



(12) **DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:
14.12.2005 Bulletin 2005/50

(51) Int Cl.7: **G10L 21/02**

(21) Numéro de dépôt: **05291254.0**

(22) Date de dépôt: **10.06.2005**

(84) Etats contractants désignés:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IS IT LI LT LU MC NL PL PT RO SE SI SK TR
 Etats d'extension désignés:
AL BA HR LV MK YU

(72) Inventeur: **Benaroya, Laurent**
92200 Neuilly (FR)

(74) Mandataire: **Blot, Philippe Robert Emile et al**
Cabinet Lavoix
2, place d'Estienne d'Orves
75441 Paris Cedex 09 (FR)

(30) Priorité: **11.06.2004 FR 0406365**

(71) Demandeur: **Mist Technologies**
75015 Paris (FR)

(54) **Procédé de séparation de signaux sources à partir d'un signal issu du mélange**

(57) L'invention concerne un procédé de détermination des signaux de séparation respectivement relatifs à des sources sonores à partir d'un signal issu du mélange de ces signaux, lesdits signaux se présentant sous forme de trames successives, ledit procédé incluant une étape de détermination d'un signal d'estimée de chacune desdites sources. Le procédé inclut de plus, pour chacune desdites sources :

- une étape de prédiction (E40) d'un signal prédit pour la trame présente basée sur le signal de séparation pour la trame précédente,
- une étape de détermination du signal de séparation (E50) pour la trame présente sur la base dudit signal prédit et dudit signal d'estimée.

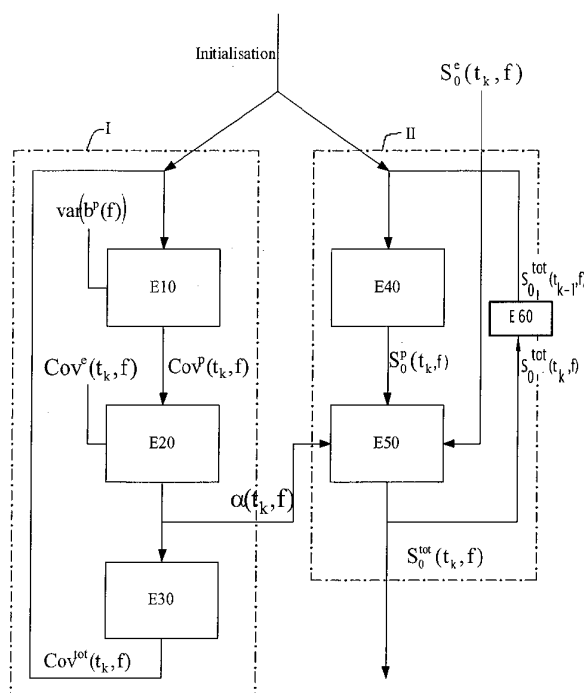


FIG.2

Description

[0001] La présente invention concerne un procédé de détermination des signaux de séparation respectivement relatifs à des sources sonores à partir d'un signal issu du mélange de ces signaux.

[0002] Le domaine de la présente invention est celui du traitement numérique de signaux relatifs à des sources sonores, dits aussi plus simplement signaux de son, audiophoniques ou audio. Dans ce domaine particulier, les traitements effectués sur les signaux de son ne sont pas dans le domaine temporel mais dans le domaine fréquentiel. Aussi, utilise-t-on fréquemment au préalable à tout traitement une transformée de Fourier à court terme qui est une transformée linéaire associant à un signal dans le domaine temporel échantillonné $\{x(t_1), \dots, x(t_N)\}$ un signal bidimensionnel temps fréquence noté ici $x(t_k, f)$, où t_k est un indice de trame du signal numérique échantillonné et f est un indice, généralement discret, de fréquence. Le signal $x(t_k, f)$ est donc un signal du domaine fréquentiel et il se présente sous forme de trames indicées en t_k .

[0003] Dans la présente description, toutes les grandeurs dont il s'agit sont décrites au moyen de variables aléatoires gaussiennes multidimensionnelles. Le mélange observé à l'instant t s'exprime sous la forme :

$$S_{\text{obs}}(t, f) = S(t, f) + b(t, f)$$

où $b(t)$ est un bruit blanc gaussien de variance σ_b^2 et $S(t, f)$ est le vecteur dont chaque composante est associée à une source :

$$S(t, f) = \begin{pmatrix} s_1(t, f) \\ \vdots \\ s_N(t, f) \end{pmatrix}$$

[0004] Pour chaque fréquence f et pour chaque source i , $s_i(t, f)$ suit une loi gaussienne centrée et de variance $\sigma_i^2(f)$

[0005] Pour désigner les variables sous forme de vecteur ou de matrice, des lettres en majuscule sont utilisées.

[0006] Par ailleurs encore, dans la présente demande la notion de signal se confond souvent à celle de la variable aléatoire qui la représente.

[0007] En ce qui concerne la séparation de signaux audio, on connaît un procédé qui est basé sur un filtre, dit filtre de Wiener, qui en définitive effectue une estimation du signal de séparation $\hat{S}_W(t, f)$ sous l'hypothèse de la stationnarité globale des signaux de mélange. Si l'on appelle $x(t_k, f)$ la variable aléatoire qui décrit le signal du domaine fréquentiel issu du mélange des signaux des sources, et qui est appliqué à l'entrée du filtre, l'espérance de la variable aléatoire décrivant le signal de sortie du filtre est conditionnée aux signaux $x(t_k, f)$. On peut donc écrire :

$$\hat{S}_W(t_k, f) = E[S(t_k, f) | x(t_k, f)]$$

[0008] Dans le cas du filtre de Wiener, chaque composante du vecteur $\hat{S}_W(t_k, f)$ peut être obtenue par la relation suivante :

$$\hat{S}_W(t, f) = \begin{pmatrix} \hat{s}_{W,1}(t, f) \\ \vdots \\ \hat{s}_{W,N}(t, f) \end{pmatrix} \text{ avec } \hat{s}_{W,i}(t, f) = \frac{e_i(f)}{\sum_j e_j(f) + \sigma_b^2} x(t, f)$$

où $e_i(f)$ est la fraction d'énergie de la source i contenue a priori dans le signal de mélange, à la fréquence d'indice f , N étant le nombre totale de sources et $x(t_k, f)$ étant le signal de mélange.

