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

(11) **EP 3 635 718 B1**

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention de la délivrance du brevet:
28.06.2023 Bulletin 2023/26

(21) Numéro de dépôt: **18737650.4**

(22) Date de dépôt: **24.05.2018**

(51) Classification Internationale des Brevets (IPC):
G10L 21/0272 *(2013.01)* ***G10L 25/84*** *(2013.01)*
G10L 21/0216 *(2013.01)* ***G10L 25/06*** *(2013.01)*
G10L 21/0208 *(2013.01)* ***G10L 21/0308*** *(2013.01)*

(52) Classification Coopérative des Brevets (CPC):
G10L 21/0272; G10L 25/84; G10L 21/0308; G10L 2021/02082; G10L 2021/02166

(86) Numéro de dépôt international:
PCT/FR2018/000139

(87) Numéro de publication internationale:
WO 2018/224739 (13.12.2018 Gazette 2018/50)

(54) **TRAITEMENT DE DONNEES SONORES POUR UNE SEPARATION DE SOURCES SONORES DANS UN SIGNAL MULTICANAL**

VERARBEITUNG VON KLANGDATEN ZUR TRENNUNG VON KLANGQUELLEN IN EINEM MEHRKANALSIGNAL

PROCESSING OF SOUND DATA FOR SEPARATING SOUND SOURCES IN A MULTICHANNEL SIGNAL

(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

(30) Priorité: **09.06.2017 FR 1755183**

(43) Date de publication de la demande:
15.04.2020 Bulletin 2020/16

(73) Titulaire: **Orange**
92130 Issy-les-Moulineaux (FR)

(72) Inventeurs:
- **BAQUÉ, Mathieu**
  92326 Châtillon Cedex (FR)
- **GUERIN, Alexandre**
  92326 Châtillon Cedex (FR)

(56) Documents cités:
US-A1- 2005 060 142 **US-A1- 2010 111 290**

- **BAQUÉ MATHIEU ET AL: "Separation of Direct Sounds from Early Reflections Using the Entropy Rate Bound Minimization Algorithm", CONFERENCE: 60TH INTERNATIONAL CONFERENCE: DREAMS (DEREVERBERATION AND REVERBERATION OF AUDIO, MUSIC, AND SPEECH); JANUARY 2016, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 27 janvier 2016 (2016-01-27), XP040680602,**
- **JOURJINE A ET AL: "BLIND SEPARATION OF DISJOINT ORTHOGONAL SIGNALS: DEMIXING N SOURCES FROM 2 MIXTURES", 2000 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). ISTANBUL, TURKEY, JUNE 5-9, 2000; [IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP)], NEW YORK, NY : IEEE, US, 5 juin 2000 (2000-06-05), pages 2985-2988, XP001035813, ISBN: 978-0-7803-6294-9 cité dans la demande**

EP 3 635 718 B1

Description

[0001] La présente invention se rapporte au domaine du traitement de signal audio ou acoustique et plus particulièrement au traitement de contenus sonores multicanal réels pour séparer les sources sonores.

[0002] La séparation de sources dans un signal sonore multicanal permet de multiples applications. Elle peut par exemple être utilisée :

- Pour le divertissement (karaoké : suppression de la voix),
- Pour la musique (mixage des sources séparées dans un contenu multicanal),
- Pour les télécommunications (rehaussement de la voix, débruitage),
- Pour la domotique (commande vocale),
- Pour le codage audio multicanal,
- Pour la localisation de sources et cartographie en imagerie.

[0003] Dans un espace E dans lequel un nombre N de sources émettent un signal s_i, une séparation aveugle des sources consiste, à partir d'un nombre M d'observations issues de capteurs répartis dans cet espace E, à dénombrer et extraire le nombre N de sources. En pratique, chaque observation est obtenue à l'aide d'un capteur qui enregistre le signal parvenu jusqu'en un point de l'espace où se situe le capteur. Le signal enregistré résulte alors du mélange et de la propagation dans l'espace E des signaux s_i et se trouve donc affecté de différentes perturbations propres au milieu traversé comme par exemple le bruit, la réverbération, les interférences, etc...

[0004] La captation multicanal d'un nombre N de sources sonores s_i se propageant en champ libre et considérées comme ponctuelles se formalise comme une opération matricielle :

$$\boldsymbol{x} = \boldsymbol{As} = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1}(\theta_1, \phi_1, r_1) & \cdots & a_{MN}(\theta_N, \phi_N, r_N) \end{bmatrix} * \mathbf{s}$$

[0005] Où **x** est le vecteur des *M* canaux enregistrés, **s** le vecteur des *N* sources et **A** une matrice dite « matrice de mélange » de dimension $M \times N$ contenant les contributions de chaque source à chaque observation, et le sigle * symbolise la convolution linéaire. Selon le milieu de propagation et le format de l'antenne, la matrice **A** peut prendre différentes formes. Dans le cas d'une antenne coïncidente (tous les microphones de l'antenne sont concentrés en un même point de l'espace) en milieu anéchoïque, **A** est une simple matrice de gains. Dans le cas d'une antenne non coïncidente, en milieu anéchoïque ou réverbérant, la matrice **A** devient une matrice de filtre. Dans ce cas, on exprime généralement la relation dans le domaine fréquentiel $\boldsymbol{x}(\boldsymbol{f}) = \boldsymbol{As}(\boldsymbol{f})$, où **A** s'exprime comme une matrice de coefficients complexes.

[0006] Dans le cas où la captation du signal sonore se fait dans un environnement anéchoïque, et si l'on se place dans l'hypothèse où le nombre de sources *N* est inférieur au nombre d'observations *M*, l'analyse (i.e. l'identification du nombre de sources et de leurs positions) et la décomposition de la scène en objets, i.e. les sources, peuvent être facilement réalisées de manière conjointe par un algorithme d'analyse en composantes indépendantes (ou « ACI » ci-après). Ces algorithmes permettent d'identifier la matrice B de séparation de dimensions *NxM,* pseudo-inverse de A, qui permet de déduire les sources à partir des observations grâce à l'équation suivante :

$$\boldsymbol{s} = \boldsymbol{Bx}$$

[0007] L'étape préalable d'estimation de la dimension du problème, i.e. l'estimation de la taille de la matrice de séparation, soit du nombre de sources N, est classiquement fait en calculant le rang de la matrice de covariance $\boldsymbol{Co} = E\{\boldsymbol{xx}^T\}$ des observations, qui est, dans ce cas anéchoïque, égal au nombre de sources :

$$N = rank(\boldsymbol{Co}).$$

[0008] Quant à la localisation des sources, elle peut être déduite de la matrice d'encodage $A = \boldsymbol{B}^{-1}$ et de la connaissance des propriétés spatiales de l'antenne utilisée, notamment la distance entre les capteurs et leurs directivités.

[0009] Parmi les algorithmes les plus connus d'ACI, on peut citer **JADE** de J.F Cardoso et A. Souloumiac. ("Blind beamforming for non-gaussian signals" dans "IEE Proceedings F - Radar and Signal Processing", volume 140, issue 6, Dec. 1993) ou **Infomax** d'Amari et. al. ("A new learning algorithm for blind signal séparation, Advances" dans "neural information processing systems", 1996)

[0010] Le document US 2005/060142 A1 décrit une étape de séparation de source d'un signal multi-canal suivie d'une étape d'identification. Ce document utilise une analyse de composantes indépendantes, ACI, basée sur une matrice de séparation, et peut identifier un signal dominant par deux signaux séparés.Le document US 2010/111290 A1 décrit une étape de séparation de sources d'un signal pouvant être réalisée à l'aide d'une technique d'analyse de composantes indépendantes, ACI, ainsi qu'une étape d'estimation d'un type de son après séparation. Le type de son estimé est stationnaire ou non-stationnaire. L'obtention d'une matrice de séparation semble implicite au vu des procédés de séparation indiqués.En outre, le document BAQUÉ MATHIEU ET AL: "Séparation of Direct Sounds from Early Reflections Using the Entropy Rate Bound Minimization Algorithm", 60ème Conférence Internationale de l'AES: DREAMS (DEREVERBERATION AND REVERBERATION OF AUDIO, MUSIC, AND SPEECH); AES, 60 EAST 42ND STREET, ROOM 2520, NEW YORK 10165-2520, USA, 27 janvier 2016 (2016-01-27), décrit une étude expérimentale utilisant une mesure de débit entropique pour séparer les sons directs de leurs réflexions.

[0011] En pratique, dans certaines conditions, l'étape de séparation $s = Bx$ revient à faire de la formation de voies sous contrainte (ou « beamforming » ci-après) : la combinaison de différents canaux donnée par la matrice **B** consiste à appliquer un filtre spatial dont la directivité revient à imposer un gain unité dans la direction de la source que l'on veut extraire, et un gain nul dans la direction des sources interférentes. Un exemple de beamforming pour extraire trois sources positionnées à respectivement 0°, 90° et -120° d'azimuth est illustré à la figure 1. Chacune des directivités formées correspond à l'extraction d'une des sources de **s.**

[0012] En présence d'un mélange de sources capté dans des conditions réelles, l'effet de salle va générer un champ sonore dit réverbéré, noté x_r , qui va s'ajouter aux champs directs des sources :

$$\boldsymbol{x} = \boldsymbol{As} + \boldsymbol{x}_r$$

[0013] Le champ acoustique total peut être modélisé comme la somme du champ direct des sources d'intérêt (représenté en 1 sur la figure 2), des premières réflexions (sources secondaires, représentées en 2 sur la figure 2) et d'un champ diffus (représenté en 3 sur la figure 2). La matrice de covariance des observations est alors de rang plein, quel que soit le nombre réel de sources actives dans le mélange : cela signifie que l'on ne peut plus utiliser le rang de **Co** pour estimer le nombre de sources.

[0014] Ainsi, lorsqu'on utilise un algorithme de SAS pour séparer des sources en milieu réverbérant, la matrice de séparation **B** de taille MxM est obtenue, générant en sortie M sources $\tilde{s}_j$, $1 \leq j \leq M$, au lieu des N désirées, les M-N dernières composantes contenant essentiellement du champ réverbéré, par l'opération matricielle :

$$\tilde{\mathbf{s}} = \mathbf{B}.\mathbf{x}$$

[0015] Ces composantes supplémentaires posent plusieurs problèmes :

- pour l'analyse de scène : on ne sait pas *a priori* quelles sont les composantes relatives aux sources et les composantes induites par l'effet de salle.
- pour la séparation des sources par formation de voies : chaque composante supplémentaire induit des contraintes sur les directivités formées et dégrade généralement le facteur de directivité avec pour conséquence un rehaussement du niveau de réverbération dans les signaux extraits.

[0016] Les méthodes existantes de comptage de sources pour des contenus multicanal sont souvent basées sur une hypothèse de parcimonie dans le domaine temps-fréquence, c'est-à-dire sur le fait que pour chaque zone temps-fréquence, une seule source ou un nombre limité de sources va avoir une contribution énergétique non-négligeable. Pour la plupart d'entre-elles, une étape de localisation de la source la plus énergétique est effectuée pour chaque zone (ou « bin » en anglais), puis les zones sont agrégées (étape dite de « clustering » en anglais) pour reconstruire la contribution totale de chaque source.

[0017] L'approche DUET (Pour « *Degenerate Unmixing Estimation Technique* ») décrite par exemple dans le document « Blind séparation of disjoint orthogonal signals: Demixing n sources from 2 mixtures.» des auteurs A. Jourjine, S. Rickard, et O. Yilmaz, publié en 2000 dans ICASSP'00, permet de localiser et extraire *N* sources en conditions anéchoïques à partir de seulement deux observations non coïncidentes, en faisant l'hypothèse que les sources ont des supports fréquentiels disjoints, soit

$$S_i(f)S_j(f) = 0$$

pour tout *f* dès lors que $i \neq j$.

[0018] Après une décomposition des observations en sous-bandes fréquentielles, typiquement réalisée via une transformée de Fourier à court-terme, une amplitude a_i et un retard t_i sont estimés pour chaque sous-bande en se basant sur l'équation de mélange théorique :

$$\begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ a_1 e^{-i\omega t_1} & \cdots & a_N e^{-i\omega t_N} \end{bmatrix} \cdot \begin{bmatrix} S_1(f) \\ \cdots \\ S_N(f) \end{bmatrix}$$

[0019] Dans chaque bande de fréquence *f*, un couple (a_i, t_i) correspondant à la source *i* active est estimée de la façon suivante :

$$\begin{cases} a_i &= \left| \frac{X_2(f)}{X_1(f)} \right| \\ t_i &= \frac{1}{2\pi f} \Im \left\{ \log \frac{X_2(f)}{X_1(f)} \right\} \end{cases}$$

[0020] Une représentation dans l'espace de l'ensemble des couples (a_i, t_i) est effectuée sous forme d'histogramme, le « *clustering* » est alors effectuée sur l'histogramme par maximum de vraisemblance, fonction de la position de la zone et de la position supposée de la source associée, en supposant une distribution gaussienne des positions estimées de chaque zone autour de la position réelle des sources.

