



(11)

EP 3 025 342 B1

(12) **FASCICULE DE BREVET EUROPEEN**

(45) Date de publication et mention de la délivrance du brevet:
13.09.2017 Bulletin 2017/37

(21) Numéro de dépôt: **14741619.2**

(22) Date de dépôt: **21.07.2014**

(51) Int Cl.:
G10L 21/02^(2013.01)

(86) Numéro de dépôt international:
PCT/EP2014/065594

(87) Numéro de publication internationale:
WO 2015/011078 (29.01.2015 Gazette 2015/04)

(54) **PROCÉDÉ DE SUPPRESSION DE LA RÉVERBÉRATION TARDIVE D'UN SIGNAL SONORE**

VERFAHREN ZUR UNTERDRÜCKUNG DES SPÄTEN NACHHALLS EINES AKUSTISCHEN SIGNALS

METHOD FOR SUPPRESSING THE LATE REVERBERATION OF AN AUDIBLE SIGNAL

(84) Etats contractants désignés:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priorité: **23.07.2013 FR 1357226**

(43) Date de publication de la demande:
01.06.2016 Bulletin 2016/22

(73) Titulaire: **Arkamys**
75017 Paris (FR)

(72) Inventeurs:
• **LOPEZ, Nicolas**
75020 Paris (FR)
• **RICHARD, Gaël**
78220 Viroflay (FR)
• **GRENIER, Yves**
78114 Magny Les Hameaux (FR)

(74) Mandataire: **Ipside**
7-9 Allées Haussmann
33300 Bordeaux Cedex (FR)

(56) Documents cités:
• **TOMOHIRO NAKATANI ET AL: "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, USA, vol. 18, no. 7, 1 septembre 2010 (2010-09-01), pages 1717-1731, XP011316583, ISSN: 1558-7916, DOI: 10.1109/TASL.2010.2052251**

- **KINOSHITA K ET AL: "Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-step Linear Prediction", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, USA, vol. 17, no. 4, 1 mai 2009 (2009-05-01), pages 534-545, XP011252305, ISSN: 1558-7916, DOI: 10.1109/TASL.2008.2009015**
- **EMANUEL A PHABETS ET AL: "Late Reverberant Spectral Variance Estimation Based on a Statistical Model", IEEE SIGNAL PROCESSING LETTERS, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 16, no. 9, 1 septembre 2009 (2009-09-01), pages 770-773, XP011262219, ISSN: 1070-9908**
- **WEIFENG LI ET AL: "Feature Denoising Using Joint Sparse Representation for In-Car Speech Recognition", IEEE SIGNAL PROCESSING LETTERS, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 20, no. 7, 1 juillet 2013 (2013-07-01), pages 681-684, XP011511246, ISSN: 1070-9908, DOI: 10.1109/LSP.2013.2245894**
- **EPHRAIM Y ET AL: "Speech Enhancement Using a- Minimum Mean- Square Error Short-Time Spectral Amplitude Estimator", IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, IEEE INC. NEW YORK, USA, vol. ASSP-32, no. 6, 1 décembre 1984 (1984-12-01), pages 1109-1121, XP002435684, ISSN: 0096-3518, DOI: 10.1109/TASSP.1984.1164453**

Il est rappelé que: Dans un délai de neuf mois à compter de la publication de la mention de la délivrance du brevet européen au Bulletin européen des brevets, toute personne peut faire opposition à ce brevet auprès de l'Office européen des brevets, conformément au règlement d'exécution. L'opposition n'est réputée formée qu'après le paiement de la taxe d'opposition. (Art. 99(1) Convention sur le brevet européen).

EP 3 025 342 B1

Description**DOMAINE TECHNIQUE**

5 **[0001]** L'invention concerne un procédé de suppression de la réverbération tardive d'un signal sonore. L'invention est plus particulièrement, mais non exclusivement, adaptée au domaine du traitement de la réverbération dans un espace fermé.

ÉTAT DE LA TECHNIQUE

10 **[0002]** La figure 1 montre une source sonore omnidirectionnelle 100 positionnée dans un espace fermé 110, tel qu'un véhicule automobile ou une salle, ainsi qu'un microphone 120. Un signal sonore émis par la source sonore omnidirectionnelle 100 se propage dans toutes les directions. Ainsi, le signal observé au niveau du microphone est formé par la superposition de plusieurs versions retardées et atténuées du signal sonore émis par la source sonore omnidirectionnelle 100. En effet, le microphone 120 capte tout d'abord le signal source 130, encore appelé signal direct 130, mais également des signaux réfléchis 140 sur les parois de l'espace fermé 110. Les différents signaux réfléchis 140 ont parcouru des chemins acoustiques de différentes longueurs et ont été atténués par l'absorption des parois de l'espace fermé 110, la phase et l'amplitude des signaux réfléchis 140 captés par le microphone 120 sont donc différentes.

