



(11) **EP 2 901 447 B1**

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention
de la délivrance du brevet:
21.12.2016 Bulletin 2016/51

(51) Int Cl.:
G10L 21/0308 ^(2013.01) **G10L 19/008** ^(2013.01)
G10L 19/018 ^(2013.01)

(21) Numéro de dépôt: **13770877.2**

(86) Numéro de dépôt international:
PCT/EP2013/069937

(22) Date de dépôt: **25.09.2013**

(87) Numéro de publication internationale:
WO 2014/048970 (03.04.2014 Gazette 2014/14)

(54) **PROCÉDÉ ET DISPOSITIF POUR SÉPARER DES SIGNAUX PAR FILTRAGE SPATIAL À
VARIANCE MINIMUM SOUS CONTRAINTE LINÉAIRE**

VERFAHREN UND VORRICHTUNG ZUR TRENNUNG VON SIGNALEN DURCH FILTERUNG
RÄUMLICHER MINDESTABWEICHUNGEN UNTER LINEAREN EINSCHRÄNKUNGEN

METHOD AND DEVICE FOR SEPARATING SIGNALS BY MINIMUM VARIANCE SPATIAL
FILTERING UNDER LINEAR CONSTRAINT

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

US-A1- 2012 099 732 US-B2- 7 917 336

(30) Priorité: **27.09.2012 FR 1259115**

(43) Date de publication de la demande:
05.08.2015 Bulletin 2015/32

(73) Titulaires:
• **Université Bordeaux 1**
33405 Talence Cedex (FR)
• **Centre National de la Recherche Scientifique**
(C.N.R.S.)
75016 Paris (FR)

- **ANTOINE LIUTKUS ET AL: "Informed source separation through spectrogram coding and data embedding", SIGNAL PROCESSING, vol. 92, no. 8, 1 août 2012 (2012-08-01), pages 1937-1949, XP055062270, ISSN: 0165-1684, DOI: 10.1016/j.sigpro.2011.09.016**
- **LUCAS C PARRA ET AL: "Geometric Source Separation: Merging Convolutional Source Separation With Geometric Beamforming", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 10, no. 6, 1 septembre 2002 (2002-09-01), XP011079661, ISSN: 1063-6676**
- **ANTOINE LIUTKUS ET AL: "Informed audio source separation: A comparative study", SIGNAL PROCESSING CONFERENCE (EUSIPCO), 2012 PROCEEDINGS OF THE 20TH EUROPEAN, IEEE, 27 août 2012 (2012-08-27), pages 2397-2401, XP032254477, ISBN: 978-1-4673-1068-0**
- **STANISLAW GORLOW ET AL: "Informed source separation: Underdetermined source signal recovery from an instantaneous stereo mixture", APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS (WASPAA), 2011 IEEE WORKSHOP ON, IEEE, 16 octobre 2011 (2011-10-16), pages 309-312, XP032011521, DOI: 10.1109/ASPAA.2011.6082312 ISBN: 978-1-4577-0692-9**

(72) Inventeurs:
• **MARCHAND, Sylvain**
29200 Brest (FR)
• **GORLOW, Stanislaw**
33000 Bordeaux (FR)

(74) Mandataire: **Balesta, Pierre**
Cabinet Beau de Loménie
Immeuble Eurocentre
179 Boulevard de Turin
59777 Lille (FR)

(56) Documents cités:

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 2 901 447 B1

Description

[0001] La présente invention concerne un procédé destiné à séparer certains des signaux sources composant un signal global numérique audio. L'invention concerne également un dispositif destiné à mettre en oeuvre ce procédé.

[0002] Le mixage de signaux consiste à sommer plusieurs signaux, appelés signaux sources, pour obtenir un ou plusieurs signaux composites, appelés signaux mixés. Dans les applications audio notamment, le mixage peut consister en une simple étape d'addition des signaux sources ou peut également comprendre des étapes de filtrage des signaux avant et/ou après l'addition. Par ailleurs, pour certaines applications telles que le compact-disc audio, les signaux sources peuvent être mixés de manière différente pour former deux signaux mixés correspondant aux deux voies ou canaux (gauche et droite) d'un signal stéréo.

[0003] La séparation de sources consiste à estimer des signaux sources à partir de l'observation d'un certain nombre de signaux mixés différents formés à partir de ces mêmes signaux sources. L'objectif est généralement de rehausser, voire si possible d'extraire complètement un ou plusieurs signaux sources cibles. La séparation de sources est notamment difficile dans les cas dits « sous-déterminés » dans lesquels on dispose d'un nombre de signaux mixés inférieur au nombre des signaux sources présents dans les signaux mixés. L'extraction est dans ce cas très difficile voire impossible en raison de la faible quantité d'information disponible dans ces signaux mixés par rapport à celle présente dans les signaux sources. Les signaux de musique sur compact-disc audio en sont un exemple particulièrement représentatif car on ne dispose que de deux canaux stéréo (c'est-à-dire deux signaux mixés gauche et droite), généralement très redondants, pour un grand nombre potentiel de signaux sources.

[0004] Il existe plusieurs types d'approches dans la séparation de signaux sources : parmi elles la séparation aveugle, l'analyse de scènes auditives computationnelle, et la séparation basée sur des modèles. La séparation aveugle est la forme la plus générale, dans laquelle aucune information sur les signaux sources ni sur la nature des signaux mixés n'est connue a priori. On fait alors un certain nombre d'hypothèses sur ces signaux sources et les signaux mixés (par exemple que les signaux sources sont statistiquement indépendants) et on estime les paramètres d'un système de séparation en maximisant un critère basé sur ces hypothèses (par exemple en maximisant l'indépendance des signaux obtenus par le dispositif de séparation). Cependant, cette méthode est utilisée généralement dans les cas où l'on dispose de nombreux signaux mixés (au moins autant que de signaux sources) et n'est donc pas applicable aux cas sous-déterminés dans lesquels le nombre de signaux mixés est inférieur au nombre de signaux sources.

[0005] L'analyse de scènes auditives computationnelle consiste généralement en une modélisation des signaux sources en partiels, mais le signal mixé n'est pas décomposé explicitement. Cette méthode se base sur les mécanismes du système auditif humain pour séparer les signaux sources de la même façon que le fait notre oreille. On peut notamment citer : D.P.W. Ellis, Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis, and its application to speech/non-speech mixture (Speech Communication, 27(3), pp. 281-298, 1999), D. Godsmark et G.J. Brown, A blackboard architecture for computational auditory scene analysis (Speech Communication, 27(3), pp. 351-366, 1999), de même que T. Kinoshita, S. Sakai, et H. Tanaka, Musical sound source identification based on frequency component adaptation (In Proc. IJCAI Workshop on CASA, pp. 18-24, 1999). Cependant, l'analyse de scènes auditives computationnelle conduit actuellement à des résultats insuffisants en terme de qualité des signaux sources séparés.

