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

(11) **EP 2 168 121 B1**

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention de la délivrance du brevet:
06.06.2018 Bulletin 2018/23

(21) Numéro de dépôt: **08806144.5**

(22) Date de dépôt: **01.07.2008**

(51) Int Cl.:
G10L 19/00 *(2013.01)* ***G10L 19/02*** *(2013.01)*
G06F 17/10 *(2006.01)* ***H03M 7/30*** *(2006.01)*
H04S 3/00 *(2006.01)* ***H04S 3/02*** *(2006.01)*

(86) Numéro de dépôt international:
PCT/FR2008/051220

(87) Numéro de publication internationale:
WO 2009/007639 (15.01.2009 Gazette 2009/03)

(54) **QUANTIFICATION APRES TRANSFORMATION LINEAIRE COMBINANT LES SIGNAUX AUDIO D'UNE SCENE SONORE, CODEUR ASSOCIE**

QUANTIFIZIERUNG NACH LINEARER UMWANDLUNG DURCH KOMBINATION VON AUDIOSIGNALEN EINER KLANGSZENE UND KODIERGERÄT DAFÜR

QUANTIFICATION AFTER LINEAR CONVERSION COMBINING AUDIO SIGNALS OF A SOUND SCENE, AND RELATED ENCODER

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

(30) Priorité: **03.07.2007 FR 0704794**

(43) Date de publication de la demande:
31.03.2010 Bulletin 2010/13

(73) Titulaire: **Orange**
75015 Paris (FR)

(72) Inventeurs:
- **MOUHSSINE, Adil**
  F-35000 Rennes (FR)
- **BENJELLOUN TOUIMI, Abdellatif**
  Ealing
  London W54UX (GB)
- **DUHAMEL, Pierre**
  F-92350 Le Plessis-robinson (FR)

(74) Mandataire: **Cabinet Plasseraud**
66, rue de la Chaussée d'Antin
75440 Paris Cedex 09 (FR)

(56) Documents cités:
- **DERRIEN O & DUHAMEL P: "Une approche statistique pour l'optimisation du MPEG-2/4 AAC (Advanced Audio Coder) en mode stéréophonique matricé (MS stéréo)" ACTES DE COLLOQUES DU GROUPE D'ETUDES DU TRAITEMENT DU SIGNAL ET DES IMAGES (GRETSI), 2003, pages 1-4, XP002464084**
- **DERRIEN O ET AL: "On the interplay between audio compression and spatialization: bit-rate reduction without quality damage" MULTIMEDIA SIGNAL PROCESSING, 2001 IEEE FOURTH WORKSHOP ON OCTOBER 3-5, 2001, PISCATAWAY, NJ, USA,IEEE, 3 octobre 2001 (2001-10-03), pages 313-318, XP010565793 ISBN: 0-7803-7025-2**
- **GERZON M A: "Hierarchical Transmission of Multispeaker Stereo" IEEE APPLICATONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 20 octobre 1991 (1991-10-20), pages 133-134, XP010255223**

Il est rappelé que: Dans un délai de neuf mois à compter de la publication de la mention de la délivrance du brevet européen au Bulletin européen des brevets, toute personne peut faire opposition à ce brevet auprès de l'Office européen des brevets, conformément au règlement d'exécution. L'opposition n'est réputée formée qu'après le paiement de la taxe d'opposition. (Art. 99(1) Convention sur le brevet européen).

EP 2 168 121 B1

Description

[0001] La présente invention concerne les dispositifs de codage de signaux audio, destinés notamment à prendre place dans des applications de transmission ou de stockage de signaux audio numérisés et compressés.

[0002] L'invention est relative plus précisément aux modules de quantification compris dans ces dispositifs de codage audio.

[0003] L'invention concerne plus particulièrement le codage de scène sonore 3D. Une scène sonore 3D, encore appelée son spatialisé, comprend une pluralité de canaux audio correspondant chacun à des signaux monophoniques.

[0004] Une technique de codage de signaux d'une scène sonore utilisée dans le codeur « MPEG Audio Surround » (cf. « Text of ISO/IEC FDIS 23003-1, MPEG Surround », ISO/IEC JTC1 / SC29 / WG11 N8324, July 2006, Klagenfurt, Austria), comprend l'extraction et le codage de paramètres spatiaux à partir de l'ensemble des signaux audio monophoniques sur les différents canaux. Ces signaux sont ensuite mélangés pour obtenir un signal monophonique ou stéréophonique, qui est alors comprimé par un codeur mono ou stéréo classique (par exemple de type MPEG-4 AAC, HE-AAC, etc). Au niveau du décodeur, la synthèse de la scène sonore 3D restituée se fait à partir des paramètres spatiaux et du signal mono ou stéréo décodé.

[0005] Le codage des signaux multicanaux nécessite dans certains cas l'introduction d'une transformation (KLT, Ambiophonique, DCT...) permettant de mieux prendre en compte les interactions qui peuvent exister entre les différents signaux de la scène sonore à coder.

[0006] Il est toujours besoin d'accroitre la qualité audio des scènes sonores restituées après une opération de codage et décodage.

[0007] DERRIEN O & DUHAMEL P: "Une approche statistique pour l'optimisation du MPEG-2/4 AAC (Advanced Audio Coder) en mode stéréophonique matricé (MS stéréo)",ACTES DE COLLOQUES DU GROUPE D'ETUDES DU TRAITEMENT DU SIGNAL ET DES IMAGES (GRETSI), 2003, pages 1-4, divulgue un procédé de quantification de composantes dans un système MS stéréo.

[0008] L'objet de l'invention est de trouver une amélioration pour la quantification dans un système multicanaux. Cet objet est résolu par les revendication indépendantes. Suivant un premier aspect, l'invention propose un procédé de quantification de composantes, certaines au moins de ces composantes étant déterminées chacune en fonction d'une pluralité de signaux audio d'une scène sonore et calculables par application d'une transformation linéaire sur lesdits signaux audio.

[0009] Selon le procédé, on détermine une fonction de quantification à appliquer audites composantes dans une bande de fréquence donnée en testant une condition relative à au moins un signal audio et dépendant au moins d'une comparaison effectuée entre un seuil de masquage psychoacoustique relatif au signal audio dans la bande de fréquence donnée, et une valeur déterminée en fonction de la transformation linéaire inverse et d'erreurs de quantification des composantes par ladite fonction sur la bande de fréquence donnée.

[0010] Un tel procédé permet donc de déterminer une fonction de quantification qui permette de masquer, dans le domaine d'écoute de restitution, le bruit introduit par rapport au signal audio de la scène sonore initiale. La scène sonore restituée après les opérations de codage et décodage présente donc une meilleure qualité audio.

[0011] En effet, l'introduction d'une transformée multicanal (par exemple de type ambiophonique) transforme les signaux réels dans un nouveau domaine différent du domaine d'écoute. La quantification des composantes résultant de cette transformée selon les méthodes de l'état de l'art, basées sur un critère perceptuel (i.e. respectant le seuil de masquage sur ces derniers), ne garantit pas une distorsion minimale sur les signaux réels restitués dans le domaine d'écoute. En effet, le calcul de la fonction de quantification selon l'invention permet de garantir que les bruits de quantification induits sur les signaux réels par la quantification des composantes transformées sont minimaux au sens d'un critère perceptuel. La condition d'une amélioration maximale de la qualité perceptuelle des signaux dans le domaine d'écoute est alors vérifiée.

[0012] Dans un mode de réalisation la condition est relative à plusieurs signaux audio et dépend de plusieurs comparaisons, chaque comparaison étant effectuée entre un seuil de masquage psychoacoustique relatif à un signal audio respectif dans la bande de fréquence donnée, et une valeur déterminée en fonction de la transformation linéaire inverse et d'erreurs de quantification des composantes par ladite fonction.

[0013] Cette disposition accroît encore la qualité audio de la scène sonore restituée.

[0014] Dans un mode de réalisation, la détermination de la fonction de quantification est réitérée lors de l'actualisation des valeurs des composantes à quantifier. Cette disposition permet également d'accroître la qualité audio de la scène sonore restituée, en adaptant la quantification dans le temps en fonction des caractéristiques des signaux.

[0015] Dans un mode de réalisation, on teste la condition relative à un signal audio au moins en comparant le seuil de masquage psychoacoustique relatif au signal audio et un élément représentant la valeur $\sum_{j=1}^{r}(h_{i,j}^2 B_j(s)^{\frac{3}{2}}\mu_{1_{2},j}(s))$,

où s est la bande de fréquence donnée, r est le nombre de composantes, $h_{i,j}$ est le coefficient de la transformée linéaire inverse relatif au signal audio et à la j^{ème} composante avec j=1 à r, $B_j(s)$ représente un paramètre de la fonction de quantification dans la bande s relative à la j^{ème} composante et $\mu_{\frac{1}{2},j}(s)$ est l'espérance mathématique dans la bande s de la racine carrée de la j^{ème} composante.

