

19 RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
COURBEVOIE

11 N° de publication : **3 148 316**
(à n'utiliser que pour les
commandes de reproduction)
21 N° d'enregistrement national : **23 04288**

51 Int Cl⁸ : **G 10 L 19/008 (2023.01), G 10 L 19/00**

12 **DEMANDE DE BREVET D'INVENTION** **A1**

22 Date de dépôt : 27.04.23.

30 Priorité :

43 Date de mise à la disposition du public de la demande : 01.11.24 Bulletin 24/44.

56 Liste des documents cités dans le rapport de recherche préliminaire : *Se reporter à la fin du présent fascicule*

60 Références à d'autres documents nationaux apparentés :

Demande(s) d'extension :

71 Demandeur(s) : **ORANGE Société anonyme** — FR.

72 Inventeur(s) : **RAGOT Stéphane, GUERIN Alexandre et LEFORT Arnaud.**

73 Titulaire(s) : **ORANGE Société anonyme.**

54 **Mot-clé(s) optimisé de réduction de canaux d'un signal audio stéréophonique.**

57 **Traitement optimisé de réduction de canaux d'un signal**

audio stéréophonique

L'invention se rapporte à un procédé de traitement de réduction de canaux d'un signal stéréophonique pour obtenir un

signal monophonique, comprenant ce qui suit :- application (610, 611), à une trame courante dudit signal, de deux (T1 ou T2, T3) traitements de réduction de canaux, l'un utilisant un filtrage du signal stéréophonique et l'autre non,

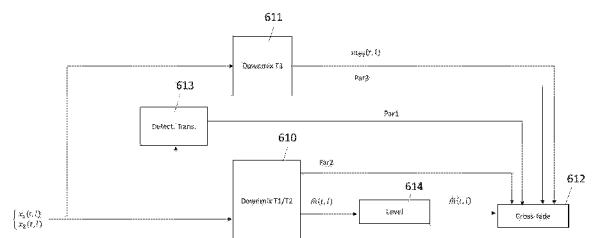
- sélection (612) d'un desdits traitements à appliquer à la trame courante, ladite sélection étant mise en œuvre en fonction de:

-- la présence ou non d'au moins une transition dans le signal stéréophonique, ou

-- un niveau de qualité du filtrage du signal stéréophonique utilisé lors de la mise en œuvre du traitement utilisant le filtrage du signal stéréophonique.

L'invention se rapporte également à un dispositif de traitement de réduction de canaux mettant en œuvre le procédé.

Figure de l'abrégé : Figure 6a



FR 3 148 316 - A1



Description

Titre de l'invention : Traitement optimisé de réduction de canaux d'un signal audio stéréophonique

Domaine technique

[0001] La présente invention se rapporte au domaine général du traitement de signaux audio. L'invention se rapporte en particulier au traitement de réduction de canaux appelé couramment « downmix », d'un signal audio multicanal. On s'intéresse en particulier au traitement de réduction d'un signal stéréophonique à un signal monophonique.

[0002] Ce type de traitement trouve des applications de manière générale dans le domaine des technologies audio, et plus précisément dans le domaine du codage audio, que ce soit à l'étape d'encodage ou du décodage.

Technique antérieure

[0003] La réduction de canaux ou downmix consiste à déduire, à partir d'une combinaison des C canaux d'un signal multicanal x , un signal y constitué d'un nombre inférieur D de canaux. En pratique, cela consiste à définir une fonction $f(\cdot)$ Telle que :

$$[0004] \quad \mathbf{m}(n) = \begin{bmatrix} m_1(n) \\ \vdots \\ m_D(n) \end{bmatrix} = f(\mathbf{x}(n)) = f\left(\begin{bmatrix} x_1(n) \\ \vdots \\ x_C(n) \end{bmatrix}\right), \text{ avec } D < C \text{ et } n \text{ représentant}$$

un indice temporel du signal d'entrée et de sortie.

[0005] Dans cette invention, on s'intéresse au cas particulier $D=1$ et $C=2$, on parlera alors de downmix stéréo à mono ou simplement de downmix mono. Dans le cas $C=2$, on parle de contenu ou signal à 2 canaux qui peut être un contenu stéréo ou un contenu binaural, où la voie 1 (x_1) est référencée comme la voie gauche (Left), et la voie 2 (x_2) comme la voie droite (pour Right). Par la suite, le cas stéréo sera vu comme un signal général à 2 canaux, incluant le cas binaural pour éviter de répéter les deux termes, même si techniquement un signal binaural a des caractéristiques spécifiques. Dans le cas du signal y obtenu après traitement de réduction, on parle de signal monophonique ou mono, que l'on notera par la suite $m(n)$, dans le domaine temporel, ou $M(k)$, dans le domaine fréquentiel.

[0006] Ces signaux stéréo peuvent être issus d'une captation à partir d'un couple de microphones stéréo ou une captation binaurale, ou encore d'un mixage artistique de pistes audio.

[0007] Il existe un certain nombre de couples de microphones permettant de créer des contenus stéréo. Parmi les plus populaires, on peut citer le couple XY composé de 2 microphones cardioïdes présentant une différence d'angle d'orientation entre 90° et 135° . Le couple MS pour « Mid-Side » est un autre couple très répandu composé d'un

microphone de directivité cardioïde, et d'un second microphone dit en « figure de 8 » orienté à 90°. Par combinaison de ces 2 micros (somme/différence), on peut créer les voies gauche et droite de contenu stéréo. Ces couples sont dits coïncidents, c'est-à-dire que les capsules microphoniques ne présentent pas de retard entre elles, la spatia- lisation étant perçue par la différence d'intensité sonore entre les voies gauche et droite.

[0008] Une autre catégorie de couples dits à stéréophonie de phase consiste à utiliser des mi- crophones distants l'un de l'autre, créant ainsi une différence de phase entre les voies pour les sources situés plus proche de l'un des micros. Le plus connu est le couple AB qui exploite 2 microphones omnidirectionnels espacés de quelques centimètres à plusieurs mètres. Un autre couple très répandu est le couple ORTF qui exploite à la fois une différence de phase par un espacement de 17cm des microphones et une différence d'amplitude par des directivités cardioïdes (les microphones présentent un angle autour de 90°). Le couple binaural consiste à placer des microphones omnidirec- tionnels dans les oreilles d'une personne ou d'une tête artificielle : ce type de dispositif permet de créer nativement des contenus binauraux que l'on peut écouter au casque.

[0009] Les contenus stéréo incluent également les signaux de parole et audio issus d'un mixage ou postproduction audio (par exemple des canaux stockés sur un CD ou un DVD, ou diffusés sur Internet).

[0010] Il existe aussi d'autres méthodes de création de contenus stéréo qui ne sont pas revues ici.

[0011] La méthode la plus simple pour créer un signal réduit par downmix est celle dite downmix passif. Elle consiste à effectuer une moyenne des voies gauche et droite du signal stéréo tel que :

$$[0012] \quad m(t) = \frac{x_1(t) + x_2(t)}{2}$$

[0013] ou dans le domaine fréquentiel :

[0014] $M(t, k) = \frac{X_1(t, k) + X_2(t, k)}{2}$ où $X_i(t, k)$, $i = 1, 2$ est la transformée de Fourier à court-terme de $x_i(t)$ (ou FFT en anglais pour « *Fast Fourier Transform* »), k étant l'indice de la fréquence et t l'indice de la trame :

$$X_i(t, k) = \sum_{l=0}^{L-1} w(l) \check{x}_i(t, l) e^{-\frac{j2\pi lk}{N}}, \text{ pour tout } k \in \{0, \dots, N-1\}$$

[0015] Où j est le nombre imaginaire tel que $j = \sqrt{-1}$, t est l'indice de la trame, l est l'indice temporel dans la trame courante, N la taille de la FFT, $w(\cdot)$ une fenêtre d'apodisation de taille L , de type sinus ou Hann ou autre, adaptée à la taille de la trame, et $\check{x}_i(t, l) = x_i(t, l)$. Cette définition s'étend au cas où $N > L$ et $\check{x}_i(\cdot)$ est une version augmentée de $x_i(t)$ avec des 0 (0-padding).

[0016] Cette méthode de downmix passif, très simple, fonctionne bien dans les situations où les voies droite et gauche sont en phase. Cependant, dans les situations où les microphones ne sont pas coïncidents et présentent des différences de phase, un tel mixage génère un filtrage en peigne, qui se manifeste par une coloration du signal d'origine. Cela est dû au fait qu'en fonction de l'espacement des microphones, de la position de la source par rapport au couple de microphones et de la fréquence, les voies gauche et droite d'une même source peuvent se trouver soit en phase comme représenté en figure 1a, soit en opposition de phase comme représenté en figure 1b. Aussi, lors du downmix, les signaux gauche et droite vont soit s'additionner (interférence constructive), soit s'annuler (interférence destructive), créant ainsi des variations de niveau et des effets de coloration du signal réduit $m(t)$. La coloration vient de l'effet du filtrage en peigne : les effets d'interférences constructives et destructives se manifestent dans des bandes de fréquences différentes, modifiant ainsi l'équilibre et donc le timbre du signal issu du downmix par rapport à l'original.

[0017] Pour corriger ce défaut de niveau d'intensité suivant les fréquences, le downmix du codec e-AAC+ implémente une correction de niveau $\gamma(k)$ qui assure que le niveau d'énergie du signal réduit, dans chaque bande de fréquence, reste comparable à celui d'origine :

$$[0018] \quad M(k) = \gamma(k) \cdot \frac{X_1(k) + X_2(k)}{2}, \text{ avec}$$

$$[0019] \quad \gamma(k) = \sqrt{\frac{|X_1(k)|^2 + |X_2(k)|^2}{0.5|X_1(k) + X_2(k)|^2}}$$

[0020] Cette approche permet de compenser dans une certaine mesure la baisse d'intensité du downmix. Cependant, cette compensation mène à des suramplifications lorsque les signaux sont proches de l'opposition de phase : aussi, en pratique, le facteur de correction $\gamma(k)$ est limité à une valeur maximum (par exemple une limite haute de 2).

[0021] La [Fig.7] décrit une méthode de réduction de canaux de stéréo à mono intégrée dans le codec candidat IVAS public disponible à l'adresse :

[0022] <https://forge.3gpp.org/rep/ivas-codec-pc/ivas-codec/-/tree/main>

[0023] Le signal à 2 canaux d'entrée $x(m)$, où m est l'indice des échantillons entrelacés, est désentrelacé (bloc 701) pour trouver les deux canaux gauche et droite. Ensuite une analyse fréquentielle avec fenêtrage et transformée de Fourier (blocs 702 et 703) est effectuée pour obtenir les spectres des 2 canaux, et l'intercorrélacion sur la base des spectres de phase est estimée (bloc 704) avant de déterminer la différence de temps entre les canaux (notée ici ITD même s'il s'agit formellement d'une ICTD entre 2 canaux) par recherche de pic (bloc 705) ; le bloc 705 fournit également le niveau de corrélation R^* correspondant à l'ITD.