[0006] Une autre forme de séparation repose sur une décomposition du mélange sur une base de fonctions adaptées. Il en existe deux grandes catégories : la décomposition parcimonieuse temporelle et la décomposition parcimonieuse en fréquence.

[0007] Pour la première, il s'agit de décomposer la forme d'onde du mélange, et pour l'autre il s'agit de décomposer sa représentation spectrale, en une somme de fonctions élémentaires appelées « atomes » éléments d'un dictionnaire. Divers algorithmes permettent de choisir le type de dictionnaire et la décomposition correspondante la plus vraisemblable. Pour le domaine temporel, on peut citer notamment : L. Benaroya, Représentations parcimonieuses pour la séparation de sources avec un seul capteur (Proc. GRETSI, 2001), ou P.J. Wolfe et S.J. Godsill, A Gabor régression scheme for audio signal analysis (Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 103-106, 2003). Dans la méthode proposée par Gribonval (R. Gribonval and E. Bacry, Harmonic Decomposition of Audio Signals With Matching Pursuit, IEEE Trans. Signal Proc., 51(1), pp. 101-112, 2003), on classe les atomes de décomposition en sous-espaces indépendants, ce qui permet d'extraire des groupes de partiels harmoniques. Une des restrictions de cette méthode est que des dictionnaires génériques d'atomes tels que les atomes de Gabor par exemple, non adaptés aux signaux, ne donnent pas de bons résultats. De plus, pour que ces décompositions soient efficaces, il faut que le dictionnaire contienne toutes les formes translatées des formes d'ondes de chaque type d'instrument. Les dictionnaires de décomposition doivent alors être extrêmement volumineux pour que la projection et donc la séparation soient efficaces.

[0008] Pour pallier à ce problème d'invariance par translation qui apparaît dans le cas temporel, il existe des approches de décomposition parcimonieuse en fréquence. On peut citer notamment M.A. Casey et A. Westner (Separation of mixed audio sources by independent subspace analysis, Proc. Int. Computer Music Conf., 2000) qui ont introduit l'analyse en sous-espaces indépendants (ISA). Cette analyse consiste à décomposer le spectre d'amplitude à court terme du

signal mixé (calculé par transformée de Fourier à court terme (TFCT)) sur une base d'atomes, et ensuite à regrouper les atomes en sous-espaces indépendants, chaque sous-espace étant propre à une source, pour ensuite resynthétiser les sources séparément. Cependant, cette approche est généralement limitée par plusieurs facteurs : la résolution de l'analyse spectrale par TFCT, la superposition des sources dans ce domaine spectral, et la restriction de la séparation spectrale à l'amplitude (la phase des signaux resynthétisés étant celle du signal mixé). Il est ainsi généralement difficile de représenter le signal mixé comme une somme de sous-espaces indépendants du fait de la complexité de la scène sonore dans le domaine spectral (imbrication forte des différentes composantes) et en raison de l'évolution, en fonction du temps, de la contribution de chaque composante dans le signal mixé. De fait, les méthodes sont souvent évaluées sur des signaux mixés « simplifiés » bien contrôlés (les signaux sources sont des instruments MIDI ou sont des instruments relativement bien séparables, en nombre restreint).

[0009] Une autre méthode de séparation de sources est la séparation de sources « informée » : des informations relatives à un ou plusieurs signaux sources sont transmises avec le signal mixé au décodeur. Le décodeur est alors capable, à partir d'algorithmes et desdites informations, de séparer au moins partiellement au moins un signal source du signal mixé. Un exemple de séparation de sources informée est décrit par M. Parvaix et L. Girin (Informed source separation of linear instantaneous under-determined audio mixtures by source index embedding, IEEE trans. Audio Speech Lang. Process., volume 19, pages 1721-1733, août 2011). L'information transmise au décodeur indique notamment les deux signaux sources prédominants dans le signal mixé, pour différentes zones fréquentielles. Cependant, une telle méthode n'est pas toujours adaptée lorsqu'il existe plus de deux signaux sources contribuant simultanément dans une même zone fréquentielle du signal mixé : dans ce cas, au moins un signal source est négligé, créant ainsi un « trou spectral » dans la reconstitution dudit signal source.

[0010] Il est également connu, notamment dans le domaine des télécommunications, de filtrer des signaux captés par une pluralité de capteurs, en fonction de la position dans l'espace desdits signaux par rapport auxdits capteurs. Il s'agit ainsi d'un filtrage spatial (ou encore « beamforming ») qui permet de privilégier le signal dans une direction de l'espace donnée et de filtrer les signaux issus d'autres directions. Un exemple de tels filtres sont les filtres spatiaux à variance minimum sous contrainte linéaire (en anglais : « *linearly constrained minimum variance (LCMV)* »). Un exemple d'un tel filtre est notamment divulgué dans le document EP 1 633 121.

[0011] Un but de la présente invention est donc de proposer un procédé permettant de séparer des signaux sources compris dans un ou plusieurs signaux mixés, de manière plus efficace.

[0012] A cet effet, dans un mode de réalisation, il est proposé un procédé pour séparer, au moins partiellement, un ou plusieurs signaux sources numériques audio particuliers contenus dans un signal mixé numérique audio multicanal (c'est-à-dire comprenant au moins deux canaux), par exemple stéréo. Le signal mixé est obtenu par mixage de plusieurs signaux sources numériques audio et comprend des valeurs représentatives du ou des signaux sources particuliers. Selon le procédé :

- on détermine le module de l'amplitude ou la puissance normalisée du ou des signaux sources particuliers à partir des valeurs représentatives dudit ou desdits signaux sources particuliers contenues dans le signal mixé, puis
- on effectue un filtrage spatial à variance minimale sous contrainte linéaire du signal mixé pour obtenir au moins partiellement chaque signal source particulier, ledit filtrage étant basé sur la répartition dudit signal source particulier entre au moins deux canaux du signal mixé, et le module de l'amplitude ou la puissance normalisée dudit signal source particulier étant utilisé comme contrainte linéaire du filtre.

[0013] Les valeurs représentatives peuvent être la répartition temporelle, spectrale ou spectro-temporelle du signal source particulier, ou la contribution temporelle, spectrale ou spectro-temporelle du signal source particulier dans le signal mixé. Les valeurs représentatives des signaux sources peuvent ainsi être en module de l'amplitude ou en puissance normalisée (c'est-à-dire en énergie, qui correspond au carré du module de l'amplitude) : les valeurs représentatives peuvent donc être les valeurs en module de l'amplitude ou les valeurs de puissance normalisée (ou d'énergie).

