



(12) **DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:
26.12.2012 Bulletin 2012/52

(51) Int Cl.:
G10L 21/02 (2006.01) H04R 3/00 (2006.01)

(21) Numéro de dépôt: **12170874.7**

(22) Date de dépôt: **05.06.2012**

(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
Etats d'extension désignés:
BA ME

(71) Demandeur: **Parrot**
75010 Paris (FR)

(72) Inventeur: **Fox, Charles**
75014 PARIS (FR)

(74) Mandataire: **Dupuis-Latour, Dominique et al**
Bardehle Pagenberg
10, boulevard Haussmann
75009 Paris (FR)

(30) Priorité: **20.06.2011 FR 1155377**

(54) **Procédé de débruitage pour équipement audio multi-microphones, notamment pour un système de téléphonie "mains libres"**

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

- estimation (22) d'une probabilité de présence de parole (p);
- estimation (46), modulée par la probabilité de présence de parole (p), d'une matrice spectrale de covariance (R_n) des bruits recueillis par les capteurs ;
- estimation (28, 38), modulée par la probabilité de présence de parole (p), de la fonction de transfert (H) des canaux acoustiques entre la source de parole et au moins

- certaines des capteurs, par rapport à une référence constituée par le signal recueilli par l'un des capteurs ;
- calcul (48) d'un projecteur linéaire optimal donnant un signal combiné unique à partir des signaux ($X_1... X_n$) recueillis par au moins certains des capteurs, de la matrice spectrale de covariance (R_n), et des fonctions de transfert estimées ($H_1...H_n$); et
- à partir de la probabilité de présence de parole (p) et du signal combiné de sortie du projecteur, réduction sélective du bruit (50) par application d'un gain variable.

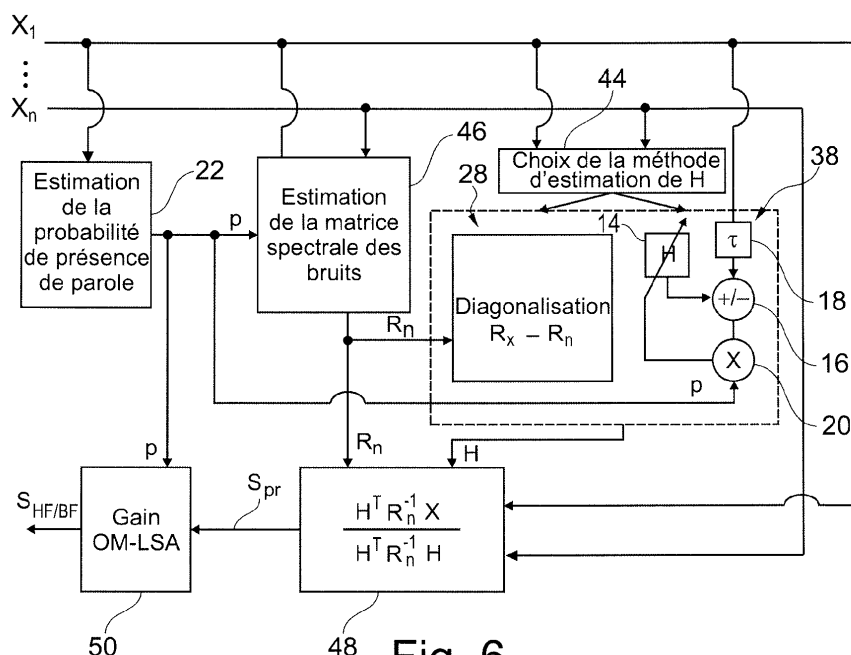


Fig. 6

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 ou plusieurs microphones ("micros") sensible non seulement à la voix de l'utilisateur, mais captant aussi le bruit environnant ainsi que l'écho dû au phénomène de réverbération par l'environnement, typiquement l'habitacle du véhicule. La composante utile (le signal de parole du locuteur proche) se trouve ainsi noyée dans une composante parasite de bruit (bruits externes et réverbération) pouvant aller, souvent, jusqu'à rendre incompréhensible pour le locuteur distant (celui qui est à l'autre bout de la voie de transmission du signal téléphonique) les paroles du locuteur proche.

[0004] 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.

[0005] 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.

[0006] Certains de ces dispositifs prévoient l'utilisation de plusieurs 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, des techniques dites de *beamforming* permettent 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). Les performances sont en outre très dégradées lorsque l'environnement est réverbérant.

[0007] Le but de l'invention est de proposer une solution de débruitage des signaux audio captés par un tel système multicanal, multi-microphones, dans un environnement très bruyant et très réverbérant, typiquement l'habitacle d'une voiture.

[0008] La principale difficulté liée aux méthodes de traitement de la parole par des systèmes multicanal est la difficulté d'estimation des paramètres utiles pour les traitements à appliquer, car les estimateurs sont fortement liés à l'environnement ambiant.

[0009] La plupart des techniques se basent sur l'hypothèse que le signal utile et/ou les bruits parasites présentent une certaine directivité, et combinent les signaux issus des différents micros de manière à améliorer le rapport signal/bruit en fonction de ces conditions de directivité.

[0010] Ainsi, le EP 2 293 594 A1 (Parrot SA) décrit un procédé à détection spatiale et filtrage des bruits non stationnaires et directifs tels que coups de klaxon, passage d'un scooter, dépassement par une voiture, etc. La technique proposée consiste à associer les propriétés de non-stationnarité temporelle et fréquentielle, d'une part, et de directivité spatiale, d'autre part, pour détecter un type de bruit qu'il est d'ordinaire difficile de discriminer de la parole, afin d'assurer un filtrage efficace de ce bruit et de déduire par ailleurs une probabilité de présence de parole qui permettra d'améliorer encore l'atténuation du bruit.

[0011] Le EP 2 309 499 A1 (Parrot SA) décrit un système à deux micros opérant une analyse de cohérence spatiale du signal capté de manière à déterminer une direction d'incidence. Le système calcule deux références de bruits selon des méthodes différentes, l'une en fonction de la cohérence spatiale des signaux captés (qui intègre les bruits non stationnaires peu directifs) et une autre en fonction de la direction principale d'incidence des signaux (qui intègre surtout les bruits non stationnaires directifs). Cette technique de débruitage repose sur l'hypothèse que la parole présente généralement une cohérence spatiale supérieure au bruit et que, par ailleurs, 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 sont tournés les micros.

[0012] Ces techniques prennent cependant mal en compte l'effet de réverbération typique de l'habitacle d'une voiture, où les réflexions puissantes et nombreuses rendent difficile le calcul d'une direction d'arrivée, avec pour conséquence une dégradation notable de l'efficacité du débruitage.

