(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 1 830 349 A1**

(12) **DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:
05.09.2007 Bulletin 2007/36

(51) Int Cl.:
G10L 21/02 *(2006.01)*

(21) Numéro de dépôt: **07290219.0**

(22) Date de dépôt: **21.02.2007**

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

(30) Priorité: **01.03.2006 FR 0601822**

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

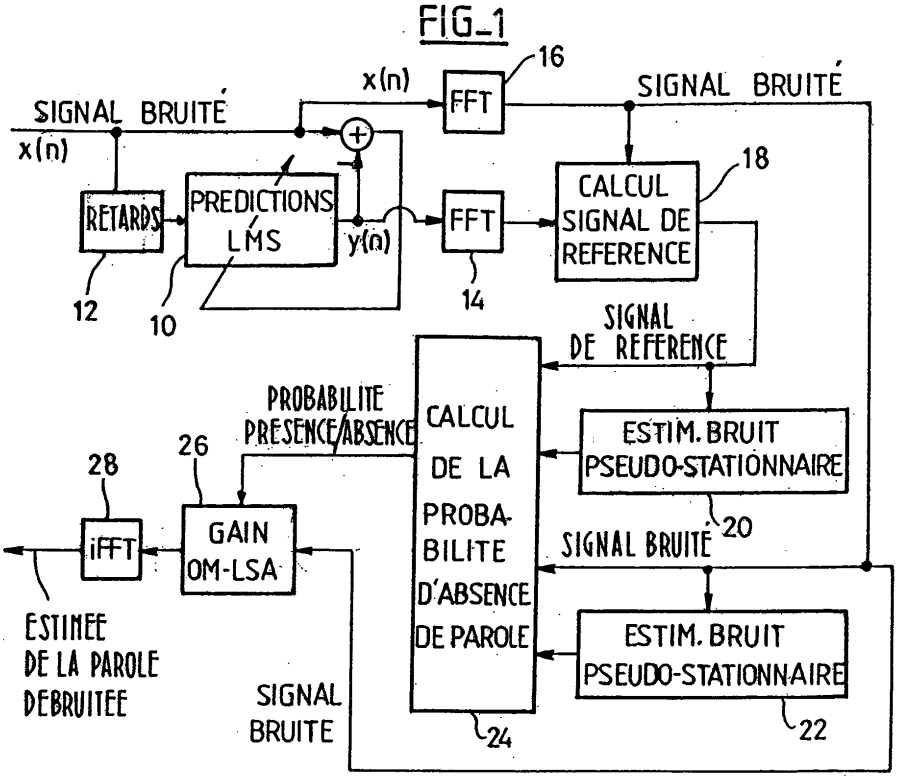
(72) Inventeur: **Pinto, Guillaume**
75004 Paris (FR)

(74) Mandataire: **Dupuis-Latour, Dominique et al**
SEP Bardehle Pagenberg Dost Altenburg Geissler
14 boulevard Malesherbes
75008 Paris (FR)

(54) **Procédé de débruitage 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.

FIG_1



EP 1 830 349 A1

Description

CONTEXTE DE L'INVENTION

Domaine de l'invention

[0001] La présente invention concerne le débruitage des signaux audio captés par un microphone dans un environnement bruité.

[0002] 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.

[0003] 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.

[0004] Il en est de même si l'on veut mettre en oeuvre 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.

[0005] 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.

Description de la technique apparentée

[0006] Diverses techniques ont été proposées pour réduire le niveau de bruit du signal capté par un microphone.

[0007] Par exemple, le WO-A-98/45997 (Parrot SA) utilise l'appui sur le bouton-poussoir d'activation d'un téléphone (par exemple lorsque le conducteur veut répondre à un appel entrant) pour détecter le début d'un signal de parole et considérer que le signal capté antérieurement à cet appui était essentiellement un signal de bruit. Ce dernier signal, mémorisé, est analysé pour donner un spectre énergétique moyen pondéré du bruit, puis soustrait du signal de parole bruité.