20 **[0003]** Deux types de réflexions existent, les réflexions précoces et la réverbération tardive. Le microphone 120 capte les signaux de réflexion précoce avec un faible retard par rapport au signal source 130, de l'ordre de zéro milliseconde à cinquante millisecondes. Lesdits signaux de réflexion précoce sont séparés temporellement et spatialement du signal source 130, mais l'oreille humaine ne perçoit pas ces signaux de réflexion précoce et le signal source 130 séparément grâce à un effet dit « effet de précedence ». Dans le cas où le signal sonore émis par la source sonore omnidirectionnelle 100 est un signal de parole, l'intégration temporelle des signaux de réflexion précoce par l'oreille humaine permet de mettre en relief certaines caractéristiques de la parole, ce qui favorise l'intelligibilité du signal sonore.

25 **[0004]** Selon la taille de la salle, la frontière entre les réflexions précoces et la réverbération tardive est comprise entre cinquante millisecondes et quatre-vingt millisecondes. La réverbération tardive comprend de nombreux signaux réfléchis rapprochés dans le temps et donc impossibles à séparer. L'ensemble de ces signaux réfléchis est donc considéré dans un cadre probabiliste comme une distribution aléatoire dont la densité augmente avec le temps. Dans le cas où le signal sonore émis par la source sonore omnidirectionnelle 100 est un signal de parole, la réverbération tardive dégrade la qualité dudit signal sonore et son intelligibilité. Ladite réverbération tardive affecte également les performances de systèmes de reconnaissance de la parole et de séparation de sources sonores.

30 **[0005]** Selon l'art antérieur, un premier procédé dit « par filtrage inverse » cherche à identifier la réponse impulsionnelle de l'espace fermé 110 pour ensuite construire un filtre inverse permettant de compenser les effets de la réverbération au niveau du signal sonore.

35 **[0006]** Ce type de procédé est par exemple décrit dans les publications scientifiques suivantes : « BWGillespie, H S Malvar, and D A F Florêncio, Speech dereverberation via maximum-kurtosis subband adaptive filtering, Proc. International Conference on Acoustics, Speech, and Signal Processing, volume 6 of ICASSP '01, pages 3701-3704. IEEE, 2001 », « M Wu and D L Wang. A two-stage algorithm for one-microphone reverberant speech enhancement, Audio, Speech, and Language Processing, IEEE Transactions on, 14(3) :774-784, 2006 », « Saeed Mosayyebpour, Abolghasem Sayyadian, Mohsen Zareian, and Ali Shahbazi, Single Channel Inverse Filtering of Room Impulse Response by Maximizing Skewness of LP Residual. ».

40 **[0007]** Ce procédé exploite dans le domaine temporel des distorsions introduites par la réverbération sur des paramètres d'un modèle de prédiction linéaire du signal sonore. Partant de l'observation que la réverbération modifie surtout le résiduel du modèle de prédiction linéaire du signal sonore, un filtre maximisant les moments d'ordre supérieur dudit résiduel est construit. Ce procédé est adapté pour des réponses impulsionnelles courtes et est surtout utilisé pour compenser les signaux de réflexion précoce.

45 **[0008]** Cependant, ce procédé suppose que la réponse impulsionnelle de l'espace fermé 110 est invariante dans le temps. De plus, ce procédé ne modélise pas la réverbération tardive. Ledit procédé doit ainsi être combiné à un autre procédé traitant la réverbération tardive. Ces deux procédés combinés nécessitent de nombreuses itérations avant d'obtenir une convergence, de sorte que lesdits procédés ne peuvent être mis en oeuvre pour une application en temps réel. En outre, le filtrage inverse introduit des artefacts tels que des pré-échos, qui doivent ensuite être compensés.

50 **[0009]** Un deuxième procédé dit « cepstral » vise à séparer l'effet de l'espace fermé 110 et du signal sonore dans le domaine cepstral. En effet, la réverbération modifie la moyenne et la variance des cepstres des signaux réfléchis par rapport à la moyenne et la variance des cepstres du signal source 130. Ainsi, lorsque la moyenne et la variance des cepstres sont normalisées, la réverbération est atténuée.