[0014] Les valeurs représentatives peuvent par exemple être la répartition temporelle, spectrale ou spectro-temporelle du signal source particulier, ou la contribution temporelle, spectrale ou spectro-temporelle du signal source particulier dans le signal mixé, pour plusieurs zones (ou points) d'un plan temps-fréquence. Dans ce cas, la détermination du module de l'amplitude ou de la puissance normalisée du ou des signaux sources particuliers peut se faire dans le plan temps-fréquence : les modules de l'amplitude et les puissances normalisées sont des valeurs spectro-temporelles.

[0015] Une transformation ou une représentation dans le plan temps-fréquence consiste à représenter, en énergie (ou puissance normalisée) ou en module de l'amplitude (c'est-à-dire la racine carrée de l'énergie), le signal source en fonction de deux paramètres, le temps et la fréquence. Cela correspond à l'évolution, en énergie ou en module, du contenu fréquentiel du signal source en fonction du temps. On obtient ainsi, pour un instant donné et une fréquence donnée, une valeur positive réelle correspondant aux composantes du signal à cette fréquence et à cet instant. Des exemples de formulations théoriques et de mises en oeuvre pratiques de représentations temps-fréquence sont déjà décrites (L. Cohen : Time-Frequency Distributions, a Review, Proceedings of the IEEE, vol. 77, N° 7, 1989 ; F. Hlawatsch,

F. Auger : Temps-fréquence, concepts et outils, Hermès Science, Lavoisier 2005 ; P. Flandrin : Temps Fréquence, Hermès Science, 1998).

[0016] Ainsi, grâce au procédé décrit, il est possible de séparer efficacement, avec un filtrage spatial amélioré par les informations contenues dans le signal mixé, les signaux sources particuliers, sans faire d'hypothèse sur ces différents signaux (à l'exception des hypothèses statistiques classiques, i.e. : indépendance des signaux sources, moyenne nulle de ces signaux sources, distribution gaussienne). En particulier, le procédé est basé sur la répartition de chaque signal source entre les différents canaux du signal mixé pour isoler lesdits signaux sources (filtrage spatial). L'utilisation d'un filtre à variance minimum sous contrainte linéaire permet d'obtenir une séparation spatiale performante, en faisant intervenir, comme contrainte, le module de l'amplitude ou la puissance normalisée du signal source. On peut ainsi décorrélérer spatialement un signal source particulier du signal mixé, et ajuster l'amplitude du signal séparé au niveau souhaité dans le même temps. On améliore donc l'étape de filtrage spatial en prenant en considération la valeur représentative du signal source particulier dont on a connaissance.

[0017] Il est notamment possible d'isoler simultanément les différents signaux sources particuliers présents dans le signal mixé, par exemple en mettant en oeuvre autant de filtres spatiaux qu'il y a de signaux sources à séparer.

[0018] Préférentiellement, le filtrage est également basé sur le module de l'amplitude ou la puissance normalisée des signaux sources particuliers. Plus précisément, l'étape de filtrage spatial peut comprendre la modélisation d'une matrice de corrélation spatiale utilisant le module de l'amplitude ou la puissance normalisée des signaux sources particuliers et la répartition dudit signal source particulier entre au moins deux canaux du signal mixé.

[0019] Préférentiellement, le signal mixé comprend des valeurs représentatives du ou des signaux sources particuliers pour au moins deux canaux du signal mixé, et, avant d'effectuer le filtrage spatial, on détermine, à partir du signal mixé et desdites valeurs représentatives des signaux sources particuliers, la répartition de chaque signal source particulier entre lesdits au moins deux canaux du signal mixé.

[0020] Alternativement, on peut recevoir en entrée, par exemple avec le signal mixé, la répartition du ou des signaux sources particuliers entre au moins deux canaux du signal mixé.

[0021] Autrement dit, la répartition des signaux sources particuliers entre les différents canaux du signal mixé peut être fournie lors de la mise en oeuvre du procédé de séparation, par exemple en même temps que les valeurs représentatives desdits signaux sources particuliers, ou bien peut être déterminée pendant le procédé de séparation, à partir du signal mixé multicanal et des valeurs représentatives des signaux sources particuliers.

[0022] Selon un mode de mise en oeuvre, la détermination du module de l'amplitude ou de la puissance normalisée du ou des signaux sources particuliers comprend l'extraction des valeurs représentatives du ou des signaux sources particuliers qui ont été insérées dans le signal mixé, par exemple par tatouage. L'extraction des valeurs représentatives découle de la transmission des valeurs représentatives des signaux sources particuliers, qui peut se faire avec le signal mixé, par exemple lorsque les informations sont tatouées ou insérées de manière inaudible, dans le signal mixé, ou bien par un canal particulier du signal mixé qui est dédié à la transmission desdites valeurs représentatives.

[0023] Selon un autre aspect, il est proposé un dispositif pour séparer, au moins partiellement, un ou plusieurs signaux sources numériques audio particuliers contenus dans un signal mixé numérique audio multicanal. Le signal mixé est obtenu par mixage de plusieurs signaux sources numériques audio et comprend des valeurs représentatives du ou des signaux sources particuliers. Le dispositif comprend :

- un moyen de détermination du module de l'amplitude ou de la puissance normalisée du ou des signaux sources particuliers à partir des valeurs représentatives dudit ou desdits signaux sources particuliers contenues dans le signal mixé, et
- un filtre spatial à variance minimale sous contrainte linéaire apte à isoler au moins partiellement, à partir du signal mixé, chaque signal source particulier, ledit filtre étant basé sur la répartition dudit signal source particulier entre au moins deux canaux du signal mixé, et le module de l'amplitude ou la puissance normalisée dudit signal source particulier étant utilisé comme contrainte linéaire.

[0024] Préférentiellement, le signal mixé est un signal stéréo.

[0025] Préférentiellement, le signal mixé comprend des valeurs représentatives du ou des signaux sources particuliers pour au moins deux canaux du signal mixé, et le dispositif comprend un moyen de détermination, à partir du signal mixé et desdites valeurs représentatives des signaux sources particuliers, de la répartition de chaque signal source particulier entre lesdits au moins deux canaux du signal mixé.

[0026] Préférentiellement, le moyen de détermination du module de l'amplitude ou de la puissance normalisée comprend un moyen d'extraction des valeurs représentatives du ou des signaux sources particuliers qui ont insérées dans le signal mixé, par exemple par tatouage.

[0027] L'invention sera mieux comprise à l'étude d'un mode de réalisation particulier, pris à titre d'exemple nullement limitatif et illustré par les dessins annexés, sur lesquels :

- la figure 1 représente schématiquement un mode de réalisation d'un dispositif de séparation selon l'invention ; et
- la figure 2 est un organigramme d'un procédé de séparation selon l'invention.

