

(19)



Europäisches
Patentamt
European
Patent Office
Office européen
des brevets



(11)

EP 2 309 499 A1

(12)

DEMANDE DE BREVET EUROPEEN

(43) Date de publication:

13.04.2011 Bulletin 2011/15

(51) Int Cl.:

G10L 21/02 (2006.01)**H04R 3/00** (2006.01)**H04B 1/12** (2006.01)(21) Numéro de dépôt: **10167065.1**(22) Date de dépôt: **23.06.2010**

(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 SE SI SK SM TR**

Etats d'extension désignés:

BA ME RS(30) Priorité: **22.09.2009 FR 0956506**(71) Demandeur: **Parrot****75010 Paris (FR)**

(72) Inventeurs:

- **Vitte, Guillaume**
75003 Paris (FR)

- **Seris, Julie**
75013 Paris (FR)

- **Pinot, Guillaume**
75004 Paris (FR)

(74) Mandataire: **Dupuis-Latour, Dominique****SEP Bardehle Pagenberg Dost Altenburg****Geissler****10, boulevard Haussmann****75009 Paris (FR)**

(54) **Procédé de filtrage optimisé des bruits non stationnaires captés par un dispositif audio multi-microphone, notamment un dispositif téléphonique "mains libres" pour véhicule automobile**

(57) Le procédé comporte dans le domaine fréquentiel les étapes suivantes :

- calcul (18) d'une première référence de bruit par analyse de cohérence spatiale des signaux captés ;
- calcul (20) d'une seconde référence de bruit par analyse des directions d'incidence des signaux captés ;
- estimation (24) d'une direction principale d'incidence ($\theta(k,l)$) des signaux captés ;
- sélection (22) comme signal de bruit référent ($Ref(k,l)$) de l'une ou l'autre des références de bruit en fonction

de la direction principale estimée ;

e) combinaison (28) des signaux captés en un signal combiné bruité ($X(k,l)$);f) calcul (26) d'une probabilité d'absence de parole ($q(k,l)$) dans le signal combiné bruité, à partir des niveaux respectifs d'énergie spectrale du signal combiné bruité ($X(k,l)$) et du signal de bruit référent ($Ref(k,l)$) ;

g) réduction sélective du bruit (34) par application d'un gain variable propre à chaque bande de fréquences et à chaque trame temporelle.

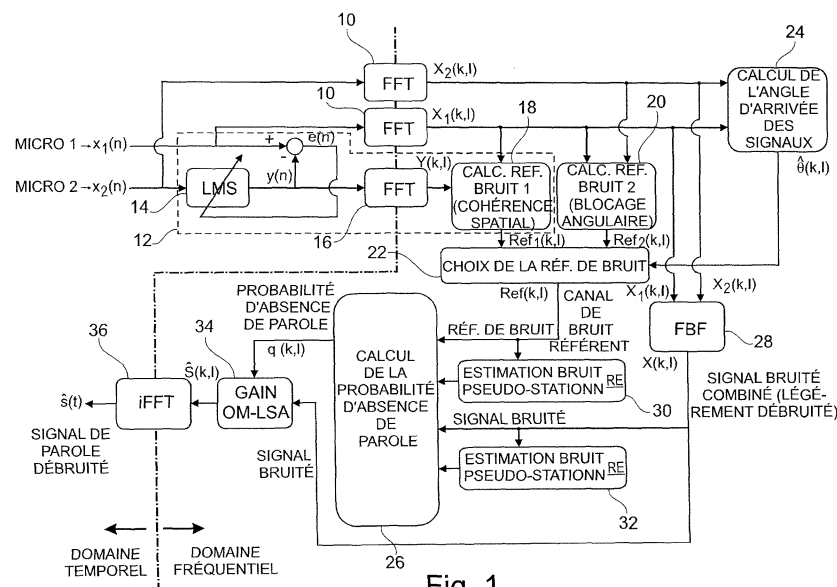


Fig. 1

Description

[0001] L'invention concerne le traitement de la parole en milieu bruité.

[0002] Elle concerne notamment, mais de façon non limitative, le traitement des signaux de parole captés par des dispositifs de téléphonie pour véhicules automobiles.

[0003] Ces appareils comportent un microphone ("micro") sensible captant non seulement la voix de l'utilisateur, mais également le bruit environnant, bruit qui constitue un élément perturbateur pouvant aller, dans certains cas, jusqu'à rendre incompréhensibles les paroles du locuteur. Il en est de même si l'on veut mettre en oeuvre des techniques de reconnaissance vocale, car il est très difficile d'opérer une reconnaissance de forme sur des mots noyés dans un niveau de bruit élevé.

[0004] Cette difficulté liée aux bruits environnants est particulièrement contraignante dans le cas des dispositifs "mains-libres". En particulier, la distance importante entre le micro et le locuteur entraîne un niveau relatif de bruit élevé qui rend difficile l'extraction du signal utile noyé dans le bruit. De plus, le milieu très bruité typique de l'environnement automobile présente des caractéristiques spectrales non stationnaires, c'est-à-dire qui évoluent de manière imprévisible en fonction des conditions de conduite : passage sur des chaussées déformées ou pavées, autoradio en fonctionnement, etc.

[0005] Certains de ces dispositifs prévoient l'utilisation de plusieurs micros, généralement deux micros, et utilisent la moyenne des signaux captés, ou d'autres opérations plus complexes, pour obtenir un signal avec un niveau de perturbations moindre. En particulier, une technique dite *beamforming* permet de créer par des moyens logiciels une directivité qui améliore le rapport signal/bruit, mais les performances de cette technique sont très limitées lorsque seulement deux micros sont utilisés (concrètement, on estime qu'une telle méthode ne fournit de bons résultats qu'à condition de disposer d'un réseau d'au moins huit micros).

[0006] Par ailleurs, les techniques classiques sont surtout adaptées au filtrage des bruits diffus, stationnaires, provenant des alentours du dispositif et se retrouvant à des niveaux comparables dans les signaux captés par les deux micros.

[0007] En revanche, un bruit non stationnaire ou "transient", c'est-à-dire un bruit évoluant de manière imprévisible en fonction du temps, ne sera pas discriminé de la parole et ne sera donc pas atténué.