[0024] Le facteur de mixage g est ensuite déterminé comme suit (bloc 706) :

$$[0025] \quad g = \begin{cases} \frac{1}{2} - \frac{R^*}{2} & \tau > 0 \\ \frac{1}{2} + \frac{R^*}{2} & \tau < 0 \\ \frac{1}{2} & \tau = 0 \end{cases}$$

[0026] où g représente le facteur de mélange (mixage) et τ est la différence de temps intercanal (appelée abusivement ITD) Ainsi, les 2 canaux sont combinés en un signal mono dans le bloc 706 selon ce facteur g , avec un lissage de la valeur du facteur appliqué échantillon par échantillon à chaque début de trame après le bloc 707:

$$[0027] \quad m(n) = \gamma(n) x_1(n) + (1-\gamma(n)) x_2(n),$$

[0028] où $\gamma(n)$ est le facteur dérivé de g après lissage temporel avec le gain de la trame précédente.

[0029] Ensuite, les énergies des canaux gauche $x_1(n)$ et droit $x_2(n)$ et du signal mono $m(n)$ sont déterminées (bloc 708, 709, 710). Les canaux gauche et droite sont séparément réinjectés (ajoutés) dans les blocs 713 et 715 au signal mono issu du bloc 706 en fonction d'une compensation d'énergie déterminée dans le bloc 711 avec des facteurs d'échelle respectifs définis dans les blocs 712 et 713.

[0030] Ces méthodes, en sus de la coloration du signal issu du downmix, liée au filtrage en peigne imparfaitement corrigé, souffrent d'un surplus de réverbération liée à la sommation de signaux lorsque les canaux d'entrée sont en déphasage, et d'un artefact de type « clics » lorsque la réinjection de canaux est effectuée ponctuellement dans une seule trame. La compensation d'un retard identique pour toutes les fréquences n'est pas réaliste, ce retard étant lié aux différentes sources composant le contenu. En pratique, cela amène, outre une coloration supplémentaire, une perte potentielle d'intelligibilité des sources, défaut que la compensation de niveau ne peut corriger.

[0031] D'autres approches de downmix cherchent à éviter les défauts précités. Le principe de ces méthodes est de remettre en phase les signaux gauche et droite avant leur sommation. Dans le dans le document intitulé "*A stereo to mono downmixing scheme for MPEG-4 parametric stereo encoder*" par Samsudin, E. Kurniawati, N. Boon Poh, F. Sattar, S. George, dans Proc. ICASSP, 2006, il est proposé une méthode qui consiste à appliquer, dans le domaine fréquentiel, un déphasage ϕ sur l'une des voies, généralement la droite, afin de la remettre en phase avec la voie gauche, tel que :

$$[0032] \quad M(k) = \frac{X_1(k) + X_2(k) \cdot e^{j\phi(k)}}{2}$$

[0033] Le déphasage idéal est donné par ce qu'on appelle ici l'IPD pour « *Inter-channel Phase Difference* » en anglais :

$$[0034] \quad IPD(k) = \angle(X_1(k), X_2^*(k))$$

[0035] où X^* est le conjugué de X et \angle indique la phase de l'opérande complexe.

[0036] Il est calculé dans le domaine fréquentiel : cela permet de remettre en phase les com-

posantes spectrales de différentes sources, une source avec son propre IPD pouvant être prépondérante dans une bande de fréquence k , tandis qu'une autre source avec un IPD différent peut être prépondérante dans une autre fréquence k' .

[0037] Dans le cas du codec MPEG4 décrit dans le document Samsudin cité ci-dessus, l'IPD appliqué est celui estimé par moyennage sur une bande de Bark et non pas l'IPD calculé pour chaque bande de fréquence, ces derniers s'avérant particulièrement bruités et variables d'une trame à l'autre. De plus, cette méthode prend comme référence de phase le canal gauche, et si la phase de ce canal est mal conditionnée, le downmix a une qualité dégradée.

[0038] Dans la demande de brevet publiée WO2017103418, il est proposé une méthode de downmix qui fait un mélange entre la méthode proposée par Samsudin et le downmix passif. Notamment, il propose un indicateur ISD pour « Inter-Spectral Distance » qui permet de choisir, bande de fréquence par bande de fréquence, quelle est la méthode la plus adaptée à la remise en phase des signaux :

$$[0039] \quad ISD(k) = \frac{|X_1(k) - X_2(k)|}{|X_1(k) + X_2(k)|}$$

[0040] Cet indicateur permet de mesurer si les signaux sont plutôt en phase ($ISD < 1$), c'est-à-dire avec une différence de phase ou IPD dans l'intervalle $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ ou plutôt en opposition de phase ($ISD > 1$), i.e. avec une IPD dans l'intervalle $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$.

Lorsqu'ils sont en opposition de phase, une grande valeur d'ISD est synonyme d'un niveau comparable entre la voie gauche et la voie droite et la remise en phase est importante pour éviter les effets de coloration et de perte de niveau. A contrario, une valeur d'ISD proche de 1 est représentative d'un canal prépondérant par rapport à l'autre ; dans ce cas dernier cas, la remise en phase n'a pas d'importance, le downmix étant à peu près égal au signal prépondérant. Il est, cependant, important d'éviter de modifier la phase du canal droit si celui-ci est prépondérant. Aussi, dans cette demande de brevet, il est proposé un mécanisme de sélection de type de downmix suivant :

$$[0041] \quad \begin{cases} M(k) = \frac{X_1(k) + X_2(k) e^{jIPD(k)}}{2}, & \text{si } ISD(k) > 1.3 \\ M(k) = \frac{X_1(k) + X_2(k)}{2}, & \text{si } ISD(k) < 1.3 \end{cases}$$

[0042] Le downmix passif proposé dans la demande de brevet WO2017103418, pour les situations de faible ISD, introduit parfois une réverbération accrue.

[0043] D'autre part, la méthode proposée dans cette demande de brevet nécessite d'appliquer un traitement dans le domaine fréquentiel, la reconstruction du signal réduit étant réalisé à partir de la transformée inverse de Fourier à court terme (TFCT ou STFT en anglais) du signal $M(t,k)$. Ce type de filtrage engendre une convolution circulaire qui crée des artefacts audibles : écho, pré-écho, grésillements. Pour masquer

ces artefacts, on peut utiliser une implémentation avec un recouvrement de trames (overlap) associé à un fenêtrage adapté (fenêtre à l'analyse et/ou à la synthèse) garantissant une reconstruction du signal filtré. De manière générale, ce recouvrement n'est pas compatible avec le fonctionnement des codeurs audio qui travaillent avec des trames adjacentes, i.e. sans recouvrement ; de plus, ce type de reconstruction induit un retard, généralement d'une demi-trame qui est le recouvrement classique, ce qui n'est pas acceptable pour un codeur qui a des contraintes fortes en termes de latence.

[0044] Une autre implémentation de type addition-recouvrement (OLA pour « *Overlap-and-Add* » en anglais) est également possible : cette méthode nécessite une étape de remplissage par des 0 (« *0-padding* ») sur les signaux et les filtres pour éviter la convolution circulaire. Cependant, les tests montrent qu'avec de tels filtres dont la phase varie très rapidement d'une trame à l'autre, une approche de type OLA ne permet pas de masquer complètement les artefacts à la transition de certaines trames, les artefacts restant alors audibles.

[0045] Il existe aussi dans l'état des méthodes de l'état de l'art où le downmix est obtenu en commutant, par exemple dans le domaine fréquentiel pour la demande de brevet publiée WO2017103418, des méthodes de downmix différentes. Dans ce cas il est important de s'assurer que la commutation se fait « sans couture », c'est-à-dire sans discontinuités ou différences de niveaux entre méthodes pour éviter des artefacts.

Exposé de l'invention

[0046] L'invention vient améliorer l'état de la technique.

[0047] A cet effet, l'invention vise un procédé de traitement de réduction de canaux d'un signal stéréophonique pour obtenir un signal monophonique, de traitement de réduction de canaux d'un signal stéréophonique pour obtenir un signal monophonique, comprenant ce qui suit :

[0048] - application, à une trame courante dudit signal, de deux traitements de réduction de canaux, l'un utilisant un filtrage du signal stéréophonique et l'autre non,

[0049] - sélection d'un desdits traitements à appliquer à la trame courante, ladite sélection étant mise en œuvre en fonction de :

[0050] -- une présence ou non d'au moins une transition dans le signal stéréophonique, ou

[0051] -- un niveau de qualité du filtrage du signal stéréophonique utilisé lors de la mise en œuvre du traitement utilisant le filtrage du signal stéréophonique.

[0052] L'invention fournit avantageusement deux critères de décision qui permettent de choisir entre deux traitements différents de réduction de canaux, dont l'un utilise un filtrage du signal stéréophonique et l'autre non.

[0053] Dans un mode de réalisation, si une présence de ladite au moins une transition est détectée, respectivement non détectée, dans le signal stéréophonique, le traitement de

réduction de canaux qui n'utilise pas un filtrage du signal stéréophonique, respectivement qui utilise un filtrage du signal stéréophonique, est sélectionné.

[0054] Ce mode de réalisation permet :

[0055] - lorsqu'une ou plusieurs transitions sont détectées, d'obtenir un meilleur rendu sur la ou les transition(s) du signal en sélectionnant le traitement de réduction de canaux qui n'utilise pas de filtrage, et qui, de ce fait, préserve la nature du signal stéréophonique,

[0056] - lorsqu'aucune transition n'est détectée, de privilégier le traitement de réduction de canaux qui utilise le filtrage, mieux adapté à des composantes du signal stéréophonique qui ne sont pas en phase.

[0057] Dans un mode de réalisation, la détection de ladite au moins une transition est mise en œuvre avant application, audit signal stéréophonique, de l'un ou l'autre des deux traitements de réduction de canaux.

[0058] Ce mode de réalisation permet, même avant l'application d'un traitement de réduction de canaux, de déduire à priori qu'au moins une transition est présente ou non dans le signal stéréophonique, sans application d'un traitement particulier supplémentaire sur ce signal.

