

12

DEMANDE DE BREVET D'INVENTION

A1

22 Date de dépôt : 01.03.06.

30 Priorité :

43 Date de mise à la disposition du public de la
demande : 07.09.07 Bulletin 07/36.

56 Liste des documents cités dans le rapport de
recherche préliminaire : *Se reporter à la fin du
présent fascicule*

60 Références à d'autres documents nationaux
apparentés :

71 Demandeur(s) : PARROT Société anonyme — FR.

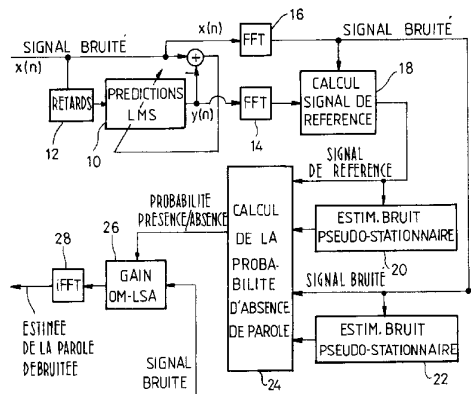
72 Inventeur(s) : PINTO GUILLAUME.

73 Titulaire(s) :

74 Mandataire(s) : SEP PAGENBERG & ASSOCIES.

54 PROCÉDE DE DEBRUITAGE D'UN SIGNAL AUDIO.

57 Ce procédé est un procédé d'analyse de cohérence temporelle du signal bruité comprenant les étapes consistant à: a) déterminer, à partir du signal bruité, un signal de référence en appliquant au signal bruité un traitement (10, 18) propre à atténuer de façon plus importante les composantes de parole que la composante de bruit, notamment au moyen d'un algorithme prédictif récursif adaptatif de type LMS; b) déterminer (24) une probabilité de présence/absence de parole à partir des niveaux d'énergie respectifs dans le domaine spectral du signal bruité et du signal de référence, et c) dériver (26) du signal bruité une estimée débruitée du signal de parole en fonction de la probabilité de présence/absence de parole ainsi déterminée.



La présente invention concerne le débruitage des signaux audio captés par un microphone dans un environnement bruité.

L'invention s'applique avantageusement, mais de façon non limitative, aux signaux de parole captés par les appareils téléphoniques de type "mains-libres" ou analogues.

Ces appareils comportent un microphone 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 œuvre des techniques de reconnaissance vocale, où il est très difficile d'opérer une reconnaissance de forme sur des mots noyés dans un niveau de bruit élevé.

Cette difficulté liée au bruit ambiant est particulièrement contraignante dans le cas des dispositifs "mains-libres" pour véhicules automobiles. En

particulier, la distance importante entre le microphone 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.

Diverses techniques ont été proposées pour réduire le niveau de bruit du signal capté par un microphone.

Le WO-A-98/45997 (Parrot SA) prévoit ainsi d'évaluer et mémoriser un niveau de bruit courant capté par le microphone en l'absence du signal de parole puis, en présence d'un signal de parole, le niveau de bruit ainsi évalué est soustrait du signal capté. Plus précisément, le bruit est capté en permanence par le microphone et analysé pour donner un spectre énergétique moyen pondéré du bruit, continûment réactualisé. Lorsque

l'on souhaite opérer une reconnaissance vocale, le dispositif analyse le signal capté par le microphone dans le domaine fréquentiel et soustrait du spectre de puissance ainsi obtenu le spectre de puissance du bruit antérieurement déterminé et mémorisé. Ce dispositif utilise l'appui sur le bouton-poussoir d'activation du téléphone (par exemple lorsque le conducteur veut répondre à un appel entrant) pour détecter le début du signal de pa-

role et considérer que le signal capté antérieurement à cet appui était essentiellement un signal de bruit.

On connaît également des mécanismes de filtrage du bruit du type dit *beamforming* ou *double-phoning* mettant en œuvre deux microphones distants, dont l'un est placé et orienté pour capter principalement la voix du locuteur, tandis que l'autre capte une composante de bruit plus importante. La comparaison des signaux captés permet d'extraire de manière efficace la voix du bruit ambiant, par des moyens logiciels relativement simples. Cette technique, fondée sur une analyse de cohérence spatiale de deux signaux, présente cependant l'inconvénient de nécessiter deux microphones distants, ce qui la cantonne généralement à des installations fixes ou semi-fixes et ne permet pas de l'intégrer à un dispositif préexistant par simple adjonction d'un module logiciel. Elle présuppose aussi que la position du locuteur par rapport aux deux microphones soit à peu près constante, ce qui est généralement le cas dans un téléphone de voiture utilisé par son conducteur. Mais cette contrainte, comme la précédente, exclut de généraliser cette technique à toutes les situations où l'on aurait besoin d'assurer un débruitage énergique d'un signal de parole noyé dans un bruit de nature complexe et imprévisible.