[0013] En outre, avec ces techniques le signal débruité obtenu en sortie restitue de façon satisfaisante l'amplitude du signal de parole initial, mais non sa phase, ce qui peut entraîner une déformation de la voix reproduite par le dispositif.

[0014] Le problème de l'invention est la prise en compte d'un environnement réverbérant ne permettant pas de calculer de façon satisfaisante une direction d'arrivée du signal utile et, subsidiairement, l'obtention d'un débruitage qui restitue à la fois l'amplitude et la phase du signal initial, en ne déformant donc pas la voix du locuteur lorsque celle-ci est reproduite par le dispositif.

[0015] L'invention propose une technique mise en oeuvre dans le domaine fréquentiel, pour une pluralité de *bins* du

signal capté (c'est-à-dire pour chaque bande de fréquences de chaque trame temporelle du signal). Le traitement consiste essentiellement à :

- calculer une probabilité de présence de parole dans le signal bruité recueilli ;
- estimer la fonction de transfert du canal acoustique entre la source de parole (le locuteur proche) et chacun des capteurs du réseau de micros ;
- calculer une projection optimale pour déterminer un canal unique à partir des fonctions de transfert des canaux multiples estimés ; et
- réduire sélectivement le bruit sur ce canal unique, pour chaque *bin*, en fonction de la probabilité de présence de parole.

[0016] Plus précisément, le procédé de l'invention est un procédé de débruitage pour un dispositif comprenant un réseau formé d'une pluralité de capteurs microphoniques disposés selon une configuration prédéterminée.

[0017] Ce procédé comporte les étapes de traitement suivantes dans le domaine fréquentiel, pour une pluralité de bandes de fréquences définies pour des trames temporelles successives de signal :

- a) estimation d'une probabilité de présence de parole dans le signal bruité recueilli ;
- b) estimation d'une matrice spectrale de covariance des bruits recueillis par les capteurs, cette estimation étant modulée par la probabilité de présence de parole ;
- c) estimation de la fonction de transfert des canaux acoustiques entre la source de parole et au moins certains des capteurs, cette estimation étant opérée par rapport à une référence de signal utile constituée par le signal recueilli par l'un des capteurs, et étant en outre modulée par la probabilité de présence de parole ;
- d) calcul d'un projecteur linéaire optimal, donnant un signal combiné débruité unique à partir des signaux recueillis par au moins certains des capteurs, de la matrice spectrale de covariance estimée à l'étape b), et des fonctions de transfert estimées à l'étape c) ; et
- e) à partir de la probabilité de présence de parole et du signal combiné donné par le projecteur calculé à l'étape d), réduction sélective du bruit par application d'un gain variable propre à chaque bande de fréquences et à chaque trame temporelle.

[0018] De préférence, le calcul du projecteur linéaire optimal de l'étape d) est opéré par un traitement de type *beam-forming* de Capon à réponse sans distorsion à variance minimale MVDR.

[0019] De préférence également, la réduction sélective de bruit de l'étape e) est opérée par un traitement de type gain à amplitude log-spectrale modifié optimisé OM-LSA.

[0020] Dans une première forme de mise en oeuvre, l'estimation de la fonction de transfert de l'étape c) est opérée par calcul d'un filtre adaptatif visant à annuler la différence entre le signal recueilli par le capteur dont on cherche à évaluer la fonction de transfert et le signal recueilli par le capteur de la référence de signal utile, avec modulation par la probabilité de présence de parole.

[0021] Le filtre adaptatif peut notamment être un filtre à algorithme de prédiction linéaire de type moindres carrés moyens LMS et la modulation par la probabilité de présence de parole, une modulation par variation du pas d'itération du filtre adaptatif.

[0022] Dans une deuxième forme de mise en oeuvre, l'estimation de la fonction de transfert de l'étape c) est opérée par un traitement de diagonalisation comprenant :

- c1) la détermination d'une matrice spectrale de corrélation des signaux recueillis par les capteurs du réseau par rapport au capteur de la référence de signal utile,
- c2) le calcul de la différence entre, d'une part, la matrice déterminée à l'étape c1) et, d'autre part, la matrice spectrale de covariance des bruits modulée par la probabilité de présence de parole, calculée à l'étape b), et
- c3) la diagonalisation de la matrice différence calculée à l'étape c2). Par ailleurs, le spectre du signal à débruiter est avantageusement divisé en une pluralité de parties de spectre distinctes, les capteurs étant regroupés en une pluralité de sous-réseaux associés chacun à l'une des parties du spectre. Le traitement de débruitage est alors opéré de façon différenciée, pour chacune des parties du spectre, sur les signaux recueillis par les capteurs du sous-réseau correspondant à la partie du spectre considérée.

[0023] En particulier, dans le cas où le réseau de capteurs est un réseau linéaire de capteurs alignés, le spectre du signal à débruiter peut être divisé en une partie basse fréquence et une partie haute fréquence. Pour la partie basse fréquence, les étapes du traitement de débruitage sont alors opérées seulement sur les signaux recueillis par les capteurs les plus éloignés du réseau.

[0024] Il est également possible, toujours avec un spectre de signal à débruiter divisé en une pluralité de parties de spectre distinctes, d'estimer de manière différenciée, à l'étape c), la fonction de transfert des canaux acoustiques par

application de traitements différents pour chacune des parties du spectre.

[0025] En particulier, dans le cas où le réseau de capteurs est un réseau linéaire de capteurs alignés et où les capteurs sont regroupés en une pluralité de sous-réseaux associés chacun à l'une des parties du spectre : pour la partie basse fréquence, le traitement de débruitage est opéré seulement sur les signaux recueillis par les capteurs les plus éloignés du réseau et l'estimation de la fonction de transfert est opérée par calcul d'un filtre adaptatif ; et pour la partie haute fréquence, le traitement de débruitage est opéré sur les signaux recueillis par tous les capteurs du réseau, et l'estimation de la fonction de transfert est opérée par un traitement de diagonalisation.

[0026] On va maintenant décrire un exemple de mise en oeuvre du dispositif de l'invention, en référence aux dessins annexés où les mêmes références numériques désignent d'une figure à l'autre des éléments identiques ou fonctionnellement semblables.

La Figure 1 est une représentation schématique des différents phénomènes acoustiques impliqués dans le recueil de signaux bruités.

La Figure 2 est une représentation schématique par blocs fonctionnels d'un filtre adaptatif pour l'estimation de la fonction de transfert d'un canal acoustique.