[0059] Dans un mode de réalisation, la détection de ladite au moins une transition est mise en œuvre comme suit :

[0060] - calcul de l'énergie du signal stéréophonique, dite énergie d'entrée, avant application du filtrage,

[0061] - calcul de l'énergie du signal monophonique, dite énergie de sortie, après application du filtrage,

[0062] - calcul d'un rapport entre l'énergie d'entrée et l'énergie de sortie,

[0063] - comparaison dudit rapport à un seuil d'énergie, et

[0064] - si ledit rapport est inférieur, respectivement supérieur, audit seuil d'énergie, le traitement de réduction de canaux qui n'utilise pas un filtrage du signal stéréophonique, respectivement qui utilise un filtrage du signal stéréophonique, est sélectionné.

[0065] Dans un mode de réalisation, ledit niveau de qualité de filtrage est comparé à un seuil de qualité de filtrage, et si ledit niveau de qualité de filtrage est inférieur, respectivement supérieur, audit seuil de qualité de filtrage, le traitement de réduction de canaux qui n'utilise pas un filtrage du signal stéréophonique, respectivement qui utilise un filtrage du signal stéréophonique, est sélectionné.

[0066] Dans un mode de réalisation, est généré un indicateur représentatif du résultat de ladite détection, de la comparaison dudit rapport à un seuil d'énergie, ou de la comparaison du niveau de qualité de filtrage à un seuil de qualité de filtrage.

[0067] Dans un mode de réalisation, deux niveaux de ladite trame courante traitée sont obtenus respectivement après application, à ladite trame courante dudit signal, des

deux traitements de réduction de canaux, l'un utilisant un filtrage du signal stéréophonique et l'autre non,

- ajustement dudit niveau de la trame courante sur laquelle a été appliqué le traitement de réduction de canaux utilisant ledit filtrage, ledit niveau ajusté étant similaire ou égal audit niveau de la trame courante sur laquelle a été appliqué le traitement de réduction de canaux n'utilisant pas ledit filtrage,

[0068] - comparaison entre la trame courante traitée ayant ledit niveau ajusté et la trame courante traitée ayant ledit niveau obtenu près application, à ladite trame courante dudit signal, du traitement de réduction de canaux n'utilisant pas un filtrage du signal stéréophonique,

[0069] - sélection d'un desdits traitements à appliquer à la trame courante, basée sur ladite comparaison.

[0070] L'invention permet avantageusement, lorsqu'un changement de traitement de réduction de canaux d'une trame courante du signal stéréophonique est mis en œuvre, de maintenir le niveau du signal stéréophonique le plus stable possible, sans entraîner de discontinuités sur le niveau perçu du signal.

[0071] L'invention vise un dispositif de de traitement de réduction de canaux comportant un circuit de traitement pour la mise en œuvre des étapes du procédé de traitement de réduction de canaux tel que décrit précédemment.

[0072] L'invention concerne également un programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé de traitement de réduction de canaux conforme à l'invention, selon l'un quelconque des modes particuliers de réalisation décrits précédemment, lorsque ledit programme est exécuté par un processeur.

[0073] De telles instructions peuvent être stockées durablement dans un support mémoire non transitoire du dispositif de de traitement de réduction de canaux mettant en œuvre le procédé de de traitement de réduction de canaux selon l'invention.

[0074] Ce programme peut utiliser n'importe quel langage de programmation, et être sous la forme de code source, code objet, ou de code intermédiaire entre code source et code objet, tel que dans une forme partiellement compilée, ou dans n'importe quelle autre forme souhaitable.

[0075] L'invention vise également un support d'enregistrement ou support d'informations lisible par un ordinateur, et comportant des instructions d'un programme d'ordinateur tel que mentionné ci-dessus.

[0076] Le support d'enregistrement peut être n'importe quelle entité ou dispositif capable de stocker le programme. Par exemple, le support peut comporter un moyen de stockage, tel qu'une ROM, par exemple un CD ROM ou une ROM de circuit microélectronique, ou encore un moyen d'enregistrement magnétique, par exemple un support mobile, un disque dur ou un SSD.

- [0077] D'autre part, le support d'enregistrement peut être un support transmissible tel qu'un signal électrique ou optique, qui peut être acheminé via un câble électrique ou optique, par radio ou par d'autres moyens, de sorte que le programme d'ordinateur qu'il contient est exécutable à distance. Le programme selon l'invention peut être en particulier téléchargé sur un réseau, par un exemple un réseau de type Internet.
- [0078] Alternativement, le support d'enregistrement peut être un circuit intégré dans lequel le programme est incorporé, le circuit étant adapté pour exécuter ou pour être utilisé dans l'exécution du procédé de de traitement de réduction de canaux précité.
- [0079] Selon un exemple de réalisation, la présente technique est mise en œuvre au moyen de composants logiciels et/ou matériels. Dans cette optique, le terme « dispositif » ou « module » peut correspondre dans ce document aussi bien à un composant logiciel, qu'à un composant matériel ou à un ensemble de composants matériels et logiciels.

Brève description des dessins

- [0080] D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description suivante de modes de réalisation particuliers, donnés à titre de simples exemples illustratifs et non limitatifs, et des dessins annexés, parmi lesquels :
- [0081] [Fig.1a] illustre des canaux d'un signal stéréophonique, en phase tel que décrit précédemment ;
- [0082] [Fig.1b] illustre des canaux d'un signal stéréophonique, en opposition de phase tel que décrit précédemment ;
- [0083] [Fig.2] illustre sous forme de schéma bloc une chaîne de traitement de réduction de canaux dans un mode de réalisation de l'invention ;
- [0084] [Fig.3] illustre un mode de réalisation d'un calcul de filtre pour le traitement de réduction de canaux, ainsi qu'une sélection de traitement de réduction de canaux ;
- [0085] [Fig.4] illustre un mode de réalisation d'une phase d'adaptation d'un filtre de traitement de réduction de canaux ;
- [0086] [Fig.5] illustre un mode de réalisation d'une transition de l'application du traitement de réduction de canaux lors d'une transition entre deux trames de signal stéréo ;
- [0087] [Fig.6a] illustre un autre mode de réalisation d'un calcul de filtre pour le traitement de réduction de canaux, ainsi qu'une sélection de traitement de réduction de canaux ;
- [0088] [Fig.6b] illustre un autre mode de réalisation d'un calcul de filtre pour le traitement de réduction de canaux, ainsi qu'une sélection de traitement de réduction de canaux ;
- [0089] [Fig.7] illustre une méthode de traitement de réduction de canaux existant décrit précédemment ;
- [0090] [Fig.8] illustre un exemple de réalisation structurelle d'un dispositif de traitement de réduction de canaux selon un mode de réalisation de l'invention.

Description des modes de réalisation

[0091] L'invention va maintenant être décrite ci-dessous en référence aux figures 2 à 8. A cet effet, les symboles suivants seront principalement utilisés dans la description qui suit.

INDICES

- [0092] k indice fréquentiel
- [0093] l indice temporel d'un signal (dans la trame courante)
- [0094] n indice temporel d'un signal
- [0095] t numéro de trame

CONSTANTES

- [0096] L nombre d'échantillons dans la trame courante
- [0097] N nombre de canaux d'entrée
- [0098] M nombre de canaux de sortie
- [0099] P longueur de filtre
- [0100] Q longueur de fondu enchaîné (crossfade)

SIGNAUX

- [0101] $x_i(n)$ canaux d'entrée
- [0102] $m(n)$ downmix (en temporel)
- [0103] $m(t,l)$ downmix dans la trame courante (en temporel)
- [0104] $M(t,k)$ downmix dans la trame courante (en fréquentiel)
- [0105] La [Fig.2] montre un exemple d'une chaîne de traitement d'un signal audio stéréophonique, le traitement comportant un traitement de réduction de canaux dit « *downmix* ».
- [0106] A l'entrée de cette chaîne de traitement, un signal stéréophonique x , composé de deux canaux ($x_1(n)$ et $x_2(n)$) encore appelés voie gauche et voie droite (respectivement), est dans un premier temps découpé en 201, en trames de L échantillons ($x_1(t,l)$ et $x_2(t,l)$, t étant l'indice de la trame et l l'indice de l'échantillon. Le bloc 202 applique un fenêtrage et une transformée de type FFT (pour « *Fast Fourier Transform* » en anglais), pour obtenir des signaux dans le domaine fréquentiel ($X_1(t,k)$ et $X_2(t,k)$, k étant l'indice de la fréquence). En 203, une sélection et une détermination d'un filtre de réduction de canaux à appliquer à une trame de signal sont ensuite effectuées. Cette étape sera décrite en référence à la [Fig.3]. Le filtre ainsi déterminé est conditionné (ou adapté) en 204 pour le rendre causal et l'optimiser afin de diminuer la complexité de traitement. Dans un exemple de réalisation, le fenêtrage et la transformée de type FFT sont similaires aux étapes 702 et 703 de la [Fig.7].
- [0107] Cette phase d'adaptation est décrite en référence à la [Fig.4].
- [0108] Une fois le filtre déterminé et adapté pour la trame courante, il est appliqué en 205 à cette trame courante. Afin d'éviter des artefacts audibles dus à un changement de

filtres entre deux trames, un fondu enchainé est effectué en prenant en compte les filtres de la trame précédente $\begin{cases} h_1(t-1, l) \\ h_2(t-1, l) \end{cases}$.

[0109] Cette étape sera décrite en référence à la [Fig.5].

[0110] La [Fig.3] décrit un mode de réalisation détaillé du bloc 203 de la [Fig.2]. Dans ce bloc de traitement, il est sélectionné et déterminé un premier ou un deuxième traitement de réduction de canaux à appliquer à une trame courante du signal stéréophonique.

[0111] Une première méthode de traitement de réduction de canaux T1 est ici définie en décalant les canaux gauche et droit du signal stéréophonique simultanément d'un angle défini par la moitié de la différence de phase (IPD) déterminée entre les deux canaux du signal stéréo. On a ainsi un signal de downmix, de la façon suivante :

$$[0112] \quad M(t, k) = \frac{X_1(t, k).e^{+j\frac{IPD(t, k)}{2}} + X_2(t, k).e^{-j\frac{IPD(t, k)}{2}}}{2}$$

[0113] Cette méthode permet d'éviter une dégradation de type surplus de réverbération généré par un downmix classique (moyenne des deux signaux) ou de type filtrage en peigne en cas de décalage temporel entre les deux voies.