[0016] Dans un mode de réalisation, on détermine une fonction de quantification à appliquer audites composantes dans la bande de fréquence donnée à l'aide d'un processus itératif générant à chaque itération un paramètre de la fonction de quantification candidat vérifiant la condition et associé à un débit correspondant, l'itération étant stoppée lorsque le débit est inférieur à un seuil donné.

[0017] Une telle disposition permet ainsi de déterminer simplement une fonction de quantification à partir des paramètres déterminés, permettant le masquage du bruit dans le domaine d'écoute de restitution tout en réduisant le débit de codage en dessous d'un seuil donné.

[0018] Dans un mode de réalisation, la transformation linéaire est une transformation ambiophonique.

[0019] Dans un mode de réalisation particulier, la transformation linéaire est une transformation ambiophonique (appelée en anglais « ambisonic »). Cette disposition permet d'une part de réduire le nombre de données à transmettre puisque, en général, les N signaux peuvent être décrits d'une manière très satisfaisante par un nombre de composantes ambiophoniques réduit (par exemple, un nombre égal à 3 ou 5), inférieur à N. Cette disposition permet en outre une adaptabilité du codage à tout type de système de rendu sonore, puisqu'il suffit au niveau du décodeur, d'appliquer une transformée ambiophonique inverse de taille Q'x(2p'+1), (où Q' est égal au nombre de haut-parleurs du système de rendu sonore utilisé en sortie du décodeur et 2p'+1 le nombre de composantes ambiophoniques reçues), pour déterminer les signaux à fournir au système de rendu sonore.

[0020] L'invention peut être mise en oeuvre avec toute transformation linéaire, par exemple la DCT ou encore la transformée KLT (en anglais « Karhunen Loeve Transform ») qui correspond à une décomposition sur des composantes principales dans un espace représentant les statistiques des signaux et permet de distinguer les composantes les plus énergétiques des composantes les moins énergétiques.

[0021] Suivant un deuxième aspect, l'invention propose un module de quantification adapté pour quantifier des composantes, certaines au moins de ces composantes étant déterminées chacune en fonction d'une pluralité de signaux audio d'une scène sonore et calculables par application d'une transformation linéaire sur lesdits signaux audio, ledit module de quantification étant adapté pour mettre en oeuvre les étapes d'un procédé suivant le premier aspect de l'invention.

[0022] Suivant un troisième aspect, l'invention propose un codeur audio adapté pour coder une scène audio comprenant plusieurs signaux respectifs en un flux binaire de sortie, comprenant :

un module de transformation adapté pour calculer par application d'une transformation linéaire sur lesdits signaux audio, des composantes dont certaines au moins sont déterminées chacune en fonction d'une pluralité des signaux audio d'une scène sonore ; et
un module de quantification suivant le deuxième aspect de l'invention adapté pour déterminer au moins une fonction de quantification sur au moins une bande de fréquence donnée et pour quantifier les composantes sur la bande de fréquence donnée en fonction d'au moins la fonction de quantification déterminée ;
le codeur audio étant adapté pour constituer un flux binaire en fonction au moins de données de quantification délivrées par le module de quantification.

[0023] Suivant un quatrième aspect, l'invention propose un programme d'ordinateur à installer dans un module de quantification, ledit programme comprenant des instructions pour mettre en oeuvre les étapes d'un procédé suivant le premier aspect de l'invention lors d'une exécution du programme par des moyens de traitement dudit module.

[0024] Suivant un cinquième aspect, l'invention propose des données de codage, déterminées suite à la mise en oeuvre d'un procédé de quantification suivant le premier aspect de l'invention.

[0025] D'autres caractéristiques et avantages de l'invention apparaîtront encore à la lecture de la description qui va suivre. Celle-ci est purement illustrative et doit être lue en regard des dessins annexés sur lesquels :

- la figure 1 représente un codeur dans un mode de réalisation de l'invention ;
- la figure 2 représente un décodeur dans un mode de réalisation de l'invention ;
- la figure 3 est un organigramme représentant des étapes d'un procédé dans un mode de réalisation de l'invention.

[0026] La figure 1 représente un codeur audio 1 dans un mode de réalisation de l'invention. Il s'appuie sur la technologie des codeurs audio perceptuels, par exemple de type MPEG-4 AAC.

[0027] Le codeur 1 comprend un module 2 de transformation temps/fréquence, un module 3 de transformation linéaire,

un module 4 de quantification, un module 5 de codage entropique de Huffman et un module 6 de calcul de courbe de masquage, en vue de la transmission d'un flux binaire Φ représentant les signaux fournis en entrée du codeur 1.

[0028] Une scène sonore 3D comprend N canaux sur chacun un signal audio respectif S_1, ..., S_N est délivré.

[0029] La figure 2 représente un décodeur audio 100 dans un mode de réalisation de l'invention.

[0030] Le décodeur 100 comprend un module 101 de lecture de séquence binaire, un module 102 de quantification inverse, un module 103 de transformation linéaire inverse, un module 104 de transformation fréquence/temps.

[0031] Le décodeur 100 est adapté pour recevoir en entrée le flux binaire Φ transmis par le codeur 1 et pour délivrer en sortie Q' signaux S'_1, ..., S'_Q, destinés à alimenter les Q' haut-parleurs H1, H2 ..., HQ' respectifs d'un système de rendu sonore 105.

Opérations réalisées au niveau du codeur :

[0032] Le module 2 de transformation temps/fréquence du codeur 1 reçoit en entrée les N signaux S_1, ..., S_N de la scène sonore 3D à coder, sous forme de blocs successifs.

[0033] Chaque bloc m reçu comporte N trames temporelles indiquant chacune différentes valeurs prises au cours du temps par un signal respectif.

[0034] Sur chaque trame temporelle de chacun des signaux, le module 2 de transformation temps/fréquence effectue une transformation temps/fréquence, dans le cas présent, une transformée en cosinus discrète modifiée (MDCT).

[0035] Ainsi, suite à la réception d'un nouveau bloc comportant une nouvelle trame pour chacun des signaux S_i, il détermine, pour chacun des signaux S_i, i=1 à N, sa représentation spectrale X_i, caractérisée par M coefficients MDCT $X_{i,t}$, avec t = 0 à M-1. Un coefficient MDCT $X_{i,t}$ représente ainsi le spectre du signal Si pour une fréquence F_i.

[0036] Les représentations spectrales X_i des signaux S_i, i= 1 à N, sont fournies en entrée du module 3 de transformation linéaire.

[0037] Les représentations spectrales X_i des signaux S_i, i= 1 à N, sont en outre fournies en entrée du module 6 de calcul des courbes de masquage.

[0038] Le codage de signaux multicanaux comporte dans le cas considéré une transformation linéaire, permettant de prendre en compte les interactions entre les différents signaux audio à coder, avant le codage monophonique, par le module 4 de quantification, des composantes résultant de la transformation linéaire.

[0039] Le module 3 de transformation linéaire est adapté pour effectuer une transformation linéaire des coefficients des représentations spectrales $(X_i)_{1\leq i\leq N}$ fournis. Dans un mode de réalisation, il est adapté pour effectuer une transformation spatiale. Il détermine alors les composantes spatiales des signaux $(X_i)_{1\leq i\leq N}$, dans le domaine fréquentiel, résultant de la projection sur un référentiel spatial dépendant de l'ordre de la transformation. L'ordre d'une transformation spatiale se rattache à la fréquence angulaire selon laquelle elle « scrute » le champ sonore.

[0040] Dans le mode de réalisation considéré, le module 3 de transformation linéaire effectue une transformation ambiophonique d'ordre p (par exemple p=1), qui donne une représentation spatiale compacte d'une scène sonore 3D, en réalisant des projections du champ sonore sur les fonctions harmoniques sphériques ou cylindriques associées.

[0041] Pour plus d'information sur les transformations ambiophoniques, on pourra se référer aux documents suivants : « Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia », Thèse de doctorat de l'université Paris 6, Jérôme DANIEL, 31 juillet 2001, « A highly scalable spherical microphone array based on an orthonormal décomposition of the sound field », Jens Meyer - Gary Elko, Vol. II - pp. 1781-1784 in Proc. ICASSP 2002.

[0042] Le module 3 de transformation spatiale délivre ainsi r (r= 2p+1) composantes ambiophoniques $(Y_j)_{1\leq j\leq r}$. Chaque composante ambiophonique Y_j considérées dans le domaine fréquentiel, comporte M paramètres spectraux $Y_{j,t}$ pour t = 0 à M-1. Le paramètre spectral $Y_{j,t}$ se rapporte à la fréquence F_t pour t = 0 à M-1.