[0021] En pratique, l'hypothèse de parcimonie des sources dans le domaine temps-fréquence est souvent mise en défaut, ce qui constitue une limitation importante de ces approches pour le dénombrement de sources, car les directions d'arrivée pointées pour chaque zone résultent alors d'une combinaison des contributions de plusieurs sources et le « *clustering* » ne s'effectue plus correctement. De plus, pour l'analyse de contenus captés en conditions réelles, la présence de réverbération peut d'une part dégrader la localisation des sources et d'autre part engendrer une sur-estimation du nombre de sources réelles lorsque des premières réflexions atteignent un niveau énergétique suffisant pour être perçues comme des sources secondaires.

[0022] La présente invention vient améliorer la situation.

Elle propose à cet effet, un procédé de traitement de données sonores pour une séparation de N sources sonores d'un signal sonore multicanal capté en milieu réel. Le procédé est tel qu'il comporte les étapes suivantes :

- application d'un traitement de séparation de sources au signal multicanal capté et obtention d'une matrice de séparation et d'un ensemble de M composantes sonores, avec $M \geq N$;
- calcul d'un ensemble de premiers descripteurs dit bi-variés, représentatifs d'une mesure de corrélation entre les composantes des couples de l'ensemble des M composantes obtenu ;
- calcul d'un ensemble de seconds descripteurs dit uni-variés représentatifs de caractéristiques d'encodage des composantes de l'ensemble des M composantes obtenu, le calcul étant fonction d'une mise en correspondance entre les caractéristiques d'encodage estimées et issues d'une matrice inverse de la matrice de séparation et des caractéristiques d'encodage théoriques d'une source de type onde plane ;
- classification des composantes de l'ensemble des M composantes, selon deux classes de composantes, une première classe de N composantes dites directes correspondant aux N sources sonores directes et une deuxième classe de M-N composantes dites réverbérées, par un calcul de probabilité d'appartenance à une des deux classes, fonction des ensembles de premiers et seconds descripteurs.

Cette méthode permet donc de discriminer les composantes issues de sources directes et les composantes issues de réverbération des sources lorsque la captation du signal sonore multicanal s'effectue dans un milieu réverbérant, c'est-à-dire avec effet de salle. Ainsi, l'ensemble de premiers descripteurs bi-variés permet de déterminer d'une part si les composantes d'un couple de l'ensemble de composantes obtenues suite à l'étape de séparation de sources font partie d'une même classe de composantes ou d'une classe différente alors que l'ensemble de seconds descripteurs uni-variés permet de définir pour une composante, si elle a plus de probabilité d'appartenir à telle ou telle classe. Ceci permet donc de déterminer la probabilité d'appartenance d'une composante à une des deux classes et ainsi de déterminer les N sources sonores directes correspondant aux N composantes classées dans la première classe.

[0023] Les différents modes particuliers de réalisation mentionnés ci-après peuvent être ajoutés indépendamment ou en combinaison les uns avec les autres, aux étapes du procédé de traitement défini ci-dessus.

[0024] Dans un mode de réalisation particulier, le calcul d'un descripteur bi-varié comporte le calcul d'un score de cohérence entre deux composantes. Ce calcul de descripteur permet de façon pertinente de savoir si un couple de composantes correspond à deux composantes directes (2 sources) ou si au moins une des composantes provient d'un effet réverbérant.

[0025] Selon un mode de réalisation, le calcul d'un descripteur bi-varié comporte la détermination d'un retard entre les deux composantes du couple.
Cette détermination du retard et du signe associé à ce retard permet de déterminer, pour un couple de composantes, quelle composante correspond plus probablement au signal direct et quelle composante correspond plus probablement au signal réverbéré.

[0026] Selon une implémentation possible de ce calcul de descripteur, le retard entre deux composantes est déterminé par la prise en compte du retard maximisant une fonction d'inter-corrélation entre les deux composantes du couple.
Cette méthode d'obtention du retard offre une détermination d'un descripteur bi-varié fiable.

[0027] Dans un mode de réalisation particulier, la détermination du retard entre deux composantes d'un couple est associée à un indicateur de fiabilité du signe du retard, fonction de la cohérence entre les composantes du couple.

[0028] Dans une variante de réalisation, la détermination du retard entre deux composantes d'un couple est associée à un indicateur de fiabilité du signe du retard, fonction du rapport du maximum d'une fonction d'inter-corrélation pour des retards de signe opposé.
Ces indicateurs de fiabilité permettent de rendre plus fiable la probabilité, pour un couple de composantes appartenant à une classe différente, que chaque composante du couple soit la composante directe ou la composante réverbérée.

[0029] Selon un mode de réalisation, le calcul d'un descripteur uni-varié est fonction d'une mise en correspondance entre des coefficients de mélange d'une matrice de mélange estimée à partir de l'étape de séparation de sources et des caractéristiques d'encodage d'une source de type onde plane.
Ce calcul de descripteur permet pour une composante seule, d'estimer la probabilité que la composante soit directe ou réverbérée.

[0030] Dans un mode de réalisation, la classification des composantes de l'ensemble des M composantes s'effectue par la prise en compte de l'ensemble des M composantes, et par le calcul de la combinaison la plus probable des classifications des M composantes.

[0031] Dans une implémentation possible de cette approche globale, le calcul de la combinaison la plus probable s'effectue par la détermination d'un maximum des valeurs de vraisemblance exprimées comme le produit des probabilités conditionnelles associées aux descripteurs, pour les combinaisons possibles de classification des M composantes.

[0032] Dans un mode de réalisation particulier, une étape de pré-selection des combinaisons possibles est effectuée en se basant sur les seuls descripteurs uni-variés avant l'étape de calcul de la combinaison la plus probable.

[0033] Cela diminue ainsi les calculs de vraisemblance à effectuer sur les combinaisons possibles puisque ce nombre de combinaisons est restreint par cette étape de pré-sélection.

[0034] Dans une variante de réalisation, une étape de pré-selection des composantes est effectuée en se basant sur les seuls descripteurs uni-variés avant l'étape de calcul des descripteurs bi-variés.

[0035] Ainsi, le nombre de descripteurs bi-variés à calculer est restreint, ce qui diminue la complexité du procédé.

[0036] Dans un exemple de réalisation, le signal multicanal est un signal ambisonique.

[0037] Cette méthode de traitement ainsi décrite s'applique parfaitement à ce type de signaux.

[0038] L'invention se rapporte également à un dispositif de traitement de données sonores mis en oeuvre pour effectuer un traitement de séparation de N sources sonores d'un signal sonore multicanal capté par une pluralité de capteurs en milieu réel. Le dispositif est tel qu'il comporte :

- une interface d'entrée pour recevoir les signaux captés par une pluralité de capteurs, du signal sonore multicanal;
- un circuit de traitement comportant un processeur et apte à mettre en oeuvre:

    - o un module de traitement de séparation de sources appliqué au signal multicanal capté pour obtenir une matrice de séparation et un ensemble de M composantes sonores, avec $M \geq N$;
    - o un calculateur apte à calculer un ensemble de premiers descripteurs dit bi-variés, représentatifs d'une mesure de corrélation entre les composantes des couples de l'ensemble des M composantes obtenu et un ensemble de seconds descripteurs dit uni-variés représentatifs de caractéristiques d'encodage des composantes de l'ensemble des M composantes obtenu, le calcul étant fonction d'une mise en correspondance entre les caractéristiques d'encodage estimées et issues d'une matrice inverse de la matrice de séparation et des caractéristiques d'encodage théoriques d'une source de type onde plane ;
    - o un module de classification des composantes de l'ensemble des M composantes, selon deux classes de composantes, une première classe de N composantes dites directes correspondant aux N sources sonores directes et une deuxième classe de M-N composantes dites réverbérées, par un calcul de probabilité d'appartenance à une des deux classes, fonction des ensembles de premiers et seconds descripteurs ;

- une interface de sortie pour délivrer l'information de classification des composantes.

[0039] L'invention s'applique également à un programme informatique comportant des instructions de code pour la mise en oeuvre des étapes du procédé de traitement tel que décrit précédemment, lorsque ces instructions sont exécutées par un processeur et à un support de stockage, lisible par un processeur, sur lequel est enregistré un programme informatique comprenant des instructions de code pour l'exécution des étapes du procédé de traitement tel que décrit. Le dispositif, programme et support de stockage présentent les mêmes avantages que le procédé décrit précédemment, qu'ils mettent en oeuvre.

[0040] D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description suivante, donnée uniquement à titre d'exemple non limitatif, et faite en référence aux dessins annexés, sur lesquels :

- la figure 1 illustre une formation de voie pour extraire trois sources selon une méthode de séparation de sources de l'état de l'art tel que décrit précédemment ;
- la figure 2 illustre une réponse impulsionnelle avec effet de salle tel que décrit précédemment ;
- la figure 3 illustre sous forme d'organigramme, les étapes principales d'un procédé de traitement selon un mode de réalisation de l'invention ;
- la figure 4 illustre en fonction de la fréquence, des fonctions de cohérence représentant des descripteurs bi-variés entre deux composantes selon un mode de réalisation de l'invention, et selon différents couples de composantes ;
- la figure 5 illustre les densités de probabilités des cohérences moyennes représentants des descripteurs bi-variés selon un mode de réalisation de l'invention et pour différents couples de composantes et différents nombres de sources ;
- la figure 6 illustre des fonctions d'inter-corrélâtion entre deux composantes de classe différentes selon un mode de réalisation de l'invention et selon le nombre de sources ;
- la figure 7 illustre les densités de probabilité d'un critère d'onde plane en fonction de la classe de la composante, de l'ordre ambisonique et du nombre de sources, pour un mode de réalisation particulier de l'invention ;
- la figure 8 illustre une représentation matérielle d'un dispositif de traitement selon un mode de réalisation de l'invention, mettant en oeuvre un procédé de traitement selon un mode de réalisation de l'invention ; et
- La figure 9 illustre un exemple de calcul de loi de probabilité pour un critère, de cohérence entre une composante directe et une composante réverbérée selon un mode de réalisation de l'invention.

[0041] La **figure 3** illustre les principales étapes d'un procédé de traitement de données sonores pour une séparation de N sources sonores d'un signal sonore multicanal capté en milieu réel dans un mode de réalisation de l'invention.

[0042] Ainsi, à partir d'un signal multicanal capté par une pluralité de capteurs placés dans un milieu réel, c'est-à-dire réverbérant, et délivrant un nombre M d'observations issues de ces capteurs (x (x_1, ..., x_M)), le procédé met en oeuvre une étape E310 de séparation aveugle de sources sonores (SAS). On suppose ici, dans ce mode de réalisation que le nombre d'observations est égal ou supérieur au nombre de sources actives.

[0043] L'utilisation d'un algorithme de séparation aveugle de sources appliqué aux M observations, permet dans le cas d'un milieu réverbérant, d'extraire par formation de voies M composantes sonores associées à une matrice de mélange estimée A_{MxM}, soit :

$\boldsymbol{s} = \boldsymbol{Bx}$ avec x le vecteur des M observations, B la matrice de séparation estimée par la séparation en aveugle de sources, de dimensions MxM et **s** le vecteur des M composantes sonores extraites. Parmi celles-ci se trouvent théoriquement N sources sonores et M-N composantes résiduelles correspondant à de la réverbération.

[0044] Pour obtenir la matrice B de séparation, l'étape de séparation aveugle de sources peut être mise en oeuvre, par exemple en utilisant un algorithme d'analyse en composantes indépendantes (ou « ACI »), ou encore un algorithme d'analyse en composantes principales.

[0045] Dans un exemple de réalisation, on s'intéresse aux signaux multicanal de type ambisonique.

[0046] L'ambisonie consiste en une projection du champ acoustique sur une base de fonctions harmoniques sphériques, pour obtenir une représentation spatialisée de la scène sonore. La fonction $Y_{mn}^{\sigma}(\theta, \phi)$ est l'harmonique sphérique d'ordre m et d'indice $n\sigma$, dépendant des coordonnées sphériques (θ, ϕ), définie avec la formule suivante :

$$Y_{mn}^{\sigma}(\theta, \phi) = \tilde{P}_{mn}(\cos\phi).\begin{cases}\cos n\theta & \sigma = 1\\ \sin n\theta & \sigma = -1 \;\; n \geq 1\end{cases}$$

où $\tilde{P}_{mn}(\cos\phi)$ est une fonction pôlaire impliquant le polynome de Legendre :

$$\tilde{P}_{mn}(\mathrm{x}) \doteq \sqrt{\epsilon_n \frac{(m-n)!}{(m+n)!}}(-1)^n(1-x^2)^{\frac{n}{2}}\frac{d^n}{dx^n}P_m(x)$$

avec $\varepsilon_0 = 1$ et $\varepsilon_0 = 2$ pour $n \geq 1$
et

$$P_m(x) = \frac{1}{2^m \cdot m!}\frac{d^n}{dx^n}(x^2-1)^m$$

[0047] En pratique, un encodage ambisonique réel se fait à partir d'un réseau de capteurs, généralement répartis sur une sphère. Les signaux capturés sont combinés pour synthétiser un contenu ambisonique dont les canaux respectent au mieux les directivités des harmoniques sphériques. On décrit ci-après les principes de base de l'encodage ambisonique.