[0008] Or, dans un environnement automobile ces bruits non stationnaires et directifs sont très fréquents : coup de klaxon, passage d'un scooter, dépassement par une voiture, etc.

[0009] Une difficulté du filtrage de ces bruits non stationnaires tient au fait que leurs caractéristiques temporelles et spatiales sont très proches de celles de la parole, d'où la difficulté d'une part, d'estimer la présence d'une parole (car le locuteur ne parle pas tout le temps) et d'autre part d'extraire le signal utile de parole dans un environnement très bruité tel qu'un habitacle de véhicule automobile.

[0010] L'un des buts de la présente invention est de proposer un dispositif mains-libres multi-microphone, notamment un système qui mette en oeuvre seulement deux microphones, permettant :

- de distinguer de façon efficace les bruits non stationnaires de la parole ; et
- d'adapter le débruitage à la présence et aux caractéristiques des bruits non stationnaires détectés, sans altérer la parole éventuellement présente, afin de traiter le signal bruité de la manière la plus efficace.

[0011] Le point de départ de l'invention consiste à associer (i) une analyse de cohérence spatiale du signal capté par les deux micros, à (ii) une analyse de la direction d'incidence de ces signaux. L'invention repose en effet sur deux constatations, à savoir que :

- la parole présente généralement une cohérence spatiale supérieure au bruit ; et par ailleurs que
- la direction d'incidence de la parole est généralement bien définie, et peut être supposée connue (dans le cas d'un véhicule automobile, elle est définie par la position du conducteur, vers lequel est tourné le micro).

[0012] Ces deux propriétés seront utilisées pour calculer deux références de bruit selon des méthodes différentes :

- une première référence de bruit calculée en fonction de la cohérence spatiale des signaux captés - une telle référence sera intéressante dans la mesure où elle intègre les bruits non stationnaires peu directifs (accrocs dans le ronronnement du moteur, etc.) ; et
- une seconde référence de bruit calculée en fonction de la direction principale d'incidence des signaux - cette caractéristique est en effet déterminable lorsque l'on utilise un réseau de plusieurs micros (au moins deux), conduisant à une référence de bruit intégrant surtout les bruits non stationnaires directifs (coups de klaxon, passage d'un scooter, dépassement par une voiture, etc.).

[0013] Ces deux références de bruit seront utilisées en alternance selon la nature du bruit présent, en fonction de la

direction d'incidence des signaux :

- de manière générale, la première référence de bruit (celle calculée par cohérence spatiale) sera utilisée par défaut ;
- en revanche, lorsque la direction principale d'incidence du signal sera éloignée de celle du signal utile (la direction du locuteur, supposée connue *a priori*) - c'est-à-dire en présence d'un bruit directif assez puissant - la seconde référence de bruit sera utilisée de façon à introduire majoritairement dans cette dernière les bruits non stationnaires directifs et puissants.

[0014] Une fois la référence de bruit ainsi sélectionnée, cette référence sera utilisée pour, d'une part, calculer une probabilité d'absence/présence de parole et pour, d'autre part, débruiter le signal capté par les micros.

[0015] Plus précisément, l'invention vise, de façon générale, un procédé de débruitage d'un signal acoustique bruité capté par deux microphones d'un dispositif audio multi-microphone opérant dans un milieu bruité, notamment un dispositif téléphonique "mains libres" pour véhicule automobile. Le signal acoustique bruité comprend une composante utile de parole issue d'une source de parole directive et une composante parasite de bruit, cette composante de bruit incluant elle-même une composante de bruit latéral non stationnaire directif.

[0016] Un tel procédé est par exemple divulgué par I. Cohen et B. Berdugo, Two-Channel Signal Detection and Speech Enhancement Based on the Transient Beam-to-Reference Ratio, Proc. ICASSP 2003, Hong-Kong, pp. 233-236, Apr. 2003.

[0017] De façon caractéristique de l'invention, ce procédé comporte, dans le domaine fréquentiel pour une pluralité de bandes de fréquences définies pour des trames temporelles successives de signal, les étapes de traitement du signal suivantes :

- a) calcul d'une première référence de bruit par analyse de cohérence spatiale des signaux captés les deux microphones, ce calcul comprenant un filtrage linéaire prédictif appliqué aux signaux captés par les deux microphones et comprenant une soustraction avec compensation du déphasage entre le signal capté et le signal de sortie du filtre prédictif ;
- b) calcul d'une seconde référence de bruit par analyse des directions d'incidence des signaux captés par les deux microphones, ce calcul comprenant le blocage spatial des composantes des signaux captés dont la direction d'incidence est située à l'intérieur d'un cône de référence défini de part et d'autre d'une direction prédéterminée d'incidence du signal utile ;
- c) estimation d'une direction principale d'incidence des signaux captés par les deux microphones ;
- d) sélection comme signal de bruit référent de l'une ou l'autre des références de bruit calculées aux étapes a) et b), en fonction de la direction principale estimée à l'étape c) ;
- e) combinaison des signaux captés par les deux microphones en un signal combiné bruité ;
- f) calcul d'une probabilité d'absence de parole dans le signal combiné bruité, à partir des niveaux respectifs d'énergie spectrale du signal combiné bruité et du signal de bruit référent ;
- g) à partir de la probabilité d'absence de parole calculée à l'étape f) et du signal combiné bruité, réduction sélective du bruit par application d'un gain variable propre à chaque bande de fréquences et à chaque trame temporelle.

[0018] Selon diverses caractéristiques subsidiaires avantageuses :

- le filtrage linéaire prédictif comprend l'application d'un algorithme de prédiction linéaire de type moindres carrés moyens LMS ;
- l'estimation de la direction principale d'incidence de l'étape c) comprend les sous-étapes successives suivantes : c1) partition de l'espace en une pluralité de secteurs angulaires ; c2) pour chaque secteur, évaluation d'un estimateur de direction d'incidence à partir des signaux captés par les deux microphones ; et c3) à partir des valeurs d'estimateurs calculées à l'étape c2), estimation de ladite direction principale d'incidence ;
- la sélection de l'étape d) est une sélection de la seconde référence de bruit comme signal de bruit référent si la direction principale estimée à l'étape c) est située hors d'un cône de référence défini de part et d'autre d'une direction prédéterminée d'incidence du signal utile ;
- la combinaison de l'étape e) comprend un préfiltrage de type *fixed beamforming* ;
- le calcul de probabilité d'absence de parole de l'étape f) comprend l'estimation de composantes de bruit pseudo-stationnaire respectives contenues dans le signal combiné bruité et dans le signal de bruit référent, la probabilité d'absence de parole étant calculée à partir également de ces composantes de bruit pseudo-stationnaire respectives ;
- la réduction sélective du bruit de l'étape g) est un traitement par application d'un gain à amplitude log-spectrale modifié optimisé OM-LSA.