[0043] Les composantes ambiophoniques sont déterminés de la façon suivante :

$$\begin{bmatrix} Y_{1,0} & . & . & Y_{1,M-1} \\ . & . & . & . \\ . & . & . & . \\ Y_{r,0} & . & . & Y_{r,M-1} \end{bmatrix} = \mathrm{R} \begin{bmatrix} X_{1,0} & . & . & X_{1,M-1} \\ . & . & . & . \\ . & . & . & . \\ X_{N,0} & . & . & X_{N,M-1} \end{bmatrix}.$$

où $\mathrm{R} = \left(R_{i,j}\right)_{\substack{1\leq i\leq r \\ 1\leq j\leq N}}$ est la matrice de transformation ambiophonique d'ordre p pour la scène sonore spatiale, avec

$R_{1,j} = 1$ $R_{i,j} = \sqrt{2}\cos\left[\left(\frac{i}{2}\right)\theta_j\right]$ si i pair et $R_{i,j} = \sqrt{2}\sin\left[\left(\frac{i-1}{2}\right)\theta_j\right]$ si i impair supérieur ou égale à 3, et θj est l'angle de propagation du signal S_j dans l'espace de la scène 3D.

[0044] Chacune des composantes ambiophoniques est donc déterminée en fonction de plusieurs signaux $(S_i)_{1\leq i\leq N}$.

[0045] Le module 6 de calcul de courbe de masquage est adapté pour déterminer la courbe de masquage spectral de chaque trame d'un signal Si considéré individuellement dans le bloc m, à l'aide de sa représentation spectrale Xi et d'un modèle psychoacoustique.

[0046] Le module 6 de calcul de courbe de masquage calcule ainsi un seuil de masquage $M_T^m(s,i)$, relatif à la trame de chaque signal $(S_i)_{1\leq i\leq N}$ dans le bloc m, pour chaque bande de fréquence s considérée lors de la quantification. Chaque bande de fréquence s est élément d'un ensemble de bandes de fréquence comprenant par exemple les bandes telles que normalisées pour le codeur MPEG-4 AAC.

[0047] Les seuils de masquage $M_T^m(s,i)$ pour chaque signal S_i et chaque bande de fréquences s sont délivrés au module 4 de quantification.

[0048] Le module 4 de quantification est adapté pour quantifier les composantes $(Y_j)_{1\leq j\leq r}$ qui lui sont fournies en entrée, de manière à réduire le débit nécessaire à la transmission. Des fonctions de quantification respectives sont déterminées par le module 4 de quantification sur chaque bande de fréquence s.

[0049] Dans une bande s quelconque, le module 4 de quantification quantifie chaque coefficient spectral $\left(Y_{j,t}\right)_{\substack{1\leq j\leq r \\ 0\leq t\leq M-1}}$ tel que la fréquence F_t est élément de la bande de fréquence s. Il détermine ainsi un indice de quantification i(k) pour chaque coefficient spectral $\left(Y_{j,t}\right)_{\substack{1\leq j\leq r \\ 0\leq t\leq M-1}}$ tel que la fréquence F_t est élément de la bande de fréquence s.

[0050] Pour une bande s considérée, k prend les valeurs de l'ensemble $\{k_{min,s}, k_{min+1,s}, \ldots k_{max,s}\}$, et $(k_{max,s} - k_{min+1,s} + 1)$ est égal au nombre de coefficients spectraux à quantifier dans la bande s pour l'ensemble des composantes ambiophoniques.

[0051] La fonction de quantification Q^m appliquée par le module 4 de quantification pour les coefficients $\left(Y_{j,t}\right)_{\substack{1\leq j\leq r \\ 0\leq t\leq M-1}}$ calculés pour un bloc m de signaux prend la forme suivante, conformément à la norme MPEG-4 AAC :

$$Q^m(Y_{j,t}) = \mathrm{Arr}\left(\left(\frac{Y_{j,t}}{B_j^m(s)}\right)^{\frac{3}{4}}\right)$$

avec la fréquence F_t élément de la bande de fréquence s, et il existe k élément de $\{k_{min,s}, k_{min+1,s}, \cdots k_{max,s}\}$ tel que $Q^m(Y_{j,t}) = i(k)$. $B_j^m(s)$, coefficient d'échelle relatif à la composante ambiophonique Y_j, prend des valeurs discrètes. Il dépend du paramètre d'échelle entier relatif $\varphi_j^m(s)$: $B_j^m(s) = 2^{\frac{1}{4}\varphi_j^{m(s)}}$.

[0052] Arr est une fonction d'arrondi délivrant une valeur entière. Arr(x) est par exemple la fonction fournissant l'entier le plus proche de la variable x, ou encore la fonction « partie entière » de la variable x, etc.

[0053] Le module 4 de quantification est adapté pour déterminer une fonction de quantification à appliquer sur une bande de fréquence s vérifiant que le seuil de masquage $M_T^m(s,i)$ de chaque signal S_i dans le domaine d'écoute, avec $1 \leq i \leq N$, est supérieur à la puissance de l'erreur apportée, sur un signal audio restitué dans le domaine d'écoute correspondant au canal i (et non pas dans le domaine de transformation linéaire), par les erreurs de quantification apportée aux composantes ambiophoniques.

[0054] Le module 4 de quantification est donc adapté pour déterminer, lors du traitement d'un bloc m de signaux, la fonction de quantification définie à l'aide des paramètres d'échelle $\left(B_j^m(s)\right)_{1\leq j\leq r}$ relatifs à chaque bande s, telle que,

pour tout i, 1 ≤ i ≤ N, l'erreur introduite sur le signal S_i dans la bande s par la quantification des composantes ambiophoniques est inférieure au seuil de masquage $M_T^m(s,i)$ du signal S_i sur la bande s.

[0055] Un problème à résoudre par le module 4 de quantification est donc de déterminer, sur chaque bande s, l'ensemble des coefficients d'échelle $\left(B_j^m(s)\right)_{1\le j\le r}$ vérifiant la formule (1) suivante :

$$\left\{B_j^m \,/\, P_e^m(s,i) \le M_T^m(s,i), 1 \le i \le N \right\}_{1\le j\le r}$$

où $P_e^m(s,i)$ est la puissance d'erreur introduite sur le signal S_i suite aux erreurs de quantification introduites par la quantification, définie par les coefficients d'échelle $\left(B_j^m(s)\right)_{1\le j\le r}$, des composantes ambiophoniques.

[0056] Ainsi, $B_j(s)$ représente un paramètre caractérisant la fonction de quantification dans la bande s relative à la jème composante. Le choix de $B_j(s)$ détermine de manière bijective la fonction de quantification utilisée.

[0057] Cette disposition a pour effet que le bruit apporté dans le domaine d'écoute par la quantification sur les composantes issues de la transformation linéaire reste masqué par le signal dans le domaine d'écoute, ce qui contribue à une meilleure qualité des signaux restitués dans le domaine d'écoute.

[0058] Dans un mode de réalisation, le problème indiqué ci-dessus par la formule (1) est traduit sous la forme de la formule (2) suivante :

$$\left\{B_j^m \,/\, \text{Probabilité}\left(P_e^m(s,i) \le M_T^m(s,i)\right) \ge \alpha, 1 \le i \le N \right\}_{1\le j\le r},$$

où α est un taux fixé de respect du seuil de masquage.

[0059] La probabilité est calculée pour la trame relative au signal S_i du bloc m considéré et sur l'ensemble des bandes de fréquence s.

[0060] La justification de cette traduction est réalisée dans le document « Optimisation de la quantification par modèles statistiques dans le codeur MPEG Advanced Audio coder (AAC) - Application à la spatialisation d'un signal comprimé en environnement MPEG-4 », Thèse de doctorat de Olivier Derrien - ENST Paris, 22 novembre 2002, nommé ci-après « document Derrien ». Selon ce document, on cherche à modifier la quantification de manière à diminuer la distorsion perçue par l'oreille d'un signal résultant d'un filtrage de spatialisation HRTF (en anglais « Head Related Transfer Function » encore appelé filtre de tête modélisant l'effet de chemin de propagation entre la position de la source sonore et l'oreille humaine et prenant en compte l'effet dû à la tête et au torse d'un auditeur, appliqué après le décodage.

[0061] Par ailleurs, $P_e^m(s,i) = \sum_{k=k_{\min}}^{k=k_{\max}} e_i^m(k)^2$, où $\left\{e_i^m(k)\right\}_{k_{\min}\le k\le k_{\max}}$ sont les erreurs introduites sur les $K_s = (k_{\max,s} - k_{\min+1,s} + 1)$ coefficients spectraux du signal S_i correspondant à des fréquences dans la bande s.

[0062] Soit $H = \left(h_{i,j}\right)_{\substack{1\le i\le N \\ 1\le j\le r}}$ la matrice inverse de la matrice de transformation ambiophonique R, alors $e_i^m(k) = \sum_{j=1}^{j=r} h_{i,j} v_j^m(k)$, où $\left\{v_j^m(k)\right\}_{k_{\min,s}\le k\le k_{\max,s}}$ sont les erreurs de quantification introduites sur les $(k_{\max,s} - k_{\min+1,s} + 1)$ coefficients spectraux de composantes ambiophoniques correspondant à des fréquences dans la bande s.