[0008] Le US-A-5 742 694 décrit une autre technique, mettant en oeuvre un mécanisme de type filtre adaptatif prédictif. Ce filtre délivre un "signal de référence" correspondant à la partie prédictible du signal bruité et un "signal d'erreur" correspondant à l'erreur de prédiction, puis atténue ces deux signaux dans des proportions variables, et les recombine pour fournir un signal débruité.

[0009] L'inconvénient majeur de cette technique de débruitage réside dans la distorsion importante introduite par le préfiltrage, donnant en sortie un signal très dégradé sur le plan de la qualité acoustique. Elle est en outre mal adaptée aux situations où l'on aurait besoin d'un débruitage énergique avec un signal de parole noyé dans un bruit de nature complexe et imprévisible, avec des caractéristiques spectrales non stationnaires.

[0010] D'autre techniques encore, dites *beamforming* ou *double-phoning*, mettent en oeuvre deux microphones distincts. Le premier est conçu et placé pour capter principalement la voix du locuteur, tandis que l'autre est conçu et placé pour capter une composante de bruit plus importante que le microphone principal. La comparaison des signaux captés permet d'extraire la voix du bruit ambiant de manière efficace, et par des moyens logiciels relativement simples.

[0011] 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. De plus, pour obtenir un débruitage à peu près satisfaisant, les signaux sont soumis à un préfiltrage important ce qui présente, ici encore, l'inconvénient d'introduire des distorsions venant dégrader la qualité du signal débruité restitué.

[0012] L'invention concerne une technique de débruitage des signaux audio captés par un microphone unique enregistrant un signal de voix dans un environnement bruité.

[0013] Une part importante des méthodes les plus efficaces mises en oeuvre dans les système à un seul microphone se fondent sur le modèle statistique établi par D. Malah et Y. Ephraim dans :

[1] Y. Ephraim 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. Ephraim 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.

[0014] 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 recombiner 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 originel (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 bruités 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.

[0015] Bien que séduisant puisque soutenu par une démonstration mathématique rigoureuse, ce procédé ne peut toutefois pas être mis en oeuvre 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 signai 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.

[0016] 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.

[0017] 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.

[0018] 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.

[0019] Pour une composante fréquentielle donnée, la connaissance des indices des 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 ces mesures. La question principale est donc de savoir quand exactement la parole du locuteur est absente du signal capté par le microphone.

[0020] 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.

[0021] Toutefois, dans une 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 comportement 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.

[0022] 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,

[0023] 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 est 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'Ephraim 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.

[0024] Cette estimation de la probabilité *a priori* d'absence de la parole se révèle efficace, mais dépend directement du modèle statistique élaboré par Ephraim et Malah et non d'une connaissance *a priori* des données.

[0025] 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 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, 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 de 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.

[0026] 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.

RÉSUMÉ DE L'INVENTION

[0027] 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.

[0028] 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 (*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 *beam-forming* à deux voies, par exemple des techniques semblables à celles de Cohen et Berdugo [4, précité]. Le calcul d'un ratio entre les niveaux 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.

[0029] En d'autres termes, la technique proposée par l'invention met en oeuvre une "soustraction intelligente" impliquant, après une prédiction linéaire opérée sur les échantillons passés du signal originel (et non d'un signal préfiltré, donc dégradé), un recalage de phase entre le signal originel et le signal prédit.

[0030] La technique de l'invention s'avère, en pratique, suffisamment performante pour assurer un débruitage extrêmement efficace directement sur le signal originel, en s'affranchissant de distorsions introduites par une chaîne de préfiltrage, devenue inutile.