[0019] On va maintenant décrire un exemple de mise en oeuvre du procédé de l'invention en référence à la figure

annexée.

[0020] La Figure 1 est un schéma par blocs montrant les différents modules et fonctions mis en oeuvre par le procédé de l'invention ainsi que leurs interactions.

[0021] Le procédé de l'invention est mis en oeuvre par des moyens logiciels, qu'il est possible de décomposer et schématiser par un certain nombre de blocs 10 à 36 illustrés Figure 1.

[0022] Ces traitements sont mis en oeuvre sous forme d'algorithmes appropriés exécutés par un microcontrôleur ou un processeur numérique de signal. Bien que, pour la clarté de l'exposé, ces divers traitements soient présentés sous forme de modules distincts, ils mettent en oeuvre des éléments communs et correspondent en pratique à une pluralité de fonctions globalement exécutées par un même logiciel.

[0023] Le signal que l'on souhaite débruiter est issu d'une pluralité de signaux captés par un réseau de micros (qui, dans la configuration minimale, peut être simplement un réseau de deux micros, comme dans l'exemple illustré) disposés selon une configuration prédéterminée. En pratique, ces deux micros peuvent par exemple être installés sur le plafonnier d'un habitacle de voiture, à environ 5 cm l'un de l'autre ; et avoir le lobe principal de leur diagramme de directivité orienté vers le conducteur. Cette direction, considérée comme *a priori* connue, sera désignée direction d'incidence du signal utile.

[0024] On appellera "bruit latéral" un bruit non stationnaire directif dont la direction d'incidence est éloignée de celle du signal utile, et on appellera "cône privilégié" la direction ou secteur angulaire de l'espace où se trouve la source de signal utile (la parole du locuteur) par rapport au réseau de micros. Lorsqu'une source sonore se manifestera en dehors du cône privilégié, il s'agira donc d'un bruit latéral, que l'on cherchera à atténuer. Comme illustré sur la Figure 1, les signaux bruités captés par les deux micros $x_1(n)$ et $x_2(n)$ font l'objet d'une transposition dans le domaine fréquentiel (blocs 10) par un calcul de transformée de Fourier à court terme (FFT) dont le résultat est noté respectivement $X_1(k, l)$ et $X_2(k, l)$, k étant l'indice de la bande de fréquence et l étant l'indice de la trame temporelle. Les signaux issus des deux micros sont également appliqués à un module 12 mettant en oeuvre un algorithme LMS prédictif schématisé par le bloc 14 et donnant, après calcul d'une transformée de Fourier à court terme (bloc 16) un signal $Y(k, l)$ qui servira au calcul d'une première référence de bruit $Ref_1(k, l)$ exécuté par un bloc 18, essentiellement sur un critère de cohérence spatiale.

[0025] Une autre référence de bruit $Ref_2(k, l)$ est calculée par un bloc 20, essentiellement sur un critère de blocage angulaire), à partir des signaux $X_1(k, l)$ et $X_2(k, l)$ directement obtenus, dans le domaine fréquentiel, à partir des signaux $x_1(n)$ et $x_2(n)$.

[0026] Un bloc 22 opère la sélection de l'une ou l'autre des références de bruit $Ref_1(k, l)$ ou $Ref_2(k, l)$ en fonction du résultat d'un calcul de l'angle d'incidence des signaux opéré par le bloc 24 à partir des signaux $X_1(k, l)$ et $X_2(k, l)$. La référence de bruit choisie, $Ref(k, l)$, est utilisée comme canal de bruit référent d'un bloc 26 de calcul d'une probabilité d'absence de parole opérée sur un signal bruité $X(k, l)$ résultant d'une combinaison, opérée par le bloc 28, des deux signaux $X_1(k, l)$ et $X_2(k, l)$. Le bloc 26 prend également en compte les composantes de bruit pseudo-stationnaire respectives du canal de bruit référent et du signal bruité, composantes estimées par les blocs 30 et 32.

[0027] Le résultat $q(k, l)$ du calcul de probabilité d'absence de parole et le signal bruité $X(k, l)$ sont appliqués en entrée d'un algorithme de contrôle de gain OM-LSA (bloc 34) dont le résultat $\hat{S}(k, l)$ est soumis (bloc 36) à une transformation de Fourier inverse (IFFT) pour obtenir dans le domaine temporel une estimée $\hat{s}(t)$ du signal de parole débruité.

[0028] On va maintenant décrire en détail chacune des étapes du traitement.

Transformée de Fourier des signaux captés par les micros (blocs 10)

[0029] Le signal dans le domaine temporel $x_n(t)$ issu de chacun des N micros ($N=1,2$ dans l'exemple illustré) est numérisé, découpé en trames de T points temporels, fenêtré temporellement par une fenêtre de type Hanning, puis la transformée de Fourier rapide FFT (transformée à court terme) $X_n(k, l)$ est calculée pour chacun de ces signaux :

$$X_n(k, l) = a_n \cdot d_n(k) \times S(k, l) + V_n(k, l)$$

avec :

$$d_n(k) = e^{-i2\pi f_k \tau_n}$$

l étant l'indice de la trame temporelle,

k étant l'indice de la bande de fréquences, et

f_k étant la fréquence centrale de la bande de fréquences indicée par k ,

$S(k,l)$ désignant la source de signal utile,
 α_n et τ_n désignant l'atténuation et le délai subis par le signal utile capté au niveau du micro n , et
 $V_n(k,l)$ désignant le bruit capté par le micro n .