[0009] On considère, à titre illustratif uniquement, le cas particulier de deux sources délivrant des signaux respectivement notés dans le domaine temporel, $s_1(t)$ et $s_2(t)$. Au départ, l'on dispose d'un signal de son, noté dans le domaine temporel $x(t)$ représentatif du mélange de ces signaux de son :

$$x(t) = s_1(t) + s_2(t).$$

[0010] Dans une phase préalable d'apprentissage, on a évalué les deux sources sonores et on a plus exactement estimé leurs formes spectrales caractéristiques respectives $\sigma_1^2(f)$ et $\sigma_2^2(f)$, qui représentent, en définitive comme il est connu, leurs répartitions énergétiques en fonction de la fréquence. Si l'on considère que les signaux dans le domaine fréquentiel relatifs à ces deux sources $s_1(t, f)$ et $s_2(t, f)$ sont des variables aléatoires gaussiennes, non stationnaires, $\sigma_1^2(f)$ et $\sigma_2^2(f)$ représentent respectivement leur variance. Le filtre de Wiener délivre une estimation du signal de son de chaque source et, ce dans le domaine fréquentiel, en accord avec les relations suivantes :

$$\hat{s}_{W,1}(t, f) = \frac{\sigma_1^2(f)}{\sigma_1^2(f) + \sigma_2^2(f)} x(t, f)$$

$$s_{W,2}(t, f) = \frac{\sigma_2^2(f)}{\sigma_1^2(f) + \sigma_2^2(f)} x(t, f)$$

qui peuvent s'écrire sous forme matricielle de la manière suivante :

$$S(t_k, f) = P \cdot x(t_k, f)$$

[0011] Où P est une matrice qui décrit les coefficients de pondération et qui est donnée ci-dessous pour N sources :

$$P = \left[\frac{\sigma_1^2(f)}{\sum_{i=1}^N \sigma_i^2(f)}, \dots, \frac{\sigma_N^2(f)}{\sum_{i=1}^N \sigma_i^2(f)} \right]$$

[0012] Dans le cadre de la séparation de signaux de sons, le filtre de Wiener présente les principaux inconvénients suivants. Il opère de manière identique sur toutes les trames du signal de son de mélange et il ne tient donc pas des changements du contenu énergétique sonore d'une trame à l'autre. En définitive, il n'est pas un filtre adaptatif. Un autre inconvénient réside dans le fait qu'il ne prend en compte qu'une forme spectrale caractéristique par source sonore alors même que les sources sonores présentent une grande variété spectrale en terme de timbre, de hauteur, d'intensité, etc.

[0013] Des améliorations du filtre de Wiener ont été proposées pour tenir compte de ces inconvénients et ont abouti à notamment deux méthodes qui sont essentiellement basées sur l'utilisation de formes spectrales multiples pour décrire chacune des sources impliquées.

[0014] La première de ces méthodes a été introduite dans le cadre de la reconnaissance de parole et a été ensuite utilisée en audio. Selon cette méthode, le signal de son de chaque source $s_i(t)$ est caractérisé par un ensemble de K_i formes spectrales $\sigma_{k_i}^2(f)$, $k_i \in [1, \dots, K_i]$. Si l'on considère N sources, leur mélange est caractérisé par un ensemble de

$K_1 \times K_2 \times \dots \times K_N$ N-uplets de formes spectrales caractéristiques $(\sigma_{k_1}^2(f), \dots, \sigma_{k_N}^2(f))$. Pour chaque trame d'indice t_k , la méthode consiste à d'abord choisir le N-uplet de formes spectrales qui correspond le mieux au signal de son du mélange. Par exemple, elle peut consister à maximiser la probabilité de correspondance entre le spectrogramme du mélange $|x(t_k, f)|^2$ et la variance résultant du couple de formes spectrales. Ensuite, elle consiste à filtrer par un filtrage de Wiener classique le mélange en utilisant le N-uplet de formes spectrales ainsi sélectionné. On peut constater que cette méthode est adaptative puisque le choix des paramètres du filtre dépend de l'indice de trame t_k considéré.

[0015] Le principal inconvénient de cette méthode réside dans sa complexité algorithmique. En effet, si K formes spectrales caractéristiques par source i et N sources i sont considérées dans le mélange, K^N N-uplets de formes spectrales caractéristiques doivent être testés pour chaque trame si bien que la complexité est en $O(K^n \times T)$ si T est

le nombre de trames du signal mélange à analyser. Cet inconvénient de complexité peut rendre cette méthode rédhibitoire, notamment lorsque le nombre de formes spectrales caractéristiques par source est relativement important.

[0016] Une autre méthode a également été proposée pour rendre adaptatif le procédé de séparation. Comme précédemment, le signal de son de chaque source $s_i(t)$ est caractérisé par un ensemble de K_i formes spectrales caractéristiques $\sigma_{k_i}^2(f)$ mais qui sont là regroupées dans un dictionnaire de formes spectrales. Ainsi, le spectrogramme du mélange $|x(t_k, f)|^2$ est décomposé sur l'union des dictionnaires en présence et il est donc possible d'écrire :

$$|x(t_k, f)|^2 \approx \sum_{k_1=1}^{K_1} a_{k_1}(t_k) \sigma_{k_1}^2(f) + \dots + \sum_{k_N=1}^{K_N} a_{k_N}(t_k) \sigma_{k_N}^2(f)$$

où les coefficients $a_{k_i}(t)$, sont nommés "facteurs d'amplitude", sont les inconnues à résoudre.

[0017] On notera que l'équation ci-dessus peut s'interpréter comme s'il y avait $K_1 + \dots + K_N$ sources élémentaires stationnaires qui sont caractérisées chacune par une forme spectrale $\sigma_{k_i}^2(f)$ et qui se mélangent entre elles avec des facteurs d'amplitude respectifs $a_{k_i}(t)$ fonction du temps. On notera que chaque facteur d'amplitude $a_{k_i}(t)$ d'une source élémentaire est caractéristique de l'enveloppe de cette source. Il est donc un nombre positif.

[0018] L'équation ci-dessus peut se réécrire de la manière suivante :

$$|x(t_k, f)|^2 \approx \sum_{i=1}^N e_i(t_k, f) \text{ avec } e_i(t_k, f) = \sum_{k=1}^{K_i} a_k(t_k) \sigma_{k,i}^2(f)$$

$e_i(t_k, f)$ représente la fraction d'énergie de la source i contenue dans le mélange à analyser.