[0028] Dans la suite de la description détaillée, on considère que le signal mixé $s_{\text{mix}}(t)$ est un signal stéréo avec un canal gauche $s_{\text{mix}}^g(t)$ et un canal droite $s_{\text{mix}}^d(t)$, et comprenant p signaux sources $s_1(t), \dots, s_p(t)$. Le signal mixé $s_{\text{mix}}(t)$ peut s'écrire comme le produit des p signaux sources par une matrice de mixage A :

$$A = \begin{bmatrix} a_1^g, \dots, a_p^g \\ a_1^d, \dots, a_p^d \end{bmatrix} = \begin{bmatrix} a_1, \dots, a_p \end{bmatrix}$$

où $a_i = [a_i^g, a_i^d]^T$ (T représentant la transposée de la matrice) et a_i^g et a_i^d représentent la répartition du signal source i dans chaque canal du signal mixé : $(a_i^g)^2 + (a_i^d)^2 = 1$.

[0029] Plus précisément, les coefficients a_i^g et a_i^d peuvent s'écrire sous la forme suivante : $a_i^g = \sin(\theta_i)$ et $a_i^d = \cos(\theta_i)$, où θ_i représente la balance du signal source i entre les deux canaux du signal mixé.

[0030] Autrement dit, on a :

$$s_{\text{mix}}(t) = A \cdot s(t)$$

avec : $s_{\text{mix}}(t) = [s_{\text{mix}}^g(t), s_{\text{mix}}^d(t)]^T$ et $s(t) = [s_1(t), \dots, s_p(t)]^T$ (T représentant la transposée).

[0031] Par ailleurs, on considère dans la suite de la description, que les signaux sont des signaux audio.

[0032] On considère, dans le cas de la présente description, la transformation de Fourier à court terme comme transformation dans le plan temps-fréquence. La transformée du signal source i dans le plan temps-fréquence s'écrit ainsi sous la forme :

$$S_i(k, m) = \sum s_i(k+n) f(n) e^{-2i\pi mn/N}$$

où N est une constante et $f(n)$ est une fonction de fenêtre de la transformée de Fourier à court terme.

[0033] On considère dans la suite de la description que la contrainte linéaire du filtre spatial est la puissance normalisée. Pour un signal source s_i donné, et un point (k, m) du plan temps-fréquence donné, on obtient donc comme énergie ou puissance normalisée $\varphi_i(k, m)$:

$$\varphi_i(k, m) = |S_i(k, m)|^2$$

[0034] La valeur représentative du signal source peut ainsi être $|S_i(k, m)|$ (valeur en module) ou bien $\varphi_i(k, m)$ (valeur en énergie égale à la valeur en puissance normalisée). La valeur représentative du signal source peut également être le logarithme de la valeur en énergie :

$$\Phi_i = 10 \log_{10}(\varphi_i(k, m))$$

[0035] La valeur représentative du signal source peut également être déterminée après avoir effectué des traitements sur le signal source, par exemple en réduisant la résolution fréquentielle du spectre en énergie ou bien encore en adaptant la quantification des valeurs représentatives à la sensibilité de l'oreille humaine. Il est alors possible d'obtenir des valeurs représentatives des signaux sources qui sont moins volumineuses en terme de taille, tout en gardant une qualité sonore voulue.

[0036] On considère dans la suite de la description que la valeur représentative des signaux sources est la valeur en puissance normalisée (ou énergie) quantifiée $\Phi_i(k, m)$.

[0037] Les valeurs représentatives des signaux sources $\Phi_i(k, m)$ sont transmises au dispositif de séparation ou décodeur. Elles peuvent l'être par un canal dédié (associé aux canaux stéréo pour former le signal mixé), ou par incorporation dans le signal mixé, par exemple par tatouage ou par utilisation de bits non-utilisés du signal mixé. Dans ce dernier cas, le dispositif de séparation peut comprendre un moyen d'extraction des valeurs représentatives, recevant en entrée le signal mixé et fournissant en sortie, les valeurs représentatives des signaux sources.

[0038] De même, le dispositif de séparation peut également recevoir les répartitions des signaux sources dans chaque

voie (ou canal) du signal mixé : $a_1^g, \dots, a_p^g, a_1^d, \dots, a_p^d$. Ces répartitions peuvent être transmises par un canal dédié (associé aux canaux stéréo pour former le signal mixé, ou indépendant des canaux stéréo), ou par incorporation dans le signal mixé, par exemple par tatouage ou par utilisation de bits non-utilisés du signal mixé. Dans ce dernier cas, le dispositif de séparation peut comprendre un moyen d'extraction des répartitions des signaux sources, recevant en entrée le signal mixé et fournissant en sortie, les répartitions des signaux sources. Le moyen d'extraction des valeurs représentatives et le moyen d'extraction des répartitions peuvent être un seul et même moyen.

[0039] Alternativement, le dispositif de séparation peut comprendre un moyen de détermination des répartitions des signaux sources : un tel moyen de détermination peut recevoir en entrée le signal mixé et les valeurs représentatives $\Phi_i(k, m)$, et fournir en sortie la répartition dudit signal source a_i^g, a_i^d . Cela est possible notamment lorsque chaque canal du signal mixé comprend les valeurs représentatives d'un signal source pour ledit canal du signal mixé : autrement dit, les valeurs représentatives d'un signal source donné ne seront pas les mêmes pour chaque canal du signal mixé, la différence entre les valeurs représentatives d'un même signal source pour les différents canaux du signal mixé permettant de déterminer la répartition dudit signal source entre les différents canaux du signal mixé.

[0040] Sur la figure 1, on a représenté schématiquement un mode de réalisation d'un dispositif de séparation 1 de signaux sources particuliers contenus dans un signal mixé s_{mix} . Le dispositif de séparation 1 reçoit en entrée les canaux stéréo s_{mix}^g et s_{mix}^d du signal mixé s_{mix} , et délivre des signaux sources particuliers séparés au moins partiellement s_i' , avec 1 variant de 1 à p. Le dispositif de séparation 1 a pour but de délivrer, au moins partiellement, plusieurs signaux sources particuliers contenus dans le signal mixé s_{mix} en utilisant les valeurs représentatives desdits signaux sources particuliers $\Phi_i(k, m)$.

[0041] On considère pour la présente description, que le dispositif de séparation 1 reçoit en entrée, les canaux du signal mixé numérique audio $s_{mix}^g(t)$ et $s_{mix}^d(t)$, dans lesquels sont insérées, par exemple par tatouage, les valeurs représentatives des signaux sources particuliers $\Phi_i(k, m)$, et éventuellement les répartitions $a_1^g, \dots, a_p^g, a_1^d, \dots, a_p^d$ des signaux sources particuliers entre les deux canaux du signal mixé numérique audio $s_{mix}^d(t)$ et $s_{mix}^g(t)$.

[0042] Le dispositif de séparation 1 comprend un moyen de transformation 2, un moyen d'extraction 3, un moyen de traitement 4, un moyen de filtrage 5, et un moyen de transformation inverse 6.

[0043] Le moyen de transformation 2 reçoit en entrée les canaux du signal mixé numérique audio $s_{mix}^g(t)$ et $s_{mix}^d(t)$ et délivre, en sortie, la transformée des canaux du signal mixé dans le plan temps-fréquence $S_{mix}^g(k, m)$ et $S_{mix}^d(k, m)$.