[0114] Ce downmix ainsi défini peut-être vu, dans le domaine fréquentiel, comme un filtrage selon l'équation suivante :

$$[0115] \quad M(t, k) = H_1(t, k).X_1(t, k) + H_2(t, k).X_2(t, k)$$

$$[0116] \quad \text{avec } H_1(t, k) = \frac{e^{+j\frac{IPD(t, k)}{2}}}{2} \text{ et } H_2(t, k) = \frac{e^{-j\frac{IPD(t, k)}{2}}}{2}.$$

[0117] Une deuxième méthode de downmix T2 est par exemple définie comme celle proposée dans le document Samsudin précédemment cité.

$$[0118] \quad \text{Dans cette deuxième méthode, on a alors } H_1(t, k) = \frac{1}{2} \text{ et } H_2(t, k) = \frac{e^{jIPD(t, k)}}{2}.$$

[0119] Les canaux du signal stéréo, dans le domaine fréquentiel ($X_1(t, k)$ et $X_2(t, k)$), servent d'une part à calculer un indicateur de phase ISD en 301, représentatif d'une mesure de degré d'opposition de phase entre les canaux du signal stéréo, et, d'autre part, la différence de phase entre les canaux (IPD) en 303.

[0120] L'indicateur de phase est par exemple défini par l'indicateur ISD pour « *Inter-Spectral Distance* » tel que défini ci-dessus.

[0121] Dans le mode de réalisation décrit ici, l'ISD est déterminé par trame de signal stéréophonique afin que la décision de sélection d'un premier traitement de réduction de canaux (premier downmix) ou d'un deuxième puisse se faire trame par trame.

[0122] Cette décision par trame permet de s'affranchir des déphasages importants entre les fréquences et d'autres d'artefacts audibles qui peuvent apparaître lorsque la décision varie d'une fréquence à l'autre comme dans la demande de brevet citée précédemment.

[0123] Pour appliquer une décision par trame, un critère de décision selon l'indicateur de

phase ISD est défini selon l'équation suivante en 302 :

$$[0124] \quad \tilde{ISD}(t) = \frac{\text{nombre total de raies fréquentielles où } ISD(t,k) > 1.3}{\text{nombre total de raies fréquentielles}}$$

[0125] Ce critère mesure le pourcentage de raies fréquentielles d'indice k pour lesquelles le critère est favorable au downmix avec remise en phase, c'est-à-dire quand $ISD(t, k) > 1.3$. La décision est prise d'appliquer le premier downmix tel que défini ci-dessus, avec remise en phase, lorsque $\tilde{ISD}(t)$ dépasse un certain seuil \tilde{ISD}_{th} , que l'on fixe par exemple à 70%.

[0126] On a alors en 305 la décision suivante de sélection de downmix entre T1 et T2 :

$$[0127] \quad \begin{cases} \tilde{ISD}(t) < \tilde{ISD}_{th} : & H_1(t, k) = \frac{e^{j\frac{IPD(t,k)}{2}}}{2}, & H_2(t, k) = \frac{e^{-j\frac{IPD(t,k)}{2}}}{2} \\ \tilde{ISD}(t) > \tilde{ISD}_{th} : & H_1(t, k) = \frac{1}{2}, & H_2(k, f) = \frac{e^{jIPD(k)}{2}}{2} \end{cases}$$

[0128] En 303, la différence de phase IPD est obtenue entre les deux canaux du signal stéréo par $IPD(t, k) = \angle(X_1(t, k)X_2^*(t, k))$ où \angle indique la phase de l'opérande complexe.

[0129] Cependant, la décision à la trame peut être parfois insuffisante car l'IPD, telle que calculée selon la formule ci-dessus, varie très rapidement dans le temps. Aussi, les filtres

[0130] $H_i(t, k)$ peuvent varier très rapidement d'une trame à l'autre, ce qui engendre des discontinuités parfois audibles lors de la transition d'une trame à l'autre. Pour limiter cet effet, on pourra utiliser une version lissée de l'IPD dans le temps pour calculer les filtres $H_i(t, k)$. Une manière de lisser l'IPD est d'appliquer un filtrage passe-bas de type IIR d'ordre 1, en 304, tel que :

$$[0131] \quad \tilde{IPD}(t, k) = a(k)\tilde{IPD}(t, k) + (1-a(k))IPD(t, k)$$

[0132] Le coefficient d'oubli $a(k)$ est choisi de manière ad-hoc. Il peut être intéressant de choisir un coefficient d'oubli important dans les basses fréquences, typiquement en dessous de 5kHz, car cela permet de garder une cohérence de phase dans la partie basse du spectre d'une trame à l'autre, le spectre des signaux de parole étant stationnaire, notamment dans sa partie basse et l'oreille particulière sensible à la phase dans cette partie du spectre. A contrario, il est intéressant de choisir un faible coefficient d'oubli dans les hautes fréquences : dans les fréquences hautes, typiquement au-dessus de 5kHz, la continuité de phase des signaux de parole est moins marquée et l'oreille est moins sensible aux discontinuités de phase dans cette partie du spectre.

[0133] On pourra choisir un coefficient d'oubli de la forme suivante :

$$[0134] \quad \alpha(k) = \begin{cases} a_{max} & k < k_{low} \\ a_{min} & k > k_{high} \\ a_{min} + \frac{k-k_{low}}{k_{high}-k_{low}}, & k_{low} < k < k_{high} \end{cases}$$

- [0135] Les valeurs a_{max} , a_{min} , k_{high} et k_{low} sont déterminées de façon expérimentale. Dans le cas de trames adjacentes de longueur $L=20$ ms, on pourra par exemple choisir le jeu de paramètres suivant :
- [0136] $a_{max} = 0.94$, $a_{min} = 0.86$, $k_{low} = 0$, k_{high} correspond à $5kHz$.
- [0137] Dans des variantes, on forcera les très basses fréquences et les très hautes fréquences à rester intactes. Ainsi dans des variantes, on force $H_1(t, k) = H_2(t, k) = \frac{1}{2}$ pour les premières raies $k = 0$ ou $k = 0, 1$ ainsi que pour les raies d'indice $k > k_{max}$ où k correspond à 8 kHz.
- [0138] La **figure 4** détaille à présent le bloc 204 de la figure 2. Les filtres $H_i(t, k)$, tels que par exemple définis dans les méthodes de downmix T1 et T2 définies ci-dessus, sont reconditionnés ou encore adaptés pour éviter la convolution circulaire.
- [0139] Dans une première étape, en 401, on calcule les réponses impulsionnelles $h_i(t, l)$ des filtres $H_i(t, k)$, tel que :
- [0140] $h_i(t, l) = FFT^{-1}\{H_i(t, k)\}$, $i = 1, 2$
- [0141] où les filtres $h_i(t, l)$ sont de taille N qui correspond à la longueur de la FFT. Ici on prend $N=L$.
- [0142] Dans le mode de réalisation privilégié, on a $L = 20$ ms, soit 320 échantillons à 16 kHz, 640 échantillons à 32 kHz et 960 échantillons à 48 kHz.
- [0143] Les réponses impulsionnelles ainsi définies sont classiquement non causales, elles correspondent à des filtres à réponse impulsionnelle finie. Une technique classique pour les rendre causales consiste à effectuer une rotation de $L/2$ échantillons pour centrer la réponse en $L/2$. L'inconvénient est que le filtre ainsi reconstruit présente alors une latence de $L/2$. Pour éviter cette latence, on choisit de tronquer le filtre de façon asymétrique, en 402, en ne conservant qu'une portion de la première moitié de la réponse impulsionnelle $h_i(t, l)$ de la manière suivante :
- [0144] $\tilde{h}_i(t, l) = h_i(t, l)$, $0 \leq l < P \leq L/2$
- [0145] Ce filtre présente l'avantage d'avoir généralement un maximum (en valeur absolue) sur son premier échantillon, ce qui garantit un traitement sans retard additionnel. En termes de réponse en fréquence, il présente une phase à peu près égale au filtre optimal de remise en phase $h_i(t, l)$.
- [0146] Le choix de P dépend de différents critères : une valeur de P proche de $N/2$ garantit une remise en phase quasi optimale, tandis qu'une valeur plus réduite permet de limiter la puissance de calcul nécessaire pour le filtrage. Une valeur réduite de P a également pour conséquence de lisser la phase du filtre. Cela présente un avantage avec des filtres définis par le traitement T2 tel que défini ci-dessus où la phase varie très rapidement d'une fréquence à l'autre : ces variations peuvent se traduire par un délai de groupe

très élevé et perceptible par l'oreille humaine. Une faible valeur de P permet de supprimer toute ou partie de ces artefacts.

[0147] Le filtre ainsi tronqué $\tilde{h}_i(t, l)$, notamment lorsque $P < N/2$, voire $P \ll N/2$, présente une queue de réponse impulsionnelle qui ne tend pas vers 0. Cela crée des discontinuités audibles de type cliquetis à la transition entre 2 trames. Aussi, pour éviter ces artefacts audibles, on pondère $\tilde{h}_i(t, l)$ avec une fenêtre $w(l)$, en 403, tel que :

$$[0148] \quad \check{h}(t, l) = w(l) \tilde{h}_i(t, l)$$

[0149] où les coefficients de $w(l)$ tendent vers 0 lorsque l tend vers P. Cela permet de réduire l'énergie des échantillons de la fin du filtre et d'éviter des discontinuités lors de la transition entre des filtres de trames consécutives.

[0150] Cette fenêtre peut être une demi-fenêtre triangulaire, ou encore une demi-fenêtre de Hann :

$$[0151] \quad w(l) = \begin{cases} 1 - l/P \\ \cos(l\pi/2P) \\ \dots \end{cases}$$

[0152] Du fait de sa troncature et du fenêtrage, les filtres $\tilde{h}_i(t, l)$ possèdent une énergie réduite par rapport à celle du filtre optimal $h_i(t, l)$: ce dernier, en tant que filtre déphaseur pur, possède une norme 1/2 par construction. Aussi, on pourra procéder à une normalisation de l'énergie du filtre ainsi tronqué et fenêtré, en 404, tel que :

$$[0153] \quad \check{h}_i(t, l) = \frac{\tilde{h}_i(t, l)}{2 \|\tilde{h}_i(t, l)\|}$$

[0154] Où $\|x(\cdot)\|$ est la norme 2 d'un vecteur.

[0155] Selon l'invention, en 405, est déterminé un niveau de qualité du filtrage qui a été mis en œuvre. A cet effet :

[0156] - une énergie E_{401} du filtre avant troncature 402 est calculée,

[0157] - une énergie E_{403} du filtre temporel après fenêtrage 403 est calculée,

[0158] - un rapport R_4 entre E_{401} et E_{403} est calculé.

[0159] On peut également déterminer une qualité de filtrage en comparant une énergie E_{401} et une énergie E_{404} du filtre temporel après normalisation.

[0160] A titre d'alternatives non exhaustives, la détermination du niveau de qualité de filtrage peut être mise en œuvre :

[0161] - directement sur le filtre $H_i(t, k)$, et/ou

[0162] - une version fréquentielle de ce même filtre avant normalisation $\check{H}_i(t, k)$ ou après normalisation $\check{H}_i(t, k)$.