[0048] Le formalisme ambisonique, initialement limité à la représentation de fonctions harmoniques sphériques d'ordre 1, a par la suite été étendu aux ordres supérieurs. Le formalisme ambisonique avec un nombre de composantes plus important est communément nommé « *Higher Order Ambisonics* » (ou « HOA » ci-après).

[0049] A chaque ordre m correspondent 2m+1 fonctions harmoniques sphériques. Ainsi, un contenu d'ordre m contient un total de $(m+1)^2$ canaux (4 canaux à l'ordre 1, 9 canaux à l'ordre 2, 16 canaux à l'ordre 3, et ainsi de suite).

[0050] On entend ci-après par « composantes ambisoniques » le signal ambisonique dans chaque canal ambisonique, en référence aux « composantes vectorielles » dans une base vectorielle qui serait formée par chaque fonction harmonique sphérique. Ainsi par exemple, on peut compter :

- une composante ambisonique pour l'ordre m=0,
- trois composantes ambisoniques pour l'ordre m=1,
- cinq composantes ambisoniques pour l'ordre m=2,
- sept composantes ambisoniques pour l'ordre m=3, etc.

[0051] Les signaux ambisoniques captés pour ces différentes composantes sont alors répartis sur un nombre M de canaux qui se déduit de l'ordre maximum m qu'il est prévu de capter dans la scène sonore. Par exemple, si une scène sonore est captée avec un microphone ambisonique à 20 capsules piézoélectriques, alors l'ordre ambisonique maximum capté est m=3, afin qu'il n'y ait pas plus de 20 canaux $M=(m+1)^2$, le nombre de composantes ambisoniques considérées est 7+5+3+1 = 16 et le nombre M de canaux est M=16, donné par ailleurs par la relation $M=(m+1)^2$, avec m=3.

[0052] Ainsi dans l'exemple d'implémentation où le signal multicanal est un signal ambisonique, l'étape E310 reçoit les signaux $\boldsymbol{x}$ ($\boldsymbol{x}_1$, ...,$\boldsymbol{x}_1$, ...,$\boldsymbol{x}_M$), captés par un microphone réel, en milieu réverbérant et qui reçoit des trames de contenus sonores ambisoniques sur $M= (m+1)^2$ canaux et contenant N sources.

[0053] La séparation en aveugle de sources est donc effectuée à l'étape E310 comme expliqué précédemment.

[0054] Cette étape permet à la fois d'extraire M composantes et la matrice de mélange estimée. Les composantes obtenues en sortie de l'étape de séparation de sources peuvent être classées selon deux classes de composantes : une première classe de composantes dites directes correspondant aux sources sonores directes et une deuxième classe de composantes dites réverbérées correspondants aux réflexions des sources.

[0055] A l'étape E320, un calcul de descripteurs des M composantes (s_1, s_2, ...s_M) issues de l'étape de séparation des sources est mis en oeuvre, descripteurs qui vont permettre d'associer à chaque composante extraite la classe qui lui correspond : composante directe ou composante réverbérée.

[0056] Deux types de descripteurs sont calculés ici : des descripteurs bi-variés qui font intervenir des couples de composantes (s_j, s_i) et des descripteurs uni-variés calculés pour une composante s_i.

[0057] Ainsi, un ensemble de premiers descripteurs bi-variés est calculé. Ces descripteurs sont représentatifs de relations statistiques entre les composantes des couples de l'ensemble des M composantes obtenu.

[0058] Trois cas de figures peuvent être modélisés en fonction des classes respectives des composantes :

- Les deux composantes sont des champs directs,
- L'une des deux composantes est directe et l'autre est réverbérée,
- Les deux composantes sont réverbérées.

Selon un mode de réalisation, on calcule ici une cohérence moyenne entre deux composantes. Ce type de descripteur représente une relation statistique entre les composantes d'un couple et fournit une indication sur la présence d'au moins une composante réverbérée dans un couple de composantes.

[0059] En effet, chaque composante directe est principalement constituée du champ direct d'une source, assimilable à une onde plane, auquel s'ajoute une réverbération résiduelle dont la contribution énergétique est inférieure à celle du champ direct. Les sources étant par nature statistiquement indépendantes, il y a donc une faible corrélation entre les composantes directes extraites.

[0060] A l'inverse, chaque composante réverbérée est constituée de premières réflexions, versions retardées et filtrées du ou des champs directs, et d'une réverbération tardive. Ainsi, les composantes réverbérées présentent une corrélation significative avec les composantes directes, et généralement un retard de groupe identifiable par rapport aux composantes directes.

[0061] La fonction de cohérence γ_{jl}^2 renseigne sur l'existence d'une corrélation entre deux signaux s_j et s_i et s'exprime suivant la formule :

$$\gamma_{jl}^2(f) = \frac{\left|\Gamma_{jl}(f)\right|^2}{\Gamma_j(f)\Gamma_l(f)}$$

où $\Gamma_{jl}(f)$ est l'interspectre entre s_j et si et $\Gamma_j(f)$ et $\Gamma_l(f)$ sont les autospectres respectifs de s_j et s_l.
La cohérence est idéalement nulle lorsque s_j et si sont les champs directs de sources indépendantes mais elle prend une valeur élevée lorsque s_j et si sont deux contributions d'une même source : le champ direct et une première réflexion ou bien deux réflexions.

[0062] Une telle fonction de cohérence indique donc une probabilité d'avoir deux composantes directes ou deux contributions d'une même source (directe/réverbérée ou première réflexion/réflexions ultérieures).

[0063] En pratique, les interspectres et autospectres pourront être calculés en segmentant les composantes extraites en K trames (adjacentes ou avec recouvrement), en appliquant une transformée à Fourier à court-terme à chaque trame *k* de ces K trames pour produire les spectres instantanés $S_j(k,f)$, et en moyennant les observations sur les K trames :

$$\Gamma_{jl}(f) = E_{k\in\{1\ldots K\}}\left\{S_j(k,f)S_l^*(k,f)\right\}$$

[0064] Le descripteur utilisé pour un signal large bande est la moyenne sur l'ensemble des fréquences de la fonction de cohérence entre deux composantes, soit :

$$d^{\gamma}(s_j, s_l) = E_f\left\{\gamma_{jl}^2(f)\right\}$$

La cohérence étant bornée entre 0 et 1, la cohérence moyenne sera également comprise dans cet intervalle, tendant vers 0 pour des signaux parfaitement indépendants et vers 1 pour des signaux fortement corrélés.
La **figure 4** donne un aperçu des valeurs de cohérence en fonction de la fréquence pour les cas suivants :

- Cas N°1 où les valeurs de cohérence sont obtenues pour deux composantes directes issues de 2 sources distinctes.
- Cas N°2 où les valeurs de cohérence sont obtenues pour un couple de composantes directes et réverbérée pour une seule source active.
- Cas N°3 où les valeurs de cohérence sont obtenues pour un couple de composantes directe et réverbérée mais lorsque deux sources sont actives simultanément.

[0065] On remarque que dans le premier cas, la valeur de cohérence d^{γ} est inférieur à 0.3 alors que dans le second cas d^{γ} atteint 0.7 en présence d'une seule source active. Ces valeurs reflètent bien à la fois l'indépendance des signaux directs et la relation liant un signal direct et le même signal réverbéré, en l'absence d'interférences. Cependant, en incorporant une seconde source active dans le mélange initial (Cas N°3), la cohérence moyenne du cas direct/réverbéré descend à 0.55 et se retrouve fortement dépendante du contenu spectral et du niveau énergétique des différentes sources. Ici, la concurrence des différentes sources fait chuter la cohérence en basses fréquences, tandis que les valeurs sont plus élevées au-dessus de 5500 Hz en raison d'une plus faible contribution de la source interférente.

[0066] On remarque donc que la détermination d'une probabilité d'appartenance à une même classe ou à une classe différente pour un couple de composante, peut dépendre du nombre de sources *a priori* actives. Pour l'étape de classification E340 décrite ultérieurement, ce paramètre pourra être pris en compte dans un mode particulier de réalisation.

[0067] A l'étape E330 de la figure 3, un calcul de probabilité est déduit du descripteur ainsi décrit.

[0068] En pratique, les densités de probabilités des figures 5 et 7 décrites ci-après, et plus généralement toutes les densités de probabilité des descripteurs, sont apprises de manière statistique sur des bases de données comprenant des conditions acoustiques variées (réverbérantes/mâtes) et différentes sources (voix d'homme/femme, langues française/anglaise/...). Les composantes sont classées de manière informée : à chaque source est associée la composante extraite la plus proche spatialement, les restantes étant classées comme composantes réverbérées. Pour calculer la position de la composante, on utilise les 4 premiers coefficients de son vecteur de mélange issu de la matrice **A** (soit l'ordre 1), inverse de la matrice de séparation **B.** En faisant l'hypothèse que ce vecteur suit la règle d'encodage d'une onde plane soit :

$$\begin{bmatrix} 1 \\ \cos\theta\cos\varphi \\ \sin\theta\cos\varphi \\ \sin\varphi \end{bmatrix}$$

où (θ, φ) représentent les coordonnées sphériques, azimuth/élévation, de la source, il est possible de déduire par simple calcul trigonométriques la position de la composantes extraite par le jeu d'équations suivant :

$$\begin{cases} \theta = arctan2\left(\dfrac{a_3}{a_2}\right) \\ \varphi = arctan2\left(\dfrac{a_4 * sign(a_1)}{\sqrt{a_2^2 + a_3^2}}\right) \end{cases}$$

où arctan2 est la fonction arctangente qui permet de lever l'ambigüité de signe de la fonction arctangente.

Une fois les signaux classés, les différents descripteurs sont calculés. Du nuage de points - issus de la base de données - pour une classe donnée est extrait un histogramme de valeurs du descripteur à partir duquel une densité de probabilité est choisie parmi une collection de densités de probabilité, sur la base d'une distance, généralement la divergence de Kullback-Leibler. La **figure 9** montre un exemple de calcul de loi pour le critère de cohérence entre une composante directe et une composante réverbérée : la loi log-normale a été sélectionnée parmi une dizaine de lois car elle minimise la divergence de Kullback-Leibler.

Pour l'exemple d'un signal ambisonique, la **figure 5** représente les distributions (densité de probabilité ou pdf pour « Probability density function » en anglais) associées à la valeur de la cohérence moyenne entre deux composantes.

[0069] Les lois de probabilité représentées ici sont présentées pour une captation microphonique ambisonique à 4 canaux (ambisonie ordre 1) ou 9 canaux (ambisonie d'ordre 2), dans le cas d'une ou deux sources actives simultanément. On observe tout d'abord que la cohérence moyenne d^{γ} prend des valeurs nettement plus faibles pour des couples de composantes directes par rapport aux cas où au moins une des composantes est réverbérée, et cette observation est d'autant plus marquée que l'ordre ambisonique est élevé. Cela est dû à une meilleure sélectivité de la formation de voies lorsque le nombre de canaux est plus important, et donc à une meilleure séparation des composantes extraites.

[0070] On constate également qu'en présence de deux sources actives, les estimateurs de cohérence se dégradent, que ce soient les couples direct/réverbéré ou réverbéré/réverbéré (en présence d'une seule source, le couple direct/direct n'existe pas).

En définitive, il apparaît que les densités de probabilité dépendent fortement du nombre de sources dans le mélange, et du nombre de capteurs à disposition.

[0071] Ce descripteur est donc pertinent pour détecter si un couple de composantes extraites correspond à deux composantes directes (2 vraies sources) ou si au moins l'une des deux composantes provient de l'effet de salle.

[0072] Dans un mode de réalisation de l'invention, un autre type de descripteur bi-varié est calculé à l'étape E320. Soit ce descripteur est calculé à la place du descripteur de type cohérence décrit précédemment, soit en complément de celui-ci.

[0073] Ce descripteur va permettre de déterminer, pour un couple (direct/réverbéré) quelle composante est plus probablement le signal direct et laquelle correspond au signal réverbéré, en se basant sur l'hypothèse simple que les premières réflexions sont des versions retardées et atténuées du signal direct.