[0044] Le moyen d'extraction 3 reçoit en entrée la transformée des canaux du signal mixé dans le plan temps-fréquence $S_{mix}^d(k, m)$ et $S_{mix}^g(k, m)$, et délivre les valeurs représentative $\Phi_i(k, m)$ des signaux sources particuliers contenues dans le signal mixé. Le moyen d'extraction 3 peut, le cas échéant, délivrer également les répartitions $a_1^g, \dots, a_p^g, a_1^d, \dots, a_p^d$ des signaux sources particuliers entre les deux canaux du signal mixé numérique audio $s_{mix}^d(t)$ et $s_{mix}^g(t)$, lorsque celles-ci sont insérées dans le signal mixé. Le moyen d'extraction 3 permet ainsi d'extraire du signal mixé les valeurs représentatives qui y ont été ajoutées a posteriori, par exemple par tatouage, et des isoler du signal mixé. Les valeurs représentatives $\Phi_i(k, m)$ sont alors transmises au moyen de traitement 4 et, le cas échéant, les répartitions $a_1^g, \dots, a_p^g, a_1^d, \dots, a_p^d$ sont transmises au moyen de filtrage 5.

[0045] Il convient de noter que le moyen d'extraction 3 peut alternativement recevoir directement en entrée, les canaux du signal mixé $s_{mix}^d(t)$ et $s_{mix}^g(t)$.

[0046] Le moyen de traitement 4 permet de traiter les valeurs représentatives $\Phi_i(k, m)$ reçues par le moyen d'extraction 3, afin de déterminer une estimation de la puissance normalisée $\varphi_i'(k, m)$ des signaux sources à séparer, dans le plan temps-fréquence. Les estimations de la puissance normalisée $\varphi_i'(k, m)$ des signaux sources à séparer sont ensuite transmises au moyen de filtrage 5.

[0047] La transformée des canaux du signal mixé dans le plan temps-fréquence $S_{mix}^d(k, m)$ et $S_{mix}^g(k, m)$ fournie par le moyen de transformation 2, les estimations des puissances normalisées des signaux sources particuliers $\varphi_i'(k, m)$, et les répartitions $a_1^g, \dots, a_p^g, a_1^d, \dots, a_p^d$ des signaux sources particuliers entre les deux canaux du signal mixé numérique audio $s_{mix}^d(t)$ et $s_{mix}^g(t)$, sont ainsi fournies au moyen de filtrage 5.

[0048] Le moyen de filtrage 5 permet d'obtenir une estimation $S_i'(k, m)$ de chaque signal source particulier, grâce à un filtrage spatial. Le moyen de filtrage 5 permet, dans le plan temps-fréquence, d'isoler le signal source particulier, grâce à un filtrage spatial à variance minimale sous contrainte linéaire. Plus particulièrement, le moyen de filtrage 5 se base sur la répartition dudit signal source particulier entre les deux canaux du signal mixé pour isoler le signal source particulier : il s'agit donc d'un filtrage spatial (en anglais : « beamforming »). Par ailleurs, afin d'améliorer le filtrage et l'estimation du signal source obtenue, le filtre spatial utilise la puissance normalisée du signal source particulier à séparer comme contrainte linéaire, afin d'obtenir une estimation plus proche du signal source d'origine.

[0049] Plus précisément, dans le plan temps-fréquence, on a :

$$S_{mix}(k, m) = A.S(k, m)$$

avec :

$$S_{mix}(k, m) = [S_{mix}^g(k, m), S_{mix}^d(k, m)]^T$$

et

$$S(k, m) = [S_1(k, m), \dots, S_p(k, m)]^T$$

[0050] Chaque signal mixé $S_{mix}^d(k, m)$ et $S_{mix}^g(k, m)$ est alors décomposé en estimations de signaux sources particuliers $S'_1(k, m), \dots, S'_p(k, m)$ en utilisant un filtrage spatial linéaire :

$$S'_i(k, m) = w_{ik}^g \cdot S_{mix}^g(k, m) + w_{ik}^d \cdot S_{mix}^d(k, m) = W_{ik}^T \cdot S_{mix}(k, m)$$

avec:

$$W_{ik} = [w_{ik}^g, w_{ik}^d]^T \text{ et } S'_i(k, m) = [S'_i{}^g(k, m), S'_i{}^d(k, m)]^T$$

[0051] W_{ik} est le filtre spatial (ou « beamformer ») permettant d'obtenir l'estimation $S'_i(k, m)$ du $i^{\text{ème}}$ signal source dans la sous-bande k à partir du signal mixé $S_{mix}(k, m)$.

[0052] Dans le cas d'un filtre spatial à variance minimum sous contrainte linéaire, on considère comme bruit la somme de tous les signaux sources interférant, à l'exception de celui à filtrer. Ainsi, le signal mixé peut être réécrit de la manière suivante :

$$S_{mix}(k, m) = a_i \cdot S_i(k, m) + r(k, m)$$

où $r(k, m)$ est la somme des autres signaux sources.

[0053] L'estimation $S'_i(k, m)$ est obtenue en minimisant la puissance moyenne du bruit ou, de manière équivalente, la puissance moyenne de sortie du filtre spatial, selon la direction du signal source à séparer :

$$P(\theta_i) = W_{ik}^T(m) \cdot R'_{Smix}(k, m) \cdot W_{ik}(m)$$

où R_{Smix} est la matrice de corrélation spatiale des deux canaux $S_{mix}^d(k, m)$ et $S_{mix}^g(k, m)$ du signal mixé $S_{mix}(k, m)$.

[0054] La solution est donnée par :

$$W_{ik}(m) = R'^{-1}_{Smix}(k, m) \cdot a_i \cdot \sqrt{\frac{\varphi'_i(k, m)}{a_i^T \cdot R'^{-1}_{Smix}(k, m) \cdot a_i}}$$

[0055] On obtient donc :

$$S'_i(k, m) = \sqrt{\frac{\varphi'_i(k, m)}{a_i^T \cdot R'^{-1}_{Smix}(k, m) \cdot a_i}} \cdot a_i^T \cdot R'^{-1}_{Smix}(k, m) \cdot S_{mix}(k, m)$$

avec : $R'^{-1}_{Smix}(k, m) = \sum \varphi'_i(k, m) \cdot a_i \cdot a_i^T$.

[0056] Une fois appliqué au signal mixé $S_{mix}(k, m)$, le filtre obtenu permet de réduire la contribution du spectre en puissance des autres signaux. Par ailleurs, grâce à la contrainte linéaire, la puissance du signal source estimé correspond à la puissance du signal source initial pour les différents points du plan temps-fréquence (ce qui peut être vérifié en réinjectant la solution W_{ik} dans l'équation définissant $P(\theta_i)$). Ainsi, le moyen de filtrage 5 permet de décorréler spatialement

the $i^{\text{ème}}$ signal source du reste du signal mixé, tout en ajustant l'amplitude dudit signal décorrélié au niveau souhaité.