[0019] Une première méthode pour estimer les signaux de son des sources 1 à N est de mettre en oeuvre un filtrage de Wiener temps fréquence classique, néanmoins adaptatif dès lors qu'il dépend de l'indice de trame t . Ce filtre est appelé filtre de Wiener généralisé. On a donc pour la source i , l'estimée $\hat{s}_{i,W_g}(t_k, f)$:

$$\hat{s}_{i,W_g}(t_k, f) = \frac{e_i(t_k, f)}{\sum_{i=1}^N e_i(t_k, f)} x(t_k, f)$$

[0020] Une autre méthode, dite de resynthèse, considère l'amplitude du signal de son de chaque source i comme étant égale à $\sqrt{e_i(t_k, f)}$ et sa phase comme étant estimée par celle du mélange. Il est donc possible d'écrire pour la source i :

$$\tilde{s}_i(t_k, f) = \sqrt{e_i(t_k, f)} \cdot \text{sign}[\tilde{x}(t_k, f)]$$

où $\text{sign}[x] = \frac{x}{|x|}$ correspond à la phase de x .

[0021] Cette seconde méthode par l'utilisation de dictionnaire de formes spectrales caractéristique présente l'avantage par rapport à la précédente méthode de diminuer la complexité algorithmique. En effet, pour n sources possédant chacune K formes spectrales, la complexité algorithmique est en $O(n \times K \times T)$ où T est le nombre de trames à analyser, donc inférieure à celle de la méthode précédente qui était en $O(K^n \times T)$.

[0022] Les trois méthodes qui viennent d'être présentées présentent néanmoins l'inconvénient majeur que la phase de chacune des sources impliquées (ou des sources élémentaires impliquées selon la méthode utilisée) est rigoureusement égale à la phase du mélange. Or, en général, les sources qui s'additionnent n'ont pas toutes la même phase si bien que, dans les méthodes présentées ci-dessus, lors de la séparation, il y a destruction de la structure de phase des sources, ce qui peut entraîner des effets gênants pour l'écoute des signaux de son des sources recouvrées. Le système auditif humain est en effet très sensible aux cohérences de phase dans les signaux audio, notamment les

cohérences inter-trames pour f fixée (phase cohérente entre $s(t_{k+1}, f)$ et $s(t_k, f)$) et les cohérences de phase pour une même trame mais pour différentes valeurs de la fréquence f (phase de $s(t_k, f)$ pour différentes valeurs de f). Ces effets de cohérence de phase sont notamment très sensibles sur les sons harmoniques, comme les sons d'un instrument de musique, ou encore les sons voisés, alors qu'ils sont moins importants sur les bruits blancs, roses, etc. ou encore les sons d'instrument de percussion.

[0023] Le but de la présente est de proposer une méthode de séparation des signaux relatifs à des sources sonores à partir d'un signal issu de mélange de ces signaux qui ne présente pas les incohérences de phase des méthodes citées ci-dessus.

[0024] Pour ce faire, un procédé de détermination des signaux de séparation respectivement relatifs à des sources sonores à partir d'un signal issu du mélange de ces signaux, lesdits signaux se présentant sous forme de trames successives, ledit procédé incluant une étape de détermination d'un signal d'estimée de chacune desdites sources. Il est caractérisé en ce qu'il inclut de plus, pour chacune desdites sources :

- une étape de prédiction (E40) d'un signal prédit pour la trame présente basée sur le signal de séparation pour la trame précédente,
- une étape de détermination du signal de séparation (E50) pour la trame présente sur la base dudit signal prédit et dudit signal d'estimée.

[0025] Ce procédé s'applique également à des signaux non sonores tels que tous signaux numériques issus de l'échantillonnage d'un transducteur permettant la transformation d'une grandeur physique en un signal électrique.

[0026] A cet effet, l'invention a pour objet un procédé de détermination des signaux de séparation respectivement relatifs à des sources non sonores à partir d'un signal issu du mélange de ces signaux, lesdits signaux se présentant sous forme de trames successives, ledit procédé incluant une étape de détermination d'un signal d'estimée de chacune desdites sources, caractérisé en ce qu'il inclut de plus, pour chacune desdites sources :

- une étape de prédiction d'un signal prédit pour la trame présente basée sur le signal de séparation pour la trame précédente,
- une étape de détermination du signal de séparation pour la trame présente sur la base dudit signal prédit et dudit signal d'estimée.

[0027] Avantagusement, ladite étape de détermination du signal de séparation consiste à sommer de manière pondérée le signal d'estimée et le signal prédit, lesdits coefficient de pondération étant déterminés de manière à minimiser la covariance du signal de séparation.

[0028] Avantagusement, le signal d'estimée est pondéré par un premier coefficient matriciel alors que le signal prédit est pondéré par un second coefficient matriciel égal à la matrice unité moins le premier coefficient matriciel, ledit premier coefficient matriciel étant déterminé de manière à minimiser la covariance du signal de séparation.

[0029] Les caractéristiques de l'invention mentionnées ci-dessus, ainsi que d'autres, apparaîtront plus clairement à la lecture de la description suivante d'un exemple de réalisation, ladite description étant faite en relation avec les dessins joints, parmi lesquels :

La Fig. 1 est un schéma synoptique d'un système de séparation des signaux relatifs à des sources sonores à partir d'un signal issu de mélange de ces signaux selon la présente invention, et

La Fig. 2 est diagramme montrant les différentes étapes mises en oeuvre par un procédé de séparation de signaux selon la présente invention.

[0030] Dans la suite de la description, on considèrera des sources sonores qui sont en elles-mêmes élémentaires, c'est-à-dire qui sont caractérisées chacune par une forme spectrale caractéristique donnée. Mais, on considèrera également des sources sonores dont la caractéristique de forme spectrale est une caractéristique parmi plusieurs caractéristiques de forme spectrale possibles, par exemple appartenant à un dictionnaire de formes spectrales caractéristiques (voir le préambule de la présente description). Comme on le mentionnait dans le préambule de la description, on peut alors considérer une source sonore comme étant une combinaison pondérée d'une pluralité de sources sonores élémentaires dont chacune présente une caractéristique de forme spectrale donnée (par exemple issue d'un dictionnaire ou déterminée).

[0031] De manière à résoudre le problème des incohérences de phase des méthodes de l'état de la technique mentionnées en préambule de la description, la présente invention prévoit des moyens de lien entre trames adjacentes. En d'autres termes, chaque source sonore élémentaire est déterminée d'une manière récursive et itérative.