Une part importante des méthodes les plus efficaces dans le cas du système comportant un seul microphone enregistrant un signal de voix bruité se fondent sur le modèle statistique établi par D. Malah et Y. Ephraïm dans :

[1] Y. Ephraïm et D. Malah, *Speech Enhancement using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No 6, pp. 1109-1121, Dec. 1984, et

[2] Y. Ephraïm et D. Malah, *Speech Enhancement using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-33, No 2, pp. 443-445, April 1985.

Faisant l'approximation que la parole et le bruit sont des processus gaussiens non corrélés et présupposant que la puissance spectrale du bruit soit une donnée connue, ces deux articles donnent une solution optimale au problème de réduction de bruit décrit plus haut. Cette solution propose

de découper le signal bruité en composantes fréquentielles indépendantes par l'utilisation de la transformée de Fourier discrète, d'appliquer un gain optimal sur chacune de ces composantes puis de recombinaer le signal ainsi traité. Les deux articles divergent sur le choix du critère d'optimalité. Dans [1], le gain appliqué est nommé *gain STSA* et permet de minimiser la distance quadratique moyenne entre le signal estimé (à la sortie de l'algorithme) et le signal de parole original (non bruité). Dans [2], l'application d'un gain nommé *gain LSA* permet quant à elle 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 bien 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 bruitées 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. Bien que séduisant puisque soutenu par une démonstration mathématique rigoureuse, ce procédé ne peut toutefois pas être mis en œuvre tout seul. En effet, comme indiqué plus haut, la puissance spectrale du bruit est inconnue et imprévisible *ex ante*. De plus, ce même procédé ne propose pas d'évaluer à quels moments la parole du locuteur est présente dans le signal capté. Il se contente simplement de supposer soit que la parole est toujours présente, soit qu'elle est présente une portion fixe du temps, ce qui peut limiter sérieusement la qualité de la réduction de bruit. Il est donc nécessaire d'utiliser un autre algorithme ayant pour fonction d'évaluer la puissance spectrale du bruit ainsi que les instants où la parole du locuteur est présente sur le signal brut capté. Il s'avère même que cette estimation constitue le facteur déterminant de la qualité de la réduction de bruit opérée, l'algorithme d'Ephraim et Malah n'étant que la manière optimale d'utiliser l'information ainsi obtenue. C'est une solution originale à ce double problème d'évaluation du bruit et des instants de présence du signal de parole qu'apporte la présente invention. Ces deux questions sont en réalité intrinsèquement liées. En effet supposons que le signal brut capté est découpé en trames de longueurs égales,

dont on calcule pour chacune la transformée de Fourier à court terme. Pour une composante fréquentielle donnée, la connaissance des indices de trames où la parole est absente permet d'évaluer la puissance du bruit ainsi que son évolution au cours du temps sur ce segment du spectre. Il suffit en effet de mesurer l'énergie du signal brut lorsque la parole est absente et de faire une moyenne continuellement mise à jour de ses mesures. La question principale est donc de savoir quand exactement la parole du locuteur est absente du signal capté par le microphone.

Si le bruit est stationnaire ou pseudo-stationnaire, ce problème peut être aisément résolu en déclarant que la parole est absente dans un segment de spectre d'une trame donnée lorsque l'énergie spectrale des données pour ce segment de spectre n'a pas évolué ou a peu évolué par rapport aux dernières trames. Inversement, on déclare que la parole est présente en cas de comportement non stationnaire.

Toutefois, dans un environnement réel, *a fortiori* un environnement automobile dont on a indiqué plus haut que le bruit comportait de nombreuses caractéristiques spectrales non stationnaires, ce procédé est aisément pris en défaut, dans la mesure où aussi bien la parole que le bruit peuvent présenter des comportements transitoires. Or, si l'on décide de conserver toutes les composantes transitoires, il restera du bruit musical résiduel dans les données débruitées ; inversement, si l'on décide de supprimer les composantes transitoires en deçà d'un seuil énergétique donné, les composantes faibles de la parole seront alors effacées, alors que ces composantes peuvent être importantes, tant pour leur contenu informatif que pour l'intelligibilité générale (faible distorsion) du signal débruité restitué après traitement.