5 *Calcul d'une première référence de bruit par cohérence spatiale (bloc 12)*

[0030] L'idée fondamentale sur laquelle repose l'invention est que, dans un environnement de télécommunications, la parole est un signal émis par une source bien localisée, relativement proche des micros et presque entièrement captée en chemin direct. À l'inverse, les bruits stationnaires et non stationnaires, qui proviennent surtout des alentours de l'utilisateur, peuvent être associés à des sources éloignées, en grand nombre et possédant une corrélation statistique inférieure à la parole entre les deux micros.

[0031] Dans un environnement de télécommunications, la parole est donc plus cohérente spatialement que le bruit.

[0032] Partant de ce principe, il est possible d'exploiter la propriété de cohérence spatiale pour construire un canal de bruit de référence plus riche et plus adapté qu'avec un *beamformer*. Le système prévoit à cet effet d'utiliser un filtre prédictif 14 de type LMS (*Least Mean Squares*, moindres carrés moyens) ayant pour entrées les signaux $x_1(n)$ et $x_2(n)$ captés par le couple de micros. On notera $y(n)$ la sortie du LMS et $e(n)$ l'erreur de prédiction.

[0033] Ce filtre prédictif est utilisé pour prédire à partir de $x_2(n)$ la composante parole qui se trouve dans $x_1(n)$. En effet, étant plus cohérente spatialement, la parole sera mieux prédite par le filtre adaptatif que le bruit.

[0034] Une première possibilité consiste à prendre pour le canal de bruit référent la transformée de Fourier de l'erreur de prédiction :

$$E(k,l) = X_1(k,l) - Y(k,l)$$

$E(k,l)$, $X_1(k,l)$ et $Y(k,l)$ étant les transformées de Fourier à court terme (TFCT) respectives de $e(k,l)$, $x_1(k,l)$ et $y(k,l)$.

[0035] On constate cependant en pratique un certain déphasage entre $X_1(k,l)$ et $Y(k,l)$ dû à une convergence imparfaite de l'algorithme LMS, ce qui empêche une bonne discrimination entre parole et bruit.

[0036] Pour pallier ce défaut, il est possible de définir le premier signal de bruit référent $Ref_1(k,l)$ par :

$$Ref_1(k,l) = X_1(k,l) - X_1(k,l) \frac{|Y(k,l)|}{|X_1(k,l)|}$$

[0037] À la différence de nombreuses méthodes classiques d'estimation du bruit, aucune hypothèse de stationnarité n'est utilisée sur le bruit pour calculer ce premier canal de bruit de référence $Ref_1(k,l)$. L'un des avantages est par conséquent que ce canal de bruit intègre une partie des bruits non stationnaires, en particulier ceux qui ont une faible corrélation statistique et qui ne sont pas prédictibles entre les deux micros.

Calcul d'une seconde référence de bruit par blocage spatial (bloc 20)

[0038] Dans un environnement de télécommunications, il est possible de rencontrer des bruits dont la source est bien localisée et relativement proche des micros. Il s'agit en général de bruits ponctuels assez puissants (passage d'un scooter, dépassement par une voiture, etc.), et qui peuvent être gênants.

[0039] Les hypothèses utilisées pour le calcul du premier canal de bruit référent ne sont pas vérifiées sur ce type de bruit ; en revanche, ces bruits ont la particularité d'avoir une direction d'incidence bien définie et distincte de la direction d'incidence de la parole.

[0040] Pour exploiter cette propriété, on supposera que l'angle d'incidence θ_S de la parole est connu, par exemple défini comme étant l'angle entre la médiatrice du couple de micros et la direction de référence correspondant à la source de parole utile.

[0041] Plus précisément, on opère une partition de l'espace en secteurs angulaires qui décrivent l'espace, et dont chacun correspond à une direction définie par un angle θ_j , $j \in [1,M]$, avec par exemple $M = 19$, donnant la collection d'angles $\{-90^\circ, -80^\circ, \dots, 0^\circ, \dots, +80^\circ, +90^\circ\}$. On notera qu'il n'y a aucun lien entre le nombre N de micros et le nombre M d'angles testés : par exemple, il est tout à fait possible de tester $M = 19$ angles avec un seul couple de micros ($N=2$).

[0042] On se donne la partition $\{A,l\}$ des angles θ_j qui sont respectivement "autorisés" et "interdits", les angles $\theta_a \in A$ étant "autorisés" en ce qu'ils correspondent à des signaux en provenance d'un cône privilégié centré sur θ_S , tandis

que les angles $\theta_j \in I$ sont "interdits" en ce qu'ils correspondent à des bruits latéraux indésirables.

[0043] Le second canal de bruit référent $Ref_2(k, l)$ est défini de la manière suivante :

$$Ref_2(k, l) = \frac{1}{|A|} \sum_{\theta_a \in A} \left(X_1(k, l) - X_2(k, l) \times e^{\frac{i 2 \pi \cdot f_k \cdot d \cdot \sin \theta_a}{c}} \right)$$

- 10 $X_1(k, l)$ étant la TFCT du signal enregistré par le micro d'indice 1,
 $X_2(k, l)$ étant la TFCT du signal enregistré par le micro d'indice 2,
 f_k étant la fréquence centrale de la bande de fréquences k ,
 l étant la trame,
 d étant la distance entre les deux micros,
15 c étant la célérité du son, et
 $|A|$ étant le nombre d'angles "autorisés" du cône privilégié.

[0044] Dans chaque terme de cette somme, on retranche au signal du micro d'indice 1 le signal du micro d'indice 2 déphasé d'un angle θ_α qui appartient à A (sous-collection des angles "autorisés"). Ainsi, dans chaque terme on bloque spatialement les signaux ayant une direction de propagation θ_α "autorisée". Ce blocage spatial est effectué pour tous les angles autorisés.

[0045] Dans ce second canal de bruit référent $Ref_2(k, l)$, on laisse donc passer les éventuels bruits latéraux (bruits non stationnaires directifs), en bloquant spatialement le signal de parole.

25 *Choix de la référence de bruit en fonction de la direction d'incidence des signaux (blocs 22 et 24)*

[0046] Cette sélection implique une estimation de l'angle d'incidence $\hat{\theta}(k, l)$ des signaux.