La Figure 3 est une caractéristique montrant les variations de la corrélation entre deux capteurs pour un champ de bruit diffus, en fonction de la fréquence.

La Figure 4 illustre de façon schématique un réseau de quatre micros utilisables de façon sélective, en fonction de la fréquence, pour la mise en oeuvre de l'invention.

La Figure 5 est un schéma d'ensemble, sous forme de blocs fonctionnels, montrant les différents traitements selon l'invention pour le débruitage des signaux recueillis par le réseau de micros de la Figure 4.

La Figure 6 illustre plus précisément, sous forme de schéma par blocs, les fonctions mises en oeuvre, dans le domaine fréquentiel, pour le traitement selon l'invention illustré Figure 5.

[0027] On va maintenant décrire en détail la technique de débruitage proposée par l'invention.

[0028] On considérera, comme illustré Figure 1, un ensemble de n capteurs microphoniques, chaque capteur pouvant être assimilé à un micro unique $M_1 \dots M_n$ captant une version réverbérée d'un signal de parole émis par une source de signal utile S (la parole d'un locuteur proche 10), signal auquel vient s'ajouter un bruit.

[0029] Chaque micro capte donc :

- une composante du signal utile (le signal de parole),
- une composante de la réverbération de ce signal de parole par l'habitable du véhicule, et
- une composante du bruit parasite environnant, sous toutes ses formes (directif ou diffus, stationnaire ou évoluant de manière imprévisible, etc.).

Modélisation des signaux captés

[0030] Il s'agit de traiter les (multiples) signaux de ces micros en opérant un débruitage (bloc 12) donnant en sortie un signal (unique) : on reconnaît dans ce schéma un modèle MISO (*Multiple Input Single Output*).

[0031] Le signal de sortie devra être le plus proche possible du signal de parole émis par le locuteur 10, c'est-à-dire :

- contenir le moins de bruit possible, et
- déformer le moins possible la voix du locuteur restituée en sortie.

[0032] Sur le capteur de rang i , le signal recueilli sera :

$$x_i(t) = h_i \otimes s(t) + b_i(t)$$

x_i étant le signal capté, h_i étant la réponse impulsionnelle entre la source de signal utile S et le capteur M_i , s étant le signal utile produit par la source S (signal de parole du locuteur proche 10) et b_i étant le bruit additif.

[0033] Pour l'ensemble des capteurs, on peut utiliser la notation vectorielle :

$$x(t) = h \otimes s(t) + b(t)$$

[0034] Dans le domaine fréquentiel, cette expression devient :

$$X(\omega) = H(\omega)S(\omega) + B(\omega)$$

[0035] On fera une première hypothèse selon laquelle la voix ainsi que le bruit sont gaussiens centrés.

[0036] Ceci se traduit dans le domaine fréquentiel par les conditions suivantes, pour toutes les fréquences ω :

- S est gaussien centré de puissance ϕ_s
- B est un vecteur gaussien centré de matrice de covariance R_n
- S et B sont décorrélés et chacun est décorrélé lorsque les fréquences sont différentes

[0037] On fera par ailleurs une deuxième hypothèse selon laquelle les bruits et la voix sont décorrélés. Cela se traduit par le fait que S est décorrélé avec toutes les composantes de B . De plus, pour des fréquences ω_i et ω_j différentes, $S(\omega_i)$ et $S(\omega_j)$ sont décorrélés. Cette hypothèse est également valable pour le vecteur de bruit B .

Calcul d'un projecteur optimal

[0038] La technique proposée consiste, sur la base des éléments que l'on vient d'exposer, à rechercher dans le domaine temporel un projecteur linéaire optimal pour chaque fréquence.

[0039] On entendra par "projecteur" un opérateur correspondant à une transformation d'une pluralité de signaux, recueillis concurremment par un dispositif multicanal, en un signal unique monocal.

[0040] Cette projection est une projection linéaire "optimale" en ce sens que la composante de bruit résiduel sur le signal monocal délivré en sortie soit minimisée (bruit et réverbération) et que la composante utile de parole soit le moins déformée possible.

[0041] Cette optimisation implique de rechercher pour chaque fréquence un vecteur A tel que :

- la projection $A^T X$ contienne le moins de bruit possible, c'est-à-dire que la puissance du bruit résiduel, qui vaut $E[A^T V V^T A] = A^T R_n A$, soit minimisée, et
- ne déforme pas la voix du locuteur, ce qui se traduit par la contrainte $A^T H = 1$.

R_n étant la matrice de corrélation entre les micros, pour chaque fréquence, et

[0042] H étant le canal acoustique considéré.

[0043] Ce problème est un problème d'optimisation sous contrainte, à savoir la recherche de $\min(A^T R_n A)$ sous la contrainte $A^T H = 1$.

[0044] Il peut être résolu en utilisant la méthode des multiplieurs de Lagrange, qui conduit à la solution :

$$A^T = \frac{H^T R_n^{-1}}{H^T R_n^{-1} H}$$

[0045] Dans le cas où les transferts H correspondent à un retard pur, on reconnaît la formule du *beamforming* MVDR (*Minimum Variance Distorsionless Response*), aussi appelé *beamforming* de Capon.

[0046] On notera que la puissance de bruit résiduel vaut, après projection :

$$\frac{1}{H^T R_n^{-1} H}$$

[0047] De plus, en écrivant des estimateurs de type *Minimum Mean-Squared Error* sur l'amplitude et la phase du signal à chaque fréquence, on constate que ces estimateurs s'écrivent comme un *beamforming* de Capon suivi d'un traitement monocal, comme décrit dans :

[1] R. C. Hendriks et al., On optimal multichannel mean-squared error estimators for speech enhancement, IEEE Signal Processing Letters, vol. 16, no. 10, 2009.

Le traitement de débruitage sélectif du bruit appliqué au signal monocanal résultant du traitement de *beamforming* est avantageusement un traitement de type gain à amplitude log-spectrale modifié optimisé OM-LSA tel que décrit par exemple dans :

[2] 1. Cohen, Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator, IEEE Signal Processing Letters, vol.9, no. 4, pp. 113-116, April 2002.

Estimation des paramètres pour le calcul du projecteur linéaire optimal

[0048] Pour la mise en oeuvre de cette technique, il est nécessaire d'estimer la fonction de transfert acoustique $H_1, H_2 \dots H_n$ entre la source de parole S et chacun des micros $M_1, M_2 \dots M_n$.

[0049] Il est également nécessaire d'estimer la matrice spectrale de covariance des bruits, notée R_n .

[0050] Pour ces estimations, on utilisera une probabilité de présence de parole, notée p .