55 **[0010]** Ce type de procédé est par exemple décrit dans la publication scientifique suivante : « D Bees, M Blostein, and P Kabal, Reverberant speech enhancement using cepstral processing, ICASSP '91 Proceedings of the Acoustics,

Speech, and Signal Processing, 1991 ».

[0011] Ce procédé est particulièrement utile pour des problèmes de reconnaissance vocale puisque les bases de données de référence des systèmes de reconnaissance peuvent également être normalisées pour se rapprocher des signaux captés par le microphone 120. Cependant, les effets de l'espace fermé 110 et du signal sonore ne sont pas complètement séparables dans le domaine cepstral. La mise en oeuvre du procédé provoque donc une distorsion du timbre du signal sonore émis par la source sonore omnidirectionnelle 100. En outre, ce procédé traite les réflexions précoces plutôt que la réverbération tardive.

[0012] Un troisième procédé dit « par estimation de la densité spectrale de puissance de la réverbération tardive » permet d'établir un modèle paramétrique de la réverbération tardive.

[0013] Ce type de procédé est par exemple décrit dans les publications scientifiques suivantes : « E.A.P Habets, Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement, PhD thesis, Technische Universiteit Eindhoven, 2007 », « T. Yoshioka, Speech Enhancement, Reverberant Environments, PhD thesis, 2010 ».

[0014] Selon ce troisième procédé, une estimation de la densité spectrale de puissance de la réverbération tardive permet de construire un filtre de soustraction spectrale pour la déréverbération. La soustraction spectrale introduit des artefacts, comme du bruit musical mais lesdits artefacts peuvent être limités en appliquant des schémas de filtrage plus complexes, utilisés par des procédés de débruitage.

[0015] Cependant, un paramètre important pour estimer la densité spectrale de puissance de la réverbération tardive dans le cadre de ce troisième procédé est le temps de réverbération. Or, le temps de réverbération est un paramètre difficile à estimer avec précision. L'estimation du temps de réverbération est faussée par le bruit de fond et d'autres signaux sonores qui interfèrent. En outre, cette estimation du temps de réverbération est chronophage et donc allonge le temps d'exécution.

[0016] Un quatrième procédé exploite la parcimonie des signaux de parole dans le plan temps/fréquence.

[0017] Ce type de procédé est par exemple décrit dans la publication scientifique suivante : « T. Yoshioka, Speech Enhancement in Reverberant Environments, PhD thesis, 2010 ».

[0018] Dans cette publication, la réverbération tardive est modélisée comme une version retardée et atténuée de l'observation courante dont le facteur d'atténuation est déterminé par résolution d'un problème de maximum de vraisemblance, avec une contrainte de parcimonie.

[0019] Ce type de procédé est en outre décrit dans la publication scientifique suivante : « H Kameoka, T Nakatani, and T Yoshioka, Robust speech dereverberation based on nonnegativity and sparse nature of speech spectrograms, Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '09, pages 45-48. IEEE Computer Society, 2009 ».

[0020] La déréverbération est abordée dans cette publication comme un problème de déconvolution par factorisation en matrices non négatives, ce qui permet de séparer la réponse de l'espace fermé 110 et le signal sonore. Cependant, ce procédé introduit beaucoup de bruit et de distorsions. En outre, ledit procédé dépend de l'initialisation des matrices pour la factorisation.

[0021] De plus, les procédés cités nécessitent une pluralité de microphones pour traiter avec précision la réverbération.

EXPOSÉ DE L'INVENTION

[0022] L'invention a notamment pour but de résoudre tout ou partie des problèmes susmentionnés.

[0023] A cette fin, l'invention concerne un procédé de suppression de la réverbération tardive d'un signal sonore caractérisé en ce qu'il comporte les étapes suivantes :

- captation d'un signal d'entrée formé par la superposition de plusieurs versions retardées et atténuées du signal sonore,
- application d'une transformation temps-fréquence au signal d'entrée afin d'obtenir une transformée temps-fréquence complexe du signal d'entrée,
- calcul d'une pluralité de vecteurs de prédiction,
- création d'une pluralité de vecteurs d'observation à partir du module de la transformée temps-fréquence complexe du signal d'entrée,
- construction d'une pluralité de dictionnaires de synthèse à partir de la pluralité de vecteurs d'observations,
- estimation d'un spectre de réverbération tardive à partir de la pluralité de dictionnaires de synthèse et de la pluralité de vecteurs de prédiction,
- filtrage de la pluralité de vecteurs d'observations afin d'éliminer le spectre de réverbération tardive et d'obtenir un module de signal déréverbéré.