[0031] Plus précisément, la présente invention propose, pour le débruitage d'un signal audio bruité comportant une composante de parole combinée à une composante de bruit comprenant elle-même une composante de bruit transitoire et une composante de bruit pseudo-stationnaire, d'opérer une analyse de cohérence temporelle du signal bruité par les étapes de :

a) détermination d'un signal de référence par application au signal bruité d'un traitement propre à atténuer de façon plus importante les composantes de parole que les composantes de bruit de ce signal bruité, ledit traitement comprenant : (a1) l'application d'un algorithme de prédiction linéaire adaptatif opérant sur une combinaison linéaire des échantillons antérieurs du signal bruité, et (a2) la détermination dudit signal de référence par une soustraction, avec compensation du déphasage, entre le signal bruité et le signal délivré par l'algorithme de prédiction linéaire ;
b) détermination d'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
c) utilisation de cette probabilité d'absence de parole *a priori* pour estimer un spectre de bruit et dériver du signal bruité une estimée débruitée du signal de parole.

[0032] Le signal de référence peut notamment être déterminé par application à l'étape a2) d'une relation du type :

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

où $X(k,l)$ et $Y(k,l)$ sont les transformées de Fourier à court terme de chaque segment de spectre k de chaque trame l, respectivement du signal bruité originel et du signal délivré par l'algorithme de prédiction linéaire.

[0033] L'algorithme prédictif est avantageusement un algorithme adaptatif récursif de type moindres carrés moyens LMS.

[0034] 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,

[0035] 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.

DESCRIPTION SOMMAIRE DES DESSINS

[0036] On va maintenant décrire un exemple de mise en oeuvre 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.

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

DESCRIPTION DÉTAILLÉE DU MODE DE MISE EN OEUVRE PRÉFÉRÉ

[0037] Le signal que l'on souhaite débruiter est un signal numérique échantillonné *x(n)*, où *n* désigne le numéro de l'échantillon (*n* est donc la variable temporelle).

[0038] Le signal capté *x(n)* est une combinaison d'un signal de parole *s(n)* et d'un bruit surajouté, non corrélé, *d(n)* :

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

[0039] Ce bruit *d(n)* a deux composantes indépendantes, à savoir une composante transitoire $d_t(n)$ et une composante pseudo-stationnaire $d_{ps}(n)$:

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

[0040] Comme illustré sur la figure 1, le signal bruité *x(n)* est appliqué en entrée 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.

[0041] On calcule ensuite la transformé 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.

[0042] 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.

[0043] 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.

[0044] On va maintenant décrire plus en détail les différentes phases de ce traitement.

[0045] L'algorithme prédictif LMS (bloc 10) est schématisé sur la figure 2.

[0046] 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.

[0047] 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 $\{x(n - \Delta - i + 1)\}_{1 \leq i \leq M}$:

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

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

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

[0048] La minimisation consiste à trouver :

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

[0049] 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, Adaptative 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

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

[0050] On peut définir un procédé récursif d'adaptation des pondérations.

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

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

[0051] 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.

[0052] 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.

[0053] Plus précisément, les signaux respectifs $x(n)$ et $y(n)$ (signal de parole bruitée et prédiction linéaire) sont dé-

coupés 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 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 *l* paire, on a :

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

[0054] Et pour le segment de spectre *k* d'une trame *l* 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.

[0055] Une première possibilité consiste à définir le signal de référence en prenant la transformée de Fourier de l'erreur de prédiction :

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

[0056] 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 :

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

[0057] On suppose que l'énergie spectrale du signal de référence peut être décrite sous la forme :

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

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.

[0058] L'étape suivante consiste à délivrer une estimation *q(k,l)* de la probabilité d'absence de parole dans le signal bruité :

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

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

[0059] 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 transitoires sur les deux voies, donné par :

$$\Omega(k,l) = \frac{SX(k,l) - MX(k,l)}{SRef(k,l) - MRef(k,l)}$$

[0060] *S* étant une estimation lissée de l'énergie instantanée :

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

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 alternatives existent dans la littérature).