À cet égard, diverses méthodes ont été proposées. Parmi les plus efficaces, on peut citer celle décrite par :

[3] I. Cohen et B. Berdugo, *Speech Enhancement for Non-Stationary Noise Environments*, Signal Processing, Elsevier, Vol. 81, pp. 2403-2418, 2001,

Comme fréquemment dans le domaine, le procédé décrit dans cet article n'a pas pour objectif d'identifier précisément sur quelles composantes fréquentielles de quelles trames la parole est absente mais plutôt de donner un indice de confiance entre 0 et 1, une valeur 1 indiquant que la parole

est absente à coup sûr (selon l'algorithme) tandis qu'une valeur 0 déclare le contraire. De par sa nature, cet indice est assimilé à la probabilité d'absence de la parole *a priori*, c'est à dire la probabilité que la parole soit absente sur une composante fréquentielle donnée de la trame considérée. Il s'agit bien sûr d'une assimilation non rigoureuse dans le sens que même si la présence de la parole soit probabiliste ex-ante, le signal capté par le microphone ne peut à chaque instant que passer par deux états distincts. Il peut soit (à l'instant considéré) comporter de la parole soit ne pas en contenir. Toutefois cette assimilation donne de bons résultats en pratique ce qui justifie son utilisation. Afin d'estimer cette probabilité d'absence, Cohen et Berdugo utilisent des moyennes sur des rapports signal à bruit *a priori* eux mêmes utilisés et calculés dans l'algorithme d'Ephraïm et Malah. Ces auteurs décrivent également la technique dite de gain OM-LSA (*Optimally-Modified Log-Spectral Amplitude*), visant à améliorer le gain LSA par l'intégration de cette probabilité d'absence de la parole. Cette estimation de la probabilité *a priori* d'absence de la parole se révèle efficace, mais dépend directement du modèle statistique élaboré par Ephraïm et Malah et non d'une connaissance *a priori* des données. Pour obtenir une estimée de la probabilité d'absence qui soit indépendante de ce modèle statistique, Cohen et Berdugo ont proposé dans :

[4] I. Cohen And 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, April 2003,

de calculer la probabilité d'absence à partir de signaux captés par deux microphones différemment placés, donnant des signaux respectifs sur deux voies différentes, dont la combinaison permet d'obtenir une voie dite de sortie et une voie dite bruit de référence. L'analyse est basée sur la constatation que les composantes de parole sont relativement plus faibles sur la voie de bruit de référence, et que les composantes de bruit transitoire présentent à peu près la même énergie sur les deux voies. Une probabilité de présence de parole pour chaque segment de spectre de chaque trame est déterminée en calculant un ratio d'énergie entre les composantes non stationnaires des signaux respectifs des deux voies.

Mais, comme pour les techniques de *beamforming* ou *double-phoning* évoquées plus haut, ce procédé est assez contraignant dans la mesure où il nécessite deux microphones.

5 L'un des buts de l'invention est de remédier aux inconvénients des méthodes proposées jusqu'à présent, grâce à un procédé perfectionné de débruitage applicable à un signal de parole considéré isolément, notamment un signal capté par un microphone unique, procédé qui soit basé sur l'analyse de la cohérence temporelle des signaux captés.

10 Le point de départ de l'invention réside dans la constatation que la parole présente généralement une cohérence temporelle supérieure au bruit et que, de ce fait, elle est nettement plus prédictible. Essentiellement, l'invention propose d'utiliser cette propriété pour calculer un signal de référence où la parole aura été plus atténuée que le bruit, en appliquant notamment un algorithme prédictif qui pourra par exemple être de type LMS
15 (*Least Mean Squares*, moindres carrés moyens). Ce signal de référence dérivé du signal de parole à débruiter pourra être utilisé de façon comparable à celle du signal du second microphone des techniques de *beamforming* à deux voies, par exemple des techniques semblables à celles de Cohen et Berdugo [4, précité]. Le calcul d'un ratio entre les niveaux
20 d'énergie respectifs du signal originel et du signal de référence ainsi obtenu permettra de discriminer entre les composantes de parole et les bruits parasites non stationnaires, et fournira une estimation de la probabilité de présence de parole de façon indépendante de tout modèle statistique.

25 Plus précisément, pour le débruitage d'un signal audio bruité comportant une composante de parole combinée à une composante de bruit, cette composante de bruit comprenant elle-même une composante de bruit transitoire et une composante de bruit pseudo-stationnaire, la présente invention propose d'opérer une analyse de cohérence temporelle du signal bruité par les étapes consistant à :

- 30 a) déterminer, à partir du signal bruité, un signal de référence en appliquant au signal bruité un traitement propre à atténuer de façon plus importante les composantes de parole que les composantes de bruit,
b) déterminer une probabilité de présence/absence de parole *a priori* à partir des niveaux d'énergie respectifs dans le domaine spectral du signal
35 bruité et du signal de référence, et

- c) utiliser cette probabilité d'absence de parole *a priori* pour estimer un spectre de bruit et dériver (26) du signal bruité une estimée débruitée du signal de parole.