[0163] - etc...

[0164] Dans des variantes, on pourra calculer :

$$[0165] \quad \sum_{l=0}^{P-1} (h_i(t, l) - \check{h}_i(t, l))^2$$

[0166] Ce critère pourra être normalisé par :

$$[0167] \sum_l h_i(t, l)^2$$

[0168] Une autre variante de qualité pourra s'intéresser à la mesure de la quantité d'énergie perdue lors de la troncature pour rendre le filtre causal. Dans cette variante, on pourra calculer E_{401} comme l'énergie « anticausale » du filtre original :

$$[0169] E_{401}(t) = \sum_{l=N/2}^{N-1} h_i^2(t, l)$$

[0170] L'énergie E_{403} (resp. E_{404}) pourra être calculée comme l'énergie du filtre synthétisé, avant normalisation (resp après normalisation) :

$$[0171] E_{403}(t) = \sum_{l=1}^{P-1} \tilde{h}_i^2(t, l)$$

$$[0172] E_{404}(t) = \sum_{l=1}^{P-1} \hat{h}_i^2(t, l)$$

[0173] Pour éviter le biais lié à la taille de la fenêtre, on pourra calculer également l'énergie de la partie anticausale sur le même nombre d'échantillons, avec fenêtrage par la fenêtre $w(l)$ ou non :

$$[0174] E_{401}(t) = \sum_{l=N/2-P}^{N-1} h_i^2(t, l)$$

$$[0175] E_{401}(t) = \sum_{l=N-P}^{N-1} w(N-l+P) h_i^2(t, l)$$

[0176] Dans des variantes, on pourra calculer la réponse en fréquence $\tilde{H}_i(t, k)$ du filtre obtenu après troncature (en sortie de 403) et comparer avec la réponse cible $H_i(t, k)$ pour définir le niveau de qualité, par exemple :

$$[0177] \sum_k |H_i(t, k) - \tilde{H}_i(t, k)|^2$$

[0178] Ce critère pourra être normalisé par :

$$[0179] \sum_k |H_i(t, k)|^2$$

[0180] Une autre variante du critère de qualité pourra consister à comparer la cohérence de phase des filtres, étant donné que l'objectif est de remettre en phase des signaux. Ainsi, on pourra calculer la cohérence $C_{H_i \tilde{H}_i}(t)$ entre les 2 filtres :

$$[0181] C_{H_i \tilde{H}_i}(t) = \frac{1}{N} \left| \sum_k \frac{H_i(t, k) \tilde{H}_i^*(t, k)}{|H_i(t, k) \tilde{H}_i^*(t, k)|} \right|$$

[0182] Une cohérence élevée (i.e proche de 1) indiquant que le filtre synthétisé est proche du filtre idéal et donc un facteur de qualité élevé, une valeur faible (proche de 0) indiquant au contraire un critère de qualité faible.

[0183] La [Fig.5] décrit à présent en détails le bloc 205 de la [Fig.2], où une implémentation de type OLS (pour « Overlap-and-Save » en anglais) du filtrage est opérée dans le domaine temporel ou fréquentiel. Le choix du domaine sera réalisé en fonction de la complexité qui dépend de la taille du filtre relativement à celle de la trame à traiter.

[0184] On présente ici une implémentation de type OLS dans le domaine temporel, avec des trames adjacentes de taille L .

[0185] Dans un premier temps, on concatène, en 501, la trame actuelle avec les $P-1$ échantillons de la trame précédente (principe du Save) :

$$[0186] \quad \tilde{x}_i(k, \cdot) = [x_i(tL-P+2), \dots, x_i(tL-1), x_i(tL), \dots, x_i((t+1)L-1)]$$

[0187] Chaque voie est filtrée par son filtre déphaseur en 503 :

$$[0188] \quad y_i(t, l) = \sum_{p=0}^{P-1} h_i(t, p) \tilde{x}_i(t, l+P-p), \quad 0 \leq l < L$$

[0189] Le downmix mono est alors créé par somme des deux voies remises en phase en 505 :

$$[0190] \quad m(t, l) = y_1(t, l) + y_2(t, l), \quad 0 \leq l < L$$

[0191] Dans la pratique, les filtres peuvent être très différents entre deux trames successives, par exemple lorsqu'une nouvelle source apparaît. Dans ce cas, la transition entre $m_{t-1}(t, l)$ et $m(t, l)$ peut occasionner des différences sous forme d'artefacts audibles. Pour éviter ces artefacts, on peut appliquer une étape de fondu enchainé ou « *cross-fade* » en anglais. Pour réaliser ce cross-fade, on applique les filtres $\check{h}_i(t-1, l)$ à la trame actuelle sur un nombre d'échantillons Q :

$$[0192] \quad y_{i,t-1}(t, l) = \sum_{p=0}^{P-1} \check{h}_i(t-1, p) \tilde{x}_i(t, l+P-p), \quad 0 \leq l < Q$$

[0193] Ceci est réalisé par le bloc 502 de la [Fig.5].

[0194] Le nombre d'échantillons nécessaires au cross-fade devra être déterminé expérimentalement, en fonction de la taille des filtres, de la fréquence d'échantillonnage, ...

[0195] On crée ensuite le downmix mono $m_{t-1}(t, l)$ en 504, à partir de filtres de la trame précédente :

$$[0196] \quad m_{t-1}(t, l) = y_{1,t-1}(t, l) + y_{2,t-1}(t, l), \quad 0 \leq l < Q$$

[0197] Le signal issu du downmix $\hat{m}(t, l)$ est créé en 506, à partir d'un cross-fade selon l'équation suivante :

$$[0198] \quad \hat{m}(t, l) = \begin{cases} \beta(l)m_{t-1}(t, l) + (1-\beta(l))m(t, l), & 0 \leq l < Q \\ m(t, l), & Q \leq l < L \end{cases}$$

[0199] Où $\beta(l)$ est une fonction monotone décroissante, avec $\beta(0) = 1$ et $\beta(Q-1) = \varepsilon$ avec ε une valeur proche de 0 ou nulle. Typiquement, on pourra choisir une demi-fenêtre de Hann :

$$[0200] \quad \beta(l) = \cos\left(\frac{l\pi}{2Q}\right), \quad 0 \leq l < Q$$

[0201] Il est à noter que, dans une intégration avec d'autres méthodes de downmix, les méthodes de downmix T1 et T2 précitées ne peuvent pas être toujours représentées par une opération de filtrage. C'est le cas de la méthode T3 de la [Fig.7] ou toute autre

méthode dont le gain dépend du niveau instantané du signal.

- [0202] Selon l'invention, dans le cas où une sélection des méthodes de downmix T1 et T2 ou du downmix T3 est mise en œuvre, comme cela sera décrit dans la suite de la description, une étape de mise à l'échelle 5040 et 5050 de chacun des signaux $m_{t-1}(t, l)$ et $m(t, l)$ respectivement est mise en œuvre. A cet effet, le signal d'(n) obtenu en sortie du downmix T3 de la figure 7 ayant un niveau N3, le niveau N1, respectivement N2, de chacun des signaux $m_{t-1}(t, l)$ et $m(t, l)$ sur lesquels a été appliqué le downmix T1, respectivement T2, est ajusté au niveau N3, de manière à être égal ou similaire à ce niveau N3. Au lieu d'être appliquée à chacun des signaux $m_{t-1}(t, l)$ et $m(t, l)$, une telle mise à l'échelle peut être effectuée également sur le signal $\hat{m}(t, l)$.
- [0203] La [Fig.6a] détaille à présent un procédé de traitement de réduction de canaux conforme à l'invention, dans lequel est mise en œuvre en 610 l'une des méthodes de downmix T1 ou T2 choisi à l'issue du procédé de sélection de la [Fig.3], et est mise en œuvre en 611 une autre méthode de downmix dont le gain dépend du niveau instantané du signal, telle que par exemple la méthode T3 de la [Fig.7].
- [0204] Dans ce cas, une solution est de calculer les downmix de chaque méthode pour la trame courante d'indice temporel l et d'opérer une transition en 612 d'un downmix à l'autre par cross-fade, ce qui génère un signal monophonique d(n). Plus précisément, dans le cas d'une transition d'un downmix T1 (ou T2) au downmix T3, le downmix peut être réalisé entre le signal $\hat{m}(t, l)$ obtenu en sortie de T1 (ou T2) et le signal $m_{T3}(t, l)$ obtenu en sortie de T3. On peut aussi opérer le downmix entre la sortie de T1 (ou T2) et le signal d(n). Dans ce dernier cas, la compensation d'énergie de la méthode T3 (adapt_gain sur la [Fig.7]) s'appliquerait en permanence à toutes les méthodes. La décision du basculement d'un downmix T1 ou T2 à un downmix T3 pourra être motivée par d'autres critères que l'ISD, par exemple des paramètres Par3 déjà calculés par la méthode T3, comme par exemple l'ITD, en choisissant d'appliquer par défaut le downmix T3 lorsque l'ITD est très important (par exemple $ITD > 600$ microsec en valeur absolue), ou toute autre critère ou combinaison de critères.
- [0205] Conformément à l'invention, la sélection du downmix T1 (ou T2) ou T3 est mise en œuvre en fonction de :
- [0206] - une présence ou non d'au moins une transition dans le signal stéréophonique, ou
- [0207] - un niveau de qualité du filtrage du signal stéréophonique utilisé lors de la mise en œuvre du downmix T1 (ou T2).
- [0208] Dans le cas d'une détection 613 de la présence ou non d'au moins une transition dans le signal stéréophonique, selon un premier mode de réalisation, une détection de transitoire est appliquée séparément à chacun des signaux d'entrée $x_1(n)$ et $x_2(n)$, ou sur la trame courante $x_1(t, l)$ et $x_2(t, l)$. La détection de transitoires étant un problème

classique en codage audio, on prend l'exemple du module déjà implémenté dans le codec EVS et décrit dans le standard 3GPP TS 26.445 clause 5.1.8. A cet effet, chacun des signaux d'entrée $x_1(n)$ et $x_2(n)$, ou la trame courante $x_1(t, l)$ et $x_2(t, l)$ est découpée en sous-blocs. L'énergie de chaque sous-bloc est calculée, puis un lissage est éventuellement mis en œuvre. L'énergie obtenue est alors comparée à un seuil d'énergie th_{61} . Si cette énergie obtenue est inférieure à th_{61} , la présence d'une transition est considérée comme non détectée. Si cette énergie obtenue est supérieure à th_{61} , la présence d'une transition est considérée comme détectée. A cet effet, un paramètre Par 1 (si T1 a été choisi à l'issue de la sélection de la [Fig.3]) ou Par2 (si T2 a été choisi à l'issue de la sélection de la [Fig.3]) est mis :