[0051] La probabilité de présence de parole p est un paramètre pouvant prendre plusieurs valeurs différentes comprises entre 0 et 100 % (et non seulement une valeur binaire 0 ou 1). Ce paramètre est calculé selon une technique en elle-même connue, dont des exemples sont notamment exposés dans :

[3] 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.

[0052] On pourra également se référer au WO 2007/099222 A1, qui décrit une technique de débruitage mettant en oeuvre un calcul de probabilité de présence de parole.

[0053] En ce qui concerne la matrice spectrale de covariance des bruits R_n , on peut utiliser un estimateur d'espérance à fenêtre exponentielle, ce qui revient à appliquer un facteur d'oubli :

$$R_n(k+1) = \alpha R_n(k) + (1-\alpha) XX^T$$

$k+1$ étant le numéro de la trame courante, et α est un facteur d'oubli compris entre 0 et 1.

[0054] Pour ne prendre en compte que les éléments où seul le bruit est présent, on module le facteur d'oubli α par la probabilité de présence de parole :

$$\alpha = \alpha_0 + (1 - \alpha_0)p$$

avec $\alpha_0 \in [0,1]$.

[0055] Pour estimer la fonction de transfert H du canal acoustique considéré, plusieurs techniques sont utilisables.

[0056] Une première technique consiste à utiliser un algorithme de type LMS dans le domaine fréquentiel.

[0057] Les algorithmes de type LMS - ou NLMS (*Normalized LMS*) qui est une version normalisée du LMS - sont des algorithmes relativement simples et peu exigeants en termes de ressources de calcul. Il s'agit d'algorithmes en eux-mêmes connus, décrits par exemple dans :

[4] B. Widrow, Adaptive Filters, Aspect of Network and System Theory, R. E. Kalman and N. De Claris Eds., New York: Holt, Rinehart and Winston, pp. 563-587, 1970 ;

[5] J. Prado et E. Moulines, Frequency-domain adaptive filtering with applications to acoustic echo cancellation, Springer, Ed. Annals of Telecommunications, 1994 ;

[6] B. Widrow et S. Stearns, Adaptive Signal Processing, Prentice-Hall Signal Processing Series, Alan V. Oppenheim Series Editor, 1985.

[0058] Le principe de cet algorithme est illustré Figure 2.

[0059] De façon caractéristique de l'invention, on prendra l'un des canaux comme référence de signal utile, par exemple le canal du micro M_1 , et l'on calculera les fonctions de transfert $H_2 \dots H_n$, pour les autres canaux. Ceci revient à contraindre $H_1 = 1$.

[0060] On notera que bien que l'on prenne comme référence de signal utile la version réverbérée (donc parasitée) du signal de parole S captée par le micro M_1 , la présence de la réverbération dans le signal capté n'est pas gênante

car à ce stade on cherche à opérer un débruitage et non une dé-réverbération.

[0061] Comme illustré Figure 2, l'algorithme LMS vise (de façon connue) à estimer un filtre H (bloc 14) au moyen d'un algorithme adaptatif, correspondant au signal x_i délivré par le micro M_i , en estimant le transfert de bruit entre le micro M_i et le micro M_1 (pris comme référence). La sortie du filtre 14 est soustraite en 16 au signal x_1 capté par le micro M_1 pour donner un signal d'erreur de prédiction permettant l'adaptation itérative du filtre 14. Il est ainsi possible de prédire à partir du signal x_i la composante de parole (réverbérée) contenue dans le signal x_1 .

[0062] Pour éviter les problèmes liés à la causalité (c'est-à-dire pour être sûr que les signaux x_i n'arrivent pas en avance par rapport à la référence x_1), on retarde légèrement (bloc 18) le signal x_1 .

[0063] Par ailleurs, on ajoute un élément 20 permettant de pondérer le signal d'erreur du filtre adaptatif 14 par la probabilité de présence de parole p délivrée en sortie du bloc 22 : il s'agit de procéder à l'adaptation du filtre seulement quand la probabilité de présence de parole est élevée. Cette pondération peut être notamment opérée par modification du pas d'adaptation en fonction de la probabilité p .

[0064] L'équation de mise à jour du filtre adaptatif s'écrit, pour chaque trame k et pour chaque capteur i :

$$H_i(k+1) = H_i(k) + \mu X(k)_i^T (X(k)_1 - H(k)_i X(k)_i)$$

[0065] Le pas μ d'adaptation de l'algorithme, modulé par la probabilité de présence de parole, s'écrit, en normalisant le LMS (le dénominateur correspondant à la puissance spectrale du signal x_1 à la fréquence considérée) :

$$\mu = \frac{p}{E[X_1^2]}$$

[0066] L'hypothèse que les bruits sont décorrélés conduit à une prédiction par l'algorithme LMS de la voix, et non du bruit, de sorte que la fonction de transfert estimée correspond effectivement au canal acoustique H entre le locuteur et les micros.

[0067] Une autre technique possible d'estimation du canal acoustique consiste à opérer par diagonalisation de matrice.

[0068] Ce mode d'estimation est basé sur l'utilisation de la matrice spectrale de corrélation du signal observé, que l'on notera $R_x = E[XX^T]$.

[0069] On estime cette matrice de la même façon que R_n :

$$R_n(k+1) = \alpha R_n(k) + (1 - \alpha) XX^T$$

α étant un facteur d'oubli (fixe, puisque l'on prend en compte tout le signal).

[0070] On peut ensuite estimer $R_x - R_n = \phi_s HH^T$: il s'agit d'une matrice de rang 1, dont la seule valeur propre non nulle est ϕ_s , qui est associée au vecteur propre H .

[0071] On peut ainsi estimer H en diagonalisant $R_x - R_n$, mais on ne peut calculer que $\text{vect}(H)$, autrement dit on n'estime H qu'à un facteur complexe près.

[0072] Pour lever cette ambiguïté, de la même manière que précédemment pour l'estimation par algorithme LMS, on choisit l'un des canaux comme canal de référence, ce qui revient à contraindre $H_1 = 1$.

Échantillonnage spatial du champ sonore

[0073] Dans le cas d'un système multi-microphone, qui réalise donc un échantillonnage spatial du champ sonore, le placement relatif des différents micros est un aspect crucial pour l'efficacité du traitement des signaux captés par ces micros.

[0074] En particulier, comme on l'a indiqué au début, on fait l'hypothèse que les bruits présents sur les micros sont décorrélés pour utiliser une identification adaptative de type LMS. Pour être au plus près de cette hypothèse, il conviendrait d'éloigner les micros les uns des autres, car la fonction de corrélation s'écrit, pour un modèle de bruit diffus, comme une fonction décroissante de la distance entre les micros, ce qui rend les estimateurs de canal acoustique plus robustes.