5 L'étape a) comprend avantageusement l'application d'un algorithme prédictif, notamment un algorithme adaptatif récursif de type moindres carrés moyens LMS.

10 L'étape b) comprend avantageusement l'application d'un algorithme d'estimation de l'énergie de la composante de bruit pseudo-stationnaire dans le signal de référence et dans le signal bruité, notamment un algorithme de type à moyennage récursif par contrôle des minima MRCA comme décrit dans :

[5] 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,

15 L'étape c) comprend avantageusement l'application d'un algorithme de gain variable fonction de la probabilité de présence/absence de parole, notamment un algorithme de type gain à amplitude log-spectrale modifié optimisé OM-LSA.

20 ◇

25 On va maintenant décrire un exemple de mise en œuvre 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 un diagramme schématique illustrant les différentes opérations effectuées par un algorithme de débruitage conformément au procédé de l'invention.

30 La figure 2 est un diagramme schématique illustrant plus particulièrement l'algorithme prédictif LMS adaptatif.

◇

Le signal que l'on souhaite débruiteur est un signal numérique échantillonné $x(n)$, où n désigne le numéro de l'échantillon (n est donc la variable temporelle).

Le signal capté $x(n)$ est une combinaison d'un signal de parole $s(n)$ et d'un
5 bruit surajouté, non corrélé, $d(n)$:

$$x(n) = s(n) + d(n)$$

Ce bruit $d(n)$ a deux composantes indépendantes, à savoir une compo-
10 sante transitoire $d_t(n)$ et une composante pseudo-stationnaire $d_{ps}(n)$:

$$d(n) = d_t(n) + d_{ps}(n)$$

Comme illustré sur la figure 1, le signal bruité $x(n)$ est appliqué en entrée
15 d'un algorithme LMS prédictif schématisé par le bloc 10, incluant l'application de retards appropriés 12. Le fonctionnement de cet algorithme LMS sera décrit plus bas, en référence à la figure 2.

On calcule ensuite la transformée de Fourier à court terme du signal capté $x(n)$ (bloc 16), ainsi que du signal $y(n)$ délivré par l'algorithme LMS prédictif (bloc 14). À partir de ces deux transformées est calculé un signal de référence (bloc 18), qui constitue l'une des variables d'entrée d'un algorithme de calcul de la probabilité d'absence de parole (bloc 24). Parallèlement, la transformée du signal bruité $x(n)$, issue du bloc 16, est également appliquée à l'algorithme de calcul de probabilité.

25 Les blocs 20 et 22 estiment le bruit pseudo-stationnaire du signal de référence et de la transformée du signal bruité est estimé, et le résultat est également appliqué à l'algorithme de calcul de probabilité.

Le résultat du calcul de probabilité d'absence de parole, ainsi que la transformée du signal bruité, sont appliqués en entrée d'un algorithme de traitement de gain OM-LSA (bloc 26), dont le résultat est soumis à une transformation inverse de Fourier (bloc 28) pour donner une estimée de la parole débruitée.

On va maintenant décrire plus en détail les différentes phases de ce traitement.

35 L'algorithme prédictif LMS (bloc 10) est schématisé sur la figure 2.

Dans la mesure où les signaux en présence sont globalement non stationnaires mais localement pseudo-stationnaires, on peut avantageusement utiliser un système adaptatif, qui pourra tenir compte des variations d'énergie du signal dans le temps et converger vers les divers optima locaux.

5

Essentiellement, si l'on applique des retards successifs Δ , la prédiction linéaire $y(n)$ du signal $x(n)$ est une combinaison linéaire des échantillons antérieurs $\{x(n - \Delta - i + 1)\}_{1 \leq i \leq M}$:

10

$$y(n) = \sum_{i=1}^M w_i x(n - \Delta - i + 1)$$

qui minimise l'erreur quadratique moyenne de l'erreur de prédiction :

$$\epsilon(n) = x(n) - y(n)$$

15

La minimisation consiste à trouver :

$$\min_{w_1, w_2, \dots, w_M} E \left[x(n) - \sum_{i=1}^M w_i x(n - \Delta - i + 1) \right]^2$$

20 Pour résoudre ce problème, il est possible d'utiliser un algorithme LMS, qui est un algorithme en lui-même connu, décrit par exemple dans :

[6] 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, et

25 [7] B. Widrow et al., *Adaptive Noise Cancelling: Principles and Applications*, Proc. IEEE, Vol. 63, No 12 pp. 1692-1716, Dec 1975.

On peut définir un procédé récursif d'adaptation des pondérations.

30

$$w_i(n+1) = w_i(n) + 2\mu\epsilon(n)x(n - \Delta - i + 1)$$

μ étant une constante de gain qui permet d'ajuster la vitesse et la stabilité de l'adaptation.

