

(19)



(11)

**EP 2 293 594 A1**

(12)

**DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:  
**09.03.2011 Bulletin 2011/10**

(51) Int Cl.:  
**H04R 3/00** (2006.01) **H04R 1/40** (2006.01)  
**G10L 21/02** (2006.01)

(21) Numéro de dépôt: **10166119.7**

(22) Date de dépôt: **16.06.2010**

(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**

(72) Inventeurs:  
• **Vitte, Guillaume**  
**75003 Paris (FR)**  
• **Seris, Julie**  
**75013 Paris (FR)**  
• **Pinto, Guillaume**  
**75004 Paris (FR)**

(30) Priorité: **23.07.2009 FR 0955133**

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

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

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

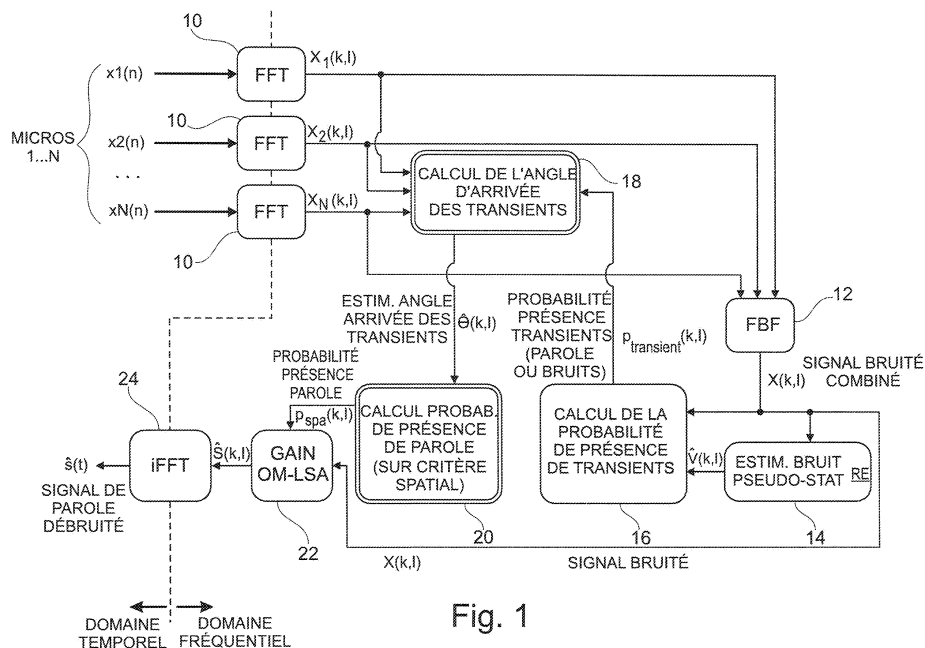
(57) Le procédé comporte dans le domaine fréquentiel les étapes suivantes :

- combinaison (12) des signaux en un signal combiné bruité ( $X(k,l)$ ) ;
- estimation (14) d'une composante de bruit pseudo-stationnaire ( $V(k,l)$ ) ;
- calcul (16) d'une probabilité de présence de transients ( $p_{Transient}(k,l)$ ) dans le signal combiné bruité ;

d) estimation (18) d'une direction principale d'arrivée des transients ( $\theta(k,l)$ ) ;

e) calcul (20) d'une probabilité de présence de parole sur un critère spatial ( $p_{spa}(k,l)$ ), propre à discriminer entre parole utile et bruit latéral parmi les transients ;

f) réduction sélective du bruit (22) 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 microphones sont utilisés.

[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, c'est-à-dire 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] L'une des difficultés 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 l'invention est de mettre à profit la structure multi-microphone du dispositif pour opérer une détection spatiale de ces bruits non stationnaires, puis de discriminer, parmi toutes les composantes non stationnaires (ci-après "transients") celles qui sont des composantes de bruit non stationnaires d'avec celles qui sont des composantes de parole, et enfin de traiter le signal capté pour le débiter de manière efficace tout en minimisant les distorsions introduites par ce traitement.

[0011] Dans la suite, on appellera "bruit latéral" un bruit non stationnaire directif dont la direction d'arrivée 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.

[0012] Le point de départ de l'invention 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, puis pour en déduire une probabilité de présence de la parole qui servira à atténuer ce bruit.

[0013] Plus précisément, l'invention a pour objet un procédé de débruitage d'un signal acoustique bruité capté par une pluralité de microphones d'un dispositif audio multi-microphone opérant dans un milieu bruité. 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.

[0014] Un tel procédé est par exemple divulgué par : 1. Cohen, Analysis of Two-Channel Generalized Sidelobe Canceller (GSC) with Post-Filtering, IEEE Transactions on Speech and Audio Processing, Vol. 11, No 6, November 2003, pp. 684-699.