[0024] Ainsi, le procédé objet de l'invention est rapide et présente une complexité réduite. Ledit procédé est donc utilisable en temps réel. De plus, ce procédé n'introduit pas d'artefacts et est robuste au bruit de fond. En outre, ledit

EP 3 025 342 B1

procédé réduit le bruit de fond et est compatible avec des procédés de réduction de bruit.

[0025] L'invention peut être mise en oeuvre selon les modes de réalisation avantageux exposés ci-après, lesquels peuvent être considérés individuellement ou selon toute combinaison techniquement opérante.

[0026] Avantageusement, le procédé comporte en outre les étapes suivantes :

- 5
- création d'un module sous échantillonné en fréquence à partir du module de la transformée temps-fréquence complexe du signal d'entrée,
 - création d'une pluralité de vecteurs d'observation sous échantillonnés à partir dudit module sous échantillonné en fréquence,
- 10
- construction d'une pluralité de dictionnaires d'analyse à partir de la pluralité de vecteurs d'observation sous échantillonnés,
 - calcul de la pluralité de vecteurs de prédiction à partir de la pluralité de vecteurs d'observation sous échantillonnés et de la pluralité de dictionnaires d'analyse.

15 **[0027]** Avantageusement, l'étape de calcul de la pluralité de vecteurs de prédiction est effectuée en minimisant, pour chaque vecteur de prédiction, l'expression $\|\bar{X}v - D^a\alpha\|_2$, qui est la norme euclidienne de la différence entre le vecteur d'observation sous échantillonné associé audit vecteur de prédiction et du dictionnaire d'analyse associé audit vecteur de prédiction multiplié par ledit vecteur de prédiction, en tenant compte de la contrainte $\|\alpha\|_1 \leq \lambda$, selon laquelle la norme 1 dudit vecteur de prédiction est inférieure ou égale à un paramètre d'intensité maximale de la réverbération tardive.

20 **[0028]** Avantageusement, la valeur du paramètre d'intensité maximale de la réverbération tardive est comprise entre 0 et 1.

[0029] Avantageusement, le procédé comporte en outre l'étape suivante :

- 25
- création d'un signal complexe déréverbéré à partir du module de signal déréverbéré et de la phase de la transformée temps-fréquence complexe du signal d'entrée.

[0030] Avantageusement, le procédé comporte en outre l'étape suivante :

- 30
- application d'une transformation fréquence-temps au signal complexe déréverbéré afin d'obtenir un signal temporel déréverbéré.

[0031] Avantageusement, le procédé comporte en outre une étape de construction d'un filtre de déréverbération selon le modèle

35

$$G = \frac{\xi}{1 + \xi} \exp\left(\int_v^\infty \frac{e^{-t}}{t} dt\right),$$

40 où ξ est le rapport signal à bruit a priori, et où la borne d'intégration v est calculée selon le modèle

$$v = \gamma \frac{\xi}{1 + \xi}$$

45

où γ est le rapport signal à bruit a posteriori.

[0032] L'invention concerne également un dispositif de suppression de la réverbération tardive d'un signal sonore caractérisé en ce qu'il comporte des moyens pour :

- 50
- capter un signal d'entrée formé par la superposition de plusieurs versions retardées et atténuées du signal sonore,
 - appliquer une transformation temps-fréquence au signal d'entrée afin d'obtenir une transformée temps-fréquence complexe du signal d'entrée,
 - calculer une pluralité de vecteurs de prédiction,
 - créer une pluralité de vecteurs d'observation à partir du module de la transformée temps-fréquence complexe du signal d'entrée,
- 55
- construire une pluralité de dictionnaires de synthèse à partir de la pluralité de vecteurs d'observations,
 - estimer un spectre de réverbération tardive à partir de la pluralité de dictionnaires de synthèse et de la pluralité de vecteurs de prédiction,

- filtrer la pluralité de vecteurs d'observations afin d'éliminer le spectre de réverbération tardive et d'obtenir un module de signal déréverbéré.

PRÉSENTATION DES FIGURES

5

[0033] L'invention sera mieux comprise à la lecture de la description suivante, donnée à titre d'exemple nullement limitatif, et faite en se référant aux figures qui représentent :