[0057] On peut également noter que, lorsque la quantité d'informations tatouées dans le signal mixé est trop importante pour que le bruit du tatouage puisse être négligé, il est également possible d'ajuster les composants des signaux sources estimés comme suit :

$$S'_i(k,m) = S_i(k,m) \cdot (\sqrt{\varphi'_i(k,m)}) / |S'_i(k,m)|$$

[0058] Les transformées des estimations des signaux sources particuliers séparés sont alors transmises au moyen de transformation inverse 6. Le moyen 6 permet de transformer les transformées des estimations des signaux sources séparés en signaux temporels $s'_1(t)$, ..., $s'_p(t)$ correspondant, au moins partiellement, aux signaux sources $s_1(t)$, ..., $s_p(t)$.

[0059] Sur la figure 2, on a représenté un organigramme représentant les différentes étapes du procédé de séparation selon l'invention.

[0060] Le procédé comprend une première étape 7 au cours de laquelle le signal mixé est transformé dans un plan temps-fréquence. Puis, dans une étape 8, on effectue une extraction des informations tatouées dans le signal mixé, notamment les valeurs représentatives et les répartitions des signaux sources entre au moins deux canaux du signal mixé. Lors d'une étape 9, on détermine les puissances normalisées des signaux sources à séparer, puis on effectue, lors de l'étape 10, un filtrage spatial à variance minimum sous contrainte linéaire, la contrainte étant la puissance normalisée du signal source à séparer. Enfin on effectue, dans une étape 11, une transformation inverse des transformées des signaux sources particuliers séparés, de manière à obtenir, au moins partiellement, les signaux sources particuliers.

[0061] Dans le cas de signaux audio, il est ainsi possible d'effectuer en sortie du système de séparation de l'invention un certain nombre de contrôles majeurs en écoute audio (volume, tonalité, effets) de façon indépendante sur les différents éléments de la scène sonore (instruments et voix obtenus par le dispositif de séparation).

Revendications

1. Procédé pour séparer, au moins partiellement, un ou plusieurs signaux sources numériques audio particuliers (s_i) contenus dans un signal mixé numérique audio multicanal (s_{mix}), le signal mixé étant obtenu par mixage de plusieurs signaux sources numériques audio (s_1 , ..., s_p) et comprenant des valeurs représentatives (Φ_i) du ou des signaux sources particuliers, dans lequel :

- on détermine (9) le module de l'amplitude ou la puissance normalisée (φ'_i) du ou des signaux sources particuliers à partir des valeurs représentatives (Φ_i) dans le plan temps-fréquence dudit ou desdits signaux sources particuliers contenues dans le signal mixé,

- le procédé étant **caractérisé en ce que** l'on effectue ensuite (10) un filtrage spatial à variance minimale sous contrainte linéaire du signal mixé pour obtenir au moins partiellement chaque signal source particulier (s'_i), ledit filtrage étant basé sur la répartition (a_i) dudit signal source particulier entre au moins deux canaux du signal mixé, et le module de l'amplitude ou la puissance normalisée (φ'_i) dudit signal source particulier étant utilisé comme contrainte linéaire du filtre.

2. Procédé selon la revendication 1 dans lequel le signal mixé comprend des valeurs représentatives (Φ_i^g , Φ_i^d) du ou des signaux sources particuliers pour au moins deux canaux du signal mixé (s_{mix}^g , s_{mix}^d), et dans lequel, avant d'effectuer le filtrage spatial, on détermine, à partir du signal mixé et desdites valeurs représentatives des signaux sources particuliers, la répartition (a_i^g , a_i^d) de chaque signal source particulier (s_i) entre lesdits au moins deux canaux du signal mixé (s_{mix}^g , s_{mix}^d).

3. Procédé selon la revendication 1 dans lequel on reçoit en entrée, par exemple avec le signal mixé, la répartition (a_i) du ou des signaux sources particuliers entre au moins deux canaux du signal mixé.

4. Procédé selon l'une des revendications précédentes dans lequel la détermination du module de l'amplitude ou de la puissance normalisée (φ'_i) du ou des signaux sources particuliers comprend l'extraction (8) des valeurs représentatives (Φ_i) du ou des signaux sources particuliers qui ont été insérées dans le signal mixé, par exemple par tatouage.

5. Procédé selon l'une des revendications précédentes dans lequel le module de l'amplitude ou la puissance normalisée (φ'_i) dudit signal source particulier sont des valeurs spectro-temporelles.

6. Dispositif (1) pour séparer, au moins partiellement, un ou plusieurs signaux sources numériques audio particuliers (s_i) contenus dans un signal mixé numérique audio multicanal (s_{mix}), le signal mixé (s_{mix}) étant obtenu par mixage de plusieurs signaux sources numériques audio ($s_1, \dots, s_p$) et comprenant des valeurs représentatives (Φ_i) du ou des signaux sources particuliers, le dispositif comprenant :

- un moyen de détermination (4) du module de l'amplitude ou de la puissance normalisée (φ'_i) du ou des signaux sources particuliers à partir des valeurs représentatives (Φ_i) dans le plan temps-fréquence dudit ou desdits signaux sources particuliers contenues dans le signal mixé,

le dispositif étant **caractérisé par**

- un filtre spatial à variance minimale sous contrainte linéaire (5) apte à isoler au moins partiellement, à partir du signal mixé (s_{mix}), chaque signal source particulier (s'_i), ledit filtre étant basé sur la répartition (a_i) dudit signal source particulier entre au moins deux canaux du signal mixé ($s_{\text{mix}}^g, s_{\text{mix}}^d$), et le module de l'amplitude ou la puissance normalisée (φ'_i) dudit signal source particulier étant utilisé comme contrainte linéaire.

7. Dispositif selon la revendication 6 dans lequel le signal mixé comprend des valeurs représentatives (Φ_i) du ou des signaux sources particuliers pour au moins deux canaux du signal mixé, le dispositif comprenant un moyen de détermination, à partir du signal mixé et desdites valeurs représentatives (Φ_i) des signaux sources particuliers, de la répartition (a_i) de chaque signal source particulier entre lesdits au moins deux canaux du signal mixé.

8. Dispositif selon la revendication 6 ou 7 comprenant également un moyen d'extraction (3) des valeurs représentatives (Φ_i) du ou des signaux sources particuliers qui ont insérées dans le signal mixé, par exemple par tatouage.