[0032] On a représenté à la Fig. 1, un système de séparation de signaux de sons issus de sources sonores selon un mode de réalisation de la présente invention qui comporte ces moyens de liens entre trames adjacentes. Ce système

est essentiellement constitué d'une unité d'estimation 10 qui, sur la base d'un signal de mélange du domaine fréquentiel notée $x(t_k, f)$ obtenu par exemple par une transformée de Fourier à court terme du signal $x(t)$ dans le domaine temporel échantillonné, délivre un signal d'estimée représentée par la variable aléatoire $S^e(t_k, f)$ dont chaque compo-

5 sante $s_i^e(t_k, f)$ est le signal d'estimée pour une source du mélange d'indice i . Si l'on dispose de N sources élémentaires, le signal de d'estimée est représenté par un vecteur dont chaque composante est relative à une source :

$$S^e(t_k, f) = \begin{pmatrix} s_1^e(t_k, f) \\ \vdots \\ s_N^e(t_k, f) \end{pmatrix}$$

[0033] L'unité d'estimation 10 est telle que l'espérance du signal en sa sortie est conditionnée aux signaux $x(t_k, f)$ qui sont réellement observés. On peut donc écrire :

$$S^e(t_k, f) = E[S(t_k, f) | x(t_k, f)]$$

[0034] L'unité d'estimation 10 est par exemple un filtre de Wiener (voir les différentes formes de ce type de filtre données dans le préambule de la présente description), une unité fonctionnant par une méthode de seuillage temps-fréquence, ou par une méthode dite Ephraïm et Malah, etc. Par exemple, dans le cas d'un filtre de Wiener, chaque composante du vecteur $S^e(t_k, f)$ peut être obtenu par la relation suivante :

$$S^e(t_k, f) = \begin{pmatrix} \hat{s}_{1, W_e}(t_k, f) \\ \vdots \\ \hat{s}_{N, W_e}(t_k, f) \end{pmatrix} \text{ avec } \hat{s}_{i, W_e}(t_k, f) = \frac{e_i(t_k, f)}{\sum_{i=1}^N e_i(t_k, f)} x(t_k, f)$$

où $e_i(t_k, f)$ est la fraction d'énergie de la source i contenue dans le signal de mélange, dans la trame d'indice t_k et de fréquence d'indice f , N étant le nombre totale de sources et $x(t_k, f)$ étant le signal de mélange.

[0035] On rappelle ici que pour une source élémentaire i , on peut écrire :

$$e_i(t_k, f) = \sum_{k_i=1}^{K_i} a_{k_i}(t_k) \sigma_{k_i}^2(f)$$

où K_i représente le nombre de sources élémentaires considérés pour la source i , $a_{k_i}(t_k)$ représente le facteur d'amplitude de la source élémentaire d'indice k_i et $\sigma_{k_i}^2(f)$ la variance de cette source élémentaire d'indice k_i .

[0036] Le système de séparation de signaux de sons de sources sonores représenté à la Fig. 1 comporte encore une unité de mise à jour 20 et une unité de prédiction 30. Ce sont ces unités 20 et 30 qui constituent les moyens de lien inter-trame qui sont mentionnés ci-dessus.

[0037] L'unité de prédiction 30 est prévue pour délivrer un signal de prédiction considéré comme une variable aléatoire correspondante $SP(t_k, f)$

[0038] On rappelle ici que si l'on dispose de N sources élémentaires, le signal de prédiction est un vecteur dont chaque composante est relative à une source :

$$S^p(t_k, f) = \begin{pmatrix} s_1^p(t_k, f) \\ \vdots \\ s_N^p(t_k, f) \end{pmatrix}$$

[0039] Comme on peut le constater sur la Fig. 1, l'unité de mise à jour 20, sur la base du signal de prédiction $S^p(t_k, f)$ délivré par l'unité de prédiction 30 et du signal d'estimée $S^e(t_k, f)$ délivré par l'unité d'estimation 10 délivre, quant à elle, le signal de séparation dont la variable aléatoire est notée $S^{tot}(t_k, f)$.

[0040] Si l'on dispose de N sources élémentaires, le signal de séparation est représenté par un vecteurs dont chaque composante est relative à une source :

$$S^{tot}(t_k, f) = \begin{pmatrix} s_1^{tot}(t_k, f) \\ \vdots \\ s_N^{tot}(t_k, f) \end{pmatrix}$$

[0041] Concernant l'unité de prédiction 30, dans le cas le plus simple elle peut revenir à introduire un terme de décalage entre deux trames successives, par son unité 32, et l'on peut donc écrire :

$$S^p(t_k, f) = H(f) \cdot S^{tot}(t_{k-1}, f)$$

[0042] Le signal prédit pour la trame présente est basé sur le signal de séparation pour la trame précédente.

[0043] L'espérance du signal de prédiction est donnée par la relation suivante :

$$\hat{S}^p(t_k, f) = H(f) \cdot \hat{S}^{tot}(t_{k-1}, f)$$

où $H(f)$ est un terme qui, dans le domaine fréquentiel, est représentatif du décalage entre deux trames successives et qui, du fait que les signaux considérés sont des signaux stationnaires, peut s'écrire :

$$H(f) = \exp\left[2\pi i \frac{fM}{T}\right]$$

où T est la longueur d'une trame, M le décalage considéré, et i le nombre complexe tel que $i^2 = -1$. Généralement, le décalage M entre trame est inférieur à la longueur T d'une trame et, même, il est souvent moitié de la longueur d'une trame :

$$M = T/2$$

[0044] Quant à l'unité de mise à jour 20, elle est prévue pour déterminer le signal de séparation $S^{tot}(t_k, f)$ en sommant de manière pondérée le signal d'estimée $S^e(t_k, f)$ et le signal prédit $S^p(t_k, f)$. Dans le mode de réalisation représenté, le signal d'estimée $S^e(t_k, f)$ est pondéré par un coefficient matriciel $\alpha(tk, f)$ alors que le signal prédit est pondéré par un coefficient $I - \alpha(tk, f)$, I étant la matrice unité.

