d'amplitude en valeurs absolues. D'autres méthodes de détermination des composantes tonales sont possibles, par exemple il serait également possible de calculer

EP 3 330 966 A1

une enveloppe du signal $env(i)$ par interpolation par splines des valeurs maximales locales (pics détectés) de $|U_{HB1}(i+240)|$, d'abaisser cette enveloppe par un certain niveau en dB pour détecter les composantes tonales comme les pics dépassant cette enveloppe et définir $y(i)$ comme

$$y(i) = \max(|U_{HB1}(i+240)| - env(i), 0)$$

Dans cette variante l'ambiance s'obtient donc par l'équation :

$$lev(i) = |U_{HB1}(i+240)| - y(i), \quad i = 0, \dots, L-1$$

[0088] Dans d'autres variantes de l'invention, la valeur absolue des valeurs spectrales sera remplacée par exemple le carré des valeurs spectrales, sans changer le principe de l'invention ; dans ce cas une racine carrée sera nécessaire pour revenir au domaine du signal, ce qui est plus complexe à réaliser.

[0089] Le module de combinaison 513 effectue une étape de combinaison par mixage adaptatif du signal d'ambiance et des composantes tonales. Pour cela, un facteur Γ de contrôle du niveau d'ambiance est défini par l'équation suivante :

$$\Gamma = \beta \frac{ener_{HB} - ener_{tonal}}{ener_{HB} - \beta ener_{tonal}}$$

β étant un facteur dont un exemple de calcul est donné ci-dessous.

Pour obtenir le signal étendu, on obtient d'abord le signal combiné en valeurs absolues pour $i=0 \dots L-1$:

$$y'(i) = \begin{cases} \Gamma y(i) + \frac{1}{\Gamma} lev(i) & y(i) > 0 \\ y(i) + \frac{1}{\Gamma} lev(i) & y(i) \leq 0 \end{cases}$$

auquel on applique les signes de $U_{HB1}(k)$:

$$y''(i) = \text{sgn}(U_{HB1}(i+240)) \cdot y'(i)$$

où la fonction $\text{sgn}(\cdot)$ donne le signe :

$$\text{sgn}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

Par définition le facteur Γ est >1 . Les composantes tonales, détectées raie par raie par la condition $y(i) > 0$, sont réduites par le facteur Γ ; le niveau moyen est amplifié par le facteur $1/\Gamma$.

Dans le bloc 513 de mixage adaptatif, un facteur de contrôle du niveau d'énergie est calculé en fonction de l'énergie totale du signal bande basse décodé (ou décodé et étendu) et des composantes tonales.

Dans un mode de réalisation préféré du mixage adaptatif, l'ajustement d'énergie est effectué de la façon suivante:

$$U_{HB2}(k) = fac \cdot y''(k-240), \quad k = 240, \dots, 319$$

$U_{HB2}(k)$ étant le signal combiné d'extension de bande.

Le facteur d'ajustement est défini par l'équation suivante:

$$fac = \gamma \sqrt{\frac{ener_{HB}}{\sum_{i=0}^{L-1} y''(i)}}$$

5

Où γ permet d'éviter une sur-estimation de l'énergie. Dans un exemple de réalisation, on calcule β de façon à garder le même niveau de signal d'ambiance par rapport à l'énergie des composantes tonales dans les bandes consécutives du signal. On calcule l'énergie des composantes tonales dans trois bandes : 2000-4000 Hz, 4000-6000 Hz et 6000-8000 Hz, avec

10

$$E_{N2-4} = \sum_{k \in N(80,159)} U^2(k)$$

15

$$E_{N4-6} = \sum_{k \in N(160,239)} U^2(k)$$

20

$$E_{N4-6} = \sum_{k \in N(240,319)} U^2(k)$$

où

25

$$U'(k) = \begin{cases} \sqrt{\frac{\sum_{k=160}^{239} U^2(k)}{\sum_{k=80}^{159} U^2(k)}} U(k) & k = 80, \dots, 159 \\ U(k) & k = 160, \dots, 239 \\ \sqrt{\frac{\sum_{k=160}^{239} U^2(k)}{\sum_{k=240}^{319} U_{HB1}^2(k)}} U_{HB1}(k) & k = 240, \dots, 319 \end{cases}$$

30

35

40

[0090] Et où $N(k_1, k_2)$ est l'ensemble des indices k pour lesquels le coefficient d'indice k est classifié comme étant associé aux composantes tonales. Cet ensemble peut être par exemple obtenu en détectant les pics locaux dans $U'(k)$ vérifiant $|U'(k)| > lev(k)$ ou $lev(k)$ est calculé comme le niveau moyen du spectre raie par raie.

On peut noter que d'autres méthodes de calcul de l'énergie des composantes tonales sont possibles, par exemple en prenant la valeur médiane du spectre sur la bande considérée.

45

On fixe β de telle sorte que le ratio entre l'énergie des composantes tonales dans les bandes 4-6 kHz et 6-8 kHz soit le même qu'entre les bandes 2-4 kHz et 4-6 kHz :

50

$$\beta = \frac{\rho - E_{N6-8}}{\sum_{k=160}^{239} U^2(k) - E_{N6-8}}$$

55

où

$$E_{N4-6} = \max(E_{N4-6}, E_{N2-4}), \quad \rho = \frac{E_{N4-6}^2}{E_{N2-4}}, \quad \rho = \max(\rho, E_{N6-8})$$

5 et $\max(.,.)$ est la fonction qui donne le maximum des deux arguments.
 Dans des variantes de l'invention, le calcul de β pourra être remplacé par d'autres méthodes. Par exemple, dans une variante, on pourra extraire (calculer) différents paramètres (ou « features » en anglais) caractérisant le signal en bande basse, dont un paramètre « tilt » similaire à celui calculé dans le codec AMR-WB, et on estimera le facteur β en fonction d'une régression linéaire à partir de ces différents paramètres en limitant sa valeur entre 0 et 1. La régression linéaire pourra par exemple être estimée de façon supervisée en estimant le facteur β en se donnant la bande haute originale dans une base d'apprentissage. On notera que le mode de calcul de β ne limite pas la nature de l'invention.