Patentansprüche

1. Verfahren zur wenigstens teilweisen Trennung eines oder mehrerer einzelner digitaler Audioquellsignale (s_i), die in einem digitalen Mehrkanal-Audiomischsignal (s_{mix}) enthalten sind, wobei das Mischsignal durch Mischen von mehreren digitalen Audioquellsignalen ($s_1, \dots, s_p$) erhalten wird und repräsentative Werte (Φ_i) des oder der einzelnen Quellsignale umfasst, bei dem:

- der Modul der Amplitude oder die normalisierte Leistung (φ'_i) des oder der einzelnen Quellsignale anhand der in dem Mischsignal enthaltenen repräsentativen Werte (Φ_i) in der Zeit-Frequenz-Ebene des oder der einzelnen Quellsignale bestimmt wird (9),

- wobei das Verfahren **dadurch gekennzeichnet ist, dass** anschließend eine räumliche Filterung mit minimaler Varianz unter linearer Einschränkung des Mischsignals durchgeführt wird (10), um jedes einzelne Quellsignal (s'_i) wenigstens teilweise zu erhalten, wobei die Filterung auf der Aufteilung (a_i) des einzelnen Quellsignals auf wenigstens zwei Kanäle des Mischsignals basiert, und wobei der Modul der Amplitude oder die normalisierte Leistung (φ'_i) des einzelnen Quellsignals als lineare Einschränkung des Filters verwendet wird.

2. Verfahren nach Anspruch 1, bei dem das Mischsignal repräsentative Werte (Φ_i^g, Φ_i^d) des oder der einzelnen Quellsignale für wenigstens zwei Kanäle des Mischsignals ($s_{\text{mix}}^g, s_{\text{mix}}^d$) umfasst, und bei dem vor der Durchführung der räumlichen Filterung anhand des Mischsignals und der repräsentativen Werte der einzelnen Quellsignale die Aufteilung (a_i^g, a_i^d) eines jeden einzelnen Quellsignals (s_i) auf die wenigstens zwei Kanäle des Mischsignals ($s_{\text{mix}}^g, s_{\text{mix}}^d$) bestimmt wird.

3. Verfahren nach Anspruch 1, bei dem am Eingang beispielsweise mit dem Mischsignal die Aufteilung (a_i) des oder der einzelnen Quellsignale auf wenigstens zwei Kanäle des Mischsignals erhalten wird.

4. Verfahren nach einem der vorhergehenden Ansprüche, bei dem die Bestimmung des Moduls der Amplitude oder der normalisierten Leistung (φ'_i) des oder der einzelnen Quellsignale die Extraktion (8) der repräsentativen Werte (Φ_i) des oder der einzelnen Quellsignale, die in das Mischsignal - beispielsweise durch Wasserzeichen - eingebettet worden sind, umfasst.

5. Verfahren nach einem der vorhergehenden Ansprüche, bei dem der Modul der Amplitude oder die normalisierte Leistung (φ'_i) des einzelnen Quellsignals spektralzeitliche Werte sind.

6. Vorrichtung (1) zur wenigstens teilweisen Trennung eines oder mehrerer einzelner digitaler Audioquellsignale (s_i), die in einem digitalen Mehrkanal-Audiomischsignal (s_{mix}) enthalten sind, wobei das Mischsignal (s_{mix}) durch Mischen von mehreren digitalen Audioquellsignalen ($s_1, \dots, s_p$) erhalten wird und repräsentative Werte (Φ_i) des oder der einzelnen Quellsignale umfasst, wobei die Vorrichtung umfasst:

- ein Mittel zur Bestimmung (4) des Moduls der Amplitude oder der normalisierten Leistung (φ'_i) des oder der einzelnen Quellsignale anhand der in dem Mischsignal enthaltenen repräsentativen Werte (Φ_i) in der Zeit-Frequenz-Ebene des oder der einzelnen Quellsignale,

wobei die Vorrichtung **gekennzeichnet ist durch**

- ein räumliches Filter mit minimaler Varianz unter linearer Einschränkung (5), das geeignet ist, jedes einzelne Quellsignal (s'_i) aus dem Mischsignal (s_{mix}) wenigstens teilweise zu isolieren, wobei das Filter auf der Aufteilung (a_i) des einzelnen Quellsignals auf wenigstens zwei Kanäle des Mischsignals ($s_{\text{mix}}^g, s_{\text{mix}}^d$) basiert, und wobei der Modul der Amplitude oder die normalisierte Leistung (φ'_i) des einzelnen Quellsignals als lineare Einschränkung verwendet wird.

7. Vorrichtung nach Anspruch 6, bei der das Mischsignal repräsentative Werte (Φ_i) des oder der einzelnen Quellsignale für wenigstens zwei Kanäle des Mischsignals umfasst, wobei die Vorrichtung ein Mittel zur Bestimmung der Aufteilung (a_i) eines jeden einzelnen Quellsignals auf die wenigstens zwei Kanäle des Mischsignals anhand des Mischsignals und der repräsentativen Werte (Φ_i) der einzelnen Quellsignale umfasst.

8. Vorrichtung nach Anspruch 6 oder 7, ebenfalls umfassend ein Mittel zur Extraktion (3) der repräsentativen Werte (Φ_i) des oder der einzelnen Quellsignale, die in das Mischsignal - beispielsweise durch Wasserzeichen - eingebettet worden sind.

Claims

1. A method of separating, at least in part, one or more particular digital audio source signals (s_i) contained in a mixed multichannel digital audio signal (s_{mix}), the mixed signal being obtained by mixing a plurality of digital audio source signals ($s_1, \dots, s_p$) and including representative values (Φ_i) of the particular source signal(s), the method comprising:

- determining (9) the modulus of the amplitude or the normalized power (φ'_i) of the particular source signal(s) from the representative values (Φ_i) in the time-frequency plane of said particular source signal(s) contained in the mixed signal;

the method being **characterized by** then performing (10) linearly constrained minimum variance spatial filtering in order to obtain, at least in part, each particular source signal (s'_i), said filtering being based on the distribution (a_i) of said particular source signal between at least two channels of the mixed signal, and the modulus of the amplitude or the normalized power (φ'_i) of said particular source signal being used as a linear constraint of the filter.

2. A method according to claim 1, wherein the mixed signal includes representative values (Φ_i^g, Φ_i^d) of the particular source signal(s) for at least two channels of the mixed signal ($s_{\text{mix}}^g, s_{\text{mix}}^d$), and wherein, prior to performing spatial filtering, the mixed signal and said representative values of the particular signals are used to determine the distribution (a_i^g, a_i^d) of each particular source signal (s_i) between said at least two channels of the mixed signal ($s_{\text{mix}}^g, s_{\text{mix}}^d$).

3. A method according to claim 1, wherein the distribution (a_i) of the particular source signal(s) between at least two channels of said mixed signal is received as input, e.g. in the mixed signal.

4. A method according to any preceding claim, wherein determining the modulus of the amplitude or the normalized power (φ'_i) of the particular source signal(s) comprises extracting (8) representative values (Φ_i) of the particular source signals that have been inserted into the mixed signal, e.g. by watermarking.