[0045] Par exemple, ceci est réalisé en additionnant, dans un additionneur 21, au signal prédit $S^p(t_k, f)$ un signal d'erreur calculé comme la différence entre le signal prédit $S^p(t_k, f)$ et le signal d'estimée $S^e(t_k, f)$, ledit signal d'erreur étant pondéré par un coefficient $\alpha(tk, f)$, la pondération étant effectuée par une unité de pondération 23. On peut donc écrire la relation :

$$S^{tot}(t_k, f) = S^p(t_k, f) + \alpha(tk, f) \cdot (S^e(t_k, f) - S^p(t_k, f))$$

[0046] Le système de séparation représenté à la Fig. 1 est prévu pour déterminer la matrice de coefficients optimale $\alpha(t_k, f)$ permettant de minimiser la variance de l'estimation du signal de séparation $S^{\text{tot}}(t_k, f)$. On peut montrer que cette valeur optimale du facteur de pondération est donnée par la relation suivante de la covariance du signal prédit $\text{Cov}^P(t_k, f)$ et de la somme de la covariance du signal prédit $\text{Cov}^P(t_k, f)$ et de la covariance du signal d'estimée $\text{Cov}^e(t_k, f)$, soit :

$$\alpha(t_k, f) = [\text{Cov}^e(t_k, f) + \text{Cov}^P(t_k, f)]^{-1} \cdot \text{Cov}^P(t_k, f)$$

[0047] La valeur du coefficient de pondération $\alpha(t_k, f)$ étant connue, il est possible de déterminer l'espérance du signal de séparation $S_0^{\text{tot}}(t_k, f)$ qui constitue alors la sortie de l'unité de mise à jour 20 :

$$S_0^{\text{tot}}(t_k, f) = S_0^P(t_k, f) + \alpha(t_k, f) \cdot (S_0^e(t_k, f) - S_0^P(t_k, f))$$

[0048] On va donc procéder conformément au diagramme de la Fig. 2. Dans ce diagramme, on peut constater qu'il présente deux branches I et II : la première I regroupe les étapes E10, E20 et E30 et correspond aux calculs des covariances des différentes variables aléatoires aboutissant essentiellement au calcul de la matrice de coefficients optimale $\alpha(t_k, f)$ alors que la seconde II qui regroupe les étapes E40 et E50 correspond aux calculs des espérances de ces variables aléatoires aboutissant au calcul de l'espérance du signal de séparation en fonction du signal d'estimation délivré par l'unité d'estimation 10.

[0049] Plus précisément, à l'étape E10, est effectuée la mise à jour de la covariance du signal prédit représentée, on le rappelle, par la variable aléatoire $\text{SP}(t_{k+1}, f)$

[0050] Du fait de l'unité 32 qui lie entre elles deux trames successives, on peut montrer facilement que la covariance du signal prédit est donnée par la relation suivante :

$$\text{Cov}^P(t_k, f) = \text{Cov}^{\text{tot}}(t_{k-1}, f) + \text{var}(b^P(t_k, f))$$

avec

$$\text{var}(b^P(t_k, f))$$

variance du bruit de prédiction.

[0051] Le module de la fonction $H(f)$ est en effet égal à 1.

[0052] La variance du bruit de prédiction $\text{var}(b^P(t_k, f))$ dépend des sources ou sous-sources considérées et de la fréquence f . Elle ne dépend pas de la trame considérée, si bien qu'elle peut également s'écrire :

$$\text{var}(b^P(t_k, f)) = \text{var}(b^P(f))$$

[0053] Cette variance est avantageusement estimée dans une phase d'apprentissage. En définitive, on a :

$$\text{Cov}^P(t_k, f) = \text{Cov}^{\text{tot}}(t_{k-1}, f) + \text{var}(b^P(f))$$

[0054] $\text{Cov}^{\text{tot}}(t_{k-1}, f)$ est une grandeur qui a été calculée à l'itération précédente (voir étape E30 ci-dessous).

[0055] A l'étape E20, on détermine la matrice de coefficients $\alpha(t_k, f)$ optimale. Pour ce faire, on utilise l'expression ci-dessus :

$$\alpha(t_k, f) = [\text{Cov}^e(t_k, f) + \text{Cov}^P(t_k, f)]^{-1} \cdot \text{Cov}^P(t_k, f)$$

[0056] La covariance du signal de séparation prédit $Cov^p(t_k, f)$ est donnée par le calcul effectué à l'étape E10. Quant à la covariance du signal d'estimée $Cov^e(t_k, f)$, elle est déterminée par les formes spectrales caractéristiques $\sigma_{k_i}^2(f)$ et les facteurs d'amplitude $a_{k_i}(t_k)$ des sources ou sources élémentaires considérées.

[0057] On rappelle que l'équation du mélange est la suivante :

$$x(t, f) = \sum_i s_i(t, f) + b(t, f)$$

où $b(t, f)$ représente l'expression d'un bruit blanc gaussien stationnaire de variance σ_b^2 . Quant aux sources élémentaires $s_i(t, f)$, elles sont considérées *a priori* comme des sources gaussiennes non stationnaires de variance $a_i(t, f)\sigma_i(f)$ mais comme stationnaires conditionnellement à $a_i(t)$.

[0058] Le signal d'estimé $S^e(t, f)$ du mélange de l'ensemble des sources élémentaires est une variable aléatoire gaussienne de variance $Cov^e(t, f)$:

[0059] On a pu montrer que cette covariance du signal d'estimée $Cov^e(t_k, f)$ pouvait s'exprimer de la manière suivante :

$$Cov^e(t_k, f) = \begin{pmatrix} a_1(t_k)\sigma_1^2(f) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_N(t_k)\sigma_N^2(f) \end{pmatrix} - \frac{1}{\sum_{j=1}^N a_j(t_k)\sigma_j^2(f) + \sigma_b^2} \begin{pmatrix} a_1(t_k)\sigma_1^2(f) \\ \vdots \\ a_N(t_k)\sigma_N^2(f) \end{pmatrix} \begin{pmatrix} a_1(t_k)\sigma_1^2(f) & \cdots & a_N(t_k)\sigma_N^2(f) \end{pmatrix}$$

expression dans laquelle :

- $a_j(t_k, f)$ est le facteur d'amplitude de la source ou de la source élémentaire d'indice j , pour la trame d'indice t_k et pour la fréquence d'indice f ,
- $\sigma_j(f)$ est la forme spectrale caractéristique de la source ou de la source élémentaire d'indice j et pour la fréquence f ,
- σ_b est la variance d'un bruit blanc gaussien, et
- N est le nombre total de sources élémentaires considérées.