On pourra trouver des indications générales sur ces aspects de l'algorithme LMS dans :

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

- 5 On peut démontrer qu'une telle prédiction linéaire adaptative permet de discriminer efficacement entre bruit et parole car les échantillons contenant de la parole seront bien mieux prédits (plus petites erreurs quadratiques entre la prédiction et le signal brut) que ceux ne contenant que du bruit.
- 10 Plus précisément, les signaux respectifs $x(n)$ et $y(n)$ (signal de parole bruitée et prédiction linéaire) sont découpés en trames de longueurs identiques, et leur transformée de Fourier à court terme (notées respectivement X et Y) est calculée pour chaque trame. Pour éviter les effets des erreurs de précision, l'algorithme prévoit un recouvrement de 50% entre trames
- 15 consécutives, et les échantillons sont multipliés par les coefficients de la fenêtre de Hanning de manière que l'addition des trames paires et impaires corresponde au signal d'origine proprement dit. Pour le segment de spectre k d'une trame / paire, on a :

$$X(k, l) = \sum_{p=1}^R h(p)x(Rl + p)e^{-j2\pi \frac{pk}{R}}$$

Et pour le segment de spectre k d'une trame / impaire :

$$X(k, l) = \sum_{p=1}^R h(p)x(\frac{R}{2}l + p)e^{-j2\pi \frac{pk}{R}}$$

h étant la fenêtre de Hanning.

Une première possibilité consiste à définir le signal de référence en prenant la transformée de Fourier de l'erreur de prédiction :

$$\hat{\epsilon}(k, l) = X(k, l) - Y(k, l)$$

Cependant, on constate en pratique un certain déphasage entre X et Y dû à une convergence imparfaite de l'algorithme LMS, empêchant une bonne

discrimination entre parole et bruit. On préfère donc adopter pour le signal de référence une autre définition qui compense ce déphasage, à savoir :

$$Ref(k, l) = X(k, l) - X(k, l) \frac{|Y(k, l)|}{|X(k, l)|}$$

5

On suppose que l'énergie spectrale du signal de référence peut être décrite sous la forme :

$$E[Ref(k, l)]^2 = E[S(k, l)]^2 \alpha_S(k) + E[D_t(k, l)]^2 \alpha_{D_t}(k) + E[D_{ps}(k, l)]^2 \alpha_{D_{ps}}(k)$$

10 où

$$\alpha_S(k) < \alpha_{D_t}(k) < \alpha_{D_{ps}}(k)$$

représentent l'atténuation sur le signal de référence des trois signaux dans chaque segment de spectre.

15 L'étape suivante consiste à délivrer une estimation $q(k, l)$ de la probabilité d'absence de parole dans le signal bruité :

$$q(k, l) = Pr\{H_0(k, l)\}$$

20 $H_0(k, l)$ indiquant l'absence de parole (et $H_1(k, l)$ la présence de parole) dans le $k^{\text{ième}}$ segment de spectre de la $l^{\text{ième}}$ trame.

La discrimination entre bruit transitoire et parole peut être opérée par une technique comparable à celle de Cohen et Berdugo [5, précité]. Plus précisément, l'algorithme de l'invention évalue un ratio des énergies transitoi-

25 res sur les deux voies, donné par :

$$\Omega(k, l) = \frac{SX(k, l) - \mathcal{M}X(k, l)}{SRef(k, l) - \mathcal{M}Ref(k, l)}$$

S étant une estimation lissée de l'énergie instantanée :

30

$$SX(k, l) = SX(k, l-1) + \sum_{i=-\omega}^{\omega} b(i) |\hat{X}(k, l)|^2$$

- b étant une fenêtre dans le domaine temporel et M étant un estimateur de l'énergie pseudo-stationnaire, qui peut être obtenu par exemple par une méthode MCRA (*Minima Controlled Recursive Averaging*) du même type que celle décrite par Cohen et Berdugo [5, précité] (cependant plusieurs
- 5 alternatives existent dans la littérature).

En présence de parole mais en l'absence de bruit transitoire, ce ratio vaut approximativement :

$$\Omega(k, l) = \frac{1}{\alpha_{D_t}(k)} = \Omega_{\max}(k)$$

10

Inversement, en l'absence de parole mais en présence de bruits transitoires :

$$\Omega(k, l) = \frac{1}{\alpha_S(k)} = \Omega_{\min}(k)$$

- 15 Si l'on suppose qu'en général :

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

- une procédure d'estimation de $q(k, l)$ est donnée par l'algorithme en méta-langage suivant :
- 20