5. A method according to any preceding claim, wherein the modulus of the amplitude or the normalized power (φ'_i) of said particular source signal are spectro-temporal values.

6. A device (1) for separating, at least in part, one or more particular digital audio source signals (s_i) contained in a

multichannel mixed digital audio signal (s_{mix}), the mixed signal (s_{mix}) being obtained by mixing a plurality of digital audio source signals ($s_1, \dots, s_p$) and including representative values (Φ_i) of the particular source signal(s), the device comprising:

- determination means (4) for determining the modulus of the amplitude or the normalized power (φ'_i) of the particular source signal(s) from the representative values (Φ_i) in the time-frequency plane of said particular source signal(s) contained in the mixed signal;

the device being **characterized by** a linearly constrained minimum variance spatial filter (5) adapted to isolate, at least in part, each particular source signal (s'_i) from the mixed signal (s_{mix}), said filter being based on the distribution (a_i) of said particular source signal between at least two channels of the mixed signal ($s_{\text{mix}}^g, s_{\text{mix}}^d$), and the modulus of the amplitude or the normalized power (φ'_i) of said particular source signal being used as a linear constraint.

7. A device according to claim 6, wherein the mixed signal includes representative values (Φ_i) of the particular source signal(s) for at least two channels of the mixed signal, the device including determination means for determining the distribution (a_i) of each particular source signal between said at least two channels of the mixed signal from the mixed signal and from said representative values (Φ_i) of the particular source signals.

8. A device according to claim 6 or claim 7, also including extractor means (3) for extracting the representative values (Φ_i) of the particular source signal(s) that have been inserted in the mixed signal, e.g. by watermarking.

FIG.1

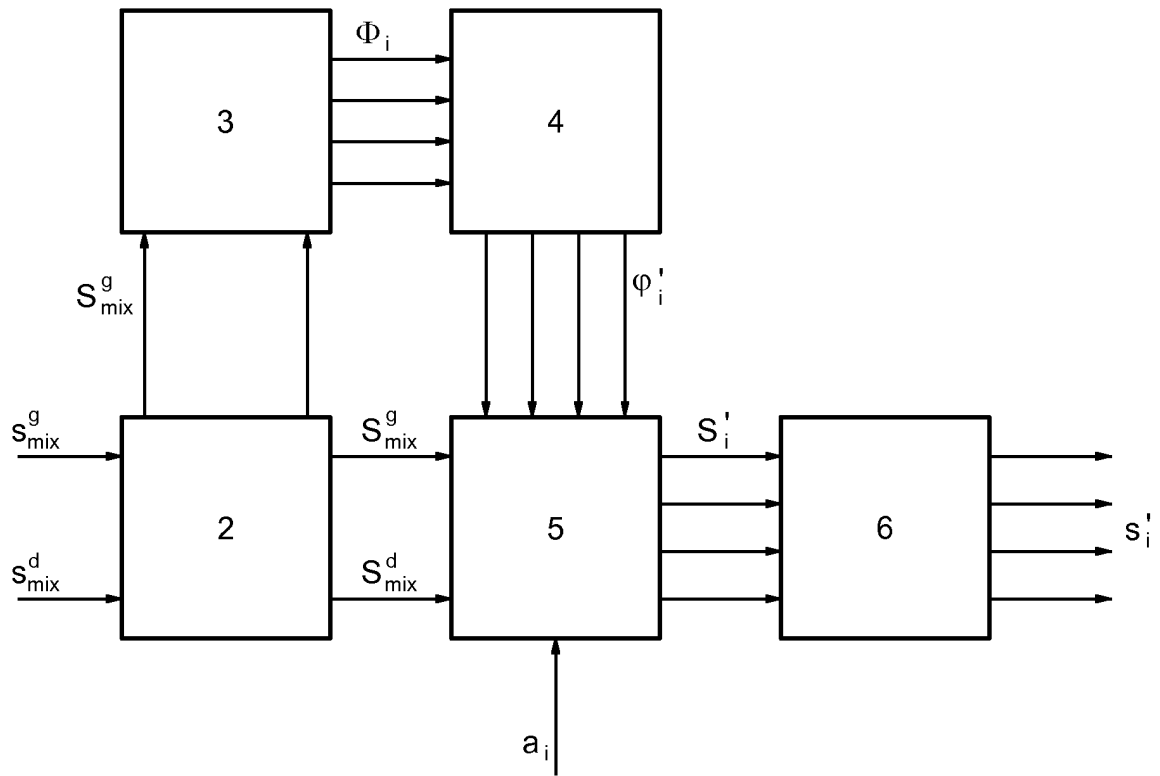
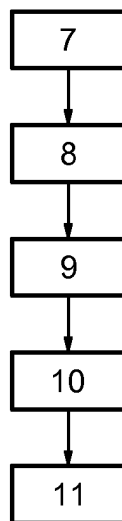


FIG.2



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.

Documents brevets cités dans la description

- EP 1633121 A [0010]

Littérature non-brevet citée dans la description

- **D.P.W. ELLIS.** Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis, and its application to speech/non-speech mixture. *Speech Communication*, 1999, vol. 27 (3), 281-298 [0005]
- **D. GODSMARK ; G.J.BROWN.** A blackboard architecture for computational auditory scene analysis. *Speech Communication*, 1999, vol. 27 (3), 351-366 [0005]
- **T. KINOSHITA ; S. SAKAI ; H. TANAKA.** Musical sound source identification based on frequency component adaptation. *Proc. IJCAI Workshop on CASA*, 1999, 18-24 [0005]
- **L. BENAROYA.** Représentations parcimonieuses pour la séparation de sources avec un seul capteur. *Proc. GRETSI*, 2001 [0007]
- **P.J. WOLFE ; S.J. GODSILL.** A Gabor régression scheme for audio signal analysis. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003, 103-106 [0007]
- **R. GRIBONVAL ; E. BACRY.** Harmonic Decomposition of Audio Signals With Matching Pursuit. *IEEE Trans. Signal Proc.*, 2003, vol. 51 (1), 101-112 [0007]
- **M.A. CASEY ; A. WESTNER.** Separation of mixed audio sources by independent subspace analysis. *Proc. Int. Computer Music Conf.*, 2000 [0008]
- **M. PARVAIX ; L. GIRIN.** Informed source separation of linear instantaneous under-determined audio mixtures by source index embedding. *IEEE trans. Audio Speech Lang. Process.*, Août 2011, vol. 19, 1721-1733 [0009]
- **L. COHEN.** Time-Frequency Distributions, a Review. *Proceedings of the IEEE*, 1989, vol. 77 (7) [0015]
- Temps-fréquence, concepts et outils. **F. HLAWATSCH ; F. AUGER.** Hermès Science. 2005 [0015]
- **P. FLANDRIN.** Temps Fréquence. *Hermès Science*, 1998 [0015]