[0074] Ce descripteur est basé sur une autre relation statistique entre les composantes, le retard entre les deux composantes du couple. On définit le retard $\tau_{jl,max}$ comme le retard qui maximise la fonction d'intercorrélation $r_{jl}(\tau) = E_t\{s_{sj}(t)s_l(t - \tau)\}$ entre les composantes d'un couple de composantes s_j et s_l :

$$\tau_{jl,max} = arg\max_{\tau} |r_{jl}(\tau)|$$

[0075] Lorsque s_j est un signal direct et s_l une réflexion associée, le tracé de la fonction d'intercorrélation fera généralement apparaître un $\tau_{jl,max}$ négatif. Ainsi, si l'on sait que l'on est en présence d'un couple de composantes direct/réverbéré, on peut ainsi théoriquement attribuer la classe à chacune des composantes grâce au signe de $\tau_{jl,max}$.

[0076] En pratique, l'estimation du signe de $\tau_{jl,max}$ est souvent très bruitée, voire même parfois inversée :

- Lorsque la scène est constituée d'une seule source, il n'y a pas forcément de délai de groupe qui émerge distinctement si le champ réverbérée est composé de multiples réflexions et de réverbération tardive. De plus les composantes directes extraites par SAS contiennent toujours un résidu d'effet de salle plus ou moins important, qui va bruiter la mesure du délai.
- Lorsque plusieurs sources sont présentes, les interférences viennent perturber la mesure, à plus forte raison si les trames d'analyse sont courtes et que tous les champs directs n'ont pas été parfaitement séparés.

[0077] Pour ces raisons, on peut choisir de fiabiliser le signe de $\tau_{jl,max}$ utilisé comme descripteur, grâce à un indicateur de robustesse ou de fiabilité.

[0078] La cohérence moyenne entre les composantes permet d'évaluer la pertinence du couple direct/réverbéré comme vu précédemment. Si celle-ci est forte, on peut espérer que le délai de groupe sera un descripteur fiable.

[0079] D'autre part, la valeur relative du pic d'intercorrélation $\tau_{jl,max}$ aux autres valeurs de la fonction d'intercorrélation $r_{jl}(\tau)$ renseigne également sur la fiabilité du délai de groupe. La **figure 6** illustre le caractère émergent du pic d'autocorrélation entre une composante directe et une composante réverbérée. Sur la partie haute (1) de la **figure 6** où une seule source est présente, le maximum d'intercorrélation émerge clairement du reste de l'intercorrélation, indiquant de manière fiable que l'une des composantes est en retard par rapport à l'autre. Il émerge notamment par rapport aux valeurs de la fonction d'autocorrélation pour des signes opposés à celui de $\tau_{jl,max}$ (celle des τ positifs sur la **figure 6)** qui sont très faibles, quelle que soit la valeur de τ.

[0080] Dans une réalisation particulière, on définit un second indicateur de fiabilité du signe du retard appelé *émergence,* en calculant le rapport entre la valeur absolue de l'intercorrélation à τ_{max} et celle du maximum de corrélation pour des τ de signe opposé à celui de $\tau_{jl,max}$:

$$emergence_{jl} = \left| \frac{r_{jl}(\tau_{jl,max})}{r_{jl}(\bar{\tau}_{jl,max})} \right|$$

où $\bar{\tau}_{jl,max}$ est défini par :

$$\bar{\tau}_{jl,max} = arg \max_{sign(\tau) \neq sign(\tau_{jl,max})} |r_{jl}(\tau)|$$

[0081] Ce ratio, que l'on nomme émergence, est un critère *ad hoc* dont la pertinence se vérifie en pratique : il prend des valeurs proches de 1 pour des signaux indépendants, i.e. 2 composantes directes, et des valeurs plus élevées pour des signaux corrélées comme une composante directe et une composante réverbérée. Dans le cas précité de la courbe (1) de la **figure 6,** la valeur d'émergence est de 4.

[0082] On a donc un descripteur d^{τ} qui détermine, pour chaque couple supposé direct/réverbéré, la probabilité pour chaque composante du couple d'être la composante directe ou la composante réverbérée. Ce descripteur est fonction du signe de $\tau_{max,}$ de la cohérence moyenne entre les composantes et de l'émergence du maximum d'intercorrélation.

[0083] Il faut noter que ce descripteur est sensible au bruit, et notamment à la présence de plusieurs sources simultanées, comme illustré sur la courbe (2) de la **figure 6** : en présence de 2 sources, même si le maximum de corrélation émerge toujours, sa valeur relative - 2.6 - est moindre du fait de la présence d'une source interférente qui réduit la corrélation entre les composantes extraites. Dans une réalisation particulière, on mesurera la fiabilité du signe du retard en fonction de la valeur de l'émergence, que l'on pondérera par le nombre *a priori* de sources à détecter.

[0084] Avec ce descripteur, on calcule à l'étape E330 une probabilité d'appartenance à une première classe de composantes directes ou une seconde classe de composants réverbérées pour un couple de composantes. Pour s_j identifiée comme étant en avance sur si, on estime la probabilité que s_j soit directe et si réverbérée par une loi à deux dimensions.

[0085] Logiquement, on estime alors la probabilité que s_j soit réverbérée et s_l directe alors même que s_j est en avance de phase comme le complément à 1 du cas direct/réverbéré :

$$p(\mathcal{C}_j = \mathcal{C}^r, \mathcal{C}_l = \mathcal{C}^d | d^\tau) = 1 - p(\mathcal{C}_j = \mathcal{C}^d, \mathcal{C}_l = \mathcal{C}^r | d^\tau)$$

où C_j et C_l sont les classes respectives des composantes s_j et s_l, C^d étant la première classe de composantes dites directes correspondant aux N sources sonores directes et C^r, la deuxième classe de M-N composantes dites réverbérées.

[0086] Ce descripteur n'est utilisable que pour les couples direct/réverbéré. Les couples direct/direct et réverbéré/réverbéré ne sont pas concernés par ce descripteur, on les considère donc comme équiprobables :

$$\begin{cases} p(\mathcal{C}_j = \mathcal{C}^d, \mathcal{C}_l = \mathcal{C}^d | d^\tau) = 0.5 \\ p(\mathcal{C}_j = \mathcal{C}^r, \mathcal{C}_l = \mathcal{C}^r | d^\tau) = 0.5 \end{cases}$$

[0087] Le signe du retard est un indicateur fiable lorsqu'à la fois la cohérence et l'émergence ont des valeurs moyennes ou élevées. Une émergence faible ou une cohérence faible vont rendre les couples direct/réverbéré ou réverbéré/direct équiprobables.

[0088] A l'étape E320, est également calculé un ensemble de seconds descripteurs dit uni-variés représentatifs de caractéristiques d'encodage des composantes de l'ensemble des M composantes obtenu.

[0089] Connaissant le système de captation utilisé, l'encodage d'une source provenant d'une direction donnée s'effectue avec des coefficients de mélange dépendant, entre autres, de la directivité des capteurs. Dans le cas où la source peut être considérée comme ponctuelle et où les longueurs d'onde sont grandes par rapport à la taille de l'antenne, on peut considérer la source comme une onde plane. Cette hypothèse se vérifie généralement dans le cas d'un microphone ambisonique qui est de petite taille, pour peu que la source soit suffisamment éloignée du microphone (en pratique, un mètre suffit).

[0090] Pour une composante s_j extraite par SAS, la $j^{ème}$ colonne de la matrice de mélange estimée **A,** obtenue par inversion de la matrice de séparation **B,** va contenir les coefficients de mélange associés à celle-ci. Si cette composante est directe, c'est-à-dire qu'elle correspond à une seule source, les coefficients de mélange de la colonne Aj vont tendre vers les caractéristiques de l'encodage microphoniques pour une onde plane. Dans le cas d'une composante réverbérée, somme de plusieurs réflexions et d'un champ diffus, les coefficients de mélange estimés seront plus aléatoires et ne correspondront pas à l'encodage d'une seule source avec une direction d'arrivée précise.

[0091] On peut donc se servir de la conformité entre les coefficients de mélange estimés et les coefficients de mélange théoriques pour une source seule pour estimer une probabilité que la composante soit directe ou réverbérée.

[0092] Dans le cas d'une captation microphonique ambisonique d'ordre 1, l'encodage d'une onde plane s_j d'incidence (θ_j, ϕ_j) au format ambisonique dit N3D s'effectue suivant la formule :

$$\mathbf{x}_j = \mathbf{A}_j s_j$$

[0093] Où

$$\mathrm{A}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ a_{3j} \\ a_{4j} \end{bmatrix} = \begin{bmatrix} 1 \\ \sqrt{3}\cos\theta_j \cos\phi_j \\ \sqrt{3}\sin\theta_j \cos\phi_j \\ \sqrt{3}\sin\phi_j \end{bmatrix}$$

[0094] Il existe en effet plusieurs formats ambisoniques, qui se différencient notamment par la normalisation des différentes composantes regroupées en ordre. On considère ici le format connu N3D. Les différents formats sont par exemple décrits au lien suivant :
https://en.wikipedia.org/wiki/Ambisonic data exchange formats.

[0095] On peut ainsi déduire des coefficients d'encodage d'une source un critère, nommé critère onde plane, qui illustre la conformité entre les coefficients de mélange estimés et l'équation théorique d'une onde plane encodée seule :

$$c_{op} = \sqrt{\frac{3a_{1j}^2}{a_{2j}^2 + a_{3j}^2 + a_{4j}^2}}$$

[0096] Le critère c_{op} est par définition égal à 1 dans le cas d'une onde plane. En présence d'un champ direct correctement identifié, le critère onde plane restera très proche de la valeur 1. A l'inverse, dans le cas d'une composante réverbérée, la multitude des contributions (premières réflexions et réverbération tardive) avec des niveaux énergétiques équivalents vont généralement éloigner le critère onde plane de sa valeur idéale.

[0097] Pour ce descripteur comme pour les autres, la distribution associée et calculé en E330, connaît une certaine variabilité, en fonction notamment du niveau de bruit présent dans les composantes extraites. Ce bruit est constitué principalement de la réverbération résiduelle et des contributions des sources interférentes qui n'auront pas été parfaitement annulées. On peut donc choisir, pour affiner l'analyse, d'estimer la distribution des descripteurs en fonction :

- Du nombre de canaux utilisés (donc ici de l'ordre ambisonique), qui influe sur la sélectivité du « *beamforming* » et donc sur le niveau de bruit résiduel,
- du nombre de sources contenues dans le mélange (comme pour les descripteurs précédents), dont l'augmentation entraine mécaniquement une hausse du niveau de bruit et une plus grande variance dans l'estimation de la matrice de séparation **B,** donc de **A.**

[0098] On peut observer sur la **figure 7** les lois de probabilités (densité de probabilité) associées à ce descripteur, en fonction du nombre de sources actives simultanément (1 ou 2) et de l'ordre ambisonique du contenu analysé (ordres 1 à 2). Conformément à l'hypothèse initiale, la valeur du critère onde plane est concentrée autour de la valeur 1 pour les composantes directes. Pour les composantes réverbérées, la distribution est plus uniforme, avec cependant une forme légèrement asymétrique, à cause du descripteur lui-même qui est asymétrique, avec une forme en 1/x.

[0099] La distance entre les distributions des deux classes permet une discrimination assez fiable entre les composantes de type ondes planes et celles plus diffuses.

[0100] Ainsi, les descripteurs calculés à l'étape E320 et exposés ici sont basés à la fois sur les statistiques des composantes extraites (cohérence moyenne et retard de groupe) et sur la matrice de mélange estimée (critère onde plane). Ceux-ci permettent de déterminer des probabilités conditionnelles d'appartenance d'une composante à une des deux classes C^d ou C^r.

[0101] A partir du calcul de ces probabilités, il est alors possible, à l'étape E340 de déterminer une classification des composantes de l'ensemble des M composantes, selon les deux classes.

[0102] Pour une composante s_j, on note C_j la classe correspondante. S'agissant de classer l'ensemble des M composantes extraites, on nomme "configuration" le vecteur des classes **C** de dimension 1xM tel que :

$$\mathbf{C} = [C_1, C_2, \ldots, C_M] \text{ avec } C_j \in \{C^d, C^r\}$$

[0103] Sachant qu'il existe deux classes possibles pour chaque composante, le problème revient finalement à choisir parmi un total de 2^M configurations potentielles supposées équiprobables. Pour ce faire, la règle du maximum *a posteriori* est appliquée : connaissant $L(\mathbf{C}_i)$ la vraisemblance de la $\mathbf{i}^{eme}$ configuration, la configuration retenue sera celle possédant la vraisemblance maximale, c'est-à-dire :

$$\mathbf{C} = \arg\max_{\mathbf{C}i} L(\mathbf{C}_i), \forall 1 \leq i \leq 2^M$$

[0104] L'approche choisie peut être exhaustive et consiste alors à estimer la vraisemblance de toutes les configurations possibles, à partir des descripteurs déterminés à l'étape E320 et des distributions qui leur sont associées et qui sont calculés à l'étape E330.