[0060] A l'étape E30, la matrice de covariance du signal de séparation est remise à jour en utilisant l'expression suivante :

$$Cov^{tot}(t_k, f) = [I - \alpha(t_k, f)]Cov^p(t_k, f)$$

expression dans laquelle :

- I est la matrice identité,
- $\alpha(t_k, f)$ est la matrice de coefficients telle que déterminée à l'étape E20 ci-dessus,
- $Cov^p(t_k, f)$ est la covariance du signal de séparation prédit telle que calculée à l'étape E10.

[0061] Après l'étape E30, pour ce qui concerne les calculs liés aux covariances, la trame suivante est considérée et le processus est repris à l'étape E10.

[0062] On considère maintenant les étapes E40 et E50 liées aux calculs des espérances. A l'étape E40, on détermine

l'espérance du signal prédit $S_0^p(t_k, f)$ laquelle est donnée par la relation suivante en fonction de l'espérance du signal de séparation $S_0^{tot}(t_{k-1}, f)$ déterminée à la trame précédente :

$$S_0^p(t_k, f) = H(f) \cdot S_0^{tot}(t_{k-1}, f)$$

[0063] A l'étape E50, l'espérance du signal de séparation est calculée au moyen de l'expression suivante :

$$S_0^{tot}(t_k, f) = S_0^p(t_k, f) + \alpha(t_k, f) \cdot (S_0^e(t_k, f) - S_0^p(t_k, f))$$

expression dans laquelle :

- $S_0^p(t_k, f)$ est l'espérance du signal de séparation prédit déterminé à l'étape E10 ci-dessus,
- $S_0^e(t_k, f)$ est l'espérance du signal d'estimée telle qu'il apparaît à la sortie du l'unité d'estimation 10, et
- $\alpha(t_k, f)$ est la matrice de coefficients telle que déterminée à l'étape E20 ci-dessus.

[0064] L'espérance du signal de séparation $S_0^{tot}(t_k, f)$ est le signal de sortie du système. Ses composantes sont les signaux de séparation de chacune des sources ou des sources élémentaires considérées.

[0065] A l'étape E60, l'espérance du signal de séparation de la trame Tr , $S_0^{tot}(t_k, f)$ est décalée d'une trame pour obtenir l'espérance du signal de séparation de la trame t_{k-1} et cette dernière espérance est utilisée au cours de l'étape E40.

[0066] Après les étapes E50 et E60, la trame suivante est considérée et le processus est repris à l'étape E40 pour ce qui concerne les étapes liées aux calculs des espérances.

[0067] Les étapes E10 et E40 sont mises en oeuvre par l'unité de prédiction 30 alors que les étapes E20, E30 et E50 sont mises en oeuvre par l'unité de mise à jour 20.

[0068] On notera qu'à l'initialisation du procédé, l'espérance et la covariance de la variable aléatoire représentant le signal de séparation sont mise à zéro puis les étapes E10 et E40 sont mises en oeuvre.

Revendications

1. Procédé de détermination des signaux de séparation respectivement relatifs à des sources sonores à partir d'un signal issu du mélange de ces signaux, lesdits signaux se présentant sous forme de trames successives, ledit procédé incluant une étape de détermination d'un signal d'estimée de chacune desdites sources, **caractérisé en ce qu'il** inclut de plus, pour chacune desdites sources :

- une étape de prédiction (E40) d'un signal prédit pour la trame présente basée sur le signal de séparation pour la trame précédente,
- une étape de détermination du signal de séparation (E50) pour la trame présente sur la base dudit signal prédit et dudit signal d'estimée.

2. Procédé de détermination des signaux de séparation respectivement relatifs à des sources non sonores à partir d'un signal issu du mélange de ces signaux, lesdits signaux se présentant sous forme de trames successives, ledit procédé incluant une étape de détermination d'un signal d'estimée de chacune desdites sources, **caractérisé en ce qu'il** inclut de plus, pour chacune desdites sources :

- une étape de prédiction (E40) d'un signal prédit pour la trame présente basée sur le signal de séparation pour la trame précédente,
- une étape de détermination du signal de séparation (E50) pour la trame présente sur la base dudit signal prédit et dudit signal d'estimée.

3. Procédé de séparation selon la revendication 1 et 2, **caractérisé en ce que** ladite étape de détermination du signal de séparation consiste à sommer de manière pondérée le signal d'estimée et le signal prédit, lesdits coefficient de pondération étant déterminés de manière à minimiser la covariance du signal de séparation.

4. Procédé de séparation selon la revendication 3, **caractérisé en ce que** le signal d'estimée est pondéré par un premier coefficient matriciel alors que le signal prédit est pondéré par un second coefficient matriciel égal à la matrice unité moins le premier coefficient matriciel, ledit premier coefficient matriciel étant déterminé de manière à minimiser la covariance du signal de séparation.

5. Procédé de séparation selon la revendication 4, **caractérisé en ce que** la valeur dudit premier coefficient matriciel est calculée au moyen de la relation suivante de la covariance du signal prédit $\text{Cov}^p(t_k, f)$ et de la somme de la covariance du signal prédit $\text{Cov}^p(t_k, f)$ et de la covariance du signal d'estimée $\text{Cov}^e(t_k, f)$, soit :

$$\alpha(t_k, f) = [\text{Cov}^e(t_k, f) + \text{Cov}^p(t_k, f)]^{-1} \cdot \text{Cov}^p(t_k, f)$$

6. Procédé de séparation selon la revendication 5, **caractérisé en ce que** la covariance du signal prédit $\text{Cov}^p(t_k, f)$ est déterminée en fonction de la covariance du signal de séparation $\text{Cov}^{\text{tot}}(t_{k-1}, f)$ pour la trame précédente au moyen de la relation suivante :

$$\text{Cov}^p(t_k, f) = \text{Cov}^{\text{tot}}(t_{k-1}, f) + \text{var}(b^p(t_k, f))$$

$\text{var}(b^p(t_k, f))$ étant la variance du bruit de prédiction qui dépend des sources ou sous-sources considérées.

7. Procédé de séparation selon la revendication 6, **caractérisé en ce que** ladite variance du bruit de prédiction $\text{var}(b^p(t_k, f))$ est estimée dans une phase d'apprentissage.