10 **[0091]** Ensuite, le paramètre β peut être utilisé pour calculer γ en tenant compte du fait qu'un signal avec un signal d'ambiance rajouté dans une bande donnée est perçu en général comme plus fort qu'un signal harmonique à la même énergie dans la même bande. Si on définit α comme la quantité de signal d'ambiance rajouté au signal harmonique :

$$\alpha = \sqrt{1 - \beta}$$

20 on pourra calculer γ comme une fonction décroissante de α , par exemple $\gamma = b - a\sqrt{\alpha}$, $b=1.1$, $a=1.2$ et γ limité de 0.3 à 1. Là encore d'autres définitions de α et γ sont possibles dans le cadre de l'invention.

[0092] A la sortie du dispositif d'extension de bande 500, le bloc 501, dans un mode de réalisation particulier réalise de façon optionnelle, une double opération d'application de réponse en fréquence de filtre passe-bande et de filtrage de désaccentuation (ou déemphasis) dans le domaine fréquentiel.

25 **[0093]** Dans une variante de l'invention, le filtrage de désaccentuation pourra être réalisé dans le domaine temporel, après le bloc 502 voire avant le bloc 510 ; cependant, dans ce cas, le filtrage passe-bande réalisé dans le bloc 501 peut laisser certaines composantes basses fréquences de très faibles niveaux qui se voient amplifier par désaccentuation, ce qui peut modifier de façon légèrement perceptible la bande basse décodée. Pour cette raison, on préfère ici réaliser la désaccentuation dans le domaine fréquentiel. Dans le mode de réalisation privilégié, les coefficients d'indice $k=0, \dots, 199$

30 sont mis à zéro, ainsi la désaccentuation est limitée aux coefficients supérieurs.
 L'excitation est d'abord désaccentuée selon l'équation suivante :

$$35 \quad U_{HB2}'(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ G_{deemph}(k-200)U_{HB2}(k) & k = 200, \dots, 255 \\ G_{deemph}(55)U_{HB2}(k) & k = 256, \dots, 319 \end{cases}$$

40 où $G_{deemph}(k)$ est la réponse en fréquence du filtre $1/(1-0.68z^{-1})$ sur une bande de fréquence discrète restreinte. En prenant en compte les fréquences discrètes (impaires) de la DCT-IV, on définit ici $G_{deemph}(k)$ comme:

$$45 \quad G_{deemph}(k) = \frac{1}{|e^{j\theta_k} - 0.68|}, \quad k = 0, \dots, 255$$

où

$$50 \quad \theta_k = \frac{256 - 80 + k + \frac{1}{2}}{256}.$$

55 Dans le cas où une autre transformation que la DCT-IV est utilisée, la définition de θ_k pourra être ajustée (par exemple pour des fréquences paires).

On note que la désaccentuation est appliquée en deux phases pour $k = 200, \dots, 255$ correspondant à la bande de fréquence 5000-6400 Hz, où la réponse $1/(1-0.68z^{-1})$ est appliquée comme à 12.8 kHz, et pour $k = 256, \dots, 319$ correspondant à la

EP 3 330 966 A1

bande de fréquence 6400-8000 Hz, où la réponse est étendue de 16 kHz ici à une valeur constante dans la bande 6.4-8 kHz.

[0094] On peut noter que dans le codec AMR-WB la synthèse HF n'est pas désaccentuée. Dans le mode de réalisation présenté ici, le signal haute-fréquence est au contraire désaccentué de manière à le ramener dans un domaine cohérent avec le signal basse-fréquence (0-6.4 kHz) qui sort du bloc 305 de la figure 3. Ceci est important pour l'estimation et l'ajustement ultérieur de l'énergie de la synthèse HF.

[0095] Dans une variante du mode de réalisation, afin de réduire la complexité, on pourra fixer $G_{deemph}(k)$ à une valeur constante indépendante de k , en prenant par exemple $G_{deemph}(k)=0.6$ ce qui correspond approximativement à la valeur moyenne de $G_{deemph}(k)$ pour $k = 200, \dots, 319$ dans les conditions du mode de réalisation décrit ci-dessus.

[0096] Dans une autre variante du mode de réalisation du décodeur, la désaccentuation pourra être réalisée de façon équivalente dans le domaine temporel après DCT inverse.

[0097] En plus de la désaccentuation, un filtrage passe-bande est appliqué avec deux parties séparées : l'une passe-haut fixe, l'autre passe-bas adaptative (fonction du débit).

[0098] Ce filtrage est effectué dans le domaine fréquentiel.

[0099] Dans le mode de réalisation privilégiée, on calcule la réponse partielle de filtre passe-bas dans le domaine fréquentiel comme suit:

$$G_p(k) = 1 - 0.999 \frac{k}{N_{lp} - 1}$$

où $N_{lp}=60$ à 6.6 kbit/s, 40 à 8.85 kbit/s, 20 aux débits >8.85 bit/s. Ensuite on applique un filtre passe-bande sous la forme :

$$U_{HB3}(k) = \begin{cases} 0 & k = 0, \dots, 199 \\ G_{hp}(k - 200)U_{HB2}'(k) & k = 200, \dots, 255 \\ U_{HB2}'(k) & k = 256, \dots, 319 - N_{lp} \\ G_{lp}(k - 320 - N_{lp})U_{HB2}'(k) & k = 320 - N_{lp}, \dots, 319 \end{cases}$$

La définition de $G_{hp}(k)$, $k = 0, \dots, 55$, est donnée par exemple au tableau 1 ci-dessous.