[0063] Ainsi $P_e^m(s,i) = \sum_{k=k_{\min,s}}^{k=k_{\max},s} \left(\sum_{j=1}^{j=r} h_{i,j} v_j^m(k)\right)^2$.

[0064] On effectue les hypothèses suivantes :

- les erreurs de quantification $e_i^m(k)$ sont des variables aléatoires indépendantes équi-distribuées selon l'indice k ;

- les erreurs de quantification $e_i^m(k)$ sont des variables aléatoires selon l'indice i ;
- le nombre d'échantillons dans une bande s est suffisamment grand ;
- le codeur 1 travaille à haute résolution.

[0065] Sous ces hypothèses et par application du théorème de la limite centrale, la puissance $P_e^m(s,i)$ de l'erreur de quantification, dans une sous-bande s et pour un signal S_i, tend, lorsque le nombre de coefficients dans une bande s augmente, vers une gaussienne dont la moyenne $m_{P_e^m(s,S_i)}$ et la variance $\sigma_{P_e^m(s,S_i)}$ sont données par les formules suivantes :

$$\begin{cases} m_{P_e^m(s,i)} = \sum_{k=k_{\min,s}}^{k_{\max,s}} \mathrm{E}[e_i^m(k)^2] \\ \sigma^2_{P_e^m(s,i)} = \sum_{k=k_{\min,s}}^{k_{\max,s}} \mathrm{E}[e_i^m(k)^4] - \mathrm{E}[e_i^m(k)^2]^2 \end{cases}$$

où la fonction E[x] délivre la moyenne de la variable x.

[0066] La contrainte « Probabilité $\left(P_e^m(s,i) \le M_T^m(s,i)\right) \ge \alpha$ » indiquée dans la formule 2 ci-dessus s'écrit alors à l'aide de la formule (3) suivante :

$$m_{P_e^m(s,i)} + \beta(\alpha)\sigma_{P_e^m(s,i)} \le M_T^m(s,i)$$

Avec :

$$\beta(\alpha) = \sqrt{2}\,\mathrm{Erf}^{-1}(2\alpha - 1)$$

et la fonction $Erf^{-1}(x)$ est l'inverse de la fonction d'erreur d'Euler.

[0067] Les variables $e_i^m(k)$ étant indépendantes selon l'indice *i,* on en déduit :

$$\mathrm{E}[e_i^m(k)^2] = \sum_{j=1}^{r} h_{i,j}^2\, \mathrm{E}[v_i^m(k)^2]$$

[0068] Par conséquent, on obtient :

$$m_{P_e^m(s,i)} = \sum_{k=k_{\min,s}}^{k_{\max,s}} \sum_{j=1}^{r} h_{i,j}^2\, \mathrm{E}[v_i^m(k)^2] = \sum_{j=1}^{r} h_{i,j}^2 \sum_{k=k_{\min,s}}^{k_{\max,s}} \mathrm{E}[v_j^m(k)^2]$$

[0069] Les variables aléatoires $e_i^m(k)$ étant indépendantes et équi-distribuées selon l'indice *k,* les variables aléatoires $v_i^m(k)$ sont également indépendantes et équi-distribuées selon l'indice k. Par conséquent :

$$m_{P_e^m(s,i)} = K_s . \sum_{j=1}^{r} h_{i,j}^2\, \mathrm{E}[\left(v_i^m(s)\right)^2]$$

avec :

$$K_s = k_{\max,s} - k_{\min,s} + 1$$

[0070] On suppose que les puissances $P_e^m(s,i)$ d'erreur de quantification tendent vers des gaussiennes, alors :

$$\mathrm{E}[e_i^m(k)^4] = 3\,\mathrm{E}[e_i^m(k)^2]^2$$

[0071] D'où :

$$\sigma^2_{P_e^m(s,i)} = 2\sum_{k=k_{\min,s}}^{k_{\max,s}} \mathrm{E}[e_i^m(k)^2]^2$$

[0072] Ainsi on peut écrire :

$$\sigma^2_{P_e^m(s,i)} = 2\sum_{k=k_{\min,s}}^{k_{\max,s}} (h_{i,j}^2 \sum_{j=1}^{r} \mathrm{E}[v_j^m(k)^2])^2$$

[0073] A partir de cette dernière équation, et en appliquant l'inégalité de Cauchy-Schwartz :

$$\sigma_{P_e^m(s,i)} = \sqrt{2}\sqrt{\sum_{k=k_{\min,s}}^{k_{\max,s}} (h_{i,j}^2 \sum_{j=1}^{r} \mathrm{E}[v_j^m(k)^2])^2} \le \sqrt{2}\sum_{k=k_{\min,s}}^{k_{\max,s}} h_{i,j}^2 \sum_{j=1}^{r} \mathrm{E}[v_j^m(k)^2]$$

[0074] Ce qui implique que :

$$\sigma_{P_e^m(s,i)} \le \sqrt{2} m_{P_e^m(s,i)}$$

[0075] Par ailleurs, en haute résolution :

$$\mathrm{E}[v_j^2] \approx \frac{16}{9}\mathrm{E}[e_R^2] B_j^m(s)^{\frac{3}{2}} \mu_{\frac{1}{2},j}(s)$$

avec $\mu_{\frac{1}{2},j}$ représentant l'espérance mathématique de $\left|Y_j^m\right|^{\frac{1}{2}}$ dans la sous bande s traitée et e_R l'erreur d'arrondi propre à la fonction d'arrondi Arr.

[0076] Si Arr(x) est par exemple la fonction fournissant l'entier le plus proche de la variable x, e_R est égale à 0,5. Si Arr(x) est la fonction « partie entière » de la variable x, e_R est égale à 1.

[0077] Ainsi la contrainte donnée par la formule (3) relative au signal S_i, i= 1 à N, sur une bande s, s'écrit sous la forme suivante :

$$K_s \frac{16}{9}\mathrm{E}[e_R^2](1+\sqrt{2}\beta(\alpha))\sum_{j=1}^{r}(h_{i,j}^2 B_j^m(s)^{\frac{3}{2}} \mu_{\frac{1}{2},j}(s)) \le M_T^m(s,i)$$

[0078] Il est ainsi possible, à partir de cette dernière équation, de déterminer si des coefficients d'échelle $\left(B_j^m(s)\right)_{1\le j\le r}$ calculés par le module 4 de quantification pour coder les composantes de la transformée, permettent ou non de respecter le seuil de masquage tel que considéré dans le domaine du signal.

[0079] Cette dernière équation représente une condition suffisante pour que le bruit correspondant au canal i soit masqué en sortie dans le domaine d'écoute.

[0080] Dans un mode de réalisation de l'invention, le module 4 de quantification est adapté pour déterminer à l'aide de cette dernière équation, pour un bloc m de trames courant, des coefficients d'échelle $\left(B_j^m(s)\right)_{1\le j\le r}$ garantissant que le bruit dans le domaine d'écoute est masqué.

[0081] Dans un mode de réalisation particulier de l'invention, le module 4 de quantification est adapté pour déterminer, pour un bloc m de trames courant, des coefficients d'échelle $\left(B_j^m(s)\right)_{1\le j\le r}$ garantissant que le bruit dans le domaine d'écoute est masqué et en outre permettant de respecter une contrainte de débit.

[0082] Dans un mode de réalisation, les conditions à respecter sont les suivantes :

- Minimiser le débit global $D^m = \sum_{j=1}^{r} D_j^m$
- Sous la contrainte :

$$K_s \frac{16}{9}\mathrm{E}[e_R^2](1+\sqrt{2}\beta(\alpha))\sum_{j=1}^{r}(h_{i,j}^2 B_j^m(s)^{\frac{3}{2}}\mu_{\frac{1}{2},j}(s)) \le M_T^m(s,i)$$

pour toute bande s, avec D_j^m le débit global attribué à la composante ambiophonique Y_j.

[0083] On peut écrire que :

$$D_j^m = \sum_s D_j^m(s)$$

où $D_j^m(s)$ est le débit attribué à la composante ambiophonique Y_j dans la bande s.

$$D^m(s) = \sum_{j=1}^{r} D_j^m(s)$$

[0084] Minimiser le débit global D^m revient donc à minimiser le débit dans chaque bande s. Dans une première approximation, on peut écrire que le débit attribué à une composante ambiophonique dans une bande s est une fonction logarithmique du coefficient d'échelle, soit :

$$D_j^m(s) = D_{j,0}^m - \gamma \ln\left(B_j^m(s)\right)$$

[0085] La nouvelle fonction à minimiser s'écrit donc sous la forme suivante :

$$\mathrm{F}(s) = -\sum_{j=1}^{r} \ln\left(B_j^m(s)\right)$$

[0086] Pour résoudre le problème de quantification par bande en minimisant le débit global sous la contrainte (3), il faut donc minimiser la fonction F sous la contrainte (3).