- Figure 1 (déjà décrite) : une représentation schématique d'une source sonore omnidirectionnelle et d'un microphone positionnés dans un espace fermé selon un exemple de réalisation de l'invention ;
- Figure 2 : une représentation schématique d'un dispositif de déréverbération d'un signal sonore selon un exemple de réalisation de l'invention ;
- Figure 3 : une représentation schématique d'une unité de déréverbération d'un dispositif de déréverbération d'un signal sonore selon un exemple de réalisation de l'invention ;
- Figure 4 : une représentation schématique d'une unité d'estimation de la réverbération tardive d'un dispositif de déréverbération d'un signal sonore selon un exemple de réalisation de l'invention ;
- Figure 5 : une représentation schématique d'un regroupement en sous bandes d'un module d'une transformée temps-fréquence complexe d'un signal d'entrée selon un exemple de réalisation de l'invention ;
- Figure 6 : une représentation schématique d'une unité de calcul de vecteurs de prédiction d'un dispositif de déréverbération d'un signal sonore selon un exemple de réalisation de l'invention ;
- Figure 7 : une représentation schématique d'une unité de calcul de vecteurs de prédiction d'un dispositif de déréverbération d'un signal sonore selon un exemple de réalisation de l'invention ;
- Figure 8 : une représentation schématique d'une unité d'évaluation de la réverbération d'un dispositif de déréverbération d'un signal sonore selon un exemple de réalisation de l'invention ;
- Figure 9 : un diagramme fonctionnel montrant différentes étapes du procédé selon un exemple de réalisation de l'invention.

30

[0034] Dans ces figures, des références identiques d'une figure à une autre désignent des éléments identiques ou analogues. Pour des raisons de clarté, les éléments représentés ne sont pas à l'échelle, sauf mention contraire.

DESCRIPTION DÉTAILLÉE DE MODES DE RÉALISATION

35

[0035] L'invention met en oeuvre un dispositif de déréverbération d'un signal sonore émis par une source sonore omnidirectionnelle 100 positionnée dans un espace fermé 110, tel qu'un véhicule automobile ou une salle, et capté par un microphone 120. Ledit dispositif de déréverbération est inséré dans la chaîne de traitement audio d'un appareil tel qu'un téléphone. Ce dispositif de déréverbération comporte une unité d'application d'une transformée temps-fréquence 200, une unité de déréverbération 210 et une unité d'application d'une transformée fréquence-temps 220 (cf. figure 2). L'unité de déréverbération 210 comporte une unité d'estimation de la réverbération tardive 300 et une unité de filtrage 310 (cf. figure 3). L'unité d'estimation de la réverbération tardive 300 comporte une unité de regroupement en sous bandes 400, une unité de calcul de vecteurs de prédiction 410 et une unité d'évaluation de la réverbération 420 (cf. figure 4). L'unité de calcul de vecteurs de prédiction 410 comporte une unité de construction d'observations 700, une unité de construction de dictionnaires d'analyse 710 et une unité de résolution du LASSO 720 (cf. figure 7). L'unité d'évaluation de la réverbération 420 comporte une unité de construction de dictionnaires de synthèse 800 (cf. figure 8).

40

45

[0036] Dans une étape 900, un microphone 120 capte un signal d'entrée $x(t)$ formé par la superposition de plusieurs versions retardées et atténuées du signal sonore émis par la source sonore omnidirectionnelle 100. En effet, le microphone 120 capte tout d'abord le signal source 130, encore appelé signal direct 130, mais également des signaux réfléchis 140 sur les parois de l'espace fermé 110. Les différents signaux réfléchis 140 ont parcouru des chemins acoustiques de différentes longueurs et ont été atténués par l'absorption des parois de l'espace fermé 110, la phase et l'amplitude des signaux réfléchis 140 captés par le microphone 120 sont donc différentes.

50

[0037] Deux types de réflexions existent, les réflexions précoces et la réverbération tardive. Le microphone 120 capte les signaux de réflexion précoce avec un faible retard par rapport au signal source 130, de l'ordre de zéro milliseconde à cinquante millisecondes. Lesdits signaux de réflexion précoce sont séparés temporellement et spatialement du signal source 130 mais l'oreille humaine ne perçoit pas ces signaux de réflexion précoce et le signal source 130 séparément grâce à un effet dit « effet de précedence ». Dans le cas où le signal sonore émis par la source sonore omnidirectionnelle 100 est un signal de parole, l'intégration temporelle des signaux de réflexion précoce par l'oreille humaine permet de mettre en relief certaines caractéristiques de la parole, ce qui favorise l'intelligibilité du signal sonore.

55

[0038] Le microphone 120 capte la réverbération tardive entre cinquante millisecondes et quatre-vingts millisecondes après l'arrivée du signal source 130. La réverbération tardive comprend de nombreux signaux réfléchis rapprochés dans

le temps et donc impossibles à séparer. L'ensemble de ces signaux réfléchis est donc considéré dans un cadre probabiliste comme une distribution aléatoire dont la densité augmente avec le temps. Dans le cas où le signal sonore émis par la source sonore omnidirectionnelle 100 est un signal de parole, la réverbération tardive dégrade la qualité dudit signal sonore et son intelligibilité. Ladite réverbération tardive affecte également les performances de systèmes de reconnaissance de la parole et de séparation de sources sonores.