Pour chaque trame l et pour chaque segment de spectre k ,

- (i) Calculer $SX(k, l)$, $MX(k, l)$, $SRef(k, l)$ et $MRef(k, l)$. Aller à (ii)
- (ii) Si $SX(k, l) > L_X MX(k, l)$ (détection de transitoires sur la voie de parole
- 25 bruitée), alors aller à (iii) sinon

$$q(k, l) = 1$$

- (iii) Si $SRef(k, l) > L_{Ref} MRef(k, l)$ (détection de transitoires sur la voie de référence), alors aller à (iv) sinon

$$q(k, l) = 0$$

- 30 (iv) Calculer $\Omega(k, l)$. aller à (v)

- (v) Calculer :

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

Les constantes L_X et L_{Ref} sont des seuils de détection des transitoires. $\Omega_{min}(k)$ et $\Omega_{max}(k)$ sont les limites supérieure et inférieure pour chaque segment de spectre. Ces divers paramètres sont choisis de manière à correspondre à des situations typiques, proches de la réalité.

- 5 L'étape suivante (correspondant au bloc 26 de la figure 1) consiste à opérer le débruitage proprement dit (renforcement de la composante de parole). L'estimateur que l'on vient de décrire sera appliqué au modèle statistique décrit par Ephraim et Malah [2, précité], qui suppose que le bruit et la parole dans chaque segment de spectre sont des processus gaussiens indépendants de variances respectives $\lambda_x(k,l)$ et $\lambda_d(k,l)$.

10 Cette étape peut avantageusement mettre en œuvre l'algorithme de gain OM-LSA (*Optimally Modified Log-Spectral Amplitude Gain*) décrit par Cohen et Berdugo [3, précité]. Le rapport signal/bruit *a priori* est défini par :

$$15 \quad \xi(k,l) = \frac{\lambda_x(k,l)}{\lambda_d(k,l)}$$

Le rapport signal/bruit *a posteriori* est défini par :

$$20 \quad \gamma(k,l) = \frac{|X(k,l)|^2}{\lambda_d(k,l)}$$

La probabilité conditionnelle de présence du signal est :

$$p(k,l) = Pr(H_1(k,l)|X(k,l))$$

- 25 Avec l'hypothèse gaussienne et les paramètres ci-dessus, il vient :

$$p(k,l) = \left\{ 1 + \frac{q(k,l)}{1 - q(k,l)} (1 + \xi(k,l)) \exp(-v(k,l)) \right\}^{-1}$$

avec :

$$30 \quad v(k,l) = \frac{\gamma(k,l)\xi(k,l)}{1 + \xi(k,l)}$$

L'estimée optimale de la parole débruitée $S(k,l)$ est donnée par :

$$\hat{S}(k, l) = G_{H_1}(k, l) p(k, l) G_{min}^{1-p(k, l)} X(k, l)$$

G_{H_1} étant le gain dans l'hypothèse où la parole est présente, qui est défini par :

5

$$G_{H_1}(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \exp\left(\frac{1}{2} \int_{v(k, l)}^{\infty} \frac{e^{-t}}{t} dt\right)$$

Le gain G_{min} dans l'hypothèse d'absence de parole est une limite inférieure pour la réduction du bruit, afin de limiter la distorsion de la parole.

10 La formule classique d'estimation du rapport signal/bruit *a priori* est :

$$\xi(k, l) = a G_{H_1}^2(k, l-1) \gamma(k, l-1) + (1-a) \max(\gamma(k, l) - 1, 0)$$

L'estimation de l'énergie du bruit est donnée par :

15

$$\hat{\lambda}_d(k, l+1) = \tilde{a}_d(k, l) \hat{\lambda}_d(k, l) + \beta(1 - \tilde{a}_d(k, l)) |X(k, l)|^2$$

Le paramètre de lissage $\tilde{a}_d$ évolue entre une limite inférieure a_d et 1, en fonction de la probabilité de présence conditionnelle :

20

$$\tilde{a}_d(k, l) = a_d + (1 - a_d) p(k, l)$$

β étant un facteur de surestimation qui compense le biais en l'absence de signal.

25 Le signal obtenu à l'issue de ce traitement est soumis à une transformée de Fourier inverse (bloc 28) pour donner l'estimée finale de la parole débruitée.

L'algorithme de la présente invention se révèle particulièrement efficace dans les environnements bruyants, parasités à la fois par des bruits mécaniques, des vibrations, etc. ainsi que par des bruits musicaux, situations caractéristiques rencontrées dans l'habitacle d'une voiture. Les spectrogrammes montrent que l'atténuation du bruit est non seulement efficace, mais se fait sans distorsion notable de la parole après débruitage.