[0015] Essentiellement, et de façon caractéristique de l'invention, le procédé comporte les étapes suivantes de traitement, exécutées dans le domaine fréquentiel:

- a) combinaison de la pluralité de signaux captés par la pluralité correspondante de microphones en un signal combiné bruité ;
- b) à partir du signal combiné bruité, estimation d'une composante de bruit pseudo-stationnaire contenue dans ce signal combiné bruité ;
- c) à partir de la composante de bruit pseudo-stationnaire estimée à l'étape b) et du signal combiné bruité, calcul

d'une probabilité de présence de transients dans le signal combiné bruité ;

d) à partir de la pluralité de signaux captés par la pluralité correspondante de microphones et de la probabilité de présence de transients calculée à l'étape c), estimation d'une direction principale d'arrivée des transients ;

e) à partir de la direction principale d'arrivée des transients estimée à l'étape d), calcul d'une probabilité de présence de parole sur un critère spatial, propre à discriminer entre parole utile et bruit latéral parmi les transients ;

f) à partir de la probabilité de présence de parole calculée à l'étape e) 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.

**[0016]** Selon diverses formes de mise en oeuvre subsidiaires avantageuses :

- le traitement de l'étape a) est un traitement de préfiltrage de type *fixed beamforming* ;
- le traitement de l'étape d) comprend les sous-étapes successives suivantes : d1) partition de l'espace en une pluralité de secteurs angulaires ; d2) pour chaque secteur, évaluation d'un estimateur de direction d'arrivée à partir de la pluralité de signaux captés par la pluralité correspondante de microphones ; d3) pondération de chaque estimateur par la probabilité de présence de transients calculée à l'étape c) ; d4) à partir des valeurs d'estimateurs pondérées calculées à l'étape d3), estimation d'une direction principale d'arrivée des transients ; et d5) validation ou invalidation de l'estimation de la direction principale d'arrivée des transients opérée à l'étape d4).
- à l'étape d5) l'estimation n'est validée que si la valeur de l'estimateur pondéré correspondant à la direction estimée est supérieure à un seuil prédéterminé, et/ou en l'absence de maximum local de l'estimateur pondéré dans le secteur angulaire d'origine du signal de parole utile, et/ou que si la valeur de l'estimateur est croissante de façon monotone sur une pluralité de trames temporelles successives ;
- le procédé comprend en outre une étape de maintien de l'estimation de la direction principale d'arrivée pendant un laps de temps minimal prédéterminé ;
- la probabilité de présence de parole calculée à l'étape e) est soit une probabilité binaire, prenant une valeur 1 ou 0 selon que la direction principale d'arrivée des transients estimée à l'étape d) est située ou non dans le secteur angulaire d'origine du signal de parole utile, soit une probabilité à valeurs multiples, fonction de l'écart angulaire entre la direction principale d'arrivée des transients estimée à l'étape d) et la direction d'origine du signal de parole utile ;
- le traitement de l'étape f) est un traitement de réduction sélective du bruit par application d'un gain à amplitude log-spectrale modifié optimisé OM-LSA.

**[0017]** On va maintenant décrire un exemple de mise en oeuvre du procédé de l'invention en référence à la figure annexée.

**[0018]** 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.

**[0019]** 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 modules 10 à 24 illustrés Figure 1.

**[0020]** 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.

**[0021]** 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) disposés selon une configuration prédéterminée.

**[0022]** Le réseau de micros capte le signal émis par la source de signal utile (signal de parole), et la différence de position entre les micros induit un ensemble de déphasages et variations d'amplitude dans l'enregistrement des signaux émis par la source de signal utile.

**[0023]** Plus précisément, le micro d'indice  $n$  délivre un signal :

$$x_n(t) = a_n \times s(t - \tau_n) + v_n(t)$$

où  $a_n$  est l'atténuation d'amplitude due à la perte d'énergie entre la position de la source sonore  $s$  et le micro,  $\tau_n$  est le déphasage entre le signal émis et reçu par le micro et  $v_n$  représente la valeur du champ de bruit diffus à la position du micro.

**[0024]** Dans la mesure où la source est éloignée d'au moins quelques centimètres des micros, on pourra faire l'approximation que la source sonore émet une onde plane. Les retards  $\tau_n$  pourront alors être calculés à partir de l'angle

$\theta_S$ , défini comme l'angle entre les médiatrices des couples de micros ( $n, m$ ) et la direction de référence correspondant à la source  $s$  de signal utile. Lorsque le système considéré comporte deux micros dont la médiatrice coupe la source, l'angle  $\theta_S$  est nul.

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

**[0025]** Le signal dans le domaine temporel  $x_n(t)$  issu de chacun des  $N$  micros 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 :

10

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

15 avec :

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

20

$l$  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 fréquence indiquée par  $k$ .

25 Constitution d'un signal combiné partiellement débruité (bloc 12)

**[0026]** Les signaux  $X_n(k, l)$  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 :

30

$$X(k, l) = \frac{1}{N} \sum_{n \in [1, N]} \overline{d_n(k)} \cdot X_n(k, l)$$

35

**[0027]** Il est 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.

**[0028]** Lorsque le système considéré comporte deux micros dont la médiatrice coupe la source, l'angle  $\theta_S$  est nul et il s'agit d'une simple moyenne qui est faite sur les deux microphones.