[0087] La résolution de ce problème d'optimisation sous contrainte est par exemple effectuée à l'aide de la méthode

des Lagrangiens.
La fonction Lagrangienne s'écrit sous la forme suivante :

$$\mathrm{L}(\mathbf{B},\boldsymbol{\lambda}) = -\sum_{j=1}^{r} \ln\left(B_j^m(s)\right) + \sum_{i=1}^{N} \lambda_i \left[K_s \frac{16}{9} \mathrm{E}[e_R^2](1+\sqrt{2}\beta(\alpha)) \sum_{j=1}^{r} (h_{i,j}^2 B_j^m(s)^{\frac{3}{2}} \mu_{\frac{1}{2},j}(s)) - M_T^m(s,i) \right]$$

$$\mathrm{L}(\mathbf{B},\boldsymbol{\lambda}) = -\sum_{j=1}^{r} \ln\left(B_j^m(s)\right) + \Delta_j^m(\boldsymbol{\lambda}) B_j^m(s)^{\frac{3}{2}} - \sum_{i=1}^{N} \lambda_i M_T^m(s,i)$$

[0088] Avec :

$$\Delta_j^m(\boldsymbol{\lambda}) = \mu_{\frac{1}{2},j}(s) K_s \frac{16}{9} \mathrm{E}[e_R^2](1+\sqrt{2}\beta(\alpha)) \sum_{i=1}^{N} h_{i,j}^2 \lambda_i$$

et les valeurs λ_j, $1 \le j \le N$, sont les coordonnées du vecteur de Lagrange λ.

[0089] La mise en oeuvre de la méthode des Lagrangiens permet d'écrire tout d'abord que, pour $1 \le j \le r$:

$$B_j^m(s) = \frac{3}{2} \frac{1}{\Delta_j^m(\boldsymbol{\lambda})}$$

[0090] On remplace par ces termes les coefficients d'échelle dans l'équation de Lagrange. Et on cherche alors à déterminer la valeur du vecteur de Lagrange λ qui maximise la fonction $\omega(\lambda) = L((B_1^m(s),...,B_r^m(s)),\lambda)$, par exemple à l'aide de la méthode du gradient de la fonction ω.

[0091] D'après la méthode du gradient d'Uzawa $\nabla w(\lambda)$, où

$$\nabla\omega(\lambda) = \begin{pmatrix} \frac{\partial\omega}{\partial\lambda_1}(\lambda) \\ \vdots \\ \frac{\partial\omega}{\partial\lambda_N}(\lambda) \end{pmatrix}$$

$$B_j^m(s) = \frac{3}{2} \frac{1}{\Delta_j^m(\lambda)}.$$

les dérivées partielles ne sont autres que les contraintes calculées pour les

[0092] On utilise la méthode itérative de gradient relatif (cf. notamment le document Derrien) pour résoudre ce système.

[0093] L'équation générale (formule (4)) de mise à jour du vecteur de Lagrange lors d'une (k+1)ième itération de la méthode s'écrit alors sous la forme suivante :

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k \otimes \left(\mathbf{1} + \rho\mathbf{m} \otimes \nabla\omega\left(\boldsymbol{\lambda}^k\right)\right)$$

avec le vecteur de Lagrange λ avec un exposant (k+1) indiquant le vecteur actualisé et le vecteur de Lagrange λ avec un exposant k indiquant le vecteur calculé précédemment lors de la kième itération, $\otimes$ désignant le produit terme à terme entre deux vecteurs de même taille, ρ désignant le pas de l'algorithme itératif et **m** étant un vecteur de pondération.

[0094] Dans un mode de réalisation, de manière à assurer la convergence de la méthode itérative, on choisit le vecteur **m** égal à :

$$\begin{pmatrix} \frac{1}{M_T^m(s,1)} \\ \vdots \\ \frac{1}{M_T^m(s,N)} \end{pmatrix}$$

[0095] Dans le mode de réalisation considéré, le module 4 de quantification est adapté pour mettre en oeuvre les étapes du procédé décrit ci-dessous en référence à la figure 3 sur chaque bande de quantification s lors de la quantification d'un bloc m de signaux $(S_i)_{1\leq i\leq N}$.

[0096] Le procédé est basé sur un algorithme itératif comprenant des instructions pour mettre en oeuvre les étapes décrites ci-dessous lors de l'exécution de l'algorithme sur des moyens de calcul du module 4 de quantification.

[0097] Dans une étape a/ d'initialisation (k=0) : on définit la valeur du pas d'itération ρ, une valeur D représentant un seuil de débit et la valeur des coordonnées $(\lambda_1, ...\lambda_N)$ du vecteur de Lagrange initial avec $\lambda_j = \lambda^0$, $1 \leq j \leq N$.

[0098] Les étapes de la boucle itérative pour une (k+1)ème itération, avec k entier supérieur ou égal à 0, sont les suivantes.

[0099] Dans une étape b/, les valeurs des coordonnées λ_j, $1 \leq j \leq N$ du vecteur de Lagrange considérées étant celles calculées précédemment lors de la kième itération, on calcule pour $1 \leq j \leq N$:

$$\Delta_j^m(\boldsymbol{\lambda}) = \mu_{\frac{1}{2},j}(s) K_s \frac{16}{9} \mathrm{E}[e_R^2](1+\sqrt{2}\beta(\alpha)) \sum_{i=1}^{N} h_{i,j}^2 \lambda_i$$

[0100] Puis dans une étape c/, on calcule les coefficients d'échelle, pour $1 \leq j \leq r$:

$$B_j^m(s) = \frac{3}{2} \frac{1}{\Delta_j^m(\boldsymbol{\lambda})}$$

[0101] Dans une étape d/, on calcule la valeur de la fonction *F* sur la bande s, représentant le débit correspondant pour la bande s :

$$\mathrm{F}(s) = -\sum_{j=1}^{r} \ln\left(B_j^m(s)\right)$$

[0102] Dans une étape e/, on compare la valeur *F*(*s*) calculée avec le seuil donné D.

[0103] Si la valeur *F*(*s*) calculée est supérieure au seuil donné D, on calcule, dans une étape f/, la valeur du vecteur de Lagrange λ pour la (k+1)ème itération à l'aide de l'équation (4) indiquée ci-dessus et du vecteur de Lagrange calculé lors de la kème itération.

[0104] Puis, dans une étape g/, on incrémente l'indice k d'une unité et on réitère les étapes b/, c/, d/ et e/.

[0105] Si la valeur *F*(*s*) calculée à l'étape e/, est inférieure au seuil donné D, on stoppe les itérations. On a alors déterminé des coefficients d'échelle $\left(B_j^m(s)\right)_{1\leq j\leq r}$ pour la bande de quantification s permettant de masquer, dans le domaine d'écoute, le bruit dû à la quantification dans la bande s, des composantes ambiophoniques $(Y_j)_{1\leq j\leq r}$, tout en garantissant que le débit nécessaire pour cette quantification dans la bande s est inférieur à une valeur déterminée, fonction de D.

[0106] On applique ensuite la fonction de quantification ainsi déterminée pour les bandes s respectives et composantes ambiophoniques respectives aux coefficients spectraux des composantes ambiophoniques. Les indices de quantification ainsi que des éléments de définition de la fonction de quantification sont fournis au module 5 de codage de Huffman.

[0107] Les données de codage délivrées par le module 5 de codage de Huffman sont ensuite transmises sous forme de flux binaire Φ au décodeur 100.

Opérations réalisées au niveau du décodeur :

[0108] Le module 101 de lecture de séquence binaire est adapté pour extraire des données de codage présentes dans le flux Φ reçu par le décodeur et en déduire, dans chaque bande s, des indices de quantification i(k) et des coefficients d'échelle $(B_j^m(s))_{1\leq j\leq r}$.

[0109] Le module de quantification inverse 102 est adapté pour déterminer les coefficients spectraux, relatifs à la bande s, des composantes ambiophoniques correspondants en fonction des indices de quantification i(k) et des coefficients d'échelles $(B_j^m(s))_{1\leq j\leq r}$ dans chaque bande s.

[0110] Ainsi un coefficient spectral $Y_{j,t}$ relatif à la fréquence F_t élément de la bande s de la composante ambiophonique Y_j et représenté par l'indice de quantification i(k) est restitué par le module 102 de quantification inverse à l'aide de la formule suivante : $Y_{j,t} = A_j^m(s)\, i(k)^{\frac{4}{3}}$

[0111] Un décodage ambiophonique est ensuite appliqué aux r composantes ambiophoniques décodées, de manière à déterminer Q' signaux S'_1, S'_2, ..., S'_Q, destinés aux Q' haut-parleurs H1, H2 ..., HQ'.