[0105] Selon une autre approche, une pré-sélection des configurations peut être effectuée pour réduire le nombre de configuration à tester, et donc la complexité de la mise en oeuvre de la solution. Cette pré-sélection peut se faire par exemple selon le critère onde plane seul en classant certaines composantes dans la catégorie C^r, dès lors que la valeur de leur critère c_{op} s'éloigne trop de la valeur théorique d'une onde plane 1 : dans le cas de signaux ambisoniques, on peut voir sur les distributions de la **figure 7** que l'on peut, quelle que soit la configuration (ordre ou nombre de sources) et *a priori* sans perte de robustesse, classer dans la catégorie C^r les composantes dont le c_{op} vérifie l'une des inégalités suivantes :

$$\begin{cases} c_{op} < 0.7 \\ c_{op} > 1.5 \end{cases}$$

[0106] Cette pré-sélection permet de réduire le nombre de configurations à tester en pré-classant certaines composantes, en excluant les configurations qui impose la classe C^d à ces composantes pré-classées.

[0107] Une autre possibilité pour réduire plus encore la complexité est d'exclure les composantes pré-classées du calcul des descripteurs bi-variés et du calcul de la vraisemblance, ce qui réduit le nombre de critères bi-variés à calculer et donc encore plus la complexité de traitement.

[0108] Pour estimer la vraisemblance de chaque configuration à l'aide des descripteurs calculés, une approche naïve bayésienne peut être utilisée. Dans ce type d'approche, on se donne un ensemble de descripteurs d_k pour chaque composante s_j Pour chaque descripteur, on formule la probabilité pour la composante s_j d'appartenir à la classe C^α (α=*d* ou *r*) grâce à la loi de Bayes :

$$p(C_j = C^\alpha | d_k) = \frac{p(C_j = C^\alpha) p(d_k | C_j = C^\alpha)}{p(d_k)}$$

[0109] Les deux classes C^r et C^d étant supposées équiprobables, il en découle :

$$p(C_j = C^\alpha) = \frac{1}{2} \quad \forall \alpha$$

ainsi que

$$p(d_k) = \frac{p(d_k | C = C^r) + p(d_k | C = C^d)}{2}$$

[0110] On obtient alors :

$$p(C^\alpha | d_k) = \frac{p(d_k | C^\alpha)}{p(d_k | C^r) + p(d_k | C^d)}$$

où le terme $C^j = C^\alpha$ est abrégé en C^α pour alléger les notations. S'agissant ici de rechercher le maximum de vraisemblance, le terme au dénominateur de chaque probabilité conditionnelle est constant quelle que soit la configuration évaluée. Aussi, on peut par la suite en simplifier l'expression:

$$p(C^\alpha | d_k) \propto p(d_k | C^\alpha)$$

[0111] Pour un descripteur bi-varié (comme par exemple la cohérence) faisant intervenir deux composantes s_j et si et leurs classes respectives supposées, on étend l'expression précédente:

$$p(C_j = C^\alpha, C_l = C^\beta | d_k) \propto p(d_k | C^\alpha, C^\beta)$$

et ainsi de suite.

[0112] La vraisemblance s'exprime comme le produit des probabilités conditionnelles associées à chacun des K

descripteurs, si l'on suppose que ceux-ci sont indépendants :

$$L(\mathbf{C}) = p(\mathbf{d}|\mathbf{C}) = \prod_{k=1}^{K} p(d_k|\mathbf{C})$$

où d est le vecteur des descripteurs et C un vecteur représentant une configuration (c'est à dire la combinaison des classes supposées des *M* composantes), comme définit ci-dessus.

[0113] Plus précisément, un nombre K_1 de descripteurs uni-variés est mis à profit pour chacune des composantes, tandis qu'un nombre K_2 de descripteurs bi-variés est utilisé pour chaque paire de composantes. Les lois de probabilités des descripteurs étant établies en fonction du nombre de sources supposé et du nombre de canaux (l'indice *m* représente l'ordre ambisonique, dans le cas d'une captation de ce type), on formule alors l'expression finale de la vraisemblance :

$$L(C) = \prod_{j=1}^{M}(\prod_{k=1}^{K_1} p(d_k(j)|\, C_j, N, m) \prod_{l=j+1}^{M} \prod_{k=1}^{K_2} p(d_k(j,l)|\, C_j, C_l, N, m))$$

où

- d_k (*j*) est la valeur du descripteur d'indice k pour la composante s_j ;
- d_k (*j, l*) est la valeur du descripteur bi-varié d'indice k pour les composantes s_j et s_l ;
- C_j *et* C_l sont les classes supposées des composantes *j* et *l* ;
- *N* est le nombre de sources actives associé à la configuration évaluée :

$$N = \sum_{j=1}^{M} (C_j = C^d)$$

[0114] Pour des raisons calculatoires, on préfère à la vraisemblance sa version logarithmique (log-vraisemblance) :

$$LL(C) = \sum_{j=1}^{M}(\sum_{k=1}^{K_1} \log p(d_k(j)|\, C_j, N, m) + \sum_{l=j+1}^{M} \sum_{k=1}^{K_2} \log p(d_k(j,l)|\, C_j, C_l, N, m))$$

[0115] Cette équation est celle utilisée en définitive pour déterminer la configuration la plus vraisemblable dans le classificateur bayésien décrit ici pour ce mode de réalisation.

[0116] Le classificateur bayésien présenté ici n'est qu'un exemple d'implémentation, il pourrait être remplacé, entre autres, par une machine à vecteurs de support ou un réseau de neurones.

[0117] Au final, la configuration présentant le maximum de vraisemblance est retenue, indiquant la classe directe ou réverbérée associée à chacune des M composantes **C**(C_1, ..., C_i, ..., C_M).

[0118] De cette combinaison, il est donc déduit les N composantes correspondant aux N sources directes actives.

[0119] Le traitement décrit ici est effectué dans le domaine temporel, mais peut aussi être, dans une variante de réalisation, appliqué dans un domaine transformé.

[0120] Le procédé tel que décrit en référence à la figure 3 étant alors mis en oeuvre par sous-bandes de fréquence après passage dans le domaine transformé des signaux captés.

[0121] Par ailleurs, la bande passante utile peut être réduite en fonction des imperfections potentielles du système de captation, en hautes fréquences (présence de repliement spatial) ou en basses fréquences (impossibilité de retrouver les directivités théoriques de l'encodage microphonique).

[0122] La **figure 8** représente ici une forme de réalisation d'un dispositif (DIS) de traitement selon un mode de réalisation de l'invention.

[0123] Des capteurs Ca_1 à Ca_M représentés ici sous la forme d'un microphone sphérique MIC permettent d'acquérir,

dans un milieu réel, donc réverbérant, M signaux de mélange $\boldsymbol{x}$ ($\boldsymbol{x}_1$, ..., $\boldsymbol{x}_i$, ..., $\boldsymbol{x}_M$), à partir d'un signal multicanal.

[0124] Bien entendu, d'autres formes de microphones ou de capteurs peuvent être prévues. Ces capteurs peuvent être intégrés au dispositif DIS ou bien en dehors du dispositif, les signaux en résultant étant alors transmis au dispositif de traitement qui les reçoit via son interface d'entrée 840. Dans une variante, ces signaux peuvent simplement être obtenus préalablement et importés en mémoire du dispositif DIS.

[0125] Ces M signaux sont alors traités par un circuit de traitement et des moyens informatiques tels qu'un processeur PROC en 860 et une mémoire de travail MEM en 870. Cette mémoire peut comporter un programme informatique comportant les instructions de code pour la mise en oeuvre des étapes du procédé de traitement tel que décrit par exemple en référence à la figure 3 et notamment les étapes d'application d'un traitement de séparation de sources au signal multicanal capté et obtention d'un ensemble de M composantes sonores, avec M≥N, de calcul d'un ensemble de premiers descripteurs dit bi-variés, représentatifs de relations statistiques entre les composantes des couples de l'ensemble des M composantes obtenu et d'un ensemble de seconds descripteurs dit uni-variés représentatifs de caractéristiques d'encodage des composantes de l'ensemble des M composantes obtenu et de classification des composantes de l'ensemble des M composantes, selon deux classes de composantes, une première classe de N composantes dites directes correspondant aux N sources sonores directes et une deuxième classe de M-N composantes dites réverbérées, par un calcul de probabilité d'appartenance à une des deux classes, fonction des ensembles de premiers et seconds descripteurs.

[0126] Ainsi, le dispositif comporte un module 810 de traitement de séparation de sources appliqué au signal multicanal capté pour obtenir un ensemble de M composantes sonores $\boldsymbol{s}$ (s_1, ..., s_i, .. s_M), avec M≥N. Les M composantes sont fournies en entrée d'un calculateur 820 apte à calculer un ensemble de premiers descripteurs dit bi-variés, représentatifs de relations statistiques entre les composantes des couples de l'ensemble des M composantes obtenu et un ensemble de seconds descripteurs dit uni-variés représentatifs de caractéristiques d'encodage des composantes de l'ensemble des M composantes obtenu.

[0127] Ces descripteurs sont utilisés par un module de classification 830 ou classificateur, apte à classer des composantes de l'ensemble des M composantes, selon deux classes de composantes, une première classe de N composantes dites directes correspondant aux N sources sonores directes et une deuxième classe de M-N composantes dites réverbérées.

[0128] Pour cela, le module de classification comporte un module 831 de calcul de probabilité d'appartenance à une des deux classes des composantes de l'ensemble M, fonction des ensembles de premiers et seconds descripteurs.

[0129] Le classificateur utilise des descripteurs liés à la corrélation entre .les composantes pour déterminer lesquelles sont des signaux directs (c'est à dire des vraies sources) et lesquelles sont des résidus de réverbération. Il utilise également des descripteurs liés aux coefficients de mélange estimés par SAS, pour évaluer la conformité entre l'encodage théorique d'une source seule et l'encodage estimé de chaque composante. Certains des descripteurs sont donc fonction d'un couple de composantes (pour la corrélation), et d'autres sont fonctions d'une composante seule (pour la conformité de l'encodage microphonique estimé).

[0130] Un module 832 de calcul de vraisemblance permet de déterminer, dans un mode de réalisation, la combinaison le plus probable des classifications des M composantes par un calcul de valeurs de vraisemblance fonction des probabilités calculées au module 831 et pour les combinaisons possibles.

[0131] Enfin, le dispositif comporte une interface de sortie 850 pour délivrer l'information de classification des composantes, par exemple à un autre dispositif de traitement qui peut utiliser cette information pour rehausser le son des sources discriminés, pour les débruiter ou bien pour effectuer un mixage de plusieurs sources discriminées. Un autre traitement possible peut également être d'analyser ou de localiser les sources pour optimiser le traitement d'une commande vocale.

[0132] Bien d'autres applications utilisant l'information de classification ainsi déterminée, sont alors possibles.

[0133] Le dispositif DIS peut être intégré dans une antenne microphonique pour effectuer par exemple des captations de scènes sonores ou pour une prise de son de commande vocale. Le dispositif peut également être intégré dans un terminal de communication apte à. traiter des signaux captés par une pluralité de capteurs intégrés ou déportés du terminal.

Revendications

1. Procédé de traitement de données sonores pour une séparation de N sources sonores d'un signal sonore multicanal capté en milieu réel, le procédé comportant les étapes suivantes :

    - application (E310) d'un traitement de séparation de sources au signal multicanal capté et obtention d'une matrice de séparation et d'un ensemble de M composantes sonores, avec M≥N ;
    - calcul (E320) d'un ensemble de premiers descripteurs dit bi-variés, représentatifs d'une mesure de corrélation

entre les composantes des couples de l'ensemble des M composantes obtenu ;
- calcul (E320) d'un ensemble de seconds descripteurs dit uni-variés représentatifs de caractéristiques d'encodage des composantes de l'ensemble des M composantes obtenu, le calcul étant fonction d'une mise en correspondance entre les caractéristiques d'encodage estimées et issues d'une matrice inverse de la matrice de séparation et des caractéristiques d'encodage théoriques d'une source de type onde plane ;
- classification (E340) des composantes de l'ensemble des M composantes, selon deux classes de composantes, une première classe de N composantes dites directes correspondant aux N sources sonores directes et une deuxième classe de M-N composantes dites réverbérées, par un calcul (E330) de probabilité d'appartenance à une des deux classes, fonction des ensembles de premiers et seconds descripteurs.