Tableau 1

K	$g_{hp}(k)$	K	$g_{hp}(k)$	K	$g_{hp}(k)$	k	$g_{hp}(k)$
0	0.001622428	14	0.114057967	28	0.403990611	42	0.776551214
1	0.004717458	15	0.128865425	29	0.430149896	43	0.800503267
2	0.008410494	16	0.144662643	30	0.456722014	44	0.823611104
3	0.012747280	17	0.161445005	31	0.483628433	45	0.845788355
4	0.017772424	18	0.179202219	32	0.510787115	46	0.866951597
5	0.023528982	19	0.197918220	33	0.538112915	47	0.887020781
6	0.030058032	20	0.217571104	34	0.565518011	48	0.905919644
7	0.037398264	21	0.238133114	35	0.592912340	49	0.923576092
8	0.045585564	22	0.259570657	36	0.620204057	50	0.939922577
9	0.054652620	23	0.281844373	37	0.647300005	51	0.954896429
10	0.064628539	24	0.304909235	38	0.674106188	52	0.968440179
11	0.075538482	25	0.328714699	39	0.700528260	53	0.980501849
12	0.087403328	26	0.353204886	40	0.726472003	54	0.991035206
13	0.100239356	27	0.378318805	41	0.751843820	55	1.000000000

EP 3 330 966 A1

On notera que dans des variantes de l'invention les valeurs de $G_{hp}(k)$ pourront être modifiées tout en gardant une atténuation progressive. De même le filtrage passe-bas à largeur de bande variable, $G_{lp}(k)$, pourra être ajusté avec des valeurs ou un support fréquentiel différents, sans changer le principe de cette étape de filtrage.

[0100] On notera aussi que le filtrage passe-bande pourra être adapté en définissant une seule étape de filtrage combinant les filtrages passe-haut et passe-bas.

[0101] Dans un autre mode de réalisation, le filtrage passe-bande pourra être réalisé de façon équivalente dans le domaine temporel (comme dans le bloc 112 de la figure 1) avec différents coefficients de filtre selon le débit, après une étape de DCT inverse. Cependant, on notera qu'il est avantageux de réaliser cette étape directement dans le domaine fréquentiel car le filtrage est effectué dans le domaine de l'excitation LPC et donc les problèmes de convolution circulaire et d'effets de bord sont très limités dans ce domaine.

[0102] Le bloc 502 de transformée inverse réalise une DCT inverse sur 320 échantillons pour trouver le signal haute-fréquence échantillonné à 16 kHz. Sa mise en oeuvre est identique au bloc 510, car la DCT-IV est orthonormée, sauf que la longueur de la transformée est de 320 au lieu de 256, et on obtient:

$$u_{HB}(n) = \sum_{k=0}^{N_{16k}-1} U_{HB3}(k) \cos\left(\frac{\pi}{N_{16k}}\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)\right)$$

où $N_{16k} = 320$ et $k = 0, \dots, 319$.

[0103] Dans le cas où le bloc 510 n'est pas une DCT, mais une autre transformation ou décomposition en sous-bandes, le bloc 502 réalise la synthèse correspondante à l'analyse réalisée dans le bloc 510.

[0104] Le signal échantillonné à 16 kHz est ensuite de façon optionnelle mis à l'échelle par des gains définis par sous-trame de 80 échantillons (bloc 504).

Dans un mode de réalisation privilégié, on calcule d'abord (bloc 503) un gain $g_{HB1}(m)$ par sous-trame par des ratios d'énergie des sous-trames tel que dans chaque sous-trame d'indice $m=0, 1, 2$ ou 3 de la trame courante:

$$g_{HB1}(m) = \sqrt{\frac{e_3(m)}{e_2(m)}}$$

où

$$e_1(m) = \sum_{n=0}^{63} u(n + 64m)^2 + \varepsilon$$

$$e_2(m) = \sum_{n=0}^{79} u_{HB}(n + 80m)^2 + \varepsilon$$

$$e_3(m) = e_1(m) \frac{\sum_{n=0}^{319} u_{HB}(n)^2 + \varepsilon}{\sum_{n=0}^{255} u(n)^2 + \varepsilon}$$

avec $\varepsilon = 0.01$. On peut écrire le gain par sous-trame $g_{HB1}(m)$ sous la forme :

$$g_{HB1}(m) = \frac{\frac{\sum_{n=0}^{63} u(n+64m)^2 + \varepsilon}{\sum_{n=0}^{255} u(n)^2 + \varepsilon}}{\frac{\sum_{n=0}^{79} u_{HB}(n+80m)^2 + \varepsilon}{\sum_{n=0}^{319} u_{HB}(n)^2 + \varepsilon}}$$

ce qui montre qu'on assure dans le signal u_{HB} le même ratio entre énergie par sous-trame et énergie par trame que dans le signal $u(n)$.

Le bloc 504 effectue la mise à l'échelle du signal combiné (comprise dans l'étape E404a de la figure 4) selon l'équation suivante:

$$u_{HB}'(n) = g_{HB1}(m)u_{HB}(n), \quad n = 80m, \dots, 80(m+1) - 1$$

[0105] On notera que la réalisation du bloc 503 diffère de celle du bloc 101 de la figure 1, car l'énergie au niveau de la trame courante est prise en compte en plus de celle de la sous-trame. Cela permet d'avoir le ratio de l'énergie de chaque sous-trame par rapport à l'énergie de la trame. On compare donc des ratios d'énergie (ou énergies relatives) plutôt que les énergies absolues entre bande basse et bande haute.

[0106] Ainsi, cette étape de mise à l'échelle permet de conserver dans la bande haute le ratio d'énergie entre la sous-trame et la trame de la même façon que dans la bande basse.