40

Estimation du bruit pseudo-stationnaire (bloc 14)

**[0029]** Cette étape a pour objet de calculer une estimation de la composante de bruit pseudo-stationnaire  $\hat{V}(k, l)$  présente sur le signal  $X(k, l)$ .

45

**[0030]** Il existe de très nombreuses publications sur ce sujet, l'estimation et la réduction du bruit pseudo-stationnaire étant en effet un problème classique assez bien résolu. Différentes méthodes sont efficaces et utilisables pour obtenir  $\hat{V}(k, l)$ , 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.

50

Calcul de la probabilité de présence des transients (bloc 16)

**[0031]** Les "transients" désignent tous les signaux non-stationnaires, incluant aussi bien la parole utile que les bruits non-stationnaires sporadiques, qui peuvent avoir une énergie équivalente ou parfois supérieure à la parole utile (passage d'un véhicule, sirène, klaxon, parole d'autres personnes etc.).

55

**[0032]** Il est possible de détecter ces transients à l'aide de l'estimation précédemment établie de la composante de bruit pseudo-stationnaire  $\hat{V}(k, l)$ , en retranchant cette dernière du signal global  $X(k, l)$ .

**[0033]** On verra plus loin (description détaillée des blocs 18 et 20) la manière dont il est possible de discriminer parmi ces transients entre ceux qui correspondent à la parole utile et ceux qui correspondent à des bruits non-stationnaires et qui ont des caractéristiques similaires à la parole utile. Le traitement opéré par le bloc 16 consiste seulement à calculer une probabilité  $p_{Transient}(k,l)$  de présence de signaux transients, sans distinction entre parole utile et bruits parasites non-stationnaires. L'algorithme est le suivant :

Pour chaque trame 1 et pour chaque bande de fréquence  $k$ ,

**[0034]**

(i) Calculer le "Transient to Stationary Ratio" :

$$TSR(k,l) = \frac{X(k,l) - \hat{V}(k,l)}{\hat{V}(k,l)}$$

(ii) Si  $TSR(k,l) \leq TSR_{min}$  :

$$p_{Transient}(k,l) = 0$$

(iii) Si  $TSR(k,l) \geq TSR_{max}$  :

$$p_{Transient}(k,l) = 1$$

(iv) Si  $TSR_{min} < TSR(k,l) < TSR_{max}$  :

$$p_{Transient}(k,l) = \frac{TSR(k,l) - TSR_{min}}{TSR_{max} - TSR_{min}}$$

**[0035]** Les constantes  $TSR_{min}$  et  $TSR_{max}$  sont choisies de manière à correspondre à des situations typiques, proches de la réalité.

Calcul de la direction d'arrivée des transients (bloc 18)

**[0036]** Ce calcul tire parti du fait que, à la différence de la composante pseudo-stationnaire du bruit qui est diffuse, les transients sont souvent directifs, c'est-à-dire issus d'une source sonore ponctuelle (comme la bouche du locuteur pour la parole utile, ou le moteur d'une motocyclette pour un bruit latéral). Il est donc judicieux de calculer la direction d'arrivée de ces signaux, qui sera en général bien définie, et de comparer cette direction d'arrivée à l'angle  $\theta_S$  correspondant à la direction d'origine parole utile), de manière à déterminer si le signal non-stationnaire considéré est utile ou parasite, et d'effectuer ainsi la discrimination entre parole utile et bruit non-stationnaire.

**[0037]** La première étape consiste à estimer la direction d'arrivée du transient. La méthode utilisée ici est basée sur l'utilisation de la probabilité de présence des transients  $p_{Transient}(k,l)$  déterminée par le bloc 18 de la manière exposée plus haut.

**[0038]** Plus précisément, on opère une partition de l'espace en secteurs angulaires, chacun correspondant à une direction définie par un angle  $\theta_i, i \in [1, M]$  (par exemple  $M=19$ , avec 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 une dizaine d'angles ( $M=10$ ) avec un seul couple de micros ( $N=2$ ).

**[0039]** Chaque angle  $\theta_i$  est testé de façon à déterminer celui qui est le plus proche de la direction d'arrivée du signal non-stationnaire étudié. Pour ce faire, on considère chaque couple de micros ( $n, m$ ) et on calcule un estimateur de direction d'arrivée  $P_{n,m}(\theta_i, k, l)$  correspondant, dont le module sera maximal lorsque l'angle  $\theta_i$  testé sera le plus proche de la direction d'arrivée du transient.

**[0040]** Cet estimateur peut par exemple s'appuyer sur un calcul d'intercorrélation et prendre la forme :

$$P_{n,m}(\theta_i, k, l) = E\left(X_m(k, l) \cdot \bar{X}_n(k, l) \cdot e^{-i2\pi f_k \tau_i}\right), \text{ avec } \tau_i = \frac{l_{n,m}}{c} \sin \theta_i$$

$l_{n,m}$  étant la distance entre les micros d'indices  $n$  et  $m$ , et  $c$  étant la célérité du son.