[0075] En effet, la corrélation entre deux capteurs pour un champ de bruit diffus s'écrit :

$$MSC(f) = \text{sinc}^2\left(\frac{fd}{c}\right)$$

5

f étant la fréquence considérée,
 d étant la distance entre les capteurs, et
 c étant la vitesse du son.

[0076] La caractéristique correspondante est illustrée Figure 3, pour une distance entre micros $d = 10$ cm.

10 **[0077]** L'éloignement des micros, qui permet de décorréliser les bruits, présente cependant l'inconvénient de se traduire, dans le domaine spatial, à un échantillonnage à une fréquence plus faible, avec pour conséquence un repliement des hautes fréquences, qui seront moins bien restituées. L'invention propose de résoudre cette difficulté en sélectionnant des configurations de capteurs différentes selon les fréquences traitées.

15 **[0078]** Ainsi, sur la Figure 4, on a illustré un réseau linéaire de quatre micros alignés $M_1 \dots M_4$ espacés chacun de $d = 5$ cm.

[0079] Pour la région inférieure du spectre (basses fréquences BF) on choisira par exemple d'utiliser seulement les deux micros extrêmes M_1 et M_4 , éloignés donc de $3d = 15$ cm, tandis que pour la partie haute du spectre (hautes fréquences HF) on utilisera les quatre micros M_1, M_2, M_3 et M_4 espacés chacun de seulement $d = 5$ cm.

20 **[0080]** En variante ou en complément, selon un autre aspect de l'invention, on peut aussi sélectionner, pour l'estimation de la fonction de transfert H du canal acoustique, des méthodes différentes en fonction des fréquences traitées. Par exemple, pour les deux méthodes exposées plus haut (traitement fréquentiel par LMS, et traitement par diagonalisation), on peut choisir l'une ou l'autre méthode en fonction de critères tels que :

- la corrélation des bruits : pour tenir compte du fait que la méthode par diagonalisation y est moins sensible, mais est moins précise, ou
- le nombre de micros employés : pour tenir compte du fait que la méthode par diagonalisation devient très coûteuse en calculs lorsque la dimension des matrices augmente, du fait de l'augmentation du nombre n de micros.

25

Description d'un mode de réalisation préférentiel

30

[0081] Cet exemple est décrit en référence aux Figures 5 et 6, et met en oeuvre les différents éléments évoqués plus haut du traitement des signaux, avec leurs différentes variantes possibles.

[0082] La Figure 5 est un schéma par blocs montrant les différentes étapes de traitement des signaux issus d'un réseau linéaire de quatre micros $M_1 \dots M_4$ tels que celui illustré Figure 4.

35 **[0083]** Des traitements différents sont opérés pour le haut du spectre (hautes fréquences HF, correspondant aux blocs 24 à 32), et pour le bas du spectre (basses fréquences BF, correspondant aux blocs 34 à 42) :

- pour le haut du spectre, sélectionné par un filtre 24, les signaux des quatre micros $M_1 \dots M_4$ sont utilisés conjointement. Ces signaux font d'abord l'objet d'une transformée rapide de Fourier FFT (bloc 26) pour passer dans le domaine fréquentiel, puis d'un traitement 28 (décrit plus bas en référence à la Figure 6) impliquant une diagonalisation de matrice. Le signal monocanal résultant S_{HF} est soumis à une transformée de Fourier rapide inverse iFFT (bloc 30) pour repasser dans le domaine temporel, puis le signal résultant s_{HF} est appliqué à un filtre de synthèse (bloc 32) pour restituer le haut du spectre du canal de sortie s ;
- pour le bas du spectre, sélectionné par le filtre 34, seuls les signaux des deux micros extrêmes M_1 et M_4 sont utilisés. Ces signaux font d'abord l'objet d'une transformée rapide de Fourier FFT (bloc 36) pour passer dans le domaine fréquentiel, puis d'un traitement 38 (décrit plus bas en référence à la Figure 6) impliquant un filtrage adaptatif LMS. Le signal monocanal résultant S_{BF} est soumis à une transformée de Fourier rapide inverse iFFT (bloc 40) pour repasser dans le domaine temporel, puis le signal résultant s_{BF} est appliqué à un filtre de synthèse (bloc 42) pour restituer le bas du spectre du canal de sortie s .

50

[0084] On va maintenant décrire en référence à la Figure 6 les traitements opérés par les blocs 28 ou 38 de la Figure 5.

[0085] Le traitement que l'on va décrire est appliqué dans le domaine fréquentiel, à chaque *bin* de fréquence, c'est-à-dire pour chaque bande de fréquences définie pour les trames temporelles successives du signal recueilli par les micros (les quatre micros M_1, M_2, M_3 et M_4 pour le haut du spectre HF, et les deux micros M_1 et M_4 pour le bas du spectre BF).

55

[0086] A ces signaux correspondent, dans le domaine fréquentiel, des vecteurs $X_1 \dots X_n$ (X_1, X_2, X_3 et X_4 et X_1, X_4 , respectivement).

[0087] Un bloc 22 produit à partir des signaux recueillis par les micros une probabilité p de présence de parole. Comme

indiqué plus haut, cette estimation est opérée selon une technique en elle-même connue, par exemple celle décrite dans le WO 2007/099222 A1, auquel on pourra se référer pour plus de détails.

[0088] Le bloc 44 schématise un sélecteur de la méthode d'estimation du canal acoustique, soit par diagonalisation sur la base des signaux recueillis par les quatre micros M_1 , M_2 , M_3 et M_4 (bloc 28 de la Figure 5, pour le haut du spectre HF), soit par filtre adaptatif LMS sur la base des signaux recueillis par les deux micros extrêmes M_1 et M_4 (bloc 38 de la Figure 5, pour le bas du spectre BF).

[0089] Le bloc 46 correspond à l'estimation de la matrice spectrale des bruits, désignée R_n , utilisée pour le calcul du projecteur linéaire optimal, et utilisée également pour le calcul de diagonalisation du bloc 28 lorsque la fonction de transfert du canal acoustique est estimée de cette manière. Le bloc 48 correspond au calcul du projecteur linéaire optimal. Comme on l'a indiqué plus haut, la projection calculée en 48 est une projection linéaire optimale, en ce sens que la composante de bruit résiduel sur le signal monocanal délivré en sortie est minimisée (bruit et réverbération). Comme on l'a également indiqué, le projecteur linéaire optimal présente la particularité de recalculer les phases des différents signaux d'entrée, ce qui permet d'obtenir en sortie un signal projeté S_{pr} qui retrouve la phase du signal initial de parole du locuteur (et également l'amplitude de ce signal, bien entendu).