2. Procédé selon la revendication 1, dans lequel le calcul d'un descripteur bi-varié comporte le calcul d'un score de cohérence entre deux composantes.

3. Procédé selon l'une des revendications 1 à 2, dans lequel le calcul d'un descripteur bi-varié comporte la détermination d'un retard entre les deux composantes du couple.

4. Procédé selon la revendication 3, dans lequel le retard entre deux composantes est déterminé par la prise en compte du retard maximisant une fonction d'inter-corrélation entre les deux composants du couple.

5. Procédé selon l'une des revendications 3 ou 4, dans lequel la détermination du retard entre deux composantes d'un couple est associée à un indicateur de fiabilité du signe du retard, fonction de la cohérence entre les composantes du couple.

6. Procédé selon l'une des revendications 3 ou 5, dans lequel la détermination du retard entre deux composantes d'un couple est associée à un indicateur de fiabilité du signe du retard, fonction du rapport du maximum d'une fonction d'inter-corrélation pour des retards de signe opposé.

7. Procédé selon l'une des revendications 1 à 6, dans lequel la classification des composantes de l'ensemble des M composantes s'effectue par la prise en compte de l'ensemble des M composantes, et par le calcul de la combinaison la plus probable des classifications des M composantes.

8. Procédé selon la revendication 7, dans lequel le calcul de la combinaison la plus probable s'effectue par la détermination d'un maximum des valeurs de vraisemblance exprimées comme le produit des probabilités conditionnelles associées aux descripteurs, pour les combinaisons possibles de classification des M composantes.

9. Procédé selon la revendication 7, dans lequel une étape de pré-selection des combinaisons possibles est effectuée en se basant sur les seuls descripteurs uni-variés avant l'étape de calcul de la combinaison la plus probable.

10. Procédé selon l'une des revendications précédentes, dans lequel une étape de pré-selection des composantes est effectuée en se basant sur les seuls descripteurs uni-variés avant l'étape de calcul des descripteurs bi-variés.

11. Procédé selon l'une des revendications précédentes, dans lequel le signal multicanal est un signal ambisonique.

12. Dispositif de traitement de données sonores mis en oeuvre pour effectuer un traitement de séparation de N sources sonores d'un signal sonore multicanal capté par une pluralité de capteurs en milieu réel, le dispositif comportant:

- une interface d'entrée pour recevoir les signaux captés par une pluralité de capteurs, du signal sonore multicanal;
- un circuit de traitement comportant un processeur et apte à controler:

◦ un module de traitement de séparation de sources appliqué au signal multicanal capté pour obtenir une matrice de séparation et un ensemble de M composantes sonores, avec $M \geq N$;
o un calculateur apte à calculer un ensemble de premiers descripteurs dit bi-variés, représentatifs d'une mesure de corrélation entre les composantes des couples de l'ensemble des M composantes obtenu et un ensemble de seconds descripteurs dit uni-variés représentatifs de caractéristiques d'encodage microphonique des composantes de l'ensemble des M composantes obtenu, le calcul étant fonction d'une mise en correspondance entre les caractéristiques d'encodage estimées et issues d'une matrice inverse de la matrice de séparation et des caractéristiques d'encodage théoriques d'une source de type onde plane ;

◦ un module de classification des composantes de l'ensemble des M composantes, selon deux classes de composantes, une première classe de N composantes dites directes correspondant aux N sources sonores directes et une deuxième classe de M-N composantes dites réverbérées, par un calcul de probabilité d'appartenance à une des deux classes, fonction des ensembles de premiers et seconds descripteurs ;

- une interface de sortie pour délivrer l'information de classification des composantes.

13. Programme informatique comportant des instructions de code pour la mise en oeuvre des étapes du procédé de traitement selon l'une des revendications 1 à 11, lorsque ces instructions sont exécutées par un processeur.

14. Support de stockage, lisible par un processeur, sur lequel est enregistré un programme informatique comprenant des instructions de code pour l'exécution des étapes du procédé de traitement selon l'un des revendications 1 à 11.

Patentansprüche

1. Verfahren zur Verarbeitung von Klangdaten zur Trennung von N Klangquellen eines Mehrkanalklangsignals, das in einer realen Umgebung erfasst wird, wobei das Verfahren die folgenden Schritte umfasst:

- Anwenden (E310) einer Verarbeitung zur Trennung von Quellen auf das erfasste Mehrkanalsignal und Erhalten einer Trennungsmatrix und eines Satzes von M Klanganteilen, mit $M \geq N$;
- Berechnen (E320) eines Satzes erster sogenannter bivariater Deskriptoren, die für ein Korrelationsmaß zwischen den Anteilen der Paare des erhaltenen Satzes der M Anteile repräsentativ sind;
- Berechnen (E320) eines Satzes zweiter sogenannter univariater Deskriptoren, die für Codierungseigenschaften der Anteile des erhaltenen Satzes der M Anteile repräsentativ sind, wobei das Berechnen von einem Abgleich zwischen den geschätzten und aus einer zu der Trennungsmatrix inversen Matrix stammenden Codierungseigenschaften und theoretischen Codierungseigenschaften einer Quelle vom Typ ebene Welle abhängt;
- Klassifizieren (E340) der Anteile des Satzes der M Anteile in zwei Klassen von Anteilen, nämlich eine erste Klasse mit N sogenannten direkten Anteilen, die den N direkten Klangquellen entsprechen, und eine zweite Klasse mit M-N sogenannten Nachhallanteilen, durch Berechnen (E330) einer Zugehörigkeitswahrscheinlichkeit zu einer der zwei Klassen, die von den Sätzen erster und zweiter Deskriptoren abhängt.

2. Verfahren nach Anspruch 1, wobei das Berechnen eines bivariaten Deskriptors das Berechnen eines Kohärenz-Scores zwischen zwei Anteilen umfasst.

3. Verfahren nach einem der Ansprüche 1 bis 2, wobei das Berechnen eines bivariaten Deskriptors das Bestimmen einer Verzögerung zwischen den zwei Anteilen des Paares umfasst.

4. Verfahren nach Anspruch 3, wobei die Verzögerung zwischen zwei Anteilen durch das Berücksichtigen der Verzögerung, die eine Kreuzkorrelationsfunktion zwischen den zwei Anteilen des Paares maximiert, bestimmt wird.

5. Verfahren nach einem der Ansprüche 3 oder 4, wobei das Bestimmen der Verzögerung zwischen zwei Anteilen eines Paares mit einem Indikator für die Zuverlässigkeit des Vorzeichens der Verzögerung assoziiert ist, der von der Kohärenz zwischen den Anteilen des Paares abhängt.

6. Verfahren nach einem der Ansprüche 3 oder 5, wobei das Bestimmen der Verzögerung zwischen zwei Anteilen eines Paares mit einem Indikator für die Zuverlässigkeit des Vorzeichens der Verzögerung assoziiert ist, der von dem Verhältnis des Maximums einer Kreuzkorrelationsfunktion für Verzögerungen mit umgekehrtem Vorzeichen abhängt.

7. Verfahren nach einem der Ansprüche 1 bis 6, wobei das Klassifizieren der Anteile des Satzes der M Anteile durch das Berücksichtigen des Satzes der M Anteile und durch das Berechnen der wahrscheinlichsten Kombination der Klassifizierungen der M Anteile erfolgt.

8. Verfahren nach Anspruch 7, wobei das Berechnen der wahrscheinlichsten Kombination durch das Bestimmen eines Maximums der Plausibilitätswerte, die als das Produkt aus den mit den Deskriptoren assoziierten bedingten Wahrscheinlichkeiten ausgedrückt werden, für die möglichen Klassifizierungskombinationen der M Anteile erfolgt.

9. Verfahren nach Anspruch 7, wobei vor dem Schritt des Berechnens der wahrscheinlichsten Kombination ein Schritt des Vorauswählens der möglichen Kombinationen nur auf Basis der univariaten Deskriptoren erfolgt.

10. Verfahren nach einem der vorhergehenden Ansprüche, wobei vor dem Schritt des Berechnens der bivariaten Deskriptoren ein Schritt des Vorauswählens der Anteile nur auf Basis der univariaten Deskriptoren erfolgt.

11. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Mehrkanalsignal ein Ambisonics-Signal ist.

12. Vorrichtung zur Verarbeitung von Klangdaten, die eingesetzt wird, um eine Verarbeitung zur Trennung von N Klangquellen eines Mehrkanalklangsignals, das von einer Vielzahl von Sensoren in einer realen Umgebung erfasst wird, durchzuführen, wobei die Vorrichtung Folgendes umfasst:

- eine Eingangsschnittstelle, um die durch eine Vielzahl von Sensoren erfassten Signale des Mehrkanalklangsignals zu empfangen;
- eine Verarbeitungsschaltung, die einen Prozessor umfasst und dazu fähig ist, Folgendes zu steuern:

  o ein Modul für eine Verarbeitung zur Trennung von Quellen, die auf das erfasste Mehrkanalsignal angewendet wird, um eine Trennungsmatrix und einen Satz von M Klanganteilen zu erhalten, mit $M \geq N$;
  o eine Recheneinheit, die dazu fähig ist, einen Satz erster sogenannter bivariater Deskriptoren, die für ein Korrelationsmaß zwischen den Anteilen der Paare des erhaltenen Satzes der M Anteile repräsentativ sind, und einen Satz zweiter sogenannter univariater Deskriptoren, die für Mikrofoncodierungseigenschaften der Anteile des erhaltenen Satzes der M Anteile repräsentativ sind, zu berechnen, wobei das Berechnen von einem Abgleich zwischen den geschätzten und aus einer zu der Trennungsmatrix inversen Matrix stammenden Codierungseigenschaften und theoretischen Codierungseigenschaften einer Quelle vom Typ ebene Welle abhängt;
  o ein Modul zum Klassifizieren der Anteile des Satzes der M Anteile in zwei Klassen von Anteilen, nämlich eine erste Klasse mit N sogenannten direkten Anteilen, die den N direkten Klangquellen entsprechen, und eine zweite Klasse mit M-N sogenannten Nachhallanteilen, durch Berechnen einer Zugehörigkeitswahrscheinlichkeit zu einer der zwei Klassen, die von den Sätzen erster und zweiter Deskriptoren abhängt;

- eine Ausgangsschnittstelle, um die Klassifizierungsinformation der Anteile zu übermitteln.

13. Computerprogramm, das Codeanweisungen zur Umsetzung der Schritte des Verarbeitungsverfahrens nach einem der Ansprüche 1 bis 11 umfasst, wenn diese Anweisungen von einem Prozessor ausgeführt werden.

14. Speicherungsmedium, das von einem Prozessor gelesen werden kann und auf dem ein Computerprogramm aufgezeichnet ist, das Codeanweisungen zur Ausführung der Schritte des Verarbeitungsverfahrens nach einem der Ansprüche 1 bis 11 beinhaltet.

Claims

1. Method for processing sound data in order to separate N sound sources of a multichannel sound signal captured in a real environment, the method comprising the following steps:

- applying (E310) source separation processing to the captured multichannel signal and obtaining a separation matrix and a set of M sound components, where $M \geq N$;
- calculating (E320) a set of what are called bivariate first descriptors, representative of a measure of correlation between the components of the pairs of the obtained set of M components;
- calculating (E320) a set of what are called univariate second descriptors, representative of encoding characteristics of the components of the obtained set of M components, the calculation being dependent on matching between the estimated encoding characteristics resulting from an inverse matrix of the separation matrix and theoretical encoding characteristics of a plane-wave source;
- classifying (E340) the components of the set of M components into two classes of components, a first class of N components called direct components corresponding to the N direct sound sources and a second class of M-N components called reverberant components, using a calculation (E330) of probability of belonging to one of the two classes, depending on the sets of first and second descriptors.

2. Method according to Claim 1, wherein calculating a bivariate descriptor comprises calculating a coherence score between two components.

3. Method according to either of Claims 1 and 2, wherein calculating a bivariate descriptor comprises determining a delay between the two components of the pair.

4. Method according to Claim 3, wherein the delay between two components is determined by taking into account the delay that maximizes an intercorrelation function between the two components of the pair.

5. Method according to either of Claims 3 and 4, wherein the determination of the delay between two components of a pair is associated with an indicator of reliability of the sign of the delay, which depends on the coherence between the components of the pair.

6. Method according to either of Claims 3 and 5, wherein the determination of the delay between two components of a pair is associated with an indicator of reliability of the sign of the delay, which depends on the ratio of the maximum of an intercorrelation function for delays of opposing sign.

7. Method according to one of Claims 1 to 6, wherein the components of the set of M components are classified by taking into account the set of M components and by calculating the most probable combination of the classifications of the M components.

8. Method according to Claim 7, wherein the most probable combination is calculated by determining a maximum of the likelihood values expressed as the product of the conditional probabilities associated with the descriptors, for the possible classification combinations of the M components.