**[0041]** Une première méthode, classique, consiste à prendre pour estimation de la direction d'arrivée l'angle qui maximise le module de cet estimateur, soit :

$$\hat{\theta}_{std}(k, l) = \arg \max_{\theta_i, i \in [1, M]} \|P_{n,m}(\theta_i, k, l)\|$$

**[0042]** Une autre méthode, utilisée ici de façon préférentielle, consiste à pondérer l'estimateur  $P_{n,m}(\theta_i, k, l)$  par la probabilité de présence de transients  $p_{Transient}(k, l)$ , et définir une nouvelle stratégie de décision. L'estimateur de direction d'arrivée correspondant sera :

$$P_{New_{n,m}}(\theta_j, k, l) = P_{n,m}(\theta_j, k, l) \times p_{Transient}(k, l)$$

**[0043]** L'estimateur peut être moyenne sur les couples de micros ( $n, m$ ) :

$$P_{New}(\theta_i, k, l) = \frac{1}{N(N-1)} \sum_{n \neq m} P_{New_{n,m}}(\theta_i, k, l)$$

**[0044]** L'intégration de la probabilité de présence de transients dans l'estimateur de direction d'arrivée présente trois avantages importants :

- l'estimation de direction est ciblée sur les parties non-stationnaires du signal (où la probabilité  $p_{Transient}(k, l)$  est proche de 1), dont la direction d'arrivée est bien définie, ce qui rend l'estimation consistante ;
- l'estimation de direction est robuste au bruit diffus (où la probabilité  $p_{Transient}(k, l)$  est proche de zéro), qui d'ordinaire perturbe les estimations de direction d'arrivée ;
- la fiabilité de l'estimateur  $P_{New_{n,m}}(\theta_i, k, l)$  permet de distinguer plusieurs signaux non-stationnaires correspondant à différentes directions et simultanément présents (on verra plus bas que cette distinction peut se faire par bande de fréquences ou par analyse des maxima angulaires locaux sur une même bande de fréquences). Ainsi, si l'on a en même temps un signal de parole utile et un bruit latéral puissant, les deux types de signaux seront détectés, évitant que le signal de parole utile concomitant soit éliminé par erreur dans la suite du processus, même si son énergie est faible.

**[0045]** On va maintenant expliciter les règles de décision permettant à partir de  $P_{New}$ .

- soit de délivrer une estimation  $\hat{\theta}(k, l)$  de la direction d'arrivée du transient,
- soit d'indiquer qu'aucune estimation de direction d'arrivée ne peut être fournie, si ces règles ne sont pas satisfaites.

1°) Significativité de  $P_{New}(\theta_{max}, k, l)$  ( $\theta_{max}$  étant l'angle qui maximise la valeur  $\|P_{New}(\theta_i, k, l)\|$ )

Règle 1 :

**[0046]**

Une estimatin de direction ne peut être fournie que si  $\|P_{New}(\theta_{max}, k, l)\|$  dépasse un seuil donné  $P_{MIN}$ .

**[0047]** Cette première règle permet de s'assurer que sur la partie  $(k, l)$  du signal considéré, la probabilité de présence d'un transient et le niveau d'inter-corrélation sont assez élevés pour que l'estimation soit consistante.

2°) Monotonie de  $P_{New}$  sur l'intervalle  $[\theta_S - \theta_{max}; \theta_{max}]$  (pour alléger les notations, dans la suite les barres de module de  $P_{New}$  seront enlevées)

Règle 2 :

**[0048]** Si  $\theta_{max}$  est en dehors du cône privilégié, une estimation d'angle ne sera validée que si  $P_{new}$  augmente de façon monotone sur l'intervalle  $[\theta_S - \theta_{max}; \theta_{max}]$ .

**[0049]** Cette deuxième règle analyse le contenu du "cône privilégié", correspondant au secteur angulaire sur lequel est centré la source  $s$  et qui présente une étendue angulaire de  $\theta_0$ . Ce cône privilégié est défini par les angles  $\theta$  tels que  $|\theta - \theta_S| \leq \theta_0$ .

**[0050]** Le "bruit latéral" correspondra à un signal dont la direction d'arrivée est extérieure au cône privilégié, et l'on considèrera donc qu'un bruit latéral est présent si  $|\theta_{max} - \theta_S|$  dépasse le seuil  $\theta_0$ .

**[0051]** Pour valider cette détection d'un bruit latéral, il faut vérifier qu'un signal de parole utile ne se trouve pas simultanément à l'entrée du système.

**[0052]** Pour cela,  $P_{New}(\theta_{max}, k, l)$  est confronté aux valeurs de  $P_{New}(\theta_i, k, l)$  obtenues pour d'autres angles, notamment ceux qui appartiennent au cône privilégié. La règle permet ainsi de s'assurer qu'il n'y a pas de maximum local dans le cône privilégié.