[0090] L'étape finale (bloc 50) consiste à opérer une réduction sélective du bruit par application d'un gain variable propre à chaque bande de fréquences et à chaque trame temporelle au signal projeté S_{pr} .

[0091] Ce débruitage est également modulé par la probabilité de présence de parole p .

[0092] Le signal $S_{HF/BF}$ délivré en sortie par le bloc 50 de débruitage fera ensuite l'objet d'une transformation de Fourier rapide inverse iFFT (blocs 30, 40 de la Figure 5) pour obtenir dans le domaine temporel le signal de parole débruitée S_{HF} ou S_{BF} recherché donnant, après reconstitution du spectre complet, le signal de parole débruitée final s .

[0093] On peut avantageusement utiliser pour le débruitage du bloc 50 une méthode de type OM-LSA (*Optimally Modified - Log Spectral Amplitude*) telle que celle décrite par la référence précitée :

[2] I. Cohen, Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator, IEEE Signal Processing Letters, Vol. 9, No 4, April 2002.

[0094] 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 original. 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.

[0095] 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 p .

[0096] Dans cette méthode, la probabilité de présence de parole p intervient à deux niveaux importants :

- pour l'estimation de l'énergie du bruit, la probabilité vient moduler le facteur d'oubli dans le sens d'une mise à jour plus rapide de l'estimation du bruit sur le signal bruité lorsque la probabilité de présence de parole est faible ;
- pour le calcul du gain final, elle joue également un grand rôle, car la réduction de bruit appliquée est d'autant plus importante (c'est-à-dire que le gain appliqué est d'autant plus faible) que la probabilité de présence de parole est faible.

Revendications

1. Un procédé de débruitage d'un signal acoustique bruité pour un dispositif audio multi-microphone opérant dans un milieu bruité, notamment un dispositif téléphonique "mains libres", le signal acoustique bruité comprenant une composante utile issue d'une source de parole (S) et une composante parasite de bruit, ledit dispositif comprenant un réseau de capteurs formé d'une pluralité de capteurs microphoniques ($M_1...M_n$) disposés selon une configuration prédéterminée et aptes à recueillir le signal bruité, caractérisé en ce qu'il comporte les étapes de traitement suivantes dans le domaine fréquentiel, pour une pluralité de bandes de fréquences définies pour des trames temporelles successives de signal :

- a) estimation (22) d'une probabilité de présence de parole (p) dans le signal bruité recueilli ;
- b) estimation (46) d'une matrice spectrale de covariance (R_n) des bruits recueillis par les capteurs, cette estimation étant modulée par la probabilité de présence de parole (p) ;
- c) estimation de la fonction de transfert ($H_1...H_n$) des canaux acoustiques entre la source de parole (S) et au

moins certains des capteurs ($M_1 \dots M_n$), cette estimation étant opérée par rapport à une référence de signal utile constituée par le signal recueilli par l'un des capteurs (M_1), et étant en outre modulée par la probabilité de présence de parole (p) ;

d) calcul (48) d'un projecteur linéaire optimal donnant un signal combiné débruité unique à partir des signaux ($X_1 \dots X_n$) recueillis par au moins certains des capteurs, de la matrice spectrale de covariance (R_n) estimée à l'étape b), et des fonctions de transfert ($H_1 \dots H_n$) estimées à l'étape c) ; et

e) à partir de la probabilité de présence de parole (p) et du signal combiné donné par le projecteur calculé à l'étape d), réduction sélective du bruit (50) 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 calcul (48) du projecteur linéaire optimal de l'étape d) est opéré par un traitement de type *beamforming* de Capon à réponse sans distorsion à variance minimale MVDR.

3. Le procédé de la revendication 1, dans lequel la réduction sélective de bruit (50) de l'étape e) est opérée par un traitement de type gain à amplitude log-spectrale modifié optimisé OM-LSA.

4. Le procédé de la revendication 1, dans lequel l'estimation de la fonction de transfert de l'étape c) est opérée par calcul (38) d'un filtre adaptatif (14) visant à annuler la différence entre le signal (X_i) recueilli par le capteur dont on cherche à évaluer la fonction de transfert et le signal (X_j) recueilli par le capteur de ladite référence de signal utile, avec modulation par la probabilité de présence de parole (p).

5. Le procédé de la revendication 4, dans lequel le filtre adaptatif (14) est un filtre à algorithme de prédiction linéaire de type moindres carrés moyens LMS.

6. Le procédé de la revendication 4, dans lequel ladite modulation par la probabilité de présence de parole (p) est une modulation par variation du pas d'itération du filtre adaptatif (14).

7. Le procédé de la revendication 1, dans lequel l'estimation de la fonction de transfert de l'étape c) est opérée par un traitement de diagonalisation (28) comprenant :

c1) la détermination d'une matrice spectrale de corrélation (R_x) des signaux recueillis par les capteurs du réseau par rapport au capteur de ladite référence de signal utile,

c2) le calcul de la différence entre, d'une part, la matrice (R_x) déterminée à l'étape c1) et, d'autre part, ladite matrice spectrale de covariance (R_n) des bruits modulée par la probabilité de présence de parole (p), calculée à l'étape b), et

c3) la diagonalisation de la matrice différence calculée à l'étape c2).

8. Le procédé de la revendication 1, dans lequel :

- le spectre du signal à débruiter est divisé en une pluralité de parties de spectre distinctes (BF, HF),
- les capteurs sont regroupés en une pluralité de sous-réseaux ($M_1 \dots M_4$; M_1, M_4) associés chacun à l'une desdites parties du spectre, et
- le traitement de débruitage est opéré de façon différenciée, pour chacune desdites parties du spectre, sur les signaux recueillis par les capteurs du sous-réseau correspondant à la partie du spectre considérée.

9. Le procédé de la revendication 8, dans lequel :

- le réseau de capteurs est un réseau linéaire de capteurs alignés ($M_1 \dots M_4$),
- le spectre du signal à débruiter est divisé en une partie basse fréquence (BF) et une partie haute fréquence (HF), et
- pour la partie basse fréquence, les étapes du traitement de débruitage sont opérées seulement sur les signaux recueillis par les capteurs les plus éloignés du réseau (M_1, M_4).

10. Le procédé de la revendication 1, dans lequel :

- le spectre du signal à débruiter est divisé en une pluralité de parties de spectre distinctes (BF, HF), et
- l'étape c) d'estimation de la fonction de transfert des canaux acoustiques est opérée de manière différenciée par application de traitements différents (28, 38) pour chacune desdites parties du spectre.