[0039] Le signal d'entrée $x(t)$ est échantillonné à une fréquence d'échantillonnage f_s . Le signal d'entrée $x(t)$ est ainsi subdivisé en échantillons. Afin de supprimer la réverbération tardive dudit signal d'entrée $x(t)$, la densité spectrale de puissance de la réverbération tardive est estimée puis un filtre de déréverbération est construit par l'unité de déréverbération 210. L'estimation de la densité spectrale de puissance de la réverbération tardive, la construction du filtre de déréverbération et l'application dudit filtre de déréverbération sont effectués dans le domaine fréquentiel. Ainsi, dans une étape 901, une transformation temps-fréquence est appliquée au signal d'entrée $x(t)$ par l'unité d'application de la Transformée de Fourier à Court Terme 200 afin d'obtenir une transformée temps-fréquence complexe du signal d'entrée $x(t)$ notée X^C (cf. figure 2). Dans un exemple, la transformation temps-fréquence est une Transformation de Fourier à Court Terme.

[0040] Chaque élément $X_{k,n}^C$ de la transformée temps-fréquence complexe X^C est calculé de la façon suivante :

$$X_{k,n}^C = \sum_{m=0}^{M-1} x(m+nR)w(m)e^{\frac{2j\pi km}{M}}$$

où k est un indice fréquentiel d'échantillonnage de valeur comprise entre 1 et un nombre K , n est un indice temporel de valeur comprise entre 1 et un nombre N , $w(m)$ est une fenêtre glissante d'analyse, m est l'indice des éléments appartenant à une trame, M est la longueur d'une trame, c'est-à-dire le nombre d'échantillons d'une trame et R est le pas d'avancement de la transformation temps-fréquence.

[0041] Le signal d'entrée $x(t)$ est analysé par trames de longueur M avec un pas d'avancement R égal à $M/4$ échantillons. Pour chaque trame du signal d'entrée $x(t)$ dans le domaine temporel une transformée temps-fréquence Discrète d'indice fréquentiel d'échantillonnage k et d'indice temporel n est ainsi calculée grâce à l'algorithme de la transformation temps-

fréquence pour obtenir un signal complexe $X_{k,n}^C$ défini par

$$X_{k,n}^C = |X_{k,n}| e^{-j\angle X_{k,n}},$$

où $|X_{k,n}|$ est le module du signal complexe $X_{k,n}^C$ et $\angle X_{k,n}$ est la phase du signal complexe $X_{k,n}^C$.

[0042] L'estimation de la densité spectrale de puissance de la réverbération tardive est réalisée sur le module de la transformée temps-fréquence complexe du signal d'entrée X^C , noté X . La phase du transformée temps-fréquence complexe X^C , notée $\angle X$ est gardée en mémoire et est utilisée pour reconstruire un signal déréverbéré dans le domaine temporel après application du filtre de déréverbération.

[0043] Le module X de la transformée temps-fréquence complexe du signal d'entrée X^C est ensuite regroupé en sous bandes. Plus précisément, ledit module X comporte le nombre K de lignes spectrales notées X_k . Le terme "ligne spectrale" désigne ici tous les échantillons du module X de la transformée temps-fréquence complexe du signal d'entrée X^C pour l'indice fréquentiel d'échantillonnage k et tous les indices temporels n . Dans une étape 903, l'unité de regroupement en sous bandes 400 regroupe les K lignes spectrales X_k en un nombre J de sous bandes, afin d'obtenir un module sous échantillonné en fréquence noté \tilde{X} comportant un nombre J de lignes spectrales notées \tilde{X}_j , où j est un indice fréquentiel de sous échantillonnage compris entre 1 et le nombre J . Le nombre J est inférieur au nombre K . Chaque sous bande comporte ainsi une pluralité de lignes spectrales X_k , l'indice fréquentiel k appartenant à un intervalle ayant une borne inférieure b_j et une borne supérieure e_j . Dans un exemple, chaque sous bande correspond à un octave afin de prendre en compte le modèle de perception sonore de l'oreille humaine. Ensuite, dans une étape 904, l'unité de regroupement en sous bandes 400 calcule, pour chaque sous bande, une moyenne Mean des lignes spectrales X_k de ladite sous bande afin d'obtenir les J lignes spectrales \tilde{X}_j du module sous échantillonné en fréquence \tilde{X} (cf. figure 5).