3°) Fiabilisation de la détection d'un bruit latéral

Règle 3 :

**[0053]** Si  $\theta_{max}$  est en dehors du cône privilégié pour la première fois sur la trame  $l$  considérée, une estimation d'angle ne sera validée que si :

$$P_{New}(\theta_{max}, k, l) \geq \alpha_1 \times P_{New}(\theta_{max}, k, l-1) ,$$

et si

$$P_{New}(\theta_{max}, k, l) \geq \alpha_2 \times \frac{1}{5} \sum_{i \in [l-5; l-1]} P_{New}(\theta_{max}, k, i) .$$

**[0054]** Si un bruit latéral est détecté, cette troisième règle tient compte des trames précédentes pour éviter les faux déclenchements. Elle ne s'applique qu'à la première trame d'un bruit latéral présumé, et vérifie que  $P_{New}(\theta_{max}, k, l)$  augmente de façon significative par rapport aux données correspondantes obtenues sur les cinq trames précédentes.

**[0055]** Les paramètres  $\alpha_1$  et  $\alpha_2$  sont choisis de manière à correspondre à des situations typiques, proches de la réalité.

**[0056]** Si les trois règles 1 à 3 ci-dessus sont vérifiées, l'estimation  $\hat{\theta}(k, l)$  de la direction d'arrivée sera donnée par :  $\hat{\theta}(k, l) = \theta_{max}$ .

4°) Stabilisation de la détection d'un bruit latéral :

**[0057]** Les deux dernières règles sont destinées à empêcher les coupures dans la détection d'un bruit latéral. Après une période de détection, elles continuent à maintenir cet état pendant un laps de temps dit de *hangover*, quand bien même les règles de décision précédentes ne seraient plus vérifiées. Cela permet de détecter les éventuelles périodes à basse énergie d'un bruit non-stationnaire.

Règle 4 :

**[0058]**

Si  $\hat{\theta}(k, l-1)$  est en dehors du cône privilégié (trame précédente),  
 si  $cpt_1 \leq HangoverTime_1$ , (i. e. la période de Hangover n'est pas terminée),  
 et si  $P_{New}(\hat{\theta}(k, l-1), k, l)$  est supérieur à un seuil donné  $P_1$  alors l'estimation d'angle est maintenue et  $cpt_1$  est incrémenté.

Règle 5 :

[0059]

Si  $\hat{\theta}(k, l-1)$  est en dehors du cône privilégié (trame précédente),

si  $cpt_2 \leq HangoverTime_2$  et si  $\frac{1}{5} \sum_{i \in [l-5; l-1]} P_{New}(\hat{\theta}(k, l-1), k, i)$  est supérieur à un seuil donné  $P_2$  alors l'estimation d'angle est maintenue et  $cpt_2$  est incrémenté.

[0060] Si l'une de ces deux dernières règles (Règle n°4 ou n°5) est vérifiée, elle est prioritaire, et il en résulte :  $\hat{\theta}(k, l) = \hat{\theta}(k, l-1)$ , donc avec correction éventuelle de la valeur de  $\hat{\theta}(k, l)$ , qui ne sera pas égale à  $\theta_{max}$  mais qui sera maintenue à sa valeur précédente.

[0061] En résumé, le calcul de  $\hat{\theta}(k, l)$  suit trois cas possibles :

- (i) si la règle n°4 ou n°5 est vérifiée, alors  $\hat{\theta}(k, l) = \hat{\theta}(k, l-1)$ ;
- (ii) dans le cas contraire (ni la règle n°4, ni la règle n°5 n'est vérifiée), si les règles n°1, n°2 et n°3 sont vérifiées, alors  $\hat{\theta}(k, l) = \theta_{max}$  ;
- (iii) sinon (ni la règle n°4, ni la règle n°5 n'est vérifiée, et l'une au moins des règles n°1, n°2 et n°3 n'est pas vérifiée), alors  $\hat{\theta}(k, l)$  n'est pas défini.

[0062] Dans une variante, l'estimateur  $P_{New}$  est moyenné sur des paquets de bandes de fréquences  $K_1, K_2, \dots, K_p$  :

$$P_{New}(\theta_i, K_j, l) = \frac{1}{N(N-1)} \frac{1}{C_j} \sum_{n \neq m} \left[ \sum_{k \in K_j} P_{New_{n,m}}(\theta_i, k, l) \right]$$

[0063]  $C_j$  désignant le cardinal de  $K_j$ .

[0064] Dans ce cas, l'estimation d'angle  $\theta_{max}$  n'est pas faite sur chaque bande de fréquences, mais sur chaque paquet  $K_j$  de bandes de fréquences.

[0065] On notera aussi qu'une approche "pleine bande" est possible ( $p=1$ , un seul angle étant estimé par trame).

[0066] On notera enfin que la méthode proposée est compatible avec l'utilisation de micros unidirectionnels. Dans ce cas il sera courant d'utiliser un réseau linéaire (micros alignés et dont les directions privilégiées sont identiques) et orienté vers le locuteur. Dans ce cas la valeur de  $\theta_S$  est donc naturellement connue et égale à zéro.

Calcul d'une probabilité de présence de parole sur critère spatial (bloc 20)

[0067] L'étape suivante, caractéristique du procédé de l'invention, consiste à calculer une probabilité de présence de parole basée sur l'estimation de direction d'arrivée  $\hat{\theta}(k', l)$  obtenue de la manière indiquée ci-dessus.