8. Procédé de séparation selon une des revendications 5 à 7, **caractérisé en ce que** ladite covariance du signal d'estimée $\text{Cov}^e(t_k, f)$ est déterminée au moyen de la relation suivante :

$$\text{Cov}^e(t_k, f) = \begin{pmatrix} a_1(t_k) \sigma_1^2(f) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_N(t_k) \sigma_N^2(f) \end{pmatrix} - \frac{1}{\sum_{j=1}^N a_j(t_k) \sigma_j^2(f) + \sigma_b^2} \begin{pmatrix} a_1(t_k) \sigma_1^2(f) \\ \vdots \\ a_N(t_k) \sigma_N^2(f) \end{pmatrix} \begin{pmatrix} a_1(t_k) \sigma_1^2(f) & \dots & a_N(t_k) \sigma_N^2(f) \end{pmatrix}$$

expression dans laquelle :

- $a_j(t_k, f)$ est le facteur d'amplitude de la source ou de la source élémentaire d'indice j , pour la trame d'indice t_k et pour la fréquence d'indice f ,
- $\sigma_j(f)$ est la forme spectrale caractéristique de la source ou de la source élémentaire d'indice j et pour la fréquence f ,
- σ_b est la variance d'un bruit blanc gaussien, et
- N est le nombre total de sources ou de sources élémentaires considérées.

9. Procédé de séparation selon une des revendications 5 à 8, **caractérisé en ce que** la matrice de covariance du signal de séparation est remise à jour en utilisant l'expression suivante :

$$\text{Cov}^{\text{tot}}(t_k, f) = [I - \alpha(t_k, f)] \text{Cov}^p(t_k, f)$$

5 Expression dans laquelle :

- I est la matrice identité ;
- $\alpha(t_k, f)$ est la matrice du premier coefficient de pondération ; et
- $\text{Cov}^p(t_k, f)$ est la covariance du signal prédit.

10

10. Procédé de séparation selon une des revendications 1 à 9, **caractérisé en ce qu'il** comporte une étape de détermination du signal d'estimée $S^e(t_k, f)$, chaque composante $\hat{s}_i^e(t_k, f)$ correspondant à l'estimation d'une source élémentaire i dudit signal d'estimée $S^e(t_k, f)$ étant obtenue à partir des formules suivantes :

15

$$\hat{s}_i^e(t_k, f) = \frac{e_i(t_k, f)}{\sum_{j=1}^N e_j(t_k, f)} x(t_k, f)$$

20

$$e_i(t_k, f) = \sum_{k_i=1}^{K_i} a_{k_i}(t_k) \sigma_{k_i}^2(f)$$

25

dans lesquelles :

30

- $e_i(t_k, f)$ étant la fraction d'énergie de la source i contenue dans le signal issu du mélange des signaux, dans une trame d'indice t_k et de fréquence d'indice f, N étant le nombre totale de sources ;
- $x(t_k, f)$ étant le signal issu du mélange des signaux ;
- K_i étant le nombre de sources élémentaires considérés pour la source i ;
- $a_{k_i}(t_k)$ étant le facteur d'amplitude de la source élémentaire d'indice k_i ; et
- $\sigma_{k_i}^2(f)$ étant la variance de cette source élémentaire d'indice k_i .