9. Method according to Claim 7, wherein a step of preselecting the possible combinations is performed on the basis of just the univariate descriptors before the step of calculating the most probable combination.

10. Method according to one of the preceding claims, wherein a step of preselecting the components is performed on the basis of just the univariate descriptors before the step of calculating the bivariate descriptors.

11. Method according to one of the preceding claims, wherein the multichannel signal is an ambisonic signal.

12. Sound data processing device implemented so as to perform separation processing of N sound sources of a multichannel sound signal captured by a plurality of sensors in a real environment, the device comprising:

- an input interface for receiving the signals captured by a plurality of sensors, of the multichannel sound signal;
- a processing circuit containing a processor and able to control:

    o a source separation processing module applied to the captured multichannel signal in order to obtain a separation matrix and a set of M sound components, where $M \geq N$;
    o a calculator able to calculate a set of what are called bivariate first descriptors, representative of a measure of correlation between the components of the pairs of the obtained set of M components and a set of what are called univariate second descriptors representative of microphonic encoding characteristics of the components of the obtained set of M components, the calculation being dependent on matching between the estimated encoding characteristics resulting from an inverse matrix of the separation matrix and theoretical encoding characteristics of a plane-wave source;
    o a module for classifying the components of the set of M components into two classes of components, a first class of N components called direct components corresponding to the N direct sound sources and a second class of M-N components called reverberant components, using a calculation of probability of belonging to one of the two classes, depending on the sets of first and second descriptors;

- an output interface for delivering the classification information of the components.

13. Computer program containing code instructions for implementing the steps of the processing method according to one of Claims 1 to 11 when these instructions are executed by a processor.

14. Storage medium able to be read by a processor and on which there is recorded a computer program comprising

code instructions for executing the steps of the processing method according to one of Claims 1 to 11.

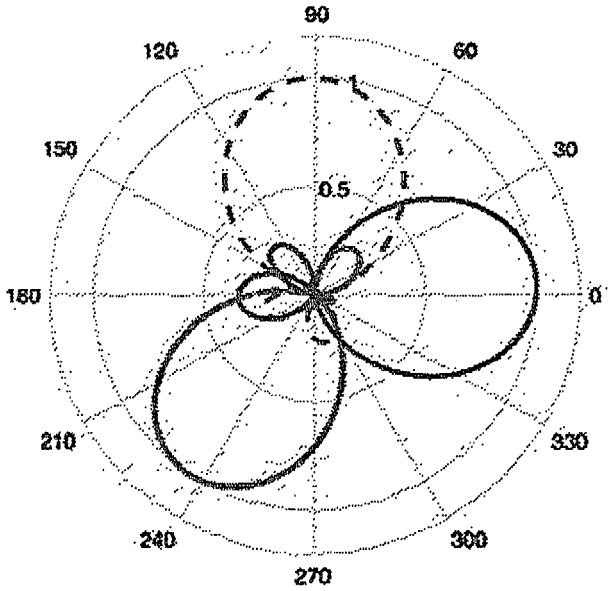
90
120
60
150
30
0.5
180
0
210
330
240
300
270

FIG. 1

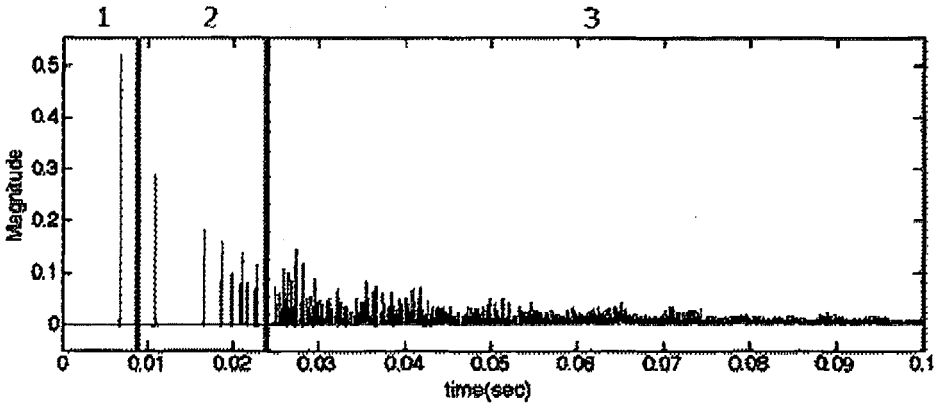
1
2
3
Magnitude
0.5
0.4
0.3
0.2
0.1
0
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
time(sec)

FIG. 2

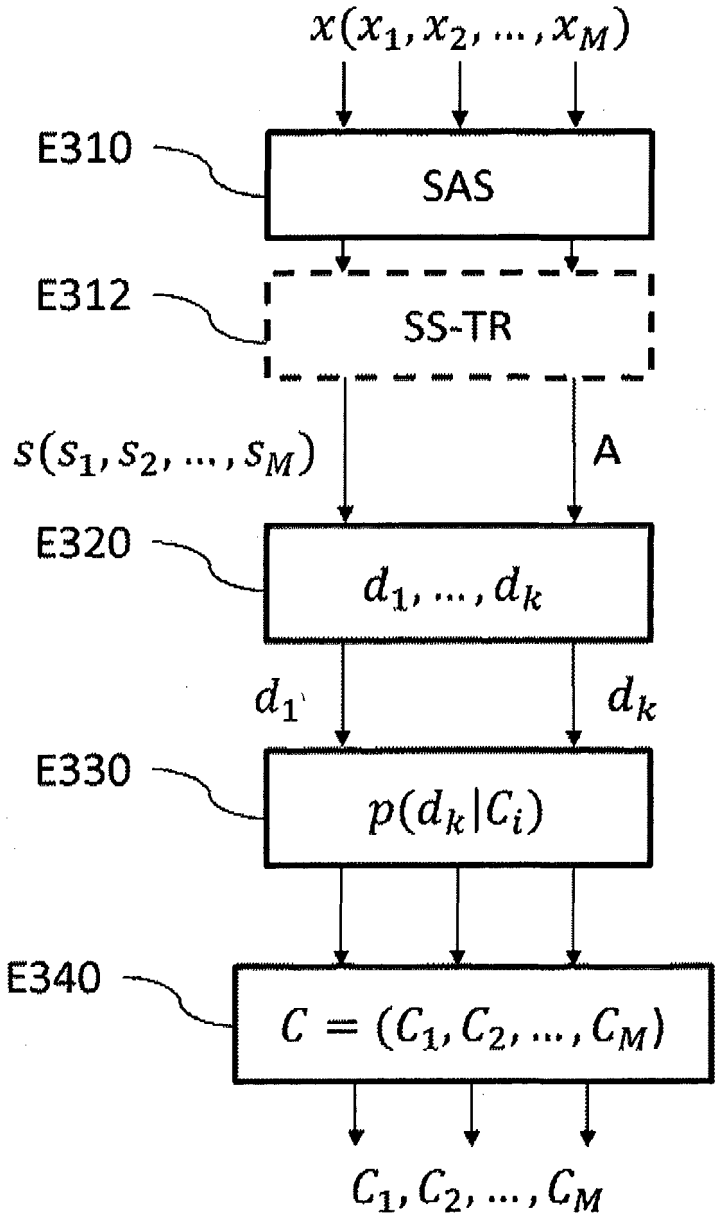
$x(x_1, x_2, \ldots, x_M)$
E310
SAS
E312
SS-TR
$s(s_1, s_2, \ldots, s_M)$
A
E320
$d_1, \ldots, d_k$
d_1
d_k
E330
$p(d_k|C_i)$
E340
$C = (C_1, C_2, \ldots, C_M)$
$C_1, C_2, \ldots, C_M$

FIG. 3

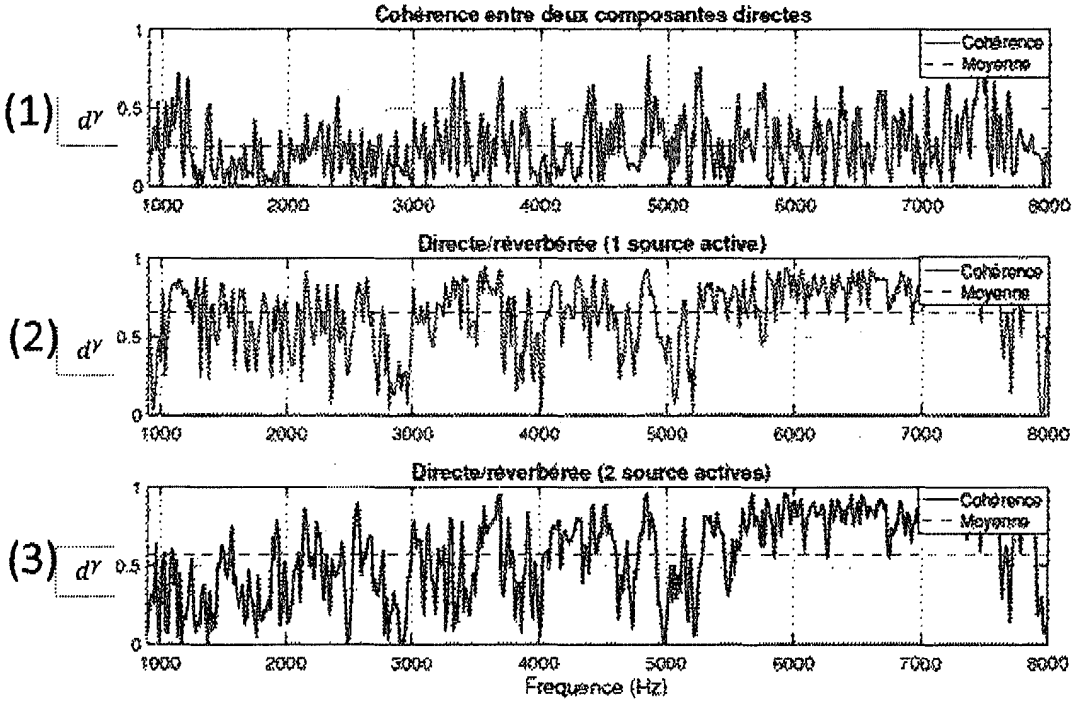
Cohérence entre deux composantes directes
Cohérence
Moyenne
(1)
d^γ
Directe/réverbérée (1 source active)
(2)
Directe/réverbérée (2 source actives)
(3)
Frequence (Hz)

FIG. 4

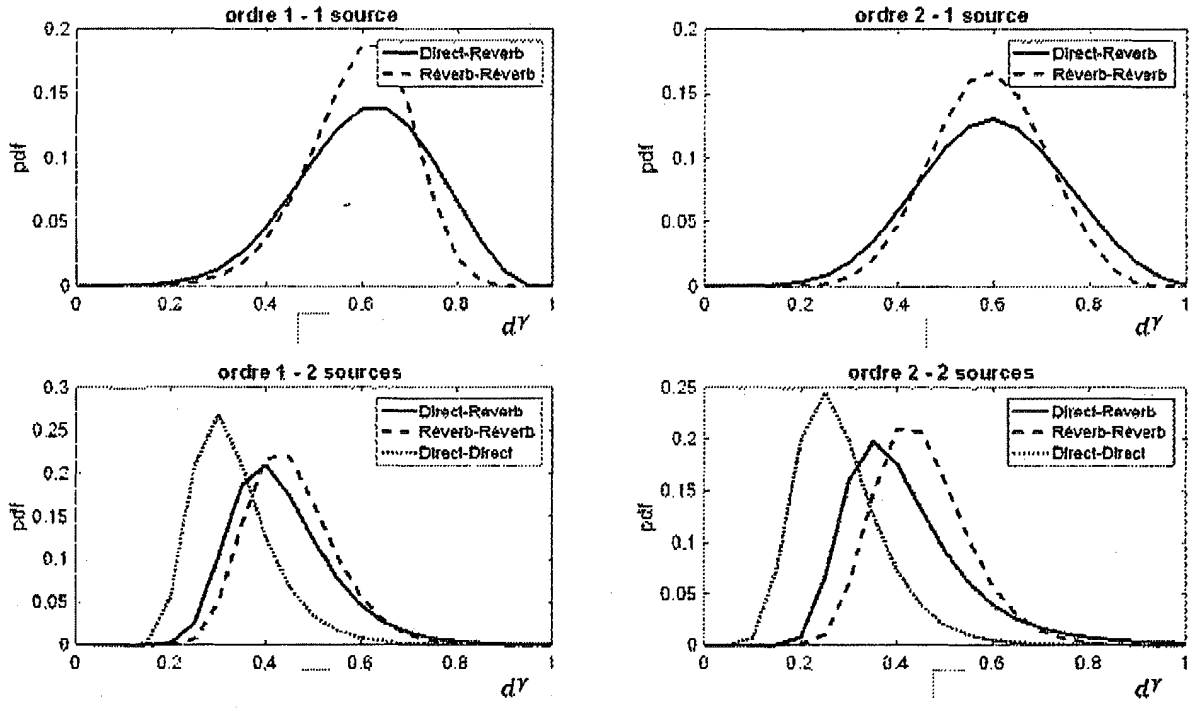
ordre 1 - 1 source
ordre 2 - 1 source
ordre 1 - 2 sources
ordre 2 - 2 sources
Direct-Reverb
Réverb-Réverb
Direct-Direct
pdf
d^γ