[0107] De façon optionnelle, le bloc 506 effectue ensuite la mise à l'échelle du signal (comprise dans l'étape E404a de la figure 4) selon l'équation suivante:

$$u_{HB}''(n) = g_{HB2}(m)u_{HB}'(n), \quad n = 80m, \dots, 80(m+1) - 1$$

où le gain $g_{HB2}(m)$ est obtenu à partir du bloc 505 en exécutant les blocs 103, 104 et 105 du codec AMR-WB (l'entrée du bloc 103 étant l'excitation décodée en bande basse, $u(n)$). Les blocs 505 et 506 sont utiles pour ajuster le niveau du filtre de synthèse LPC (bloc 507), ici en fonction du tilt du signal. D'autres méthodes de calcul du gain $g_{HB2}(m)$ sont possibles sans changer la nature de l'invention.

[0108] Enfin, le signal, $u_{HB}'(n)$ ou $u_{HB}''(n)$, est filtré par le module de filtrage 507 qui peut être réalisé ici en prenant comme fonction de transfert $1/\hat{A}(z/\gamma)$, où $\gamma=0.9$ à 6.6 kbit/s et $\gamma=0.6$ aux autres débits, ce qui limite l'ordre du filtre à l'ordre 16.

Dans une variante, ce filtrage pourra être réalisé de la même façon que ce qui est décrit pour le bloc 111 de la figure 1 du décodeur AMR-WB, cependant l'ordre du filtre passe à 20 au débit de 6.6, ce qui ne change pas de façon significative la qualité du signal synthétisé. Dans une autre variante, on pourra effectuer le filtrage de synthèse LPC dans le domaine fréquentiel, après avoir calculé la réponse en fréquence du filtre mis en oeuvre dans le bloc 507.

[0109] Dans des variantes de réalisation de l'invention, le codage de la bande basse (0-6.4 kHz) pourra être remplacé par un codeur CELP autre que celui utilisé dans AMR-WB, comme par exemple le codeur CELP dans G.718 à 8 kbit/s. Sans perte de généralité d'autres codeurs en bande élargie ou fonctionnant à des fréquences supérieures à 16 kHz, dans lesquels le codage de la bande basse fonctionne à une fréquence interne à 12.8 kHz pourraient être utilisés. Par ailleurs, l'invention peut être adaptée de façon évidente à d'autres fréquences d'échantillonnage que 12.8 kHz, lorsqu'un codeur de basses fréquences fonctionne à une fréquence d'échantillonnage inférieure à celle du signal original ou reconstruit. Lorsque le décodage en bande basse n'utilise pas de prédiction linéaire, on ne dispose pas d'un signal d'excitation à étendre, dans ce cas on pourra réaliser une analyse LPC du signal reconstruit dans la trame courante et on calculera une excitation LPC de manière à pouvoir appliquer l'invention.

[0110] Enfin, dans une autre variante de l'invention, l'excitation ou le signal bande basse ($u(n)$) est ré-échantillonnée, par exemple par interpolation linéaire ou "spline" cubique, de 12.8 à 16 kHz avant transformation (par exemple DCT-IV) de longueur 320. Cette variante a le défaut d'être plus complexe, car la transformée (DCT-IV) de l'excitation ou du signal est alors calculée sur une plus grande longueur et le ré-échantillonnage n'est pas effectué dans le domaine de la transformée.

[0111] De plus, dans des variantes de l'invention, tous les calculs nécessaires à l'estimation des gains (G_{HBN} , $g_{HB1}(m)$,

$g_{HB2}(m)$, g_{HBN} , ...) pourront être effectués dans un domaine logarithmique.

[0112] La figure 6 représente un exemple de réalisation matérielle d'un dispositif de d'extension de bande 600 selon l'invention. Celui-ci peut faire partie intégrante d'un décodeur de signal audiofréquence ou d'un équipement recevant des signaux audiofréquences décodés ou non.

[0113] Ce type de dispositif comporte un processeur PROC coopérant avec un bloc mémoire BM comportant une mémoire de stockage et/ou de travail MEM.

Un tel dispositif comporte un module d'entrée E apte à recevoir un signal audio décodé ou extrait dans une première bande de fréquence dite bande basse ramené dans le domaine fréquentiel ($U(k)$). Il comporte un module de sortie S apte à transmettre le signal d'extension dans une deuxième bande de fréquence ($U_{HB2}(k)$) par exemple à un module de filtrage 501 de la figure 5.

[0114] Le bloc mémoire peut avantageusement comporter un programme informatique comportant des instructions de code pour la mise en oeuvre des étapes du procédé de d'extension de bande au sens de l'invention, lorsque ces instructions sont exécutées par le processeur PROC, et notamment les étapes d'extraction (E402) de composantes tonales et d'un signal d'ambiance à partir d'un signal issu du signal bande basse décodé ($U(k)$), de combinaison (E403) des composantes tonales ($y(k)$) et du signal d'ambiance ($U_{HBA}(k)$) par mixage adaptatif utilisant des facteurs de contrôle de niveau d'énergie pour obtenir un signal audio, dit signal combiné ($U_{HB2}(k)$), d'extension (E401a) sur au moins une deuxième bande de fréquence supérieure à la première bande de fréquence du signal décodé bande basse avant l'étape d'extraction ou du signal combiné après l'étape de combinaison.

[0115] Typiquement, la description de la figure 4 reprend les étapes d'un algorithme d'un tel programme informatique. Le programme informatique peut également être stocké sur un support mémoire lisible par un lecteur du dispositif ou téléchargeable dans l'espace mémoire de celui-ci.

[0116] La mémoire MEM enregistre de manière générale, toutes les données nécessaires à la mise en oeuvre du procédé.

[0117] Dans un mode possible de réalisation, le dispositif ainsi décrit peut également comporter les fonctions de décodage bande basse et autre fonctions de traitement décrites par exemple en figure 5 et 3 en plus des fonctions d'extension de bande selon l'invention.