[0112] Le bruit de quantification à la sortie du décodeur 100 est une constante qui ne dépend que de la transformée R utilisée et du module 4 de quantification car les données psychoacoustiques utilisées lors du codage ne prennent pas en considération les traitements effectués lors de la restitution par le décodeur. En effet, le modèle psychoacoustique ne prend pas en compte les interactions acoustiques entre les différents signaux, mais calcule la courbe de masquage d'un signal comme s'il était le seul écouté. L'erreur calculée sur ce signal reste donc constante et masquée pour toute matrice de décodage ambiophonique utilisée. Cette matrice de décodage ambiophonique va simplement modifier la distribution de l'erreur sur les différents haut-parleurs en sortie.

Revendications

1. Procédé de quantification de composantes, certaines au moins desdites composantes ($(Y_j)_{1\leq j\leq r}$) étant déterminées chacune en fonction d'une pluralité de signaux audio ($(S_j)_{1\leq j\leq N}$) d'une scène sonore et calculées par application d'une transformation linéaire multicanal, ambiophonique et à plus de deux canaux, sur lesdits signaux audio, selon lequel on détermine une fonction de quantification (Q^m) à appliquer audites composantes dans une bande de fréquence donnée (s) en testant une condition relative à au moins un signal audio (S_i) et dépendant au moins d'une comparaison effectuée entre :

   - un seuil de masquage psychoacoustique $(M_T^m(s,i))$ relatif au signal audio dans la bande de fréquence donnée, et
   - une valeur déterminée en fonction de la transformation linéaire multicanal inverse et d'erreurs de quantification des composantes par ladite fonction sur la bande de fréquence donnée.

2. Procédé selon la revendication 1, selon lequel la condition est relative à plusieurs signaux audio et dépend de plusieurs comparaisons, chaque comparaison étant effectuée entre un seuil de masquage psychoacoustique relatif à un signal audio respectif dans la bande de fréquence donnée, et une valeur déterminée en fonction de la transformation linéaire multicanal inverse et d'erreurs de quantification des composantes par ladite fonction.

3. Procédé selon la revendication 1 ou la revendication 2, selon laquelle la détermination de la fonction de quantification (Q^m) est réitérée lors de l'actualisation des valeurs des composantes à quantifier.

4. Procédé selon l'une quelconque des revendications précédentes, selon lequel on teste la condition relative à un signal audio au moins en comparant le seuil de masquage psychoacoustique relatif au signal audio et un élément représentant la valeur mathématique $\sum_{j=1}^{r}(h_{i,j}^2 B_j(s)^{\frac{3}{2}}\mu_{\frac{1}{2},j}(s))$, où s est la bande de fréquences donnée, r est le nombre de composantes, $h_{i,j}$ est le coefficient de la transformée linéaire multicanal inverse relatif au signal audio (Si) et à la j^{ième} composante avec j=1 à r, $B_j(s)$ est un coefficient d'échelle relatif à la composante ambiophonique Yj prenant des valeurs discrètes et caractérisant la fonction de quantification (Q^m) dans la bande s relative à la j^{ième}

composante et $\mu_{\frac{1}{2},j}(s)$ est l'espérance mathématique dans la bande s de la racine carrée de la j[ième] composante.

5. Procédé selon l'une quelconque des revendications précédentes, selon lequel on détermine une fonction de quantification à appliquer audites composantes dans la bande de fréquence donnée à l'aide d'un processus itératif générant à chaque itération un paramètre de la fonction de quantification candidat vérifiant la condition et associé à un débit correspondant, l'itération étant stoppée lorsque le débit est inférieur à un seuil donné.

6. Module (4) de quantification adapté pour quantifier au moins des composantes ($(Y_j)_{1\leq j\leq r}$) déterminées chacune en fonction d'une pluralité de signaux audio ($(S_j)_{1\leq j\leq N}$) d'une scène sonore et calculées par application d'une transformation linéaire multicanal sur lesdits signaux audio, ledit module de quantification étant adapté pour mettre en oeuvre les étapes d'un procédé selon l'une quelconque des revendications 1 à 5.

7. Codeur audio (1) adapté pour coder une scène audio comprenant plusieurs signaux audio respectifs ($(S_j)_{1\leq j\leq N}$) en un flux binaire de sortie (Φ), comprenant :

- un module (3) de transformation adapté pour calculer par application d'une transformation linéaire multicanal sur lesdits signaux audio, des composantes ($(Y_j)_{1\leq j\leq r}$) dont au moins certaines sont déterminées chacune en fonction d'une pluralité des signaux audio ; et
- un module (4) de quantification selon la revendication 6 adapté pour déterminer au moins une fonction de quantification (Q^m) sur au moins une bande de fréquence donnée (s) et pour quantifier les composantes sur la bande de fréquence donnée en fonction d'au moins la fonction de quantification déterminée ;

ledit codeur étant adapté pour constituer un flux binaire en fonction au moins de données de quantification délivrées par le module de quantification.

8. Programme d'ordinateur à installer dans un module (4) de quantification, ledit programme comprenant des instructions pour mettre en oeuvre les étapes d'un procédé selon l'une quelconque des revendications 1 à 5 lors d'une exécution du programme par des moyens de traitement dudit module.

9. Données de codage (Φ), déterminées suite à la mise en oeuvre d'un procédé de quantification selon l'une quelconque des revendications 1 à 5.

Patentansprüche

1. Verfahren zur Quantifizierung von Komponenten, wobei wenigstens einige dieser Komponenten ($(Y_j)_{1\leq j\leq r}$) jeweils in Abhängigkeit von mehreren Audiosignalen ($(S_j)_{1\leq j\leq N}$) einer Klangszene bestimmt werden und durch Anwendung einer Mehrkanal-Lineartransformation, ambisonisch und mit mehr als zwei Kanälen, auf diese Audiosignale berechnet werden,
wobei eine Quantifizierungsfunktion (Q^m) bestimmt wird, die auf die Komponenten in einem gegebenen Frequenzband (s) anzuwenden ist, indem eine Bedingung geprüft wird, die sich auf wenigstens ein Audiosignal (S_j) bezieht und von wenigstens einem Vergleich abhängt, der durchgeführt wird zwischen:

- einem Schwellenwert der psychoakustischen Maskierung $(M_T^m(s,i))$, der sich auf das Audiosignal in dem gegebenen Frequenzband bezieht, und
- einem Wert, der in Abhängigkeit von der inversen Mehrkanal-Lineartransformation und von Fehlern der Quantifizierung der Komponenten durch die Funktion auf dem gegebenen Frequenzband bestimmt wird.

2. Verfahren nach Anspruch 1, wobei sich die Bedingung auf mehrere Audiosignale bezieht und von mehreren Vergleichen abhängt, wobei jeder Vergleich zwischen einem Schwellenwert der psychoakustischen Maskierung, der sich auf ein jeweiliges Audiosignal in dem gegebenen Frequenzband bezieht, und einem Wert, der in Abhängigkeit von der inversen Mehrkanal-Lineartransformation und von Fehlern der Quantifizierung der Komponenten durch die Funktion bestimmt wird, durchgeführt wird.

3. Verfahren nach Anspruch 1 oder Anspruch 2, wobei die Bestimmung der Quantifizierungsfunktion (Q^m) bei der Aktualisierung der Werte der zu quantifizierenden Komponenten wiederholt wird.

4. Verfahren nach einem der vorhergehenden Ansprüche, wobei die Bedingung, die sich auf wenigstens ein Audiosignal bezieht, geprüft wird, indem der Schwellenwert der psychoakustischen Maskierung, der sich auf das Audiosignal bezieht, und ein Element, das den mathematischen Wert $\sum_{j=1}^{r}(h_{i,j}^2 B_j(s)^{\frac{3}{2}}\mu_{\frac{1}{2},j}(s))$ repräsentiert, verglichen werden, wobei s das gegebene Frequenzband ist, r die Anzahl der Komponenten ist, $h_{i,j}$ der Koeffizient der inversen Mehrkanal-Lineartransformation in Bezug auf das Audiosignal (Si) und auf die j-te Komponente ist, mit j=1 bis r, $B_j(s)$ ein Maßstabsfaktor ist, der sich auf die ambisonische Komponente Y_j bezieht, diskrete Werte annimmt und die Quantifizierungsfunktion (Q^m) im Band s in Bezug auf die j-te Komponente charakterisiert, und $\mu_{\frac{1}{2},j}(s)$ der Erwartungswert im Band s der Quadratwurzel aus der j-ten Komponente ist.

5. Verfahren nach einem der vorhergehenden Ansprüche, wobei eine Quantifizierungsfunktion, die auf die Komponenten in dem gegebenen Frequenzband anzuwenden ist, mithilfe eines iterativen Prozesses bestimmt wird, der in jeder Iteration einen Parameter-Kandidaten der Quantifizierungsfunktion bestimmt, der die Bedingung erfüllt und einer entsprechenden Rate zugeordnet ist, wobei die Iteration gestoppt wird, wenn die Rate kleiner als ein vorgegebener Schwellenwert ist.