30

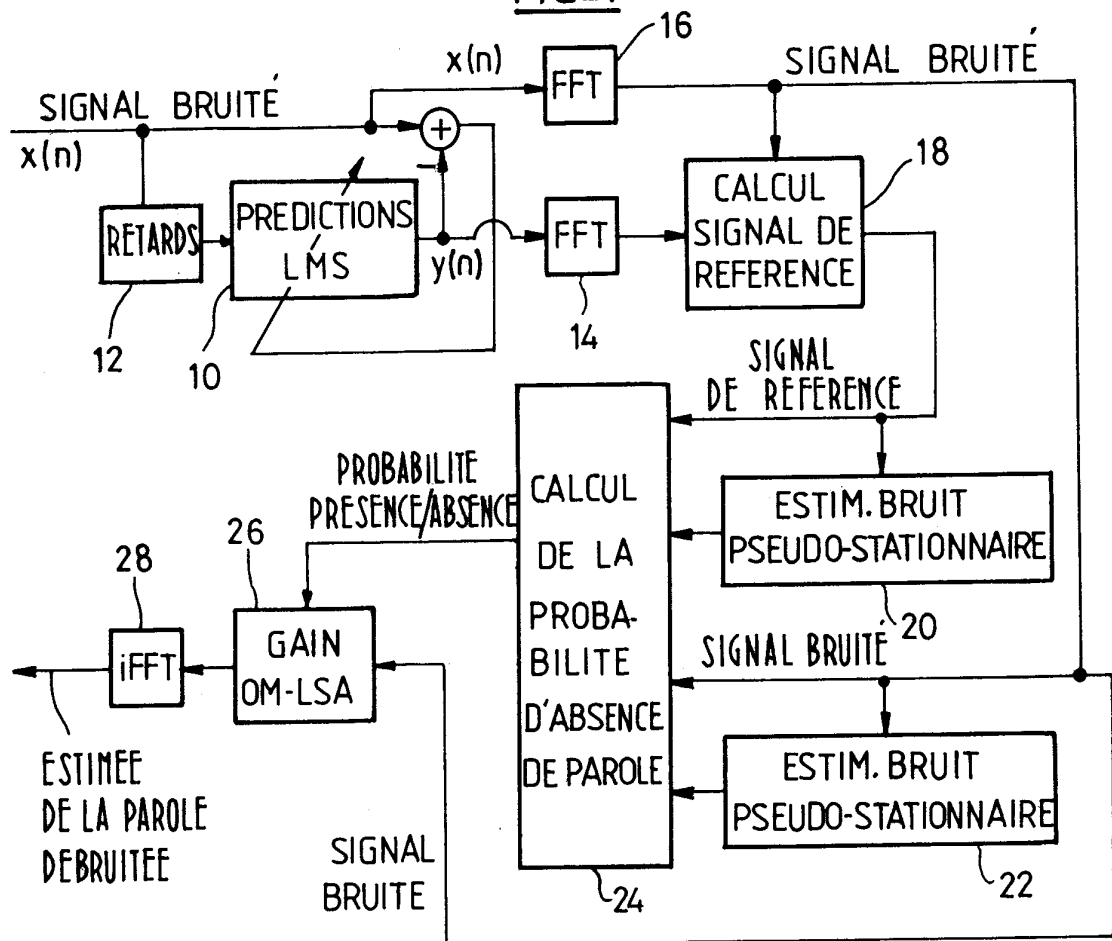
REVENDICATIONS

1. Un procédé de traitement d'un signal audio, pour le débruitage d'un signal bruité comportant une composante de parole combinée à une composante de bruit, cette composante de bruit comprenant elle-même une composante de bruit transitoire et une composante de bruit pseudo-stationnaire, caractérisé en ce que ce procédé est un procédé d'analyse de cohérence temporelle du signal bruité comprenant les étapes consistant à :
 - 5 a) déterminer, à partir du signal bruité, un signal de référence en appliquant au signal bruité un traitement (10, 18) propre à atténuer de façon plus importante les composantes de parole que les composantes de bruit,
 - b) déterminer (24) une probabilité de présence/absence de parole *a priori* à partir des niveaux d'énergie respectifs dans le domaine spectral du signal bruité et du signal de référence, et
 - 15 c) utiliser cette probabilité d'absence de parole *a priori* pour estimer un spectre de bruit et dériver (26) du signal bruité une estimée débruitée du signal de parole.
- 20 2. Le procédé de la revendication 1, dans lequel l'étape a) comprend l'application d'un algorithme prédictif (10).
3. Le procédé de la revendication 2, dans lequel l'algorithme prédictif (10)
 - 25 est un algorithme de type moindres carrés moyens LMS.
4. Le procédé de la revendication 2, dans lequel l'algorithme prédictif (10) est un algorithme adaptatif récursif.
- 30 5. Le procédé de la revendication 1, dans lequel l'étape b) comprend l'application d'un algorithme d'estimation de l'énergie de la composante de bruit pseudo-stationnaire dans le signal de référence et dans le signal bruité.