Revendications

1. Procédé d'extension de bande de fréquence d'un signal audiofréquence lors d'un processus de décodage ou d'amélioration comportant une étape d'obtention du signal décodé dans une première bande de fréquence dite bande basse, le procédé étant **caractérisé en ce qu'il** comporte les étapes suivantes :

- Extraction (E402) de composantes tonales et d'un signal d'ambiance à partir du signal issu du signal bande basse décodé ;
- Combinaison (E403) des composantes tonales et du signal d'ambiance par mixage adaptatif utilisant des facteurs de contrôle de niveau d'énergie pour obtenir un signal audio, dit combiné ;
- Extension (E401a) sur au moins une deuxième bande de fréquence supérieure à la première bande de fréquence du signal décodé bande basse avant l'étape d'extraction pour former un signal bande basse décodé étendu $U_{HB1}(k)$,

et selon lequel l'étape d'extraction (E402) des composantes tonales et du signal d'ambiance comporte les opérations suivantes :

- (a) calcul de l'énergie tonale du signal de bande basse décodé étendu ;
- (b) calcul de l'ambiance en valeur absolue correspondant au niveau moyen du spectre raie par raie et calcul de l'énergie des parties tonales dominantes dans le spectre hautes fréquences.

2. Procédé selon la revendication 1, selon lequel l'opération (a) de calcul de l'énergie tonale du signal de bande basse décodé étendu comprend le calcul de :

$$ener_{HB} = \sum_{k=240}^{319} U_{HB1}(k)^2 + \varepsilon$$

où $\varepsilon=0.1$.

3. Procédé selon la revendication 1 ou 2, selon lequel le niveau moyen du spectre raie par raie est obtenu par l'équation :

$$lev(i) = \frac{1}{fn(i) - fb(i) + 1} \sum_{j=fb(i)}^{fn(i)} |U_{HB1(j+240)}|$$

où

$$fb(i) = 0 \text{ et } fn(i) = i+7 \text{ pour } i=0, \dots, 6$$

$$fb(i) = i-7 \text{ et } fn(i) = i+7 \text{ pour } i=7, \dots, L-8$$

$$fb(i) = i-7 \text{ et } fn(i) = L-1 \text{ pour } i=L-7, \dots, L-1.$$

4. Procédé selon la revendication 1, 2 ou 3, selon lequel le calcul de l'énergie des composantes tonales dominantes comprend le calcul du signal résiduel :

$$y(i) = |U_{HB1}(i + 240)| - lev(i), i = 0, \dots, L - 1.$$

5. Procédé selon la revendication 4, comprenant une étape de détection des composantes tonales sur la base d'une condition de détection sur le signal résiduel $y(i)$.

6. Procédé selon la revendication 5, selon lequel la condition de détection est $y(i) > 0$.

7. Procédé selon la revendication 6, selon lequel l'énergie des composantes tonales dominantes est définie par

$$ener_{tonal} = \sum_{i=0 \dots 7 | y(i) > 0} y(i)^2$$

8. Dispositif d'extension de bande de fréquence d'un signal audiofréquence le signal ayant été décodé dans une première bande de fréquence dite bande basse, le dispositif étant **caractérisé en ce qu'il** comporte :

- un module d'extraction (512) de composantes tonales et d'un signal d'ambiance à partir d'un signal bande basse décodé ;
- un module de combinaison (513) des composantes tonales et du signal d'ambiance par mixage adaptatif utilisant des facteurs de contrôle de niveau d'énergie pour obtenir un signal audio, dit signal combiné ;

- un module d'extension (511) sur au moins une deuxième bande de fréquence supérieure à la première bande de fréquence mis en oeuvre sur le signal décodé de bande basse avant le module d'extraction pour former un signal bande basse décodé étendu $U_{HB1}(k)$,

et dans lequel le module d'extraction (512) des composantes tonales et du signal d'ambiance est adapté pour effectuer les opérations suivantes :

- (a) calcul de l'énergie tonale du signal de bande basse décodé étendu ;
- (b) calcul de l'ambiance en valeur absolue correspondant au niveau moyen du spectre raie par raie et calcul de l'énergie des parties tonales dominantes dans le spectre hautes fréquences.

9. Décodeur de signal audiofréquence **caractérisé en ce qu'il** comporte un dispositif d'extension de bande de fréquence conforme à la revendication 8.