[0044] Ensuite, l'unité de calcul de vecteurs de prédiction 410 calcule pour chaque ligne spectrale \tilde{X}_j du module sous échantillonné en fréquence \tilde{X} et pour chaque indice temporel n un vecteur de prédiction $\alpha_{j,n}$ (cf. figure 6). Plus précisément, dans une étape 905, l'unité de construction d'observation 700 construit, pour chaque indice temporel n et indice fréquentiel j de sous échantillonnage, un vecteur d'observation sous échantillonné $\tilde{X}_{j,n}$ à partir de l'ensemble des échantillons $\tilde{X}_{j,n_1:n}$ appartenant à la j -ième ligne spectrale \tilde{X}_j du module sous échantillonné en fréquence \tilde{X} et compris entre les instants $n_1=n-N+1$ et n , où n est l'indice de l'instant courant et $n-n_1$ est la taille de la mémoire du dispositif de déréver-

bération. Chaque vecteur d'observation sous échantillonné $\tilde{X}_{V_{j,n}}$ est défini par

5

$$\tilde{X}_{V_{j,n}} := \left[\tilde{X}_{j,n} \dots \tilde{X}_{j,n-N+1} \right]^T,$$

10

[0045] Chaque vecteur d'observation $\tilde{X}_{V_{j,n}}$ est de taille $N \times 1$, où le nombre N est la longueur de l'observation. La longueur de l'observation N est le nombre de trames de la transformation temps-fréquence nécessaires pour l'estimation de la réverbération tardive. La longueur de l'observation N permet de définir la résolution temporelle de l'estimation. Quand la longueur de l'observation N augmente, la complexité du système diminue. Le sous-échantillonnage du module X de la transformée temps-fréquence complexe du signal d'entrée X^C permet entre autre l'application du procédé en temps réel.

15

[0046] Dans une étape 906, l'unité de construction de dictionnaires d'analyse 710 construit des dictionnaires d'analyse D^a . Plus précisément, pour chaque indice temporel n et indice fréquentiel de sous échantillonnage j , un dictionnaire d'analyse $D_{j,n}^a$ est construit en concaténant un nombre L de vecteurs d'observations passées déterminés à l'étape 905.

Le dictionnaire d'analyse $D_{j,n}^a$ se définit ainsi comme la matrice

20

$$D_{j,n}^a := \begin{bmatrix} \tilde{X}_{j,n-\delta} & \tilde{X}_{j,n-\delta-1} & \dots & \tilde{X}_{j,n-\delta-L+1} \\ \tilde{X}_{j,n-\delta-1} & \tilde{X}_{j,n-\delta-2} & \dots & \tilde{X}_{j,n-\delta-L} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{X}_{j,n-\delta-N+1} & \tilde{X}_{j,n-\delta-N} & \dots & \tilde{X}_{j,n-\delta-L-N+2} \end{bmatrix}$$

25

où L est le nombre de vecteurs d'observations passées et donc la taille du dictionnaire d'analyse $D_{j,n}^a$, et $\delta \in \mathbb{R}^*$

30

est le retard du dictionnaire d'analyse $D_{j,n}^a$. Plus précisément, le retard δ est le retard de trames entre le vecteur d'observation courante sous échantillonné $\tilde{X}_{V_{j,n}}$ et les autres vecteurs d'observations sous échantillonnés appartenant

35

au dictionnaire d'analyse $D_{j,n}^a$. Ledit retard δ permet de réduire les distorsions introduites par le procédé. Ce retard δ permet en outre de d'améliorer la séparation de la réverbération tardive et des réflexions précoces. Pour calculer le

vecteur d'observation courante $\tilde{X}_{V_{j,n}}$ et le dictionnaire d'analyse $D_{j,n}^a$ et donc le vecteur de prédiction $\alpha_{j,n}$ pour chaque ligne spectrale \tilde{X}_j et pour chaque indice temporel n , un nombre $L+N+\delta$ de trames doit être gardé en mémoire.