6. Modul (4) zur Quantifizierung, welches dafür ausgelegt ist, wenigstens Komponenten $((Y_j)_{1\leq j\leq r})$ zu quantifizieren, die jeweils in Abhängigkeit von mehreren Audiosignalen $((S_j)_{1\leq j\leq N})$ einer Klangszene bestimmt werden und durch Anwendung einer Mehrkanal-Lineartransformation auf diese Audiosignale berechnet werden, wobei das Quantifizierungsmodul dafür ausgelegt ist, die Schritte eines Verfahrens nach einem der Ansprüche 1 bis 5 durchzuführen.

7. Audiocoder (1), welcher dafür ausgelegt ist, eine Audioszene, die mehrere jeweilige Audiosignale $((S_j)_{1\leq j\leq N})$ umfasst, in einen Ausgangsbitstrom (Φ) zu codieren, und umfasst:

   - ein Transformationsmodul (3), das dafür ausgelegt ist, durch Anwendung einer Mehrkanal-Lineartransformation auf die Audiosignale Komponenten $((Y_j)_{1\leq j\leq N})$ zu berechnen, von denen wenigstens einige jeweils in Abhängigkeit von mehreren Audiosignalen bestimmt werden; und
   - ein Modul (4) zur Quantifizierung nach Anspruch 6, das dafür ausgelegt ist, wenigstens eine Quantifizierungsfunktion (Q^m) auf wenigstens einem gegebenen Frequenzband (s) zu bestimmen und die Komponenten auf dem gegebenen Frequenzband in Abhängigkeit von wenigstens der bestimmten Quantifizierungsfunktion zu quantifizieren;

   wobei der Coder dafür ausgelegt ist, einen Bitstrom in Abhängigkeit wenigstens von Quantifizierungsdaten, die von dem Quantifizierungsmodul geliefert werden zu erstellen.

8. Computerprogramm, welches in einem Modul (4) zur Quantifizierung zu installieren ist, wobei das Programm Anweisungen zum Durchführen der Schritte eines Verfahrens nach einem der Ansprüche 1 bis 5 bei einer Ausführung des Programms durch Verarbeitungsmittel des Moduls umfasst.

9. Codierungsdaten (Φ), welche im Anschluss an die Durchführung eines Verfahrens zur Quantifizierung nach einem der Ansprüche 1 bis 5 bestimmt werden.

Claims

1. Method for quantizing components, some at least of said components $((Y_j)_{1\leq j\leq r})$ each being determined as a function of a plurality of audio signals $((S_j)_{1\leq j\leq N})$ of a sound scene and computed by applying an ambiophonic multichannel linear transformation with more than two channels to said audio signals,
   according to which a quantization function (Q^m) to be applied to said components in a given frequency band (s) is determined by testing a condition relating to at least one audio signal (S_i) and depending at least on a comparison performed between:

   - a psychoacoustic masking threshold $(M^m_T(s,i))$ relating to the audio signal in the given frequency band, and
   - a value determined as a function of the inverse multichannel linear transformation and of errors of quantization

of the components by said function on the given frequency band.

2. Method according to Claim 1, according to which the condition relates to several audio signals and depends on several comparisons, each comparison being performed between a psychoacoustic masking threshold relating to a respective audio signal in the given frequency band, and a value determined as a function of the inverse multichannel linear transformation and of errors of quantization of the components by said function.

3. Method according to Claim 1 or Claim 2, according to which the determination of the quantization function (Q^m) is repeated during the updating of the values of the components to be quantized.

4. Method according to any one of the preceding claims, according to which the condition relating to an audio signal at least is tested by comparing the psychoacoustic masking threshold relating to the audio signal and an element representing the mathematical value $\sum_{j=1}^{r}(h_{i,j}^2 B_j(s)^{\frac{3}{2}}\mu_{\frac{1}{2},j}(s))$, where s is the given band of frequencies, r is the number of components, $h_{i,j}$ is that coefficient of the inverse multichannel linear transform relating to the audio signal (Si) and to the j^{th} component with j=1 to r, $B_j(s)$ is a scale coefficient relating to the ambiophonic component Yj taking discrete values and characterizing the quantization function (Q^m) in the band s relating to the j^{th} component and $\mu_{12,j}(s)$ is the mathematical expectation in the band s of the square root of the j^{th} component.

5. Method according to any one of the preceding claims, according to which a quantization function to be applied to said components in the given frequency band is determined with the aid of an iterative process generating at each iteration a parameter of the candidate quantization function satisfying the condition and associated with a corresponding bit rate, the iteration being halted when the bit rate is below a given threshold.

6. Quantization module (4) adapted for quantizing at least components $((Y_j)_{1\leq j\leq r})$ each determined as a function of a plurality of audio signals $((S_j)_{1\leq j\leq N})$ of a sound scene and computed by applying a multichannel linear transformation to said audio signals, said quantization module being adapted for implementing the steps of a method according to any one of Claims 1 to 5.

7. Audio coder (1) adapted for coding an audio scene comprising several respective audio signals $((S_j)_{1\leq j\leq N})$ as a binary output stream (Φ), comprising:

- a transformation module (3) adapted for computing by applying a multichannel linear transformation to said audio signals, components $((Y_j)_{1\leq j\leq r})$ at least some of which are each determined as a function of a plurality of the audio signals; and
- a quantization module (4) according to Claim 6 adapted for determining at least one quantization function (Q^m) on at least one given frequency band (s) and for quantizing the components on the given frequency band as a function of at least the determined quantization function;

said coder being adapted for constructing a binary stream as a function at least of quantization data delivered by the quantization module.

8. Computer program to be installed in a quantization module (4), said program comprising instructions for implementing the steps of a method according to any one of Claims 1 to 5 during execution of the program by processing means of said module.

9. Coding data (Φ), determined following the implementation of a quantization method according to any one of Claims 1 to 5.

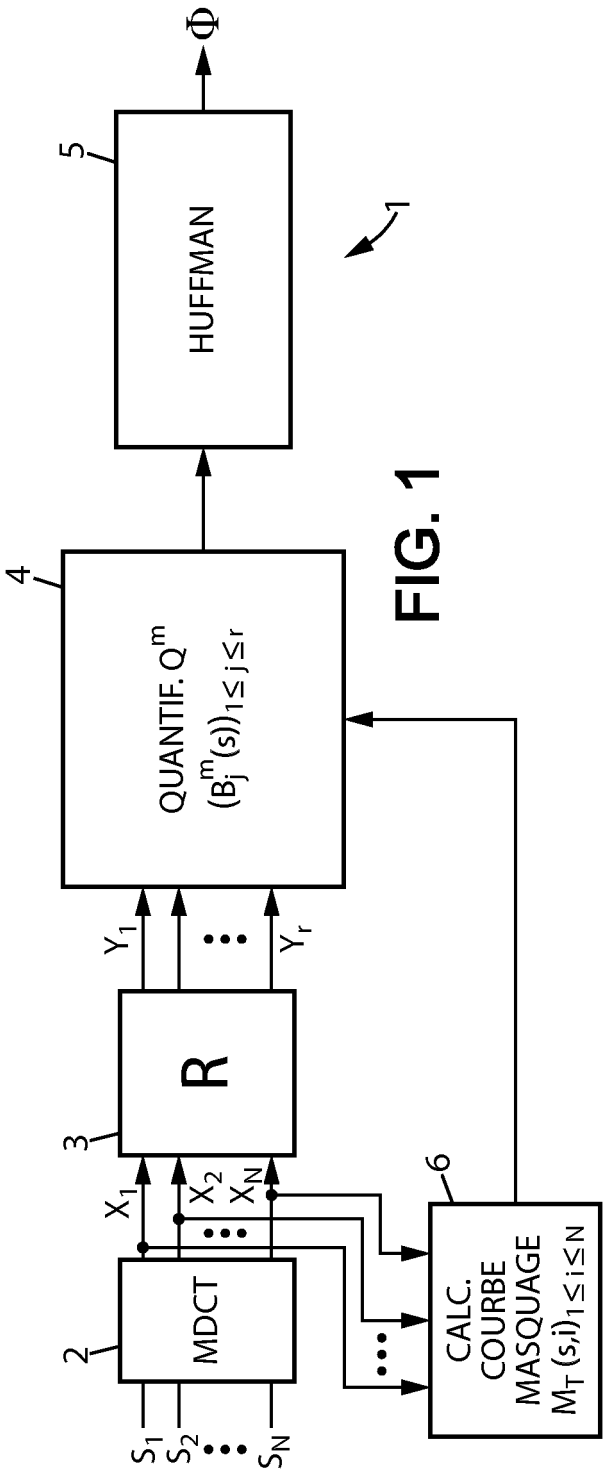

Φ
5
HUFFMAN
1
4
QUANTIF. Q^m
$(B_j^m(s))_{1\le j\le r}$
Y_1
Y_r
3
R
X_1
X_2
X_N
2
MDCT
S_1
S_2
S_N
6
CALC.
COURBE
MASQUAGE
$M_T(s,i)_{1\le i\le N}$