- [0209] - à une première valeur, par exemple 0, pour indiquer que la présence d'une transition n'a pas été détectée,
- [0210] - à une deuxième valeur, par exemple 1, pour indiquer que la présence d'une transition a été détectée.
- [0211] Un tel paramètre Par1/Par2 est transmis, en tant que critère de décision, au bloc 612 de cross-fading.
- [0212] Dans le cas où la sélection du downmix T1 (ou T2) ou du downmix T3 est basée sur un niveau de qualité du filtrage du signal stéréophonique utilisé lors de la mise en œuvre du downmix T1 (ou T2), en référence à la [Fig.4], les étapes suivantes sont mises en œuvre :
- [0213] - calcul de l'énergie E_{61} du signal $\hat{y}_i(t, l)$, issu du filtrage $\hat{x}_1(t, l)$ par $h_i(t, l)$
- [0214] - calcul de l'énergie E_{62} du signal $y_i(t, l)$,
- [0215] - calcul d'un rapport entre E_{61} et E_{62} ,
- [0216] - comparaison dudit rapport à un seuil d'énergie SE.
- [0217] Si le rapport entre E_{61} et E_{62} est inférieur à SE, la présence d'une transition est considérée comme détectée. Si le rapport entre E_{61} et E_{62} est supérieur à SE, la présence d'une transition est considérée comme détectée. A cet effet, un paramètre Par 1 (si T1 a été choisi à l'issue de la sélection de la [Fig.3]) ou Par2 (si T2 a été choisi à l'issue de la sélection de la [Fig.3]) est mis :
- [0218] - à une première valeur, par exemple 0, pour indiquer que la présence d'une transition n'a pas été détectée,
- [0219] - à une deuxième valeur, par exemple 1, pour indiquer que la présence d'une transition a été détectée.
- [0220] Un tel paramètre Par1/Par2 est transmis, en tant que critère de décision, au bloc 612 de cross-fading.
- [0221] Ainsi, selon l'invention, lorsqu'un transitoire est détecté dans la trame courante dans l'un ou les deux canaux, on applique à la trame courante le downmix T3 plutôt que le

downmix T1 (ou T2). En effet les downmix T1 et T2 utilisent un filtrage par une réponse impulsionnelle qui a parfois tendance à étaler l'enveloppe du signal, ce problème est plus prononcé sur des sons transitoires comme des castagnettes, des claquements de ciseaux sur un enregistrement binaural simulant une coupe chez un coiffeur, etc.

[0222] Afin de limiter les artefacts potentiels dus au fondu enchaîné (bloc 612) entre méthodes de downmix, l'invention peut prévoir d'appliquer, en 614, une mise à l'échelle du signal en sortie de T1 (ou T2). Selon l'invention, une telle mise à l'échelle 614 met en œuvre les étapes suivantes :

[0223] - déterminer l'énergie E' du signal de downmix à la sortie de T1 (ou T2), de façon préférentielle avec un lissage, en effectuant les mêmes étapes que les étapes 708, 709, 710 de la [Fig.7],

[0224] - déterminer le facteur d'échelle :

$$[0225] \quad g' = \sqrt{\frac{E_1 + E_2}{2E'}}$$

[0226] - appliquer le facteur g' audit signal de downmix en reprenant le principe de fonctionnement (avec lissage en début de trame) des blocs 712 et 714 de la [Fig.7].

[0227] A la différence du downmix à la [Fig.7], ici on applique une compensation directe du niveau du downmix en sortie de T1 ou T2, sans réinjecter de signal d'entrée.

[0228] La [Fig.6b] illustre une alternative du procédé de traitement de réduction de canaux décrite en référence à la [Fig.6a].

[0229] Le mode de réalisation illustré en [Fig.6b] se distingue de celui de la [Fig.6a] uniquement par le fait que la détection d'une transition est mise en œuvre selon un deuxième mode de réalisation. Selon ce deuxième mode, une telle détection est mise en œuvre au cours de l'étape 610 de downmix T1 (ou T2) au lieu d'être appliquée séparément à chacun des signaux d'entrée $x_1(n)$ et $x_2(n)$, ou sur la trame courante $x_1(t,l)$ et $x_2(t,l)$. A cet effet, en référence à la [Fig.5], on compare l'énergie de chaque sous-bloc de 2 ms entre le signal d'entrée et de sortie des blocs 502 et 503. Si l'énergie a baissé un seuil, par exemple 3 dB, on considère qu'une transitoire a été fortement atténuée, et dans ce cas on applique le downmix T3 dans la trame courante. Si l'énergie n'a pas baissé un seuil, par exemple 3 dB, on considère qu'une transitoire n'a pas été atténuée, et dans ce cas on applique le downmix T1 (ou T2) dans la trame courante.

[0230] On a illustré sur la [Fig.8], un dispositif de traitement de réduction de canaux 800, au sens de l'invention.

[0231] Le dispositif 800 comporte un circuit de traitement incluant typiquement :

- une mémoire MEM1 pour stocker des données d'instructions d'un programme informatique au sens de l'invention ;
- une interface INT1 de réception d'un signal audio stéréo :

$$\begin{cases} x_1(n) \\ x_2(n) \end{cases} ;$$

- un processeur PROC1 pour recevoir ce signal et le traiter en exécutant les instructions de programme informatique que stocke la mémoire MEM1, en vue d'effectuer un traitement de réduction de canaux ; en particulier, le processeur étant apte à piloter les modules de traitement tels que décrits en référence aux figures 2 à 5; et
- une interface de communication COM 1 pour transmettre les signaux réduits, issus du traitement de réduction de canaux, $\hat{m}(t, l)$, à un autre module de traitement, par exemple un module de codage ou de décodage d'un codeur ou d'un décodeur de signal audio.

- [0232] Bien entendu, cette [Fig.8] illustre un exemple d'une réalisation structurelle d'un dispositif de traitement de réduction de canaux au sens de l'invention.
- [0233] Les figures 2 à 7 commentées ci-dessus décrivent en détails des réalisations fonctionnelles de ce dispositif.
- [0234] Les applications de ce type de traitement de réduction de canaux se trouvent par exemple dans le codage audio, par exemple lorsque le distant n'a pas les capacités de restituer sur son terminal un son stéréo. Dans ce cas, il n'est pas nécessaire de transporter un signal stéréo et cela permet d'économiser de la bande passante. Ce type de procédé peut intervenir lors de conversations en point à point si l'un des participants fait une prise son stéréo ou binaurale. Ce type de procédé peut également être présent lors de conférence à plusieurs : le pont de conférence spatialise la scène en créant une scène stéréo, mais tous les participants n'ont pas nécessairement des capacités de restitution stéréo. Il convient alors d'effectuer un downmix de la scène en mono pour ces participants.
- [0235] L'invention peut également s'appliquer pour un décodage audio : dans ce cas, c'est le renderer (module de restitution) de l'utilisateur qui va effectuer le downmix du contenu stéréo pour s'adapter aux capacités réduites de l'utilisateur (simple haut-parleur par exemple).

Revendications

- [Revendication 1] Procédé de traitement de réduction de canaux d'un signal stéréophonique pour obtenir un signal monophonique, comprenant ce qui suit :
- application (610, 611), à une trame courante dudit signal, de deux (T1 ou T2, T3) traitements de réduction de canaux, l'un utilisant un filtrage du signal stéréophonique et l'autre non,
 - sélection (612) d'un desdits traitements à appliquer à la trame courante, ladite sélection étant mise en œuvre en fonction de :
 - une présence ou non d'au moins une transition dans le signal stéréophonique, ou
 - un niveau de qualité du filtrage du signal stéréophonique utilisé lors de la mise en œuvre du traitement utilisant le filtrage du signal stéréophonique.
- [Revendication 2] Procédé de traitement de réduction de canaux selon la revendication 1, dans lequel si une présence de ladite au moins une transition est détectée (613), respectivement non détectée (613), dans le signal stéréophonique, le traitement de réduction de canaux qui n'utilise pas un filtrage du signal stéréophonique, respectivement qui utilise un filtrage du signal stéréophonique, est sélectionné (612).
- [Revendication 3] Procédé de traitement de réduction de canaux selon la revendication 1, dans lequel la détection de ladite au moins une transition est mise en œuvre avant application, audit signal stéréophonique, de l'un ou l'autre des deux traitements de réduction de canaux.
- [Revendication 4] Procédé de traitement de réduction de canaux selon la revendication 1, dans lequel la détection de ladite au moins une transition est mise en œuvre comme suit :
- calcul de l'énergie du signal stéréophonique, dite énergie d'entrée, avant application du filtrage,
 - calcul de l'énergie du signal monophonique, dite énergie de sortie, après application du filtrage,
 - calcul d'un rapport entre l'énergie d'entrée et l'énergie de sortie,
 - comparaison dudit rapport à un seuil d'énergie, et
 - si ledit rapport est inférieur, respectivement supérieur, audit seuil d'énergie, le traitement de réduction de canaux qui n'utilise pas un filtrage du signal stéréophonique, respec-

tivement qui utilise un filtrage du signal stéréophonique, est sélectionné.

- [Revendication 5] Procédé de traitement de réduction de canaux selon la revendication 1, dans lequel ledit niveau de qualité de filtrage est comparé à un seuil de qualité de filtrage, et si ledit niveau de qualité de filtrage est inférieur, respectivement supérieur, audit seuil de qualité de filtrage, le traitement de réduction de canaux qui n'utilise pas un filtrage du signal stéréophonique, respectivement qui utilise un filtrage du signal stéréophonique, est sélectionné.
- [Revendication 6] Procédé de traitement de réduction de canaux selon la revendication 2, 4 ou 5, dans lequel est généré un indicateur (Par1 ou Par2, Par3) représentatif du résultat de ladite détection, de la comparaison dudit rapport à un seuil d'énergie, ou de la comparaison du niveau de qualité de filtrage à un seuil de qualité de filtrage.
- [Revendication 7] Procédé de de traitement de réduction de canaux selon l'une quelconque des revendications 1 à 6, dans lequel deux niveaux (E' , E_1+E_2) de ladite trame courante traitée sont obtenus respectivement après application, à ladite trame courante dudit signal, des deux (T1 ou T2, T3) traitements de réduction de canaux, l'un utilisant un filtrage du signal stéréophonique et l'autre non,
 - ajustement (614) dudit niveau (E') de la trame courante sur laquelle a été appliqué le traitement de réduction de canaux utilisant ledit filtrage, ledit niveau (E') ajusté étant similaire ou égal audit niveau (E_1+E_2) de la trame courante sur laquelle a été appliqué le traitement de réduction de canaux n'utilisant pas ledit filtrage,
 - comparaison entre la trame courante traitée ayant ledit niveau ajusté et la trame courante traitée ayant ledit niveau obtenu près application, à ladite trame courante dudit signal, du traitement de réduction de canaux n'utilisant pas un filtrage du signal stéréophonique,
 - sélection d'un desdits traitements à appliquer à la trame courante, basée sur ladite comparaison.
- [Revendication 8] Dispositif de traitement de réduction de canaux comportant un circuit de traitement pour la mise en œuvre des étapes du procédé de traitement de réduction de canaux selon l'une des revendications 1 à 7.
- [Revendication 9] Support de stockage, lisible par un processeur, mémorisant un programme informatique comportant des instructions pour l'exécution

du procédé de selon l'une des revendications 1 à 7.