[0047] Cet estimateur (bloc 24) peut par exemple s'appuyer sur un calcul d'intercorrélation, en prenant comme direction d'incidence l'angle qui maximise le module de l'estimateur, soit :

30

$$\hat{\theta}(k, l) = \arg \max_{\theta_j, j \in [1, M]} \|P_{1,2}(\theta_j, k, l)\|$$

35

avec :

40

$$P_{1,2}(\theta_j, k, l) = E(X_1(k, l) \cdot \bar{X}_2(k, l) \cdot e^{-i 2 \pi f_k \tau_j})$$

et

45

$$\tau_j = \frac{d}{c} \sin \theta_j$$

50

[0048] Le canal de bruit référent sélectionné $Ref(k, l)$ va dépendre de la détection d'un angle "autorisé" ou "interdit" pour la trame l et la bande de fréquence k :

- 55 - si $\hat{\theta}(k, l)$ est "autorisé" ($\hat{\theta}(k, l) \in A$), alors $Ref(k, l) = Ref_1(k, l)$
- si $\hat{\theta}(k, l)$ est "interdit" ($\hat{\theta}(k, l) \in I$), alors $Ref(k, l) = Ref_2(k, l)$
- si $\hat{\theta}(k, l)$ n'est pas défini, alors $Ref(k, l) = Ref_1(k, l)$

[0049] Ainsi, dans le cas d'un angle "autorisé" détecté, ou en l'absence de signaux directifs à l'entrée des micros, le canal de bruit référent $Ref(k,l)$ est calculé par cohérence spatiale, ce qui permet d'intégrer les bruits non stationnaires peu directifs.

[0050] En revanche si un angle "interdit" est détecté, cela signifie qu'un bruit directif et assez puissant est présent. Dans ce cas, le canal de bruit référent $Ref(k,l)$ est calculé suivant une méthode différente, par blocage spatial, de façon à introduire efficacement dans ce canal les bruits non stationnaires directifs et puissants.

Constitution d'un signal combiné partiellement débruité (bloc 28)

[0051] Les signaux $X_n(k,l)$ (les TFCT des signaux captés par les micros) peuvent être combinés entre eux par une technique simple de préfiltrage par *beamforming* du type *Delay and Sum*, qui est appliquée pour obtenir un signal combiné $X(k,l)$ partiellement débruité :

$$X(k,l) = \frac{1}{2} [X_1(k,l) + \overline{d_2(k)} \cdot X_2(k,l)]$$

avec :

$$d_2(k) = e^{i2\pi f_k \tau_s} \text{ avec } \tau_s = \frac{d}{c} \sin \theta_s$$

[0052] Lorsque le système considéré comporte, comme dans le présent exemple, deux micros dont la médiatrice coupe la source, l'angle θ_s est nul et il s'agit d'une simple moyenne qui est faite sur les deux micros. Il est par ailleurs à noter que, concrètement, le nombre de micros étant limité, ce traitement ne procure qu'une faible amélioration du rapport signal/bruit, de l'ordre de 1 dB seulement.

Estimation du bruit pseudo-stationnaire (blocs 30 et 32)

[0053] Cette étape a pour objet de calculer une estimation de la composante de bruit pseudo-stationnaire présente dans la référence de bruit $Ref(k,l)$ (bloc 30) et, de la même manière, la composante de bruit pseudo-stationnaire présente dans le signal à débruiter $X(k,l)$ (bloc 32).

[0054] Il existe de très nombreuses publications sur ce sujet, l'estimation de la composante de bruit pseudo-stationnaire étant en effet un problème classique assez bien résolu. Différentes méthodes sont efficaces et utilisables à cet effet, notamment un algorithme d'estimation de l'énergie de la composante de bruit pseudo-stationnaire à moyennage récursif par contrôle des minima (MCRA) comme celui décrit par I. Cohen et B. Berdugo, Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement, IEEE Signal Processing Letters, Vol. 9, No 1, pp. 12-15, Jan. 2002.

Calcul de la probabilité d'absence de parole (bloc 26)

[0055] Une méthode efficace et reconnue pour estimer la probabilité d'absence de parole dans un environnement bruité est celle du rapport des transients, décrite par I. Cohen et B. Berdugo, Two-Channel Signal Detection and Speech Enhancement Based on the Transient Beam-to-Reference Ratio, Proc. ICASSP 2003, Hong-Kong, pp. 233-236, Apr. 2003.

[0056] Le rapport des transients est défini de la manière suivante :

$$\Omega(k,l) = \frac{S[X(k,l)] - M[X(k,l)]}{S[Ref(k,l)] - M[Ref(k,l)]}$$

$X(k,l)$ étant le signal combiné partiellement débruité,

$Ref(k,l)$ étant le canal de bruit référent calculé dans la partie précédente,
 k étant la bande de fréquences, et
 l étant la trame

5 **[0057]** L'opérateur S est une estimation de l'énergie instantanée, et l'opérateur M est une estimation de l'énergie pseudo-stationnaire (estimation effectuée par les blocs 30 et 32). $S - M$ fournit une estimation des parties transitoires du signal analysé, appelés aussi transients.

[0058] Les deux signaux analysés sont ici le signal bruité combiné $X(k,l)$ et le signal du canal de bruit référent $Ref(k,l)$. Le numérateur va donc mettre en évidence les transients de parole et de bruits, alors que le dénominateur va extraire
10 uniquement les transients de bruits se trouvant dans le canal de bruit référent.

[0059] Ainsi, en présence de parole mais en l'absence de bruit non stationnaire, le ratio $\Omega(k,l)$ va tendre vers une limite haute $\Omega_{\max}(k)$, alors qu'à l'inverse, en l'absence de parole mais en présence de bruit non stationnaire, ce ratio va se rapprocher de la limite basse $\Omega_{\min}(k)$, k étant la bande de fréquences. Ceci va permettre de réaliser la différenciation entre parole et bruits non stationnaires.

15 **[0060]** Dans le cas général, on a :

$$\Omega_{\min}(k) \leq \Omega(k,l) \leq \Omega_{\max}(k)$$

20

[0061] La probabilité d'absence de parole, notée ici $q(k,l)$, va être calculée de la manière suivante.