35

40

45

50

55

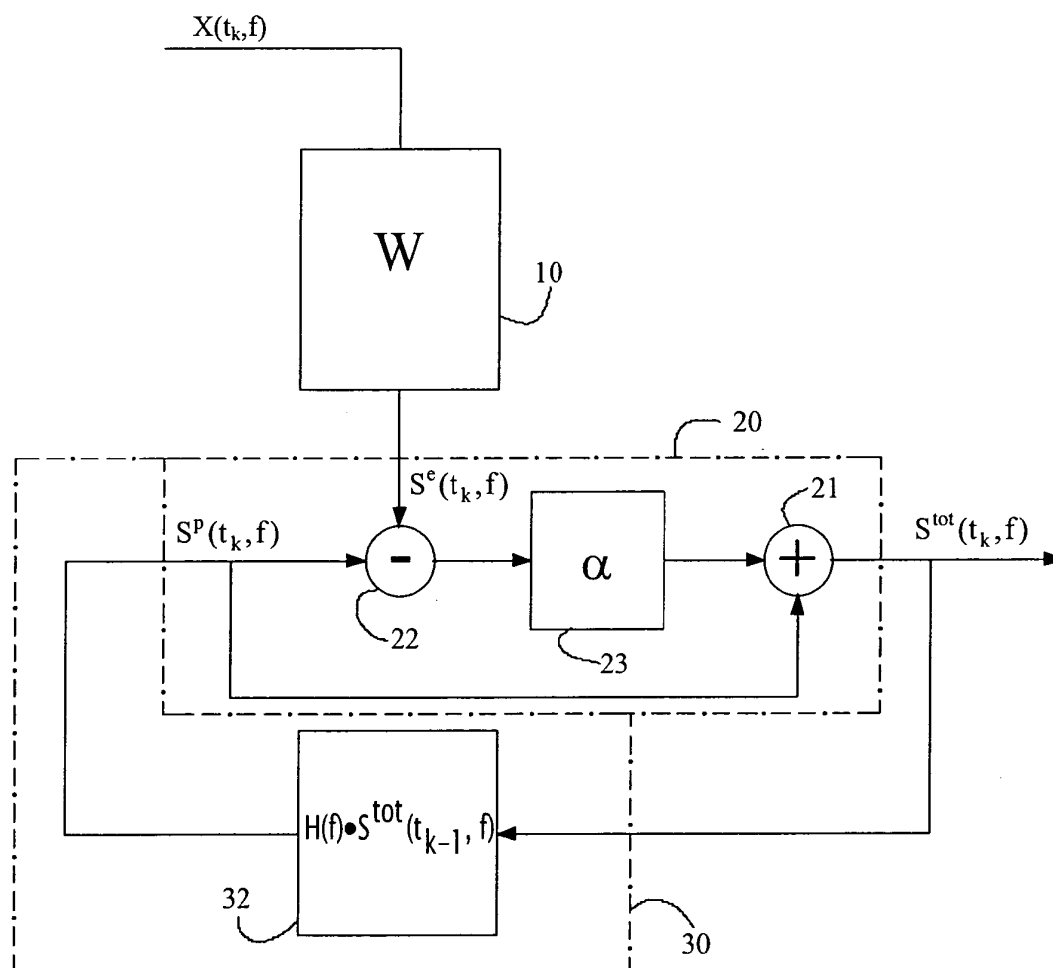
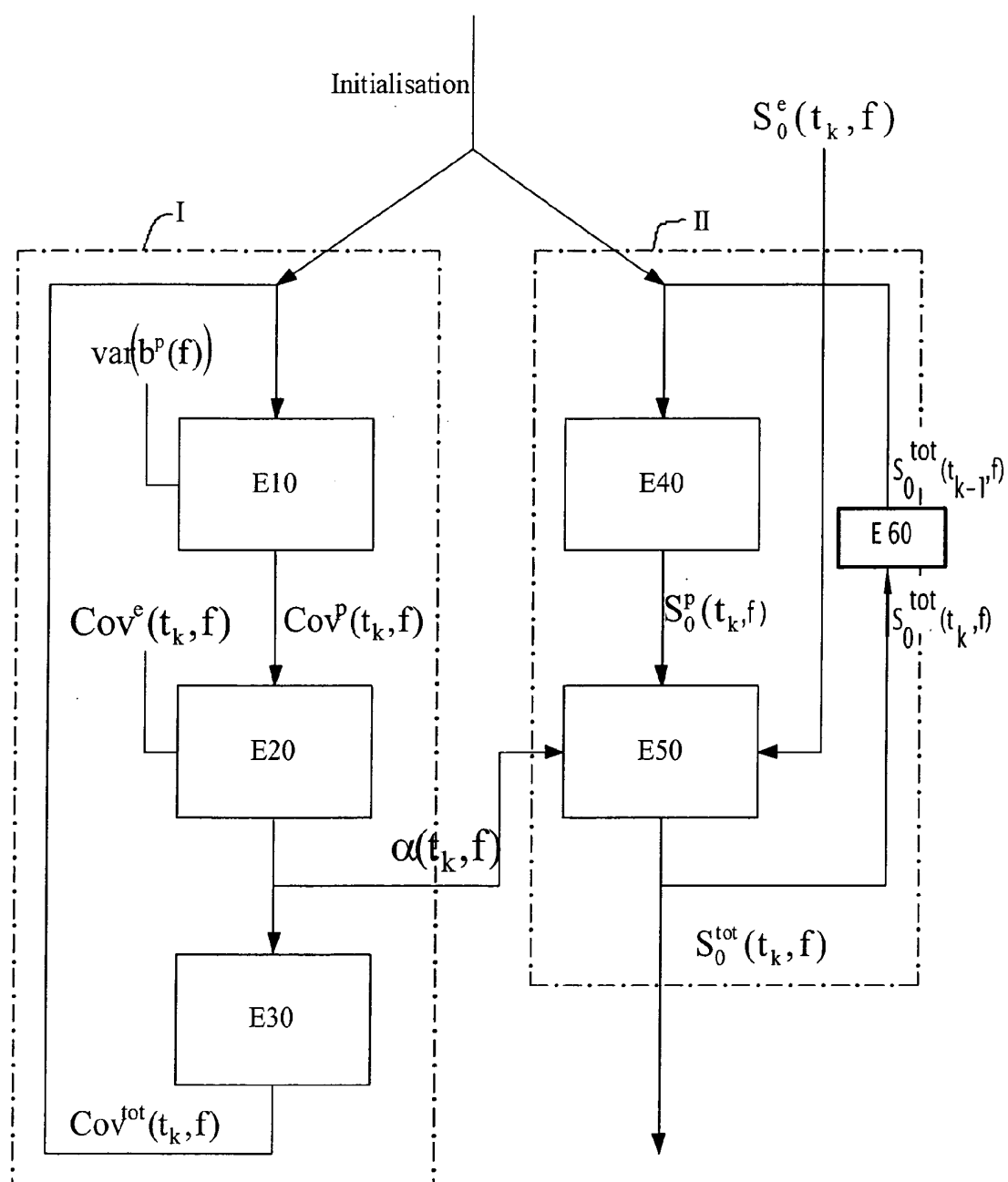


FIG.1

**FIG.2**



Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande
EP 05 29 1254

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (Int.Cl.7)
X	ELIE LAURENT BENNAROYA: "Séparation de plusieurs sources sonores avec un seul microphone" 26 juin 2003 (2003-06-26), UNIVERSITE DE RENNES 1, RENNES 2874, XP002346340 * le document en entier *	1,3-9	G10L21/02
X	STONE J V: "Blind source separation using temporal predictability" NEURAL COMPUTATION MIT PRESS USA, vol. 13, no. 7, 2001, pages 1559-1574, XP002303769 ISSN: 0899-7667 * abrégé *	1,3,4	
X	MANDIC D P ET AL: "An on-line algorithm for blind source extraction based on nonlinear prediction approach" NEURAL NETWORKS FOR SIGNAL PROCESSING, 2003. NNSP'03. 2003 IEEE 13TH WORKSHOP ON TOULOUSE, FRANCE SEPT. 17-19, 2003, PISCATAWAY, NJ, USA, IEEE, 17 septembre 2003 (2003-09-17), pages 429-438, XP010712478 ISBN: 0-7803-8177-7 * abrégé *	1	
			DOMAINES TECHNIQUES RECHERCHES (Int.Cl.7)
			G10L
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche La Haye		Date d'achèvement de la recherche 23 septembre 2005	Examineur Quélavoine, R
CATEGORIE DES DOCUMENTS CITES X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

EPO FORM 1503 03.82 (P04C02)



Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande
EP 05 29 1254

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (Int.Cl.7)
A	<p>BENAROYA L ET AL: "Non negative sparse representation for wiener based source separation with a single sensor"</p> <p>2003 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). HONG KONG, APRIL 6 - 10, 2003, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), NEW YORK, NY : IEEE, US,</p> <p>vol. VOL. 1 OF 6,</p> <p>6 avril 2003 (2003-04-06), pages VI613-VI616, XP010640826</p> <p>ISBN: 0-7803-7663-3</p> <p>* le document en entier *</p> <p>-----</p>	1-9	
			DOMAINES TECHNIQUES RECHERCHES (Int.Cl.7)
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche		Date d'achèvement de la recherche	Examineur
La Haye		23 septembre 2005	Quélavoine, R
<p>CATEGORIE DES DOCUMENTS CITES</p> <p>X : particulièrement pertinent à lui seul</p> <p>Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie</p> <p>A : arrière-plan technologique</p> <p>O : divulgation non-écrite</p> <p>P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention</p> <p>E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date</p> <p>D : cité dans la demande</p> <p>L : cité pour d'autres raisons</p> <p>& : membre de la même famille, document correspondant</p>			

2
EPO FORM 1503 03.82 (F04C02)