[0068] Il s'agit d'une probabilité, notée  $p_{spa}(k, l)$ , qui a donc pour originalité d'être calculée sur un critère spatial (à partir de  $\hat{\theta}(k, l)$ ), et qui permettra de distinguer parmi les signaux non-stationnaires la parole utile des bruits parasites. Cette probabilité sera ensuite utilisée dans une structure classique de débruitage (bloc 22, décrit ci-après).

[0069] La probabilité  $p_{spa}(k, l)$  peut être calculée de différentes manières, donnant une valeur binaire ou bien des valeurs multiples. On donnera ci-dessous deux exemples de calcul  $p_{spa}(k, l)$ , sachant que d'autres lois peuvent être utilisées pour exprimer  $p_{spa}(k, l)$  à partir de  $\hat{\theta}(k, l)$ .

1°) Calcul d'une probabilité  $p_{spa}(k, l)$  binaire :

[0070]

La probabilité de présence de parole prendra les valeurs '0' ou '1' :

- elle sera mise à '0' lorsqu'un bruit latéral, c'est-à-dire un transient provenant d'une direction extérieure au cône privilégié, est détecté ;
- elle sera mise à '1' lorsque la direction d'arrivée du transient est à l'intérieur du cône privilégié, ou lorsqu'aucune estimation fiable n'a pu être faite sur cette direction.

**[0071]** L'algorithme correspondant est le suivant :

Si  $\hat{\theta}(k,l)$  est à l'intérieur du cône privilégié ( $|\hat{\theta}(k,l) - \theta_S| \leq \theta_0$ ),  
 alors  $p_{spa}(k,l) = 1$   
 Si  $\hat{\theta}(k,l)$  est à l'extérieur du cône privilégié ( $|\hat{\theta}(k,l) - \theta_S| > \theta_0$ ),  
 alors  $p_{spa}(k,l) = 0$   
 Si  $\hat{\theta}(k,l)$  n'est pas défini,  
 alors  $p_{spa}(k,l) = 1$

2°) Calcul d'une probabilité  $p_{spa}(k,l)$  à valeurs continues dans  $[0;1]$  :

**[0072]** Il est possible d'utiliser pour  $p_{spa}(k,l)$  un calcul progressif, par exemple selon l'algorithme suivant :

Si  $\hat{\theta}(k,l)$  est à l'intérieur du cône privilégié ( $|\hat{\theta}(k,l) - \theta_S| \leq \theta_0$ ),  
 alors  $p_{spa}(k,l) = 1$   
 Si  $\hat{\theta}(k,l)$  est à l'extérieur du cône privilégié ( $|\hat{\theta}(k,l) - \theta_S| > \theta_0$ ),

$$alors \quad p_{spa}(k,l) = 1 - \frac{|\hat{\theta}(k,l) - \theta_0|}{\frac{\pi}{2} - \theta_0}$$

Si  $\hat{\theta}(k,l)$  n'est pas défini,  
 alors  $p_{spa}(k,l) = 1$

Réduction de bruit latéral (bloc 22)

**[0073]** La probabilité  $p_{spa}(k,l)$  de présence de parole calculée au bloc 20, dépendant elle-même de la probabilité  $p_{Transient}(k,l)$  de présence de transients calculée au bloc 16, va être utilisée comme paramètre d'entrée dans une technique classique de débruitage.

**[0074]** On sait que la probabilité de présence de parole est un estimateur crucial pour le bon fonctionnement d'un algorithme de débruitage, car elle sous-tend la bonne estimation du bruit et le calcul d'un gain optimal efficace. 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.

**[0075]** 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.

**[0076]** 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.

**[0077]** Dans cette méthode, la probabilité de présence 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  $p_{spa}(k,l)$  sera utilisée à ces deux niveaux.

**[0078]** 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 :

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

**[0079]** On peut noter ici que la probabilité  $p_{spa}(k, l)$  module 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 de parole est faible, ce mécanisme conditionnant entièrement la qualité de  $\lambda_{Bruit}(k, l)$ .

**[0080]** Le gain de débruitage  $G_{OM-LSA}(k, l)$  est donné par :

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

**[0081]**  $G_{H1}(k, l)$  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.

**[0082]** On note ici que la probabilité  $p_{spa}(k, l)$  joue 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.

**[0083]** Le signal débruité  $\hat{S}(k, l)$  en sortie du bloc 22 est donné par :

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

**[0084]** 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 la présente invention est d'éliminer efficacement ces bruits non-stationnaires.

**[0085]** Par ailleurs, il est possible d'utiliser dans les expressions ci-dessus une probabilité de présence de parole hybride  $p_{hybrid}(k, l)$ , c'est-à-dire calculée à l'aide de  $p_{spa}(k, l)$  combinée à une autre probabilité de présence de parole  $p(k, l)$ , par exemple calculée selon la méthode décrite dans le WO 2007/099222 A1 (Parrot SA). Il vient :

$$p_{hybrid}(k, l) = \min(p(k, l), p_{spa}(k, l))$$

**[0086]** Cette probabilité hybride permet de bénéficier du repérage des bruits non-stationnaires associé aux petites valeurs de  $p_{spa}(k, l)$ , et de compléter l'estimation de la probabilité  $p_{hybrid}(k, l)$  sur les parties  $(k, l)$  où l'estimation de direction d'arrivée  $\theta(k, l)$  n'a pas été définie (produisant une probabilité  $p_{spa}(k, l)$  forcée à la valeur 1 par sécurité).