FIG. 1

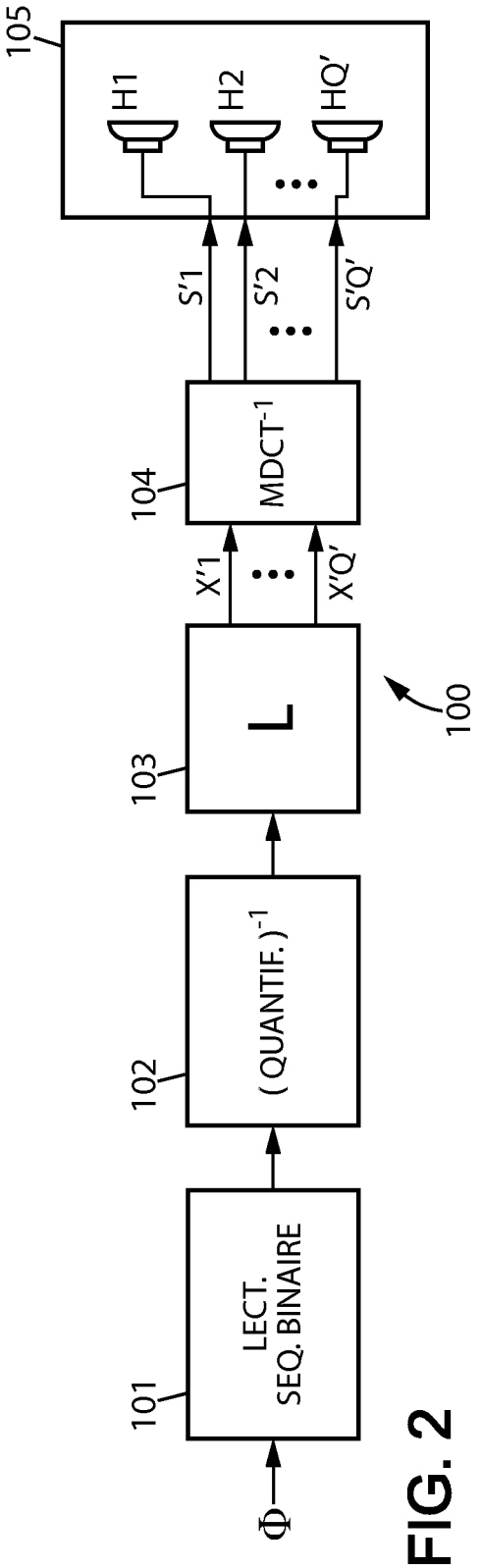

105
H1
H2
HQ'
S'1
S'2
S'Q'
104
$MDCT^{-1}$
X'1
X'Q'
103
L
100
102
$(QUANTIF.)^{-1}$
101
LECT.
SEQ. BINAIRE
Φ

FIG. 2

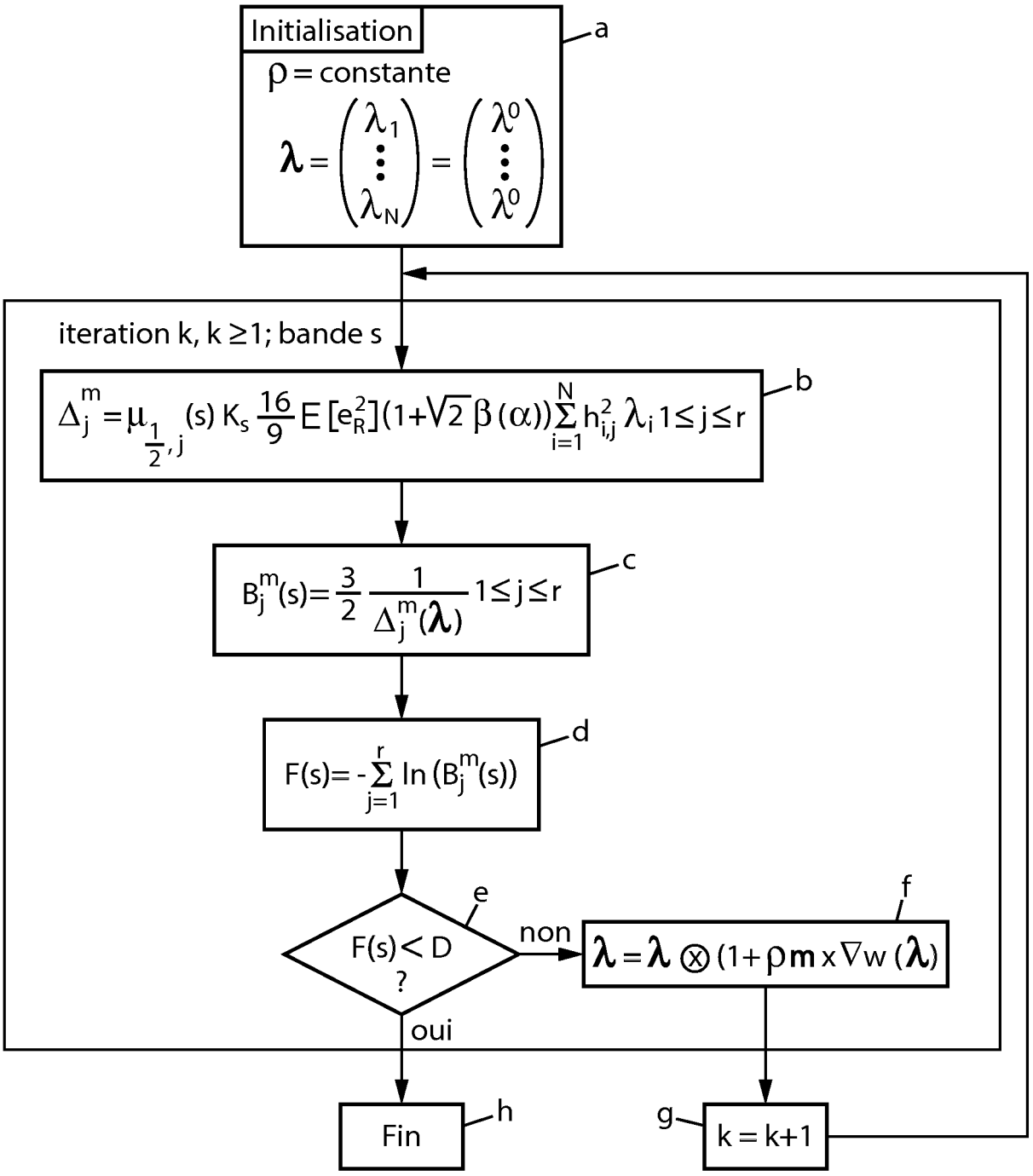

Initialisation
a
ρ = constante
λ = (λ1 ⋮ λN) = (λ0 ⋮ λ0)
iteration k, k ≥1; bande s
b
Δ_j^m = μ_{1/2,j}(s) K_s 16/9 E[e_R^2](1+√2 β(α)) Σ_{i=1}^{N} h_{i,j}^2 λ_i 1≤j≤r
c
B_j^m(s) = 3/2 · 1/Δ_j^m(λ) 1≤j≤r
d
F(s) = - Σ_{j=1}^{r} ln(B_j^m(s))
e
F(s) < D ?
non
f
λ = λ ⊗ (1+ρm x ∇w(λ))
oui
h
Fin
g
k = k+1

FIG. 3

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

- **DERRIEN O ; DUHAMEL P.** Une approche statistique pour l'optimisation du MPEG-2/4 AAC (Advanced Audio Coder) en mode stéréophonique matricé (MS stéréo. *ACTES DE COLLOQUES DU GROUPE D'ETUDES DU TRAITEMENT DU SIGNAL ET DES IMAGES (GRETSI),* 2003, 1-4 **[0007]**
- **JÉRÔME DANIEL.** Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. *Thèse de doctorat de l'université Paris,* 31 Juillet 2001, vol. 6 **[0041]**
- **JENS MEYER ; GARY ELKO.** A highly scalable spherical microphone array based on an orthonormal décomposition of the sound field. *Proc. ICASSP,* 2002, vol. II, 1781-1784 **[0041]**
- Optimisation de la quantification par modèles statistiques dans le codeur MPEG Advanced Audio coder (AAC) - Application à la spatialisation d'un signal comprimé en environnement MPEG-4. *Thèse de doctorat de Olivier Derrien - ENST Paris,* 22 Novembre 2002 **[0060]**