11. Le procédé de la revendication 10, dans lequel :

- le réseau de capteurs est un réseau linéaire de capteurs alignés ($M_1 \dots M_4$),
- les capteurs sont regroupés en une pluralité de sous-réseaux ($M_1 \dots M_4$; M_1, M_4) associés chacun à l'une desdites parties du spectre,
- pour la partie basse fréquence (BF), le traitement de débruitage est opéré seulement sur les signaux recueillis par les capteurs (M_1, M_4) les plus éloignés du réseau, et l'estimation de la fonction de transfert est opérée par calcul d'un filtre adaptatif (38), et
- pour la partie haute fréquence, le traitement de débruitage est opéré sur les signaux recueillis par tous les capteurs du réseau ($M_1 \dots M_4$), et l'estimation de la fonction de transfert est opérée par un traitement de diagonalisation (28).

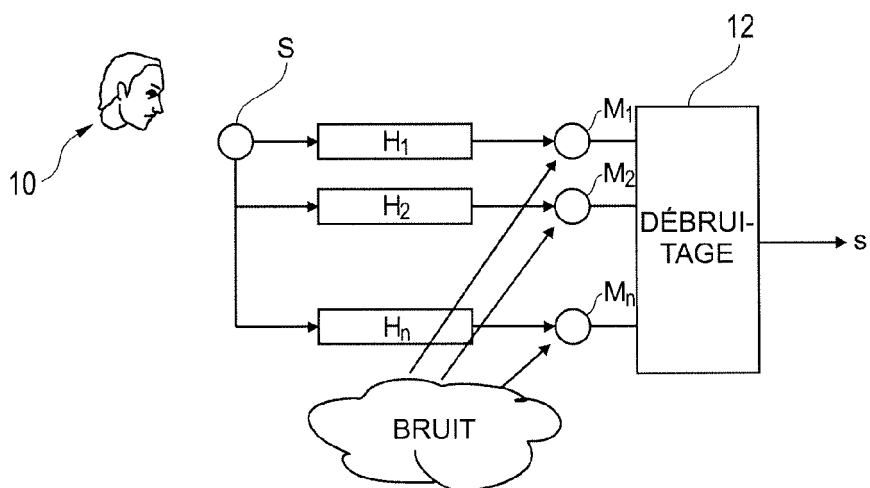


Fig. 1

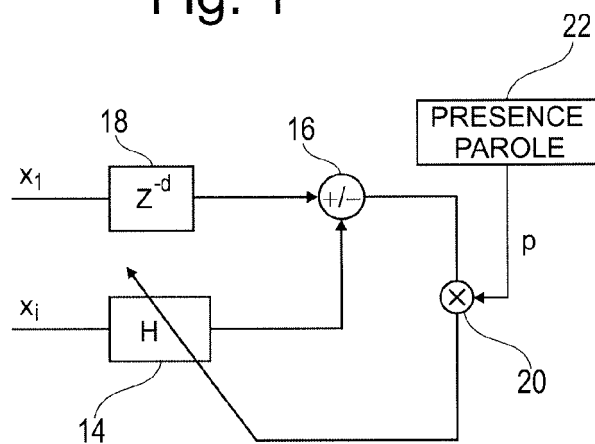


Fig. 2

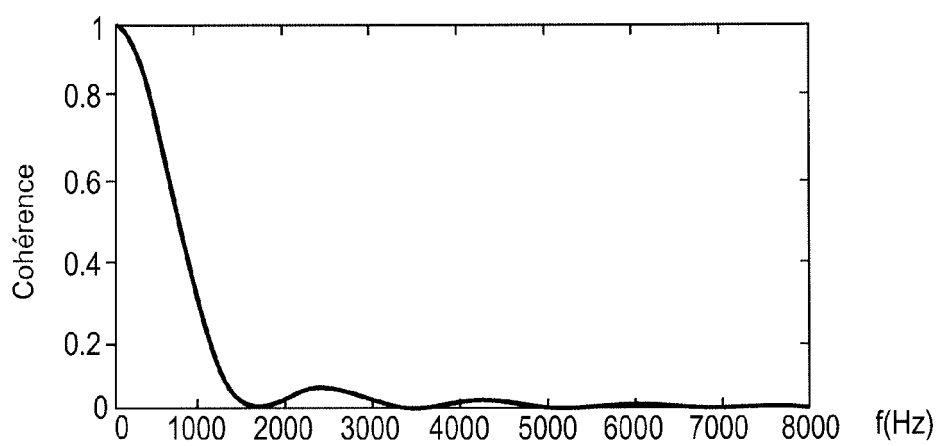


Fig. 3

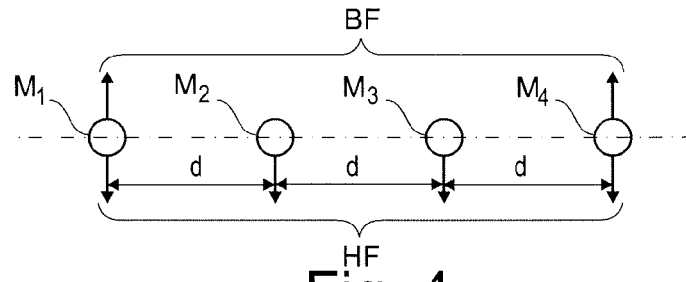


Fig. 4

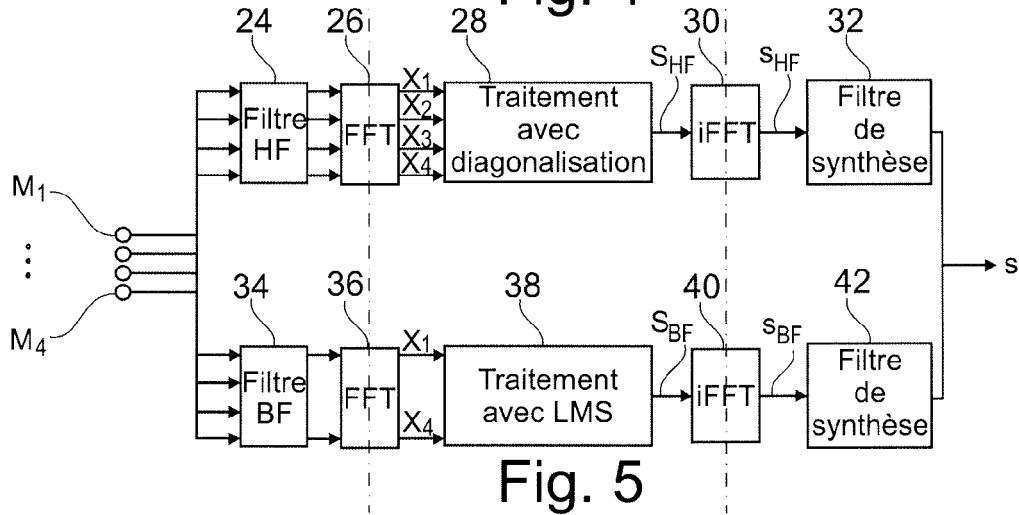


Fig. 5

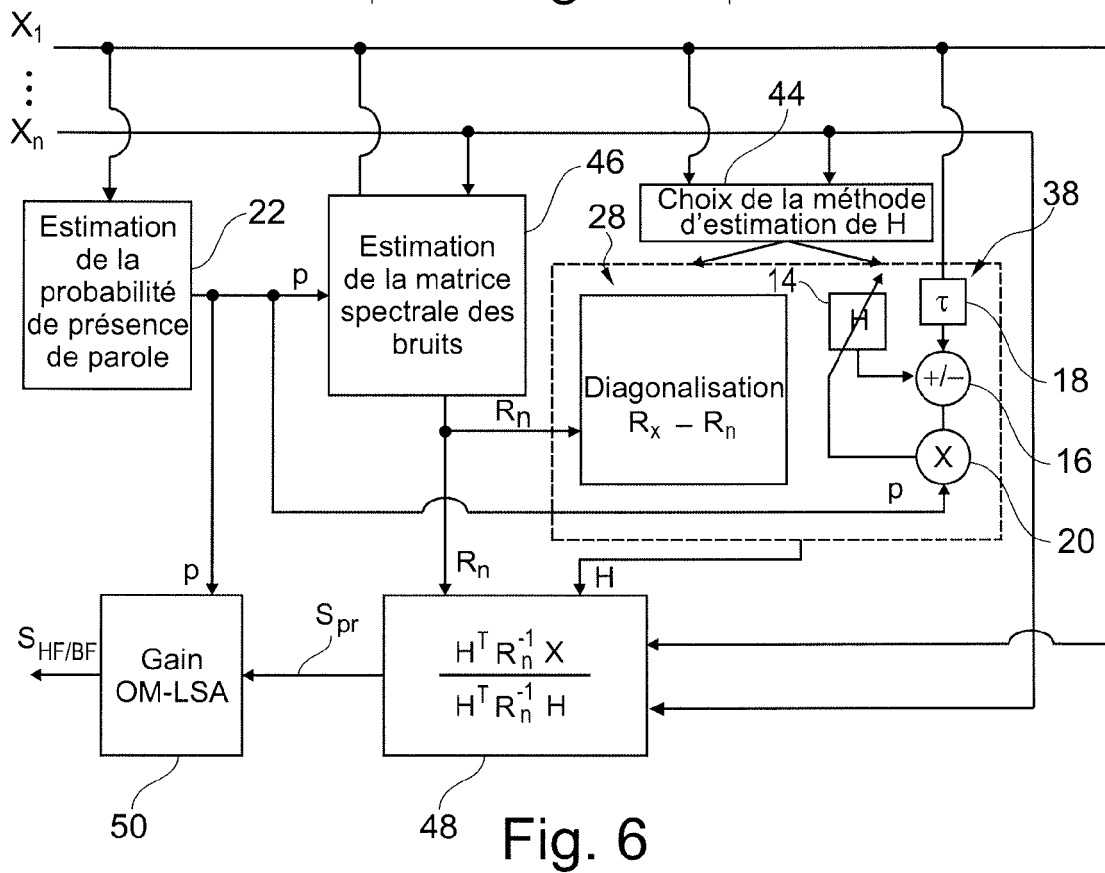


Fig. 6



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande

EP 12 17 0874

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	EP 2 309 499 A1 (PARROT [FR]) 13 avril 2011 (2011-04-13) * figure 1 *	1-11	INV. G10L21/02 H04R3/00
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 *	1-11	
A,D	HENDRIKS R C ET AL: "On Optimal Multichannel Mean-Squared Error Estimators for Speech Enhancement", IEEE SIGNAL PROCESSING LETTERS, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 16, no. 10, 1 octobre 2009 (2009-10-01), pages 885-888, XP011269912, ISSN: 1070-9908, DOI: 10.1109/LSP.2009.2026205 * le document en entier *	1-11	DOMAINES TECHNIQUES RECHERCHES (IPC) G10L H04R
A	US 2004/150558 A1 (LI JIAN [US] ET AL) 5 août 2004 (2004-08-05) * le document en entier *	1-11	