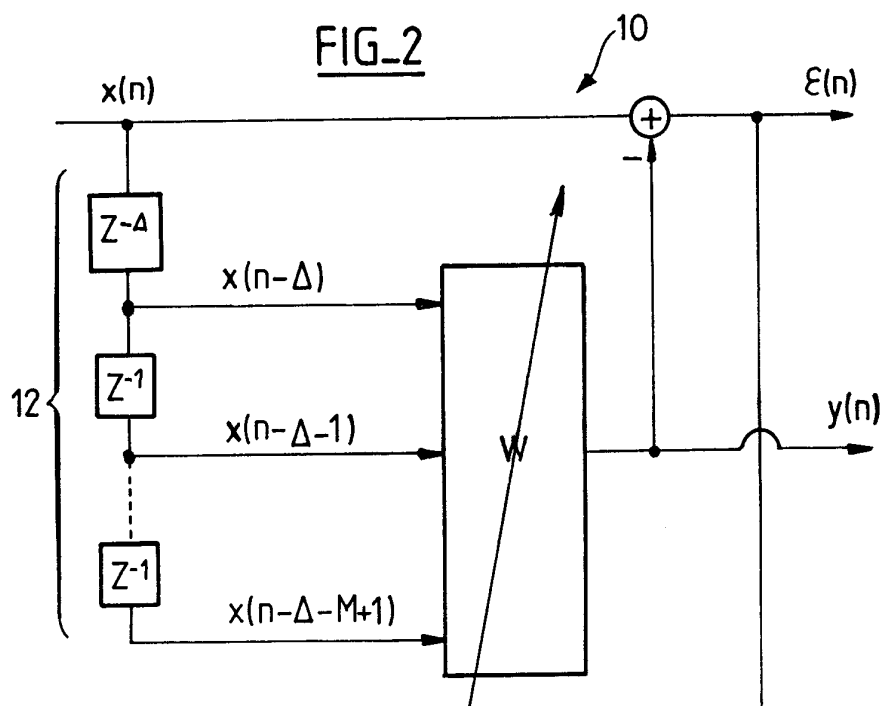
6. Le procédé de la revendication 5, dans lequel l'algorithme d'estimation de l'énergie de la composante de bruit pseudo-stationnaire est un algorithme de type à moyennage récursif par contrôle des minima MRCA.
- 5 7. Le procédé de la revendication 1, dans lequel l'étape c) comprend l'application d'un algorithme de gain variable fonction de la probabilité de présence/absence de parole.
- 10 8. Le procédé de la revendication 7, dans lequel l'algorithme de gain variable est un algorithme de type gain à amplitude log-spectrale modifié optimisé OM-LSA.
-

1/1

FIG_1



FIG_2





RAPPORT DE RECHERCHE PRÉLIMINAIRE

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 677528
FR 0601822

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
D,Y	COHEN I ET AL: "Two-channel signal detection and speech enhancement based on the transient beam-to-reference ratio" 2003 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). HONG KONG, APRIL 6 - 10, 2003, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), NEW YORK, NY : IEEE, US, vol. VOL. 1 OF 6, 6 avril 2003 (2003-04-06), pages V233-V236, XP010639251 ISBN: 0-7803-7663-3 * le document en entier *	1-8	G10L15/20 H04M1/60
Y	US 5 742 694 A (EATWELL ET AL) 21 avril 1998 (1998-04-21) * figures 2,4 * * colonne 2, ligne 28-53 * * colonne 4, ligne 56 - colonne 6, ligne 57 *	1-8	
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, 2002, pages 4-6, XP010631024 ISBN: 0-7803-7693-5 * le document en entier *	1-8	DOMAINES TECHNIQUES RECHERCHÉS (IPC) G10L
A	US 4 658 426 A (CHABRIES ET AL) 14 avril 1987 (1987-04-14) * figures 1,2 * * colonne 4, ligne 41-55 *	1-8	
Date d'achèvement de la recherche		Examineur	
2 octobre 2006		Bensa, Julien	
CATÉGORIE DES DOCUMENTS CITÉS 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 bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant			

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 0601822 FA 677528

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

Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **02-10-2006**

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

Document brevet cité au rapport de recherche		Date de publication		Membre(s) de la famille de brevet(s)		Date de publication
US 5742694	A	21-04-1998	EP	0920751 A1		09-06-1999
			WO	9802983 A1		22-01-1998

US 4658426	A	14-04-1987	AU	582018 B2		09-03-1989
			AU	6365386 A		16-04-1987
			CA	1250348 A1		21-02-1989
			DE	3685474 D1		02-07-1992
			DK	485986 A		11-04-1987
			EP	0220563 A1		06-05-1987
			JP	62155606 A		10-07-1987