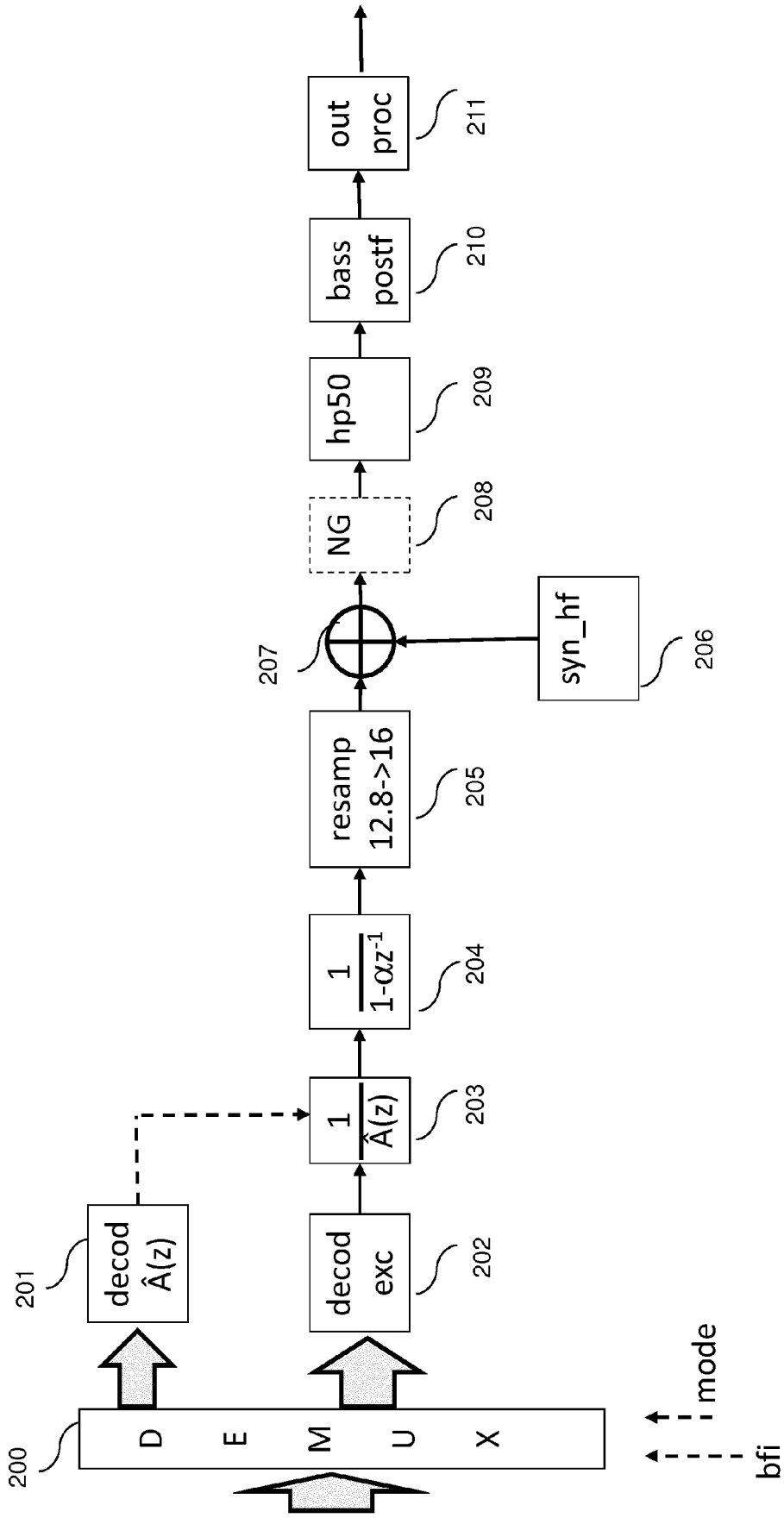


Fig.2

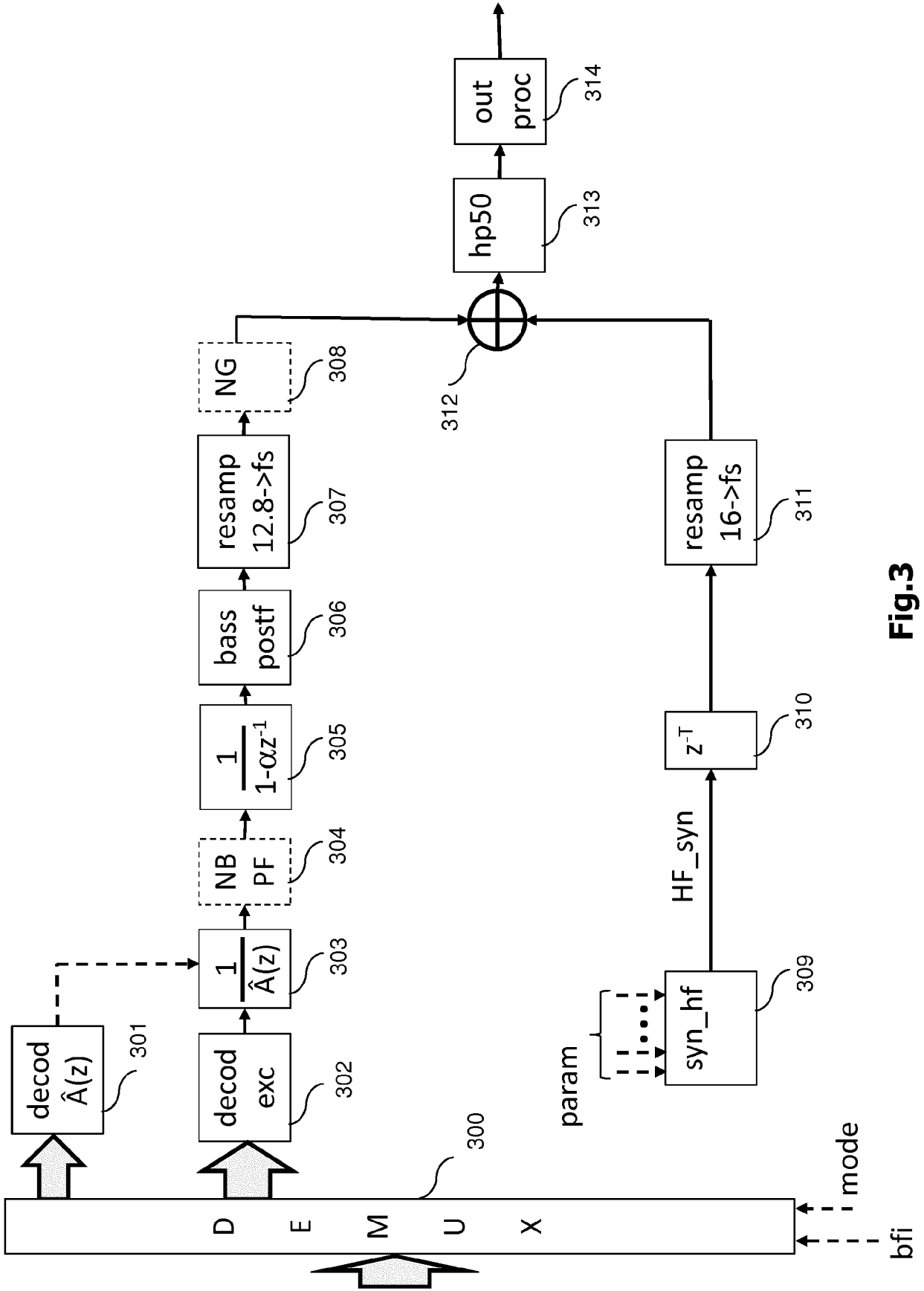


Fig.3

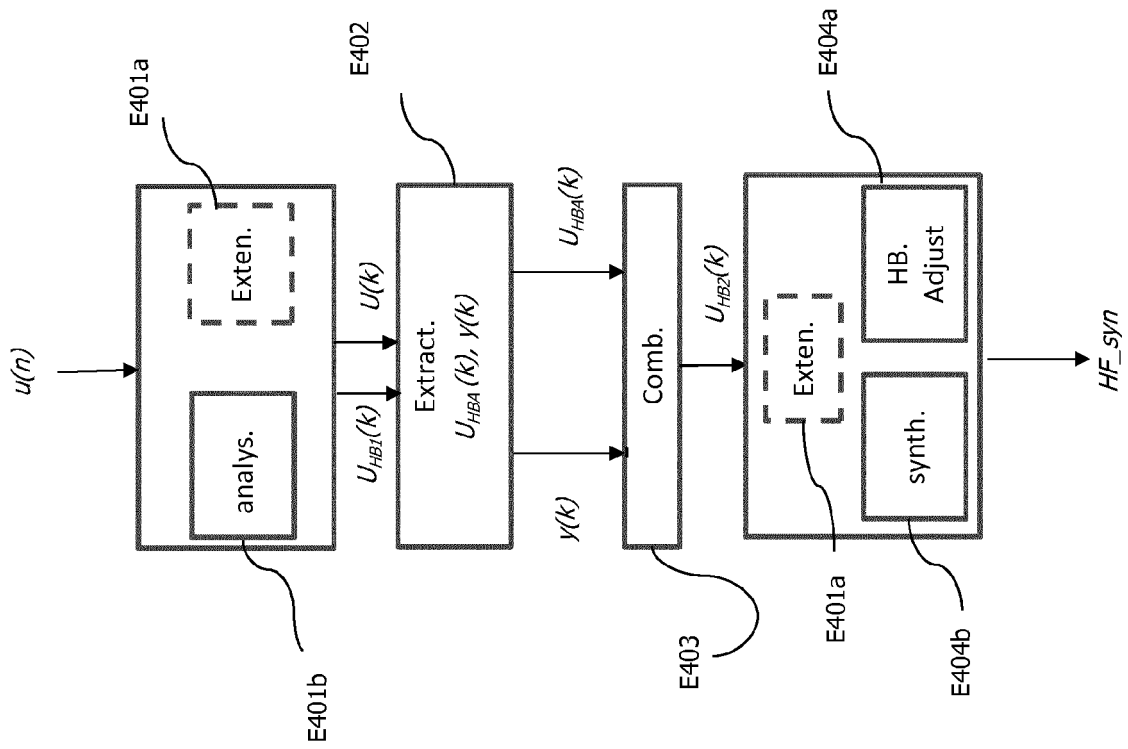


Fig.4

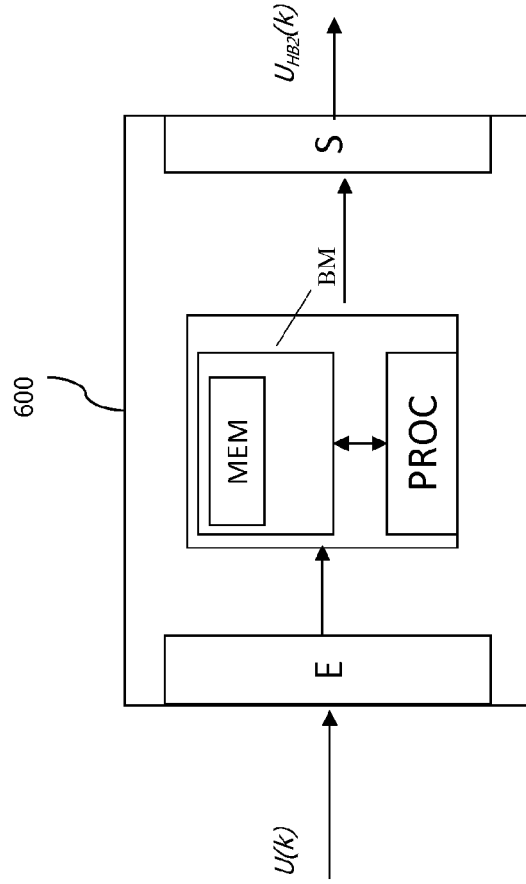


Fig.6

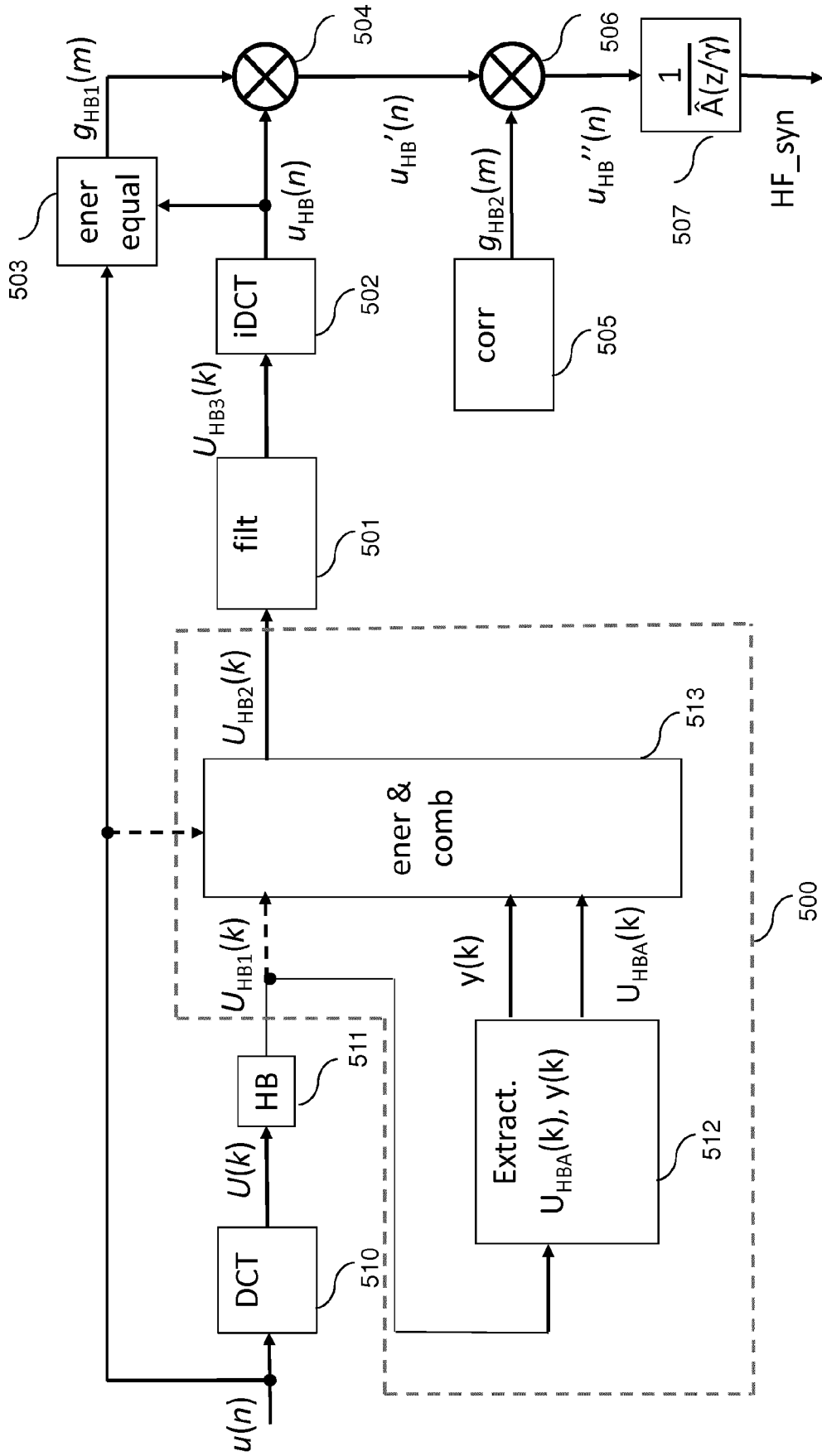


Fig.5



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande
EP 17 20 6563

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
A	ANNADANA RAGHURAM ET AL: "New Enhancements to the Audio Bandwidth Extension Toolkit (ABET)", AES CONVENTION 124; MAY 2008, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 1 mai 2008 (2008-05-01), XP040508704, * section 3; premier paragraphe de la section 6. *	1-9	INV. G10L21/038
A	US 2001/044722 A1 (GUSTAFSSON HARALD [SE] ET AL) 22 novembre 2001 (2001-11-22) * alinéa [0040] * * alinéa [0050] - alinéa [0051]; figure 5 *	1-9	
			DOMAINES TECHNIQUES RECHERCHES (IPC)
			G10L