[Fig. 1a]

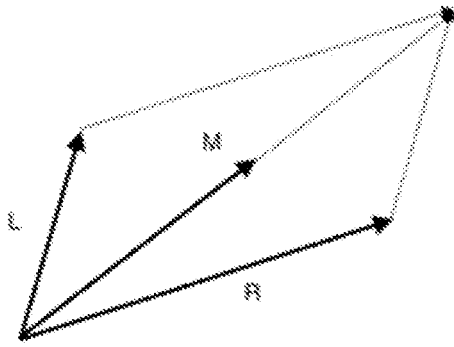


FIG. 1a

[Fig. 1b]



FIG. 1b

[Fig. 2]

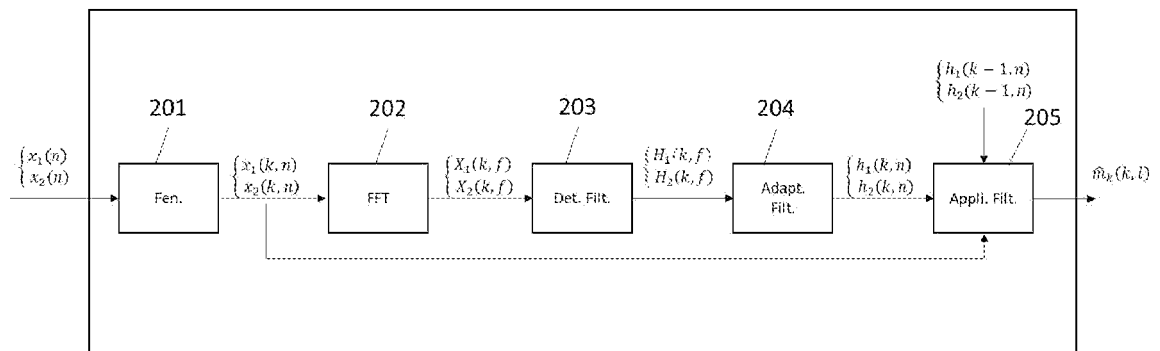


FIG. 2

[Fig. 3]

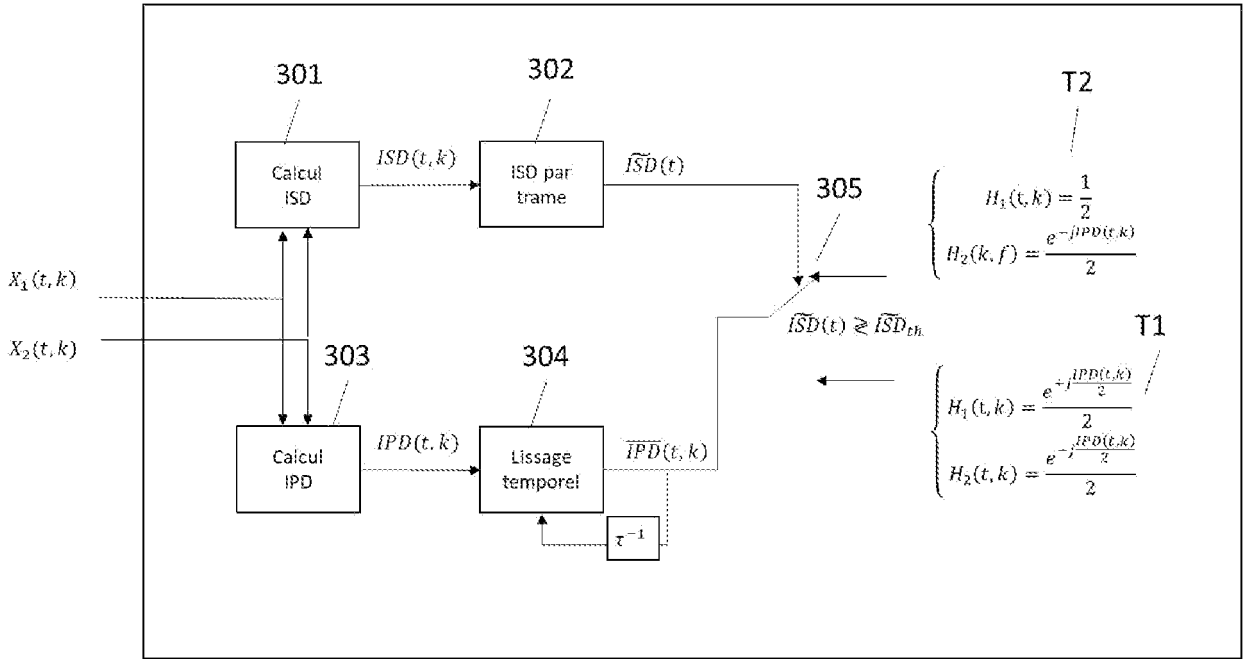


FIG.3

[Fig. 4]

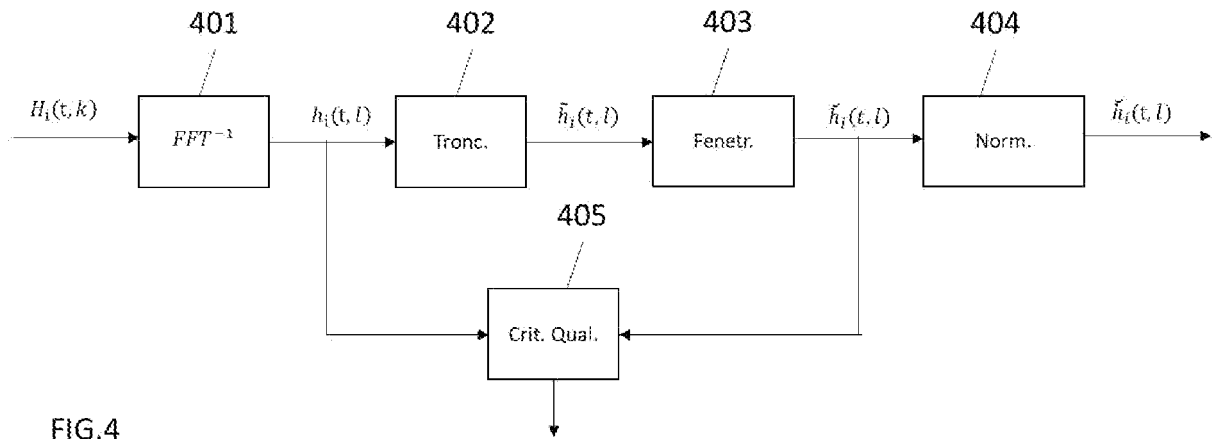


FIG.4

[Fig. 5]

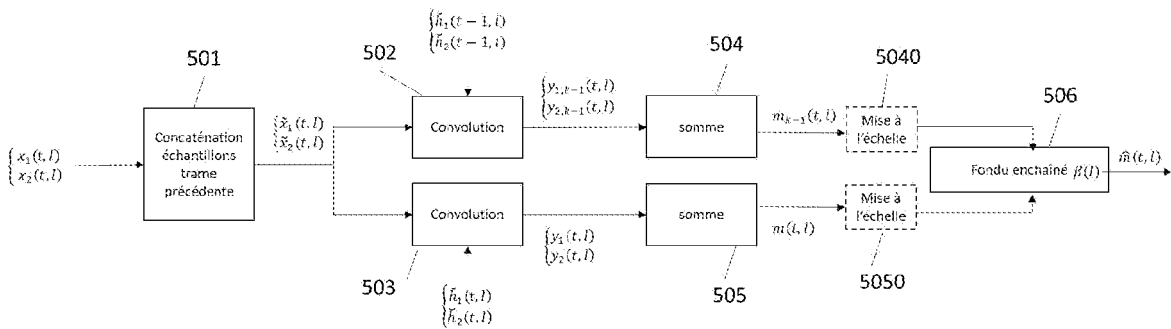


FIG.5

[Fig. 6a]

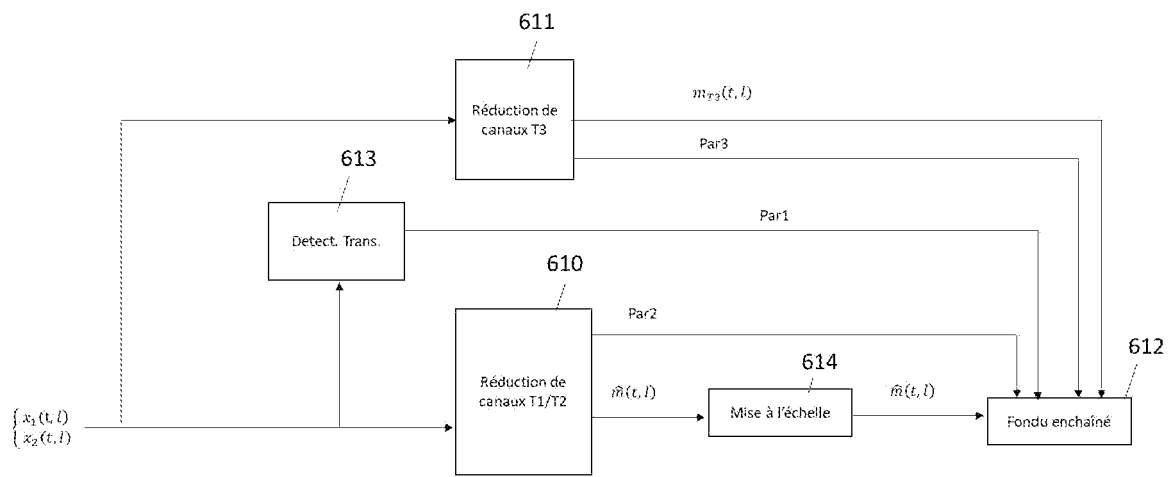


FIG.6a

[Fig. 6b]