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche Munich		Date d'achèvement de la recherche 1 août 2012	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 (P04C02)

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

EP 12 17 0874

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.

01-08-2012

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP 2309499 A1	13-04-2011	AT 529860 T	15-11-2011
		EP 2309499 A1	13-04-2011
		ES 2375844 T3	06-03-2012
		FR 2950461 A1	25-03-2011
		US 2011070926 A1	24-03-2011

US 2004150558 A1	05-08-2004	AUCUN	

EPO FORM P0480

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

- EP 2293594 A1 [0010]
- EP 2309499 A1 [0011]
- WO 2007099222 A1 [0052] [0087]

Littérature non-brevet citée dans la description

- **R. C. HENDRIKS et al.** On optimal multichannel mean-squared error estimators for speech enhancement. *IEEE Signal Processing Letters*, 2009, vol. 16 (10) [0047]
- **COHEN.** Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator. *IEEE Signal Processing Letters*, Avril 2002, vol. 9 (4), 113-116 [0047]
- **I. COHEN ; B. BERDUGO.** Two-Channel Signal Detection and Speech Enhancement Based on the Transient Beam-to-Reference Ratio. *Proc. ICASSP 2003, Hong-Kong*, Avril 2003, 233-236 [0051]
- **B. WIDROW.** Adaptive Filters, Aspect of Network and System Theory. Rinehart and Winston, 1970, 563-587 [0057]
- Frequency-domain adaptive filtering with applications to acoustic echo cancellation. **J. PRADO ; E. MOULINES.** *Annals of Telecommunications*. Springer, 1994 [0057]
- **B. WIDROW ; S. STEARNS.** Adaptive Signal Processing, Prentice-Hall Signal Processing Series. *Alan V. Oppenheim Series Editor*, 1985 [0057]
- **I. COHEN.** Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator. *IEEE Signal Processing Letters*, Avril 2002, vol. 9 (4) [0093]