**[0087]** La probabilité hybride  $p_{hybrid}(k, l)$  intègre ainsi à la fois les bruits non-stationnaires détectés par  $p_{spa}(k, l)$  et les autres bruits (par exemple pseudo-stationnaires) détectés par  $p(k, l)$ .

*Reconstitution temporelle du signal (bloc 24)*

**[0088]** 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  $\hat{s}(t)$ .

## Revendications

1. Un procédé de débruitage d'un signal acoustique bruité capté par une pluralité de 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) combinaison (12) de la pluralité de signaux captés par la pluralité correspondante de microphones en un signal combiné bruité ( $X(k,l)$ ) ;  
b) à partir du signal combiné bruité, estimation (14) d'une composante de bruit pseudo-stationnaire ( $\hat{V}(k,l)$ ) contenue dans ce signal combiné bruité ;  
c) à partir de la composante de bruit pseudo-stationnaire estimée à l'étape b) et du signal combiné bruité, calcul (16) d'une probabilité de présence de transients ( $P_{Transient}(k,l)$ ) dans le signal combiné bruité ;  
d) à partir de la pluralité de signaux captés par la pluralité correspondante de microphones et de la probabilité de présence de transients calculée à l'étape c), estimation (18) d'une direction principale d'arrivée des transients ( $\hat{\theta}(k,l)$ ) ;  
e) à partir de la direction principale d'arrivée des transients estimée à l'étape d), calcul (20) d'une probabilité de présence de parole sur un critère spatial ( $p_{spa}(k,l)$ ), propre à discriminer entre parole utile et bruit latéral parmi les transients ;  
f) à partir de la probabilité de présence de parole calculée à l'étape e) et du signal combiné bruité, réduction sélective du bruit (22) 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 traitement de l'étape a) est un traitement de préfiltrage de type *fixed beamforming*.

3. Le procédé de la revendication 1, dans lequel le traitement de l'étape d) comprend les sous-étapes successives suivantes :

d1) partition de l'espace en une pluralité de secteurs angulaires ;  
d2) pour chaque secteur, évaluation d'un estimateur de direction d'arrivée à partir de la pluralité de signaux captés par la pluralité correspondante de microphones ;  
d3) pondération de chaque estimateur par la probabilité de présence de transients calculée à l'étape c) ;  
d4) à partir des valeurs d'estimateurs pondérées calculées à l'étape d3), estimation d'une direction principale d'arrivée des transients ;  
d5) validation ou invalidation de l'estimation de la direction principale d'arrivée des transients opérée à l'étape d4).

4. Le procédé de la revendication 3, dans lequel à l'étape d5) l'estimation n'est validée que si la valeur de l'estimateur pondéré correspondant à la direction estimée est supérieure à un seuil prédéterminé.

5. Le procédé de la revendication 3, dans lequel à l'étape d5) l'estimation n'est validée qu'en l'absence de maximum local de l'estimateur pondéré dans le secteur angulaire d'origine du signal de parole utile.