FIG. 5

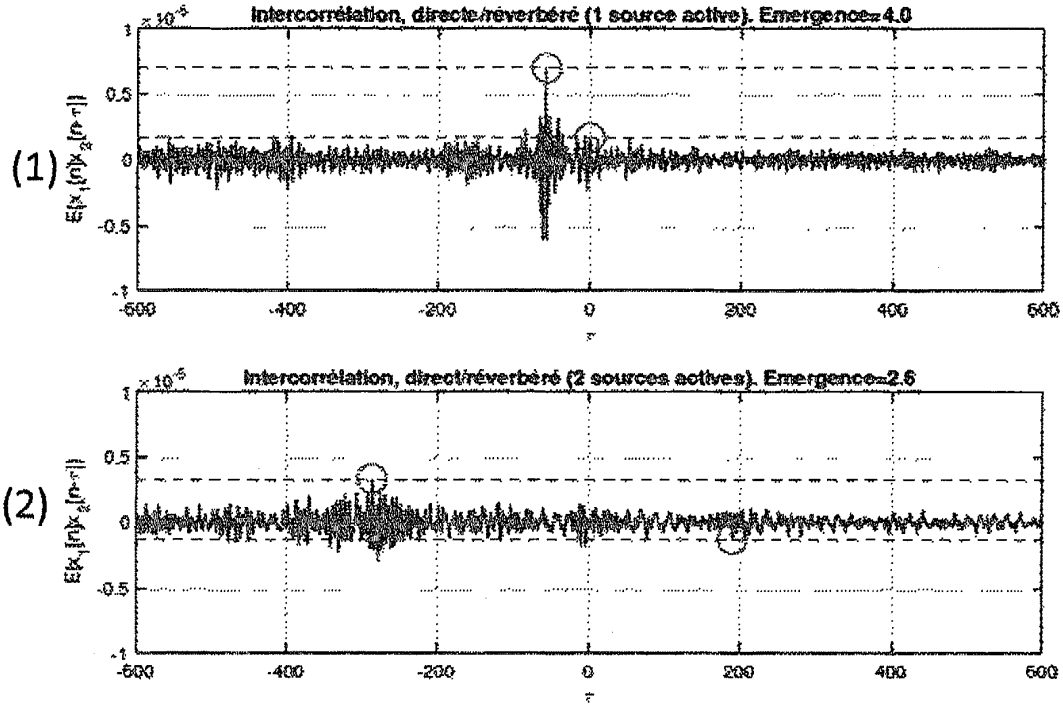
(1)
Intercorrélation, directe/réverbéré (1 source active). Emergence=4.0
(2)
Intercorrélation, direct/réverbéré (2 sources actives). Emergence=2.6

FIG. 6

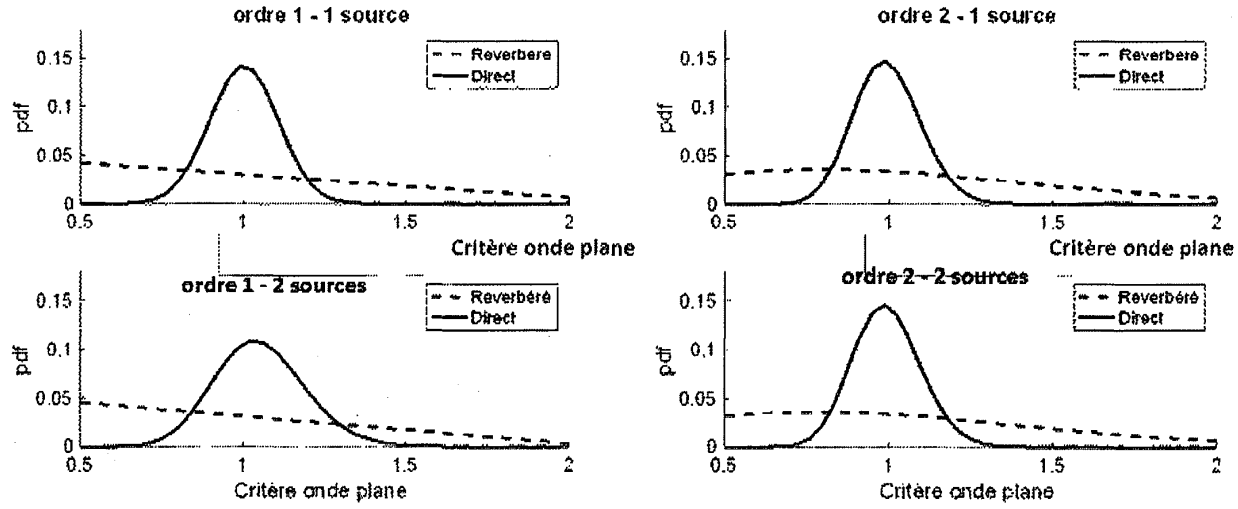
ordre 1 - 1 source
ordre 2 - 1 source
ordre 1 - 2 sources
ordre 2 - 2 sources
Reverbere
Direct
Reverbéré
pdf
Critère onde plane

FIG. 7

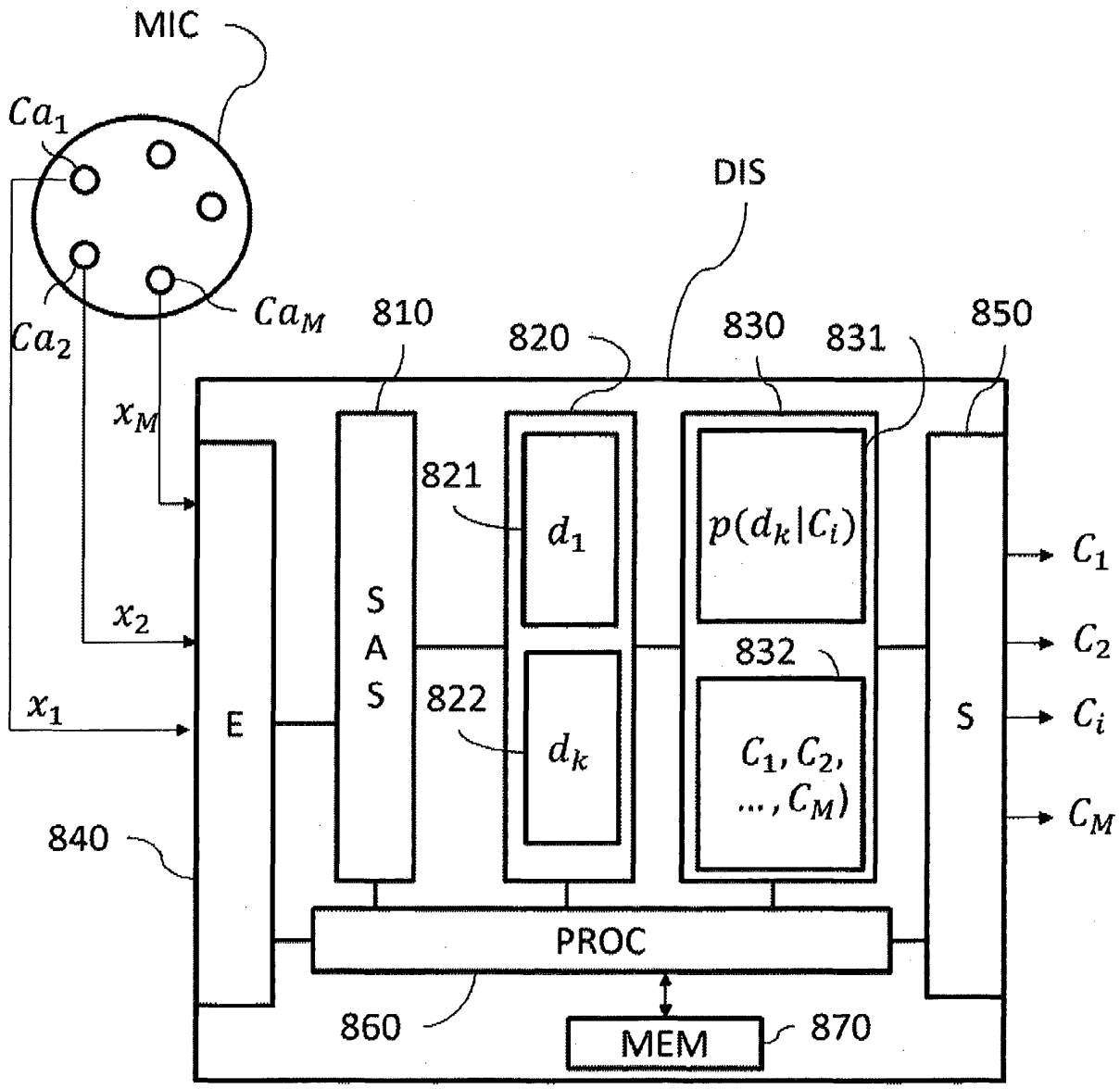
MIC
Ca_1
Ca_2
Ca_M
DIS
810
820
830
831
850
x_M
x_2
x_1
E
S
A
S
821
d_1
822
d_k
$p(d_k|C_i)$
832
$C_1, C_2, \ldots, C_M)$
S
C_1
C_2
C_i
C_M
840
PROC
860
MEM
870

FIG. 8

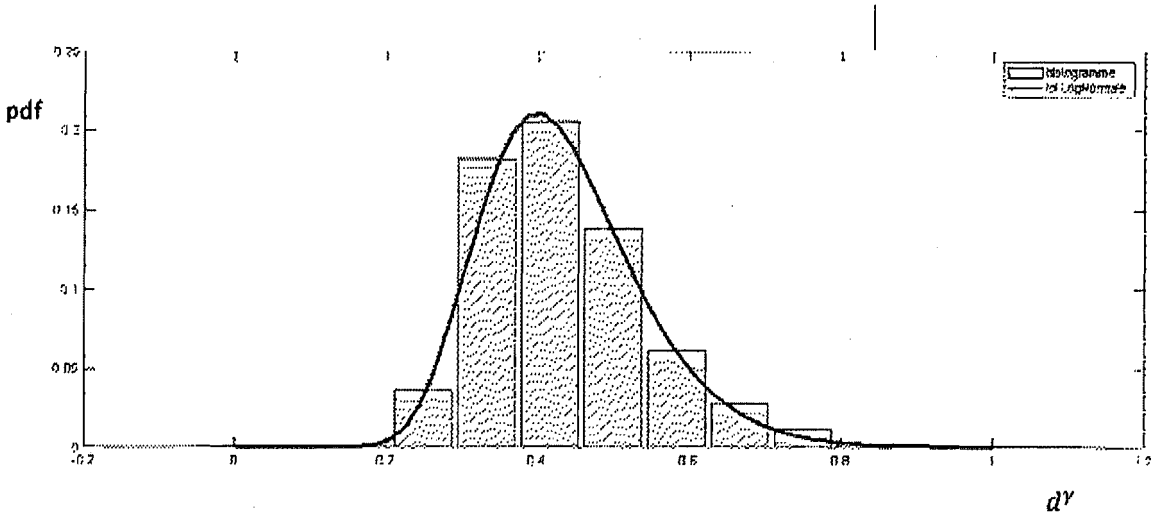

FIG. 9

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

- US 2005060142 A1 **[0010]**
- US 2010111290 A1 **[0010]**

Littérature non-brevet citée dans la description

- **J.F CARDOSO ; A. SOULOUMIAC.** Blind beamforming for non-gaussian signals. *IEE Proceedings F - Radar and Signal Processing,* Décembre 1993, vol. 140 (6 **[0009]**
- **D'AMARI.** A new learning algorithm for blind signal séparation, Advances. *neural information processing systems,* 1996 **[0009]**
- **BAQUÉ MATHIEU et al.** Séparation of Direct Sounds from Early Reflections Using the Entropy Rate Bound Minimization Algorithm. *60ème Conférence Internationale de l'AES: DREAMS (DEREVERBERATION AND REVERBERATION OF AUDIO, MUSIC, AND SPEECH),* 27 Janvier 2016 **[0010]**
- **A. JOURJINE ; S. RICKARD ; O. YILMAZ.** Blind séparation of disjoint orthogonal signals: Demixing n sources from 2 mixtures. *ICASSP'00,* 2000 **[0017]**