[0062] Pour chaque trame l et chaque bande de fréquences k :

25 i) Calcul de $S[X(k,l)]$, $S[Ref(k,l)]$, $M[X(k,l)]$ et $M[Ref(k,l)]$;

ii) Si $S[X(k,l)] \geq \alpha_x M[X(k,l)]$, la parole est susceptible d'être présente, l'analyse est poursuivie à l'étape (iii), dans le cas contraire, la parole est absente : alors $q(k,l) = 1$;

30 iii) Si $S[Ref(k,l)] \geq \alpha_{Ref} M[Ref(k,l)]$, du bruit transitoire est susceptible d'être présent, l'analyse est poursuivie à l'étape (iv), dans le cas contraire, ceci signifie les transients trouvés dans $X(k,l)$ sont tous des transients de parole : alors $q(k,l) = 0$;

35 iv) Calcul du ratio

v) Détermination de la probabilité d'absence de parole :

40

$$q(k,l) = \max \left(\min \left(\frac{\Omega_{\max}(k,l) - \Omega(k,l)}{\Omega_{\max}(k,l) - \Omega_{\min}(k,l)}, 1 \right), 0 \right).$$

45 **[0063]** Les constantes α_x et α_{Ref} utilisées dans cet algorithme sont en fait des seuils de détection des parties transitoires. Les paramètres α_x , α_{Ref} ainsi que $\Omega_{\min}(k)$ et $\Omega_{\max}(k)$, sont tous choisis de manière à correspondre à des situations typiques, proches de la réalité.

Réduction de bruit par application d'un gain OM-LSA (bloc 34)

50

[0064] La probabilité $q(k,l)$ d'absence de parole calculée au bloc 26 va être utilisée comme paramètre d'entrée dans une technique (en elle-même connue) de débruitage. Elle présente l'avantage de permettre d'identifier les périodes d'absence de parole même en présence d'un bruit non stationnaire, peu directif ou directif. La probabilité d'absence de parole est un estimateur crucial pour le bon fonctionnement d'une structure de débruitage telle que nous allons utiliser, car elle sous-tend la bonne estimation du bruit et le calcul d'un gain de débruitage efficace.

55 **[0065]** On peut avantageusement utiliser une méthode de débruitage de type OM-LSA (*Optimally Modified - Log Spectral Amplitude*) telle que celle décrite par : I. Cohen, Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator, IEEE Signal Processing Letters, Vol. 9, No 4, April 2002.

[0066] Essentiellement, l'application d'un gain nommé "gain LSA" (*Log-Spectral Amplitude*) permet de minimiser la distance quadratique moyenne entre le logarithme de l'amplitude du signal estimé et le logarithme de l'amplitude du signal de parole originel. Ce second critère se montre supérieur au premier car la distance choisie est en meilleure adéquation avec le comportement de l'oreille humaine et donne donc qualitativement de meilleurs résultats. Dans tous les cas, l'idée essentielle est de diminuer l'énergie des composantes fréquentielles très parasitées en leur appliquant un gain faible, tout en laissant intactes (par l'application d'un gain égal à 1) celles qui le sont peu ou pas du tout.

[0067] L'algorithme "OM-LSA" (*Optimally-Modified Log-Spectral Amplitude*) améliore le calcul du gain LSA à appliquer en le pondérant par la probabilité conditionnelle de présence de parole.

[0068] Dans cette méthode, la probabilité d'absence de parole intervient à deux moments importants, pour l'estimation de l'énergie du bruit et pour le calcul du gain final, et la probabilité $q(k,l)$ sera utilisée à ces deux niveaux. Si l'on note $\hat{\lambda}_{Bruit}(k,l)$ l'estimation de la densité spectrale de puissance du bruit, cette estimation est donnée par :

$$\hat{\lambda}_{Bruit}(k,l) = \alpha_{Bruit}(k,l) \cdot \hat{\lambda}_{Bruit}(k,l-1) + [1 - \alpha_{Bruit}(k,l)] |X(k,l)|^2$$

avec :

$$\alpha_{Bruit}(k,l) = \alpha_B + (1 - \alpha_B) \cdot p_{spa}(k,l)$$

[0069] On peut noter ici que la probabilité $q(k,l)$ vient moduler le facteur d'oubli dans l'estimation du bruit, qui est mise à jour plus rapidement sur le signal bruité $X(k,l)$ lorsque la probabilité d'absence de parole est forte, ce mécanisme conditionnant entièrement la qualité de $\hat{\lambda}_{Bruit}(k,l)$.

[0070] Le gain de débruitage $G_{OM-LSA}(k,l)$ est donné par :

$$G_{OM-LSA}(k,l) = \{G_{H1}(k,l)\}^{1-q(k,l)} \cdot G_{min}^{q(k,l)}$$

[0071] $G_{H1}(k,l)$ étant un gain de débruitage (dont le calcul dépend de l'estimation du bruit $\hat{\lambda}_{Bruit}$) décrit dans l'article précité de Cohen, et G_{min} étant une constante correspondant au débruitage appliqué lorsque la parole est considérée comme absente.

[0072] On notera que la probabilité $q(k,l)$ joue ici un grand rôle dans la détermination du gain $G_{OM-LSA}(k,l)$. Notamment, lorsque cette probabilité est nulle, le gain est égal à G_{min} et une réduction de bruit maximale est appliquée : si par exemple une valeur de 20 dB est choisie pour G_{min} , les bruits non stationnaires précédemment détectés sont atténués de 20 dB.

[0073] Le signal débruité $\hat{S}(k,l)$ en sortie du bloc 34 est donné par :

$$\hat{S}(k,l) = G_{OM-LSA}(k,l) \cdot X(k,l)$$

[0074] On notera que d'ordinaire une telle structure de débruitage produit un résultat peu naturel et agressif sur les bruits non stationnaires, qui sont confondus avec la parole utile. L'un des intérêts majeurs de l'invention est, au contraire, d'éliminer efficacement ces bruits non stationnaires.

[0075] D'autre part, dans une variante avantageuse, il est possible d'utiliser dans les expressions données plus haut une probabilité d'absence de parole hybride $q_{hybrid}(k,l)$, qui sera calculée à l'aide de $q(k,l)$ et d'une autre probabilité d'absence de parole $q_{std}(k,l)$, par exemple évaluée selon la méthode décrite dans le WO 2007/099222 A1 (Parrot SA). On a alors :

$$q_{\text{hybrid}}(k, l) = \max(q(k, l), q_{\text{std}}(k, l))$$

Reconstitution temporelle du signal (bloc 36)

[0076] La dernière étape consiste à appliquer au signal $\hat{S}(k, l)$ une transformée de Fourier rapide inverse iFFT pour obtenir dans le domaine temporel le signal de parole débruité $\hat{s}(t)$ recherché.