[0061] En présence de parole mais en l'absence de bruit transitoire, ce ratio vaut approximativement :

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

[0062] Inversement, en l'absence de parole mais en présence de bruits transitoires :

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

[0063] 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 métalangage suivant :

[0064] *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_XMX(k,l) (détection de transitoires sur la voie de parole 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

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

(iv) Calculer Ω(k,l). aller à (v)

(v) Calculer :

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

[0065] 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.

[0066] 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 $\lambda_d(k,l)$.

[0067] Cette étape peut avantageusement mettre en oeuvre 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 :

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

[0068] Le rapport signal/bruit *a posteriori* est défini par :

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

[0069] La probabilité conditionnelle de présence du signal est :

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

[0070] 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 :

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

[0071] L'estimée optimale de la parole débruitée $S(k,l)$ est donnée par :

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

[0072] G_{H1} étant le gain dans l'hypothèse où la parole est présente, qui est défini par:

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

[0073] 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.

[0074] La formule classique d'estimation du rapport signal/bruit *a priori* est :

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

[0075] L'estimation de l'énergie du bruit est donnée par :

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

[0076] Le paramètre de lissage $\tilde{a}_d$ évolue entre une limite inférieure a_d et 1, en fonction de la probabilité de présence conditionnelle :

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

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

[0077] 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.

[0078] 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.

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 échantillonné comprenant les étapes de :

   a) détermination d'un signal de référence par application au signal bruité d'un traitement (10, 18) propre à atténuer de façon plus importante les composantes de parole que les composantes de bruit de ce signal bruité, ledit traitement comprenant :

      a1) l'application d'un algorithme de prédiction linéaire adaptatif opérant sur une combinaison linéaire des échantillons antérieurs du signal bruité, et
      a2) la détermination dudit signal de référence par une soustraction, avec compensation du déphasage, entre le signal bruité et le signal délivré par l'algorithme de prédiction linéaire ;

   b) détermination (24) d'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
   c) utilisation de 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.

2. Le procédé de la revendication 1, dans lequel ledit signal de référence est déterminé par application à l'étape a2) d'une relation du type :

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

où $X(k,l)$ et $Y(k,l)$ sont les transformées de Fourier à court terme de chaque segment de spectre k de chaque trame l, respectivement du signal bruité originel et du signal délivré par l'algorithme de prédiction linéaire.

3. Le procédé de la revendication 1, dans lequel l'algorithme de prédiction linéaire (10) est un algorithme de type moindres carrés moyens LMS.

4. Le procédé de la revendication 1, dans lequel l'algorithme de prédiction linéaire (10) est un algorithme adaptatif récursif.

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.

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.

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.