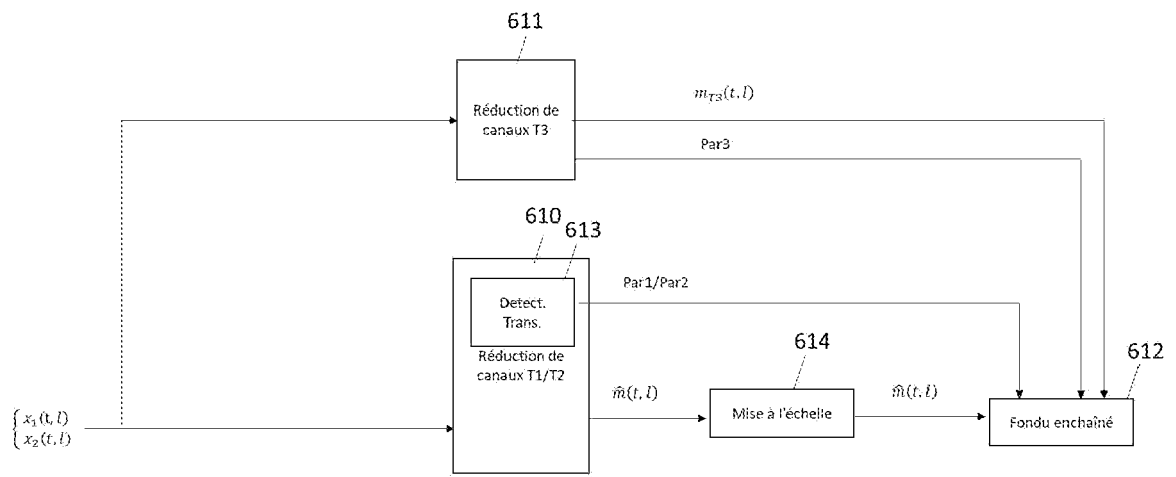


FIG.6b

[Fig. 7]

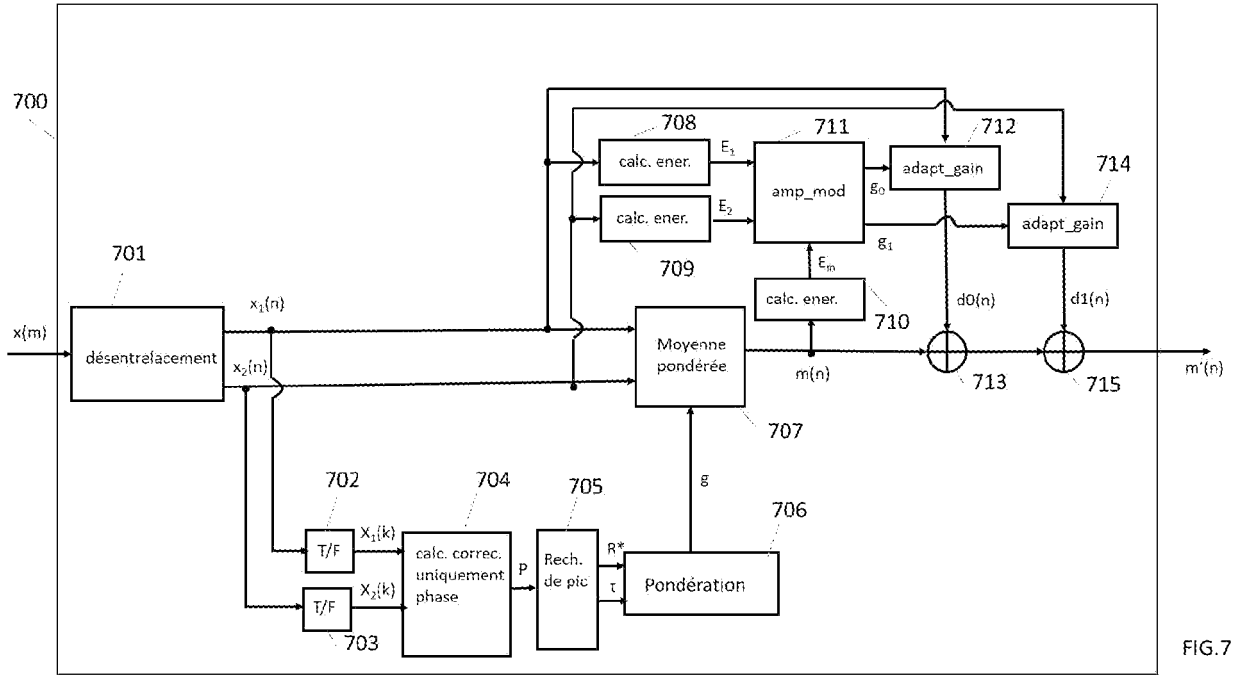


FIG.7

[Fig. 8]

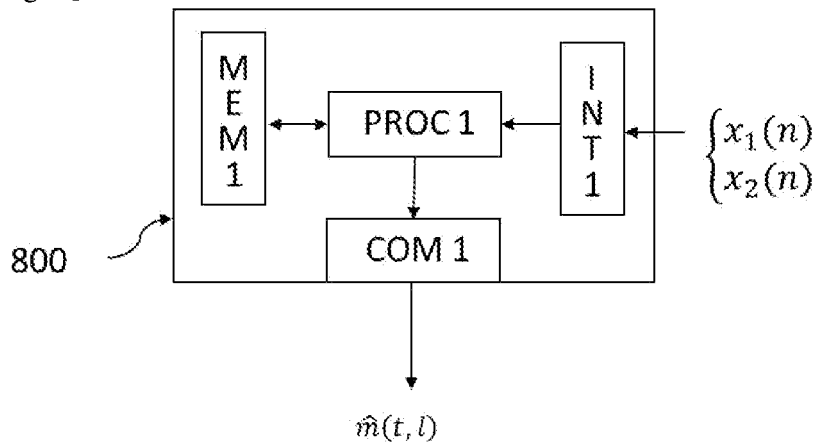


FIG.8

**RAPPORT DE RECHERCHE
PRÉLIMINAIRE**

N° d'enregistrement
national

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

FR 918374
FR 2304288

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X, D A	WO 2017/103418 A1 (ORANGE [FR]) 22 juin 2017 (2017-06-22) * figures 3, 4a *	1-3, 5, 6, 8, 9 4, 7	G10L 19/00 G10L 19/008
X	EP 2 360 683 A1 (DOLBY LAB LICENSING CORP [US]; DOLBY INT AB [NL]) 24 août 2011 (2011-08-24) * figure 7 *	1, 8, 9	
A	EP 3 703 050 A1 (HUAWEI TECH CO LTD [CN]) 2 septembre 2020 (2020-09-02) * figure 4 *	1-9	
A	WO 98/43466 A1 (SONY ELECTRONICS INC [US]) 1 octobre 1998 (1998-10-01) * figure 1 *	1-9	
			DOMAINES TECHNIQUES RECHERCHÉS (IPC)
			G10L
Date d'achèvement de la recherche		Examineur	
24 octobre 2023		Chétry, Nicolas	
CATÉGORIE DES DOCUMENTS CITÉS			
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 bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

**ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 2304288 FA 918374**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.
Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **24-10-2023**
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication	
WO 2017103418	A1	22-06-2017	CN 108369810 A	03-08-2018
			EP 3391370 A1	24-10-2018
			FR 3045915 A1	23-06-2017
			US 2019156841 A1	23-05-2019
			WO 2017103418 A1	22-06-2017

EP 2360683	A1	24-08-2011	AP 3147 A	31-03-2015
			AR 080183 A1	21-03-2012
			AR 089918 A2	01-10-2014
			AU 2011218351 A1	20-10-2011
			BR PI1105248 A2	03-05-2016
			CA 2757643 A1	25-08-2011
			CA 2794029 A1	25-08-2011
			CA 2794047 A1	25-08-2011
			CN 102428514 A	25-04-2012
			CN 103400581 A	20-11-2013
			CO 6501169 A2	15-08-2012
			DK 2360683 T3	16-06-2014
			EA 201171268 A1	30-03-2012
			EC SP11011358 A	31-01-2012
			EP 2360683 A1	24-08-2011
			EP 2698789 A2	19-02-2014
			ES 2467290 T3	12-06-2014
			GE P20146086 B	13-05-2014
			GT 201100246 A	04-04-2014
			HK 1160282 A1	10-08-2012
			HK 1170059 A1	15-02-2013
			HN 2011002584 A	26-01-2015
			HR P20140506 T1	04-07-2014
			IL 215254 A	31-10-2013
			IL 227701 A	31-12-2014
			IL 227702 A	29-01-2015
			JP 5501449 B2	21-05-2014
			JP 5863858 B2	17-02-2016
			JP 2012527021 A	01-11-2012
			JP 2014146040 A	14-08-2014
KR 20120031937 A	04-04-2012			
KR 20130055033 A	27-05-2013			
MA 33270 B1	02-05-2012			
ME 01880 B	20-12-2014			
MY 157229 A	13-05-2016			
NI 201100175 A	14-06-2012			
NZ 595739 A	29-08-2014			
PE 20121261 A1	14-09-2012			
PL 2360683 T3	29-08-2014			
PT 2360683 E	27-05-2014			

EPO FORM P0465

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

**ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 2304288 FA 918374**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.
Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **24-10-2023**
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
		RS 53336 B	31-10-2014
		SG 174552 A1	28-10-2011
		SI 2360683 T1	31-07-2014
		TW 201142826 A	01-12-2011
		TW 201443876 A	16-11-2014
		US 2012016680 A1	19-01-2012
		US 2012237039 A1	20-09-2012
		US 2016035355 A1	04-02-2016
		WO 2011102967 A1	25-08-2011
		ZA 201106950 B	27-12-2012

EP 3703050	A1	02-09-2020	BR 112020010850 A2
			10-11-2020
			CN 109859766 A
			07-06-2019
			EP 3703050 A1
			02-09-2020
			JP 7088450 B2
			21-06-2022
			JP 2021504759 A
			15-02-2021
			KR 20200090856 A
			29-07-2020
			TW 201926318 A
			01-07-2019
			US 2020294513 A1
			17-09-2020
			WO 2019105436 A1
			06-06-2019

WO 9843466	A1	01-10-1998	AT E283621 T1
			15-12-2004
			AU 6658598 A
			20-10-1998
			CN 1257639 A
			21-06-2000
			DE 69827775 T2
			22-12-2005
			EP 0968625 A1
			05-01-2000
			JP 2001518267 A
			09-10-2001
			KR 20000076214 A
			26-12-2000
			US 6005948 A
			21-12-1999
			WO 9843466 A1
			01-10-1998