Revendications

1. Un procédé de débruitage d'un signal acoustique bruité capté par deux microphones d'un dispositif audio multi-microphone opérant dans un milieu bruité, notamment un dispositif téléphonique "mains libres" pour véhicule automobile, le signal acoustique bruité comprenant une composante utile de parole issue d'une source de parole directive et une composante parasite de bruit, cette composante de bruit incluant elle-même une composante de bruit latéral non stationnaire directif, procédé **caractérisé en ce qu'il** comporte, dans le domaine fréquentiel pour une pluralité de bandes de fréquences définies pour des trames temporelles successives de signal, les étapes de traitement du signal suivantes :

a) calcul (18) d'une première référence de bruit par analyse de cohérence spatiale des signaux captés les deux microphones, ce calcul comprenant un filtrage linéaire prédictif appliqué aux signaux captés par les deux microphones et comprenant une soustraction avec compensation du déphasage entre le signal capté et le signal de sortie du filtre prédictif ;

b) calcul (20) d'une seconde référence de bruit par analyse des directions d'incidence des signaux captés par les deux microphones, ce calcul comprenant le blocage spatial des composantes des signaux captés dont la direction d'incidence est située à l'intérieur d'un cône de référence défini de part et d'autre d'une direction prédéterminée d'incidence du signal utile ;

c) estimation (24) d'une direction principale d'incidence ($\hat{\theta}(k, l)$) des signaux captés par les deux microphones ;
d) sélection (22) comme signal de bruit référent ($Ref(k, l)$) de l'une ou l'autre des références de bruit calculées aux étapes a) et b), en fonction de la direction principale estimée à l'étape c) ;

e) combinaison (28) des signaux captés par les deux microphones en un signal combiné bruité ($X(k, l)$);

f) calcul (26) d'une probabilité d'absence de parole ($q(k, l)$) dans le signal combiné bruité, à partir des niveaux respectifs d'énergie spectrale du signal combiné bruité ($X(k, l)$) et du signal de bruit référent ($Ref(k, l)$) ;

g) à partir de la probabilité d'absence de parole ($q(k, l)$) calculée à l'étape f) et du signal combiné bruité ($X(k, l)$), réduction sélective du bruit (34) par application d'un gain variable propre à chaque bande de fréquences et à chaque trame temporelle.

2. Le procédé de la revendication 1, dans lequel le filtrage prédictif comprend l'application d'un algorithme de prédiction linéaire de type moindres carrés moyens LMS.

3. Le procédé de la revendication 1, dans lequel l'estimation (24) de la direction principale d'incidence de l'étape c) comprend les sous-étapes successives suivantes :

c1) partition de l'espace en une pluralité de secteurs angulaires ;

c2) pour chaque secteur, évaluation d'un estimateur de direction d'incidence à partir des deux signaux captés par les deux microphones correspondants ; et

c3) à partir des valeurs d'estimateurs calculées à l'étape c2), estimation de ladite direction principale d'incidence.

4. Le procédé de la revendication 1, dans lequel la sélection (22) de l'étape d) est une sélection de la seconde référence de bruit comme signal de bruit référent si la direction principale estimée à l'étape c) est située hors d'un cône de référence défini de part et d'autre d'une direction prédéterminée d'incidence du signal utile.

5. Le procédé de la revendication 1, dans lequel la combinaison (28) de l'étape e) comprend un préfiltrage de type *fixed beamforming*.

6. Le procédé de la revendication 1, dans lequel le calcul (26) de probabilité d'absence de parole de l'étape f) comprend l'estimation (30, 32) de composantes de bruit pseudo-stationnaire respectives contenues dans le signal combiné bruité et dans le signal de bruit référent, la probabilité d'absence de parole ($q(k, l)$) étant calculée à partir également de ces composantes de bruit pseudo-stationnaire respectives.

7. Le procédé de la revendication 1, dans lequel la réduction sélective du bruit (34) de l'étape g) est un traitement par application d'un gain à amplitude log-spectrale modifié optimisé OM-LSA.