FIG_1

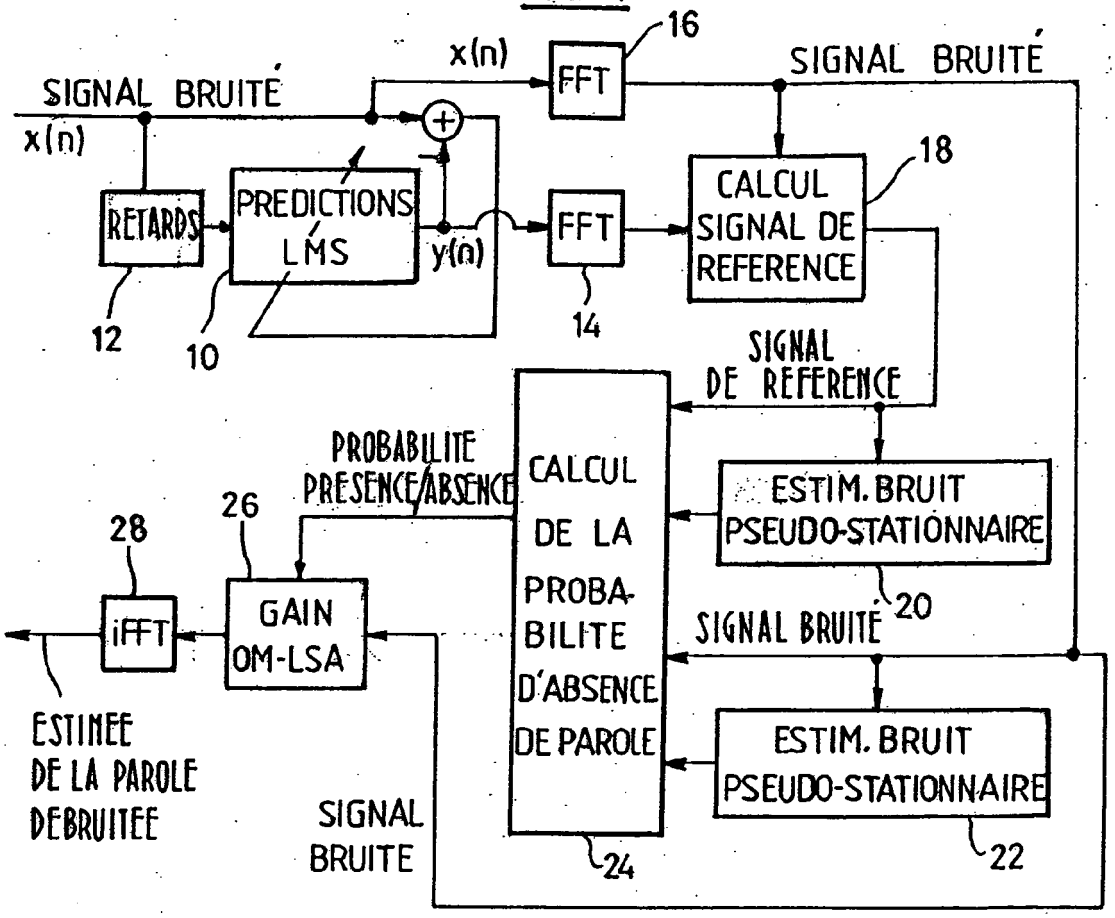
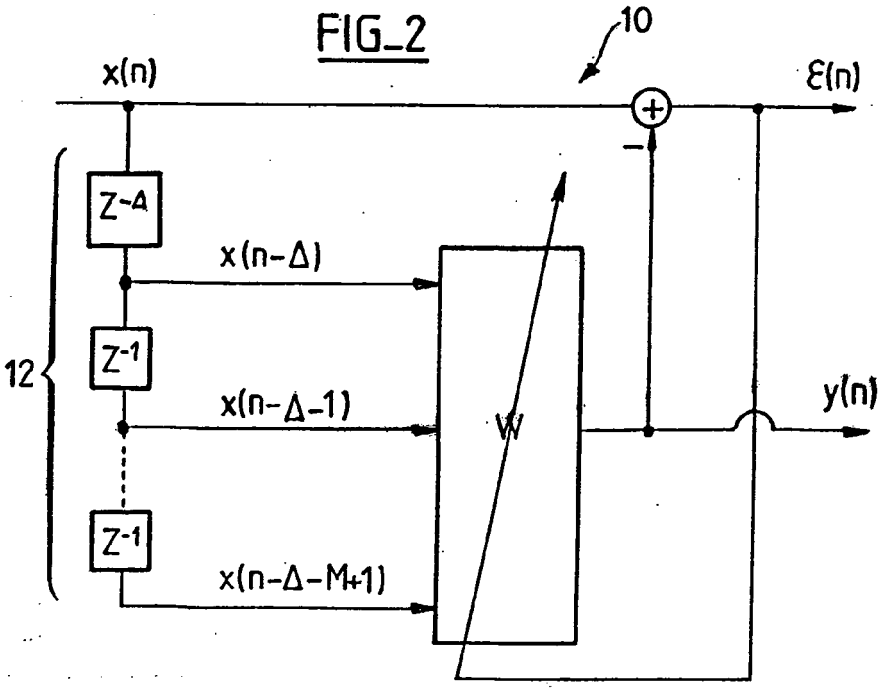