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche La Haye		Date d'achèvement de la recherche 19 avril 2018	Examineur De Meuleneire, M
CATEGORIE DES DOCUMENTS CITES X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

1
EPO FORM 1503 03 82 (P04C02)

**ANNEXE AU RAPPORT DE RECHERCHE EUROPEENNE
RELATIF A LA DEMANDE DE BREVET EUROPEEN NO.**

EP 17 20 6563

5 La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche européenne visé ci-dessus.
Lesdits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

19-04-2018

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 2001044722 A1	22-11-2001	AT 253766 T	15-11-2003
		AU 3019001 A	07-08-2001
		CN 1397064 A	12-02-2003
		DE 60101148 D1	11-12-2003
		DE 60101148 T2	27-05-2004
		EP 1252621 A1	30-10-2002
		US 2001044722 A1	22-11-2001
		WO 0156021 A1	02-08-2001

EPO FORM P0480

Pour tout renseignement concernant cette annexe : voir Journal Officiel de l'Office européen des brevets, No.12/82

RÉFÉRENCES CITÉES DANS LA DESCRIPTION

Cette liste de références citées par le demandeur vise uniquement à aider le lecteur et ne fait pas partie du document de brevet européen. Même si le plus grand soin a été accordé à sa conception, des erreurs ou des omissions ne peuvent être exclues et l'OEB décline toute responsabilité à cet égard.

Littérature non-brevet citée dans la description

- Speech Coding and Synthesis. Elsevier, 1995 [0005]
- **M. BOSI ; R.E. GOLDBERG.** Introduction to Digital Audio Coding and Standards. Springer, 2002 [0005]
- Handbook of Speech Processing. Springer, 2008 [0005]
- **B. BESSETTE et al.** The adaptive multirate wide-band speech codec (AMR-WB). *IEEE Transactions on Speech and Audio Processing*, 2002, vol. 10 (8), 620-636 [0009]
- **D.M. ZHANG ; H.T. LI.** A Low Complexity Transform - Evolved DCT. *IEEE 14th International Conference on Computational Science and Engineering (CSE)*, Août 2011, 144-149 [0069]