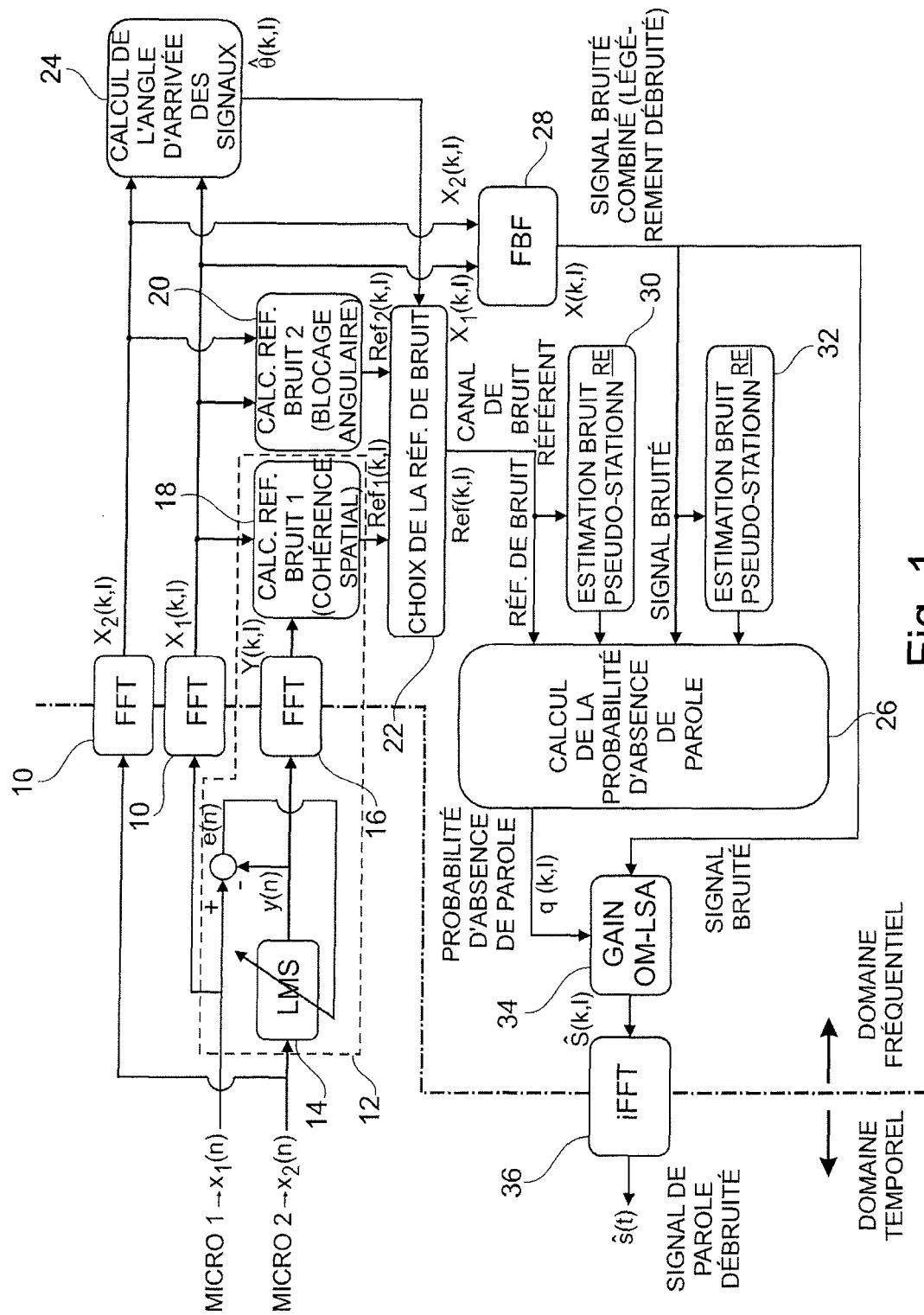


Fig. 1



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande

EP 10 16 7065

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
A,D	WO 2007/099222 A1 (PARROT [FR]) 7 septembre 2007 (2007-09-07) * page 6, ligne 20 - page 8, ligne 22 * * page 9, ligne 1 - page 16, ligne 2 * * figure 1 *	1-7	INV. G10L21/02 H04R3/00 H04B1/12
A	HOHMANN V ET AL: "Binaural noise reduction for hearing aids" 2002 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). ORLANDO, FL, MAY 13 - 17, 2002; [IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP)], NEW YORK, NY : IEEE, US, vol. 4, 13 mai 2002 (2002-05-13), pages IV-4000, XP010804508 ISBN: 978-0-7803-7402-7 * p. 4000, section 2. Directional Filtering and Dereverberation * * figure 1 *	1-7	
			DOMAINES TECHNIQUES RECHERCHES (IPC)
A,D	COHEN I ET AL: "Two-channel signal detection and speech enhancement based on the transient beam-to-reference ratio" PROCEEDINGS OF INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP'03) 6-10 APRIL 2003 HONG KONG, CHINA; [IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP)], 2003 IEEE INTERNATIONAL CONFERENCE, vol. 5, 6 avril 2003 (2003-04-06), pages V 233-V 236, XP010639251 ISBN: 978-0-7803-7663-2 * le document en entier *	1-7	G10L H04R H04B
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche Munich		Date d'achèvement de la recherche 20 août 2010	Examineur Chétry, Nicolas
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 (PC4C02)



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande

EP 10 16 7065

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
A	SIOW YONG LOW ET AL: "Robust microphone array using subband adaptive beamformer and spectral subtraction" COMMUNICATION SYSTEMS, 2002. ICCS 2002. THE 8TH INTERNATIONAL CONFERENCE ON NOV. 25-28, 2002, PISCATAWAY, NJ, USA, IEEE, vol. 2, 25 novembre 2002 (2002-11-25), pages 1020-1024, XP010629375 ISBN: 978-0-7803-7510-9 * le document en entier * -----	1-7	DOMAINES TECHNIQUES RECHERCHES (IPC)
A	COHEN I ET AL: "Speech enhancement based on a microphone array and log-spectral amplitude estimation" ELECTRICAL AND ELECTRONICS ENGINEERS IN ISRAEL, 2002. THE 22ND CONVENTION OF DEC. 1, 2002, PISCATAWAY, NJ, USA, IEEE, 1 janvier 2002 (2002-01-01), pages 4-6, XP010631024 ISBN: 978-0-7803-7693-9 * le document en entier * -----	1-7	
A	MOHAMMED J R ED - ZOOLFAKAR A S ET AL: "Intelligent Method for Designing Exact Orthogonal Blocking Matrix into Robust Wideband Beamformer Using Auxiliary Sensors" MODELING&SIMULATION, 2008. AICMS 08. SECOND ASIA INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 13 mai 2008 (2008-05-13), pages 511-515, XP031265054 ISBN: 978-0-7695-3136-6 * le document en entier * -----	1-7	
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche		Date d'achèvement de la recherche	Examineur
Munich		20 août 2010	Chétry, Nicolas
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)

**ANNEXE AU RAPPORT DE RECHERCHE EUROPEENNE
RELATIF A LA DEMANDE DE BREVET EUROPEEN NO.**

EP 10 16 7065

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche européenne visé ci-dessus.

Lesdits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du

Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

20-08-2010

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
W0 2007099222 A1	07-09-2007	EP 1830349 A1	05-09-2007
		FR 2898209 A1	07-09-2007
		US 2007276660 A1	29-11-2007

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

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

- WO 2007099222 A1 [0075]

Littérature non-brevet citée dans la description

- **I. Cohen ; B. Berdugo.** Two-Channel Signal Detection and Speech Enhancement Based on the Transient Beam-to-Reference Ratio. *Proc. ICASSP 2003*, Avril 2003, 233-236 [0016] [0055]
- **I. Cohen ; B. Berdugo.** Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement. *IEEE Signal Processing Letters*, Janvier 2002, vol. 9 (1), 12-15 [0054]
- **I. Cohen.** Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator. *IEEE Signal Processing Letters*, Avril 2002, vol. 9 (4) [0065]