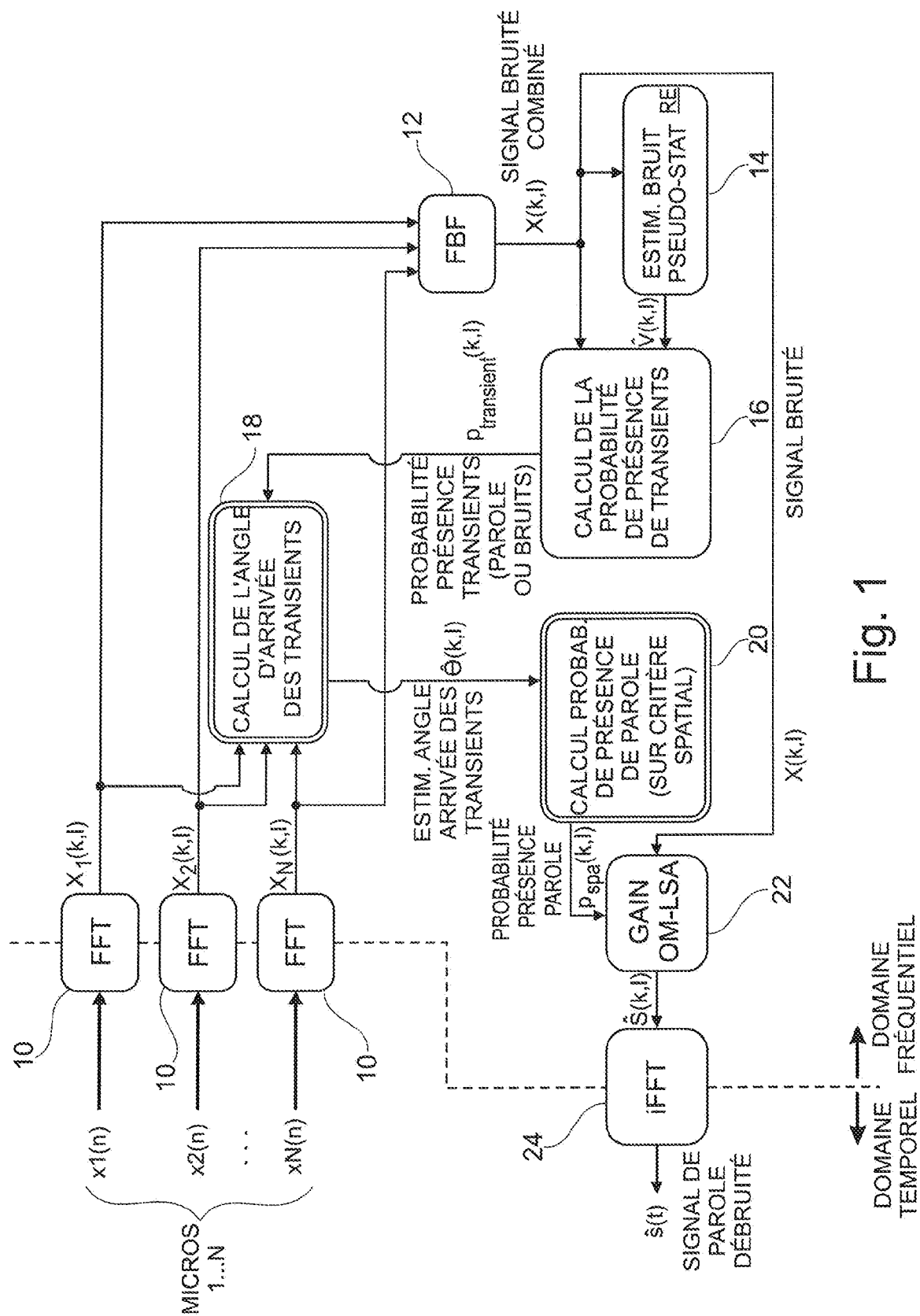
6. Le procédé de la revendication 3, dans lequel à l'étape d5) l'estimation n'est validée que si la valeur de l'estimateur est croissante de façon monotone sur une pluralité de trames temporelles successives.

7. Le procédé de la revendication 3, comprenant en outre une étape de maintien de l'estimation de la direction principale d'arrivée pendant un laps de temps minimal prédéterminé.

8. Le procédé de la revendication 1, dans lequel la probabilité de présence de parole calculée à l'étape e) est une probabilité binaire, prenant une valeur 1 ou 0 selon que la direction principale d'arrivée des transients estimée à l'étape d) est située ou non dans le secteur angulaire d'origine du signal de parole utile.

9. Le procédé de la revendication 1, dans lequel la probabilité de présence de parole calculée à l'étape e) est une probabilité à valeurs multiples, fonction de l'écart angulaire entre la direction principale d'arrivée des transients estimée à l'étape d) et la direction d'origine du signal de parole utile.

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





## RAPPORT DE RECHERCHE EUROPEENNE

 Numéro de la demande  
EP 10 16 6119

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	COHEN I: "Analysis of two-channel generalized sidelobe canceller (GSC) with post-filtering" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 11, no. 6, 1 novembre 2003 (2003-11-01), pages 684-699, XP011104541 ISSN: 1063-6676 * abrégé; figures 1, 2 *	1-10	INV. H04R3/00 H04R1/40 G10L21/02
A	EP 1 473 964 A2 (SAMSUNG ELECTRONICS CO LTD [KR]) 3 novembre 2004 (2004-11-03) * abrégé * * alinéas [0046], [0048]; figures 6,8 *	1-10	
A	WO 02/32356 A1 (LEAR CORP [US]) 25 avril 2002 (2002-04-25) * abrégé; figures 2,5,21-25 * * page 48, ligne 21 - page 52, ligne 12 *	1-10	
			DOMAINES TECHNIQUES RECHERCHES (IPC)
			H04R G10L
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche Munich		Date d'achèvement de la recherche 15 septembre 2010	Examineur Zimmermann, Elko
CATEGORIE DES DOCUMENTS CITES		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	
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			

 1  
EPO FORM 1503 03.82 (P04C02)

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

EP 10 16 6119

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.

15-09-2010

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP 1473964	A2	03-11-2004	JP 4248445 B2	02-04-2009
			JP 2004334218 A	25-11-2004
			US 2004220800 A1	04-11-2004
-----				
WO 0232356	A1	25-04-2002	AU 2441302 A	29-04-2002
-----				

EPO FORM P0460

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 [0085]

### Littérature non-brevet citée dans la description

- **Cohen.** Analysis of Two-Channel Generalized Sidelobe Canceller (GSC) with Post-Filtering. *IEEE Transactions on Speech and Audio Processing*, Novembre 2003, vol. 11 (6), 684-699 [0014]
- **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 [0030]
- **I. Cohen.** Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator. *IEEE Signal Processing Letters*, Avril 2002, vol. 9 (4) [0074]