SIGNAL BRUITÉ
x(n)
RETARDS
PREDICTIONS LMS
y(n)
x(n)
FFT
16
SIGNAL BRUITÉ
FFT
CALCUL SIGNAL DE REFERENCE
18
12
10
14
SIGNAL DE REFERENCE
PROBABILITE PRESENCE/ABSENCE
CALCUL DE LA PROBA-BILITE D'ABSENCE DE PAROLE
ESTIM. BRUIT PSEUDO-STATIONNAIRE
20
26
28
iFFT
GAIN OM-LSA
SIGNAL BRUITÉ
ESTIMEE DE LA PAROLE DEBRUITEE
SIGNAL BRUITE
ESTIM. BRUIT PSEUDO-STATIONNAIRE
24
22

FIG_2


10
x(n)
ε(n)
Z-Δ
x(n-Δ)
12
Z-1
x(n-Δ-1)
W
y(n)
Z-1
x(n-Δ-M+1)

Office européen des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande
EP 07 29 0219

DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
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	INV. G10L21/02
D,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	DOMAINES TECHNIQUES RECHERCHES (IPC) G10L
D,A	COHEN I ET AL: "Speech enhancement for non-stationary noise environments" SIGNAL PROCESSING, AMSTERDAM, NL, vol. 81, no. 11, novembre 2001 (2001-11), pages 2403-2418, XP004308517 ISSN: 0165-1684 * page 2407, colonne de droite, ligne 8 - page 2408, colonne de gauche, ligne 6 *	1-8	

-/--

Le présent rapport a été établi pour toutes les revendications

Lieu de la recherche	Date d'achèvement de la recherche	Examinateur
La Haye	31 mai 2007	Bensa, Julien

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

8

EPO FORM 1503 03.82 (P04C02)

Office européen des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande
EP 07 29 0219

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 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	
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	
A	EPHRAIM Y., MALAH D.: "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, décembre 1984 (1984-12), pages 1109-1121, XP002435684 * abrégé * * page 1109, colonne de gauche, ligne 34-36 * * page 1110, colonne de droite, ligne 30-39 * *section IV*	1-8	DOMAINES TECHNIQUES RECHERCHES (IPC)

Le présent rapport a été établi pour toutes les revendications

Lieu de la recherche	Date d'achèvement de la recherche	Examinateur
La Haye	31 mai 2007	Bensa, Julien

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 07 29 0219

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 members 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.

31-05-2007

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

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 9845997 A **[0007]**
- US 5742694 A **[0008]**

Littérature non-brevet citée dans la description

- **Y. EPHRAIM ; D. MALAH.** Speech Enhancement using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* Décembre 1984, vol. ASSP-32 (6), 1109-1121 **[0013]**
- **Y. EPHRAIM ; D. MALAH.** Speech Enhancement using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* Avril 1985, vol. ASSP-33 (2), 443-445 **[0013]**
- Speech Enhancement for Non-Stationary Noise Environments. **I. COHEN ; B. BERDUGO.** Signal Processing. Elsevier, 2001, vol. 81, 2403-2418 **[0022]**
- **I. COHEN ; B. BERDUGO.** Two Channel Signal Detection and Speech Enhancement Based on the Transient Beam-to-Reference Ratio. *Proc. ICASSP 2003,* Avril 2003, 233-236 **[0025]**
- **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 **[0034]**
- Adaptative Filters. **B. WIDROW.** Aspect of Network and System Theory. Holt, Rinehart and Winston, 1970, 563-587 **[0049]**
- **B. WIDROW et al.** Adaptative Noise Cancelling: Principles and Applications. *Proc. IEEE,* Décembre 1975, vol. 63 (12), 1692-1716 **[0049]**
- **B. WIDROW ; S. STEARNS.** Adaptative Signal Processing. Prentice-Hall Signal Processing Series, 1985 **[0051]**