40

[0047] Dans une étape 907, l'unité de résolution du LASSO 720 résout un problème appelé "LASSO" qui est de

minimiser la norme euclidienne $\left\| \tilde{X}_{V_{j,n}} - D_{j,n}^a \alpha_{j,n} \right\|_2$ en tenant compte de la contrainte $\|\alpha_{j,n}\| \leq \lambda$ où λ est un paramètre

45

d'intensité maximale. Pour résoudre ledit problème, la meilleure combinaison linéaire des L vecteurs du dictionnaire permettant d'approcher l'observation courante doit être trouvée. Dans un exemple, un procédé connu, appelé LARS, selon l'acronyme anglo-saxon de "Least Angle Regression" permet de résoudre ledit problème. La contrainte $\|\alpha_{j,n}\|_1 \leq$

50

λ permet de privilégier les solutions ayant peu d'éléments non nuls, c'est-à-dire les solutions parcimonieuses. Le paramètre d'intensité maximale λ permet de régler l'intensité maximale estimée de la réverbération tardive. Ce paramètre d'intensité maximale λ dépend *a priori* de l'environnement acoustique, c'est-à-dire dans un exemple de l'espace fermé

55

110. Pour chaque espace fermé 110, une valeur optimale du paramètre d'intensité maximale λ existe. Cependant, des essais ont montré que ledit paramètre d'intensité maximale λ peut être fixé à une valeur identique pour tous les espaces fermés 110, sans que ladite valeur introduise de dégradations par rapport à la valeur optimale. Ainsi le procédé fonctionne

dans une grande variété d'espaces fermés 110 sans nécessiter de réglage particulier, ce qui permet de s'affranchir des erreurs d'estimation du temps de réverbération de l'espace fermé 110. En outre, le procédé selon l'invention ne nécessite pas de paramètre devant être estimé, ce qui permet l'application dudit procédé en temps réel. La valeur du paramètre

d'intensité maximale λ est comprise entre 0 et 1. Dans un exemple, la valeur du paramètre d'intensité maximale λ est égale à 0,5, ce qui est un bon compromis entre la réduction de la réverbération et la qualité globale du procédé.

[0048] Dans une étape 908, pour chaque indice temporel n et chaque indice fréquentiel d'échantillonnage k , un vecteur d'observation courante $X_{V_{k,n}}$ est créé à partir de l'ensemble des échantillons appartenant à la k -ième ligne spectrale X_k

du module X de la transformée temps-fréquence complexe et compris entre les instants n_1 et n , noté $X_{k,n1:n}$ où n est l'indice d'instant courant et $n - n_1$ est la taille de la mémoire du dispositif de déréverbération. Chaque vecteur d'observation $X_{k,n}$ est défini par la formule $X_{k,n} := [X_{k,n} \dots X_{k,n-N+1}]^T$, et est de taille $N \times 1$, où N est la longueur de l'observation.

[0049] Dans une étape 909, l'unité de construction d'un dictionnaire de synthèse 800 construit un dictionnaire de synthèse D^s . Plus précisément, pour chaque indice temporel n et chaque indice fréquentiel d'échantillonnage k , le

dictionnaire de synthèse $D_{k,n}^s$ est construit en concaténant un nombre L de vecteurs d'observations passées déterminés à l'étape 908. Le dictionnaire de synthèse $D_{k,n}^s$ se définit ainsi comme la matrice

$$D_{k,n}^s := \begin{bmatrix} X_{k,n-\delta} & X_{k,n-\delta-1} & \cdots & X_{k,n-\delta-L+1} \\ X_{k,n-\delta-1} & X_{k,n-\delta-2} & \cdots & X_{k,n-\delta-L} \\ \vdots & \vdots & \vdots & \vdots \\ X_{k,n-\delta-N+1} & X_{k,n-\delta-N} & \cdots & X_{k,n-\delta-L-N+2} \end{bmatrix}$$

où L et δ sont les mêmes paramètres que pour le dictionnaire d'analyse $D_{j,n}^a$.

[0050] Dans une étape 910, pour chaque indice temporel n et chaque indice fréquentiel d'échantillonnage k , une estimation de la densité spectrale de puissance de la réverbération tardive ou du spectre de la réverbération tardive

$X_{k,n}^\ell$ est construit par multiplication du dictionnaire de synthèse $D_{k,n}^s$ avec le vecteur de prédiction $\alpha_{j,n}$ selon la formule

$$X_{k,n}^\ell = D_{k,n}^s \alpha_{j,n} \quad \forall k \in [b_j, e_j] \quad j = 1, \dots, J$$

[0051] Le vecteur de prédiction $\alpha_{j,n}$ indique donc les colonnes du dictionnaire de synthèse qui ont été retenues pour l'estimation de la réverbération, ainsi que la contribution de chacune d'elles à la réverbération. Le spectre de la réverbération tardive X^ℓ est considéré dans la suite du procédé comme un signal de bruit à éliminer.

[0052] A cette fin, un filtrage de la réverbération est effectué par l'unité de filtrage 310. Plus précisément, dans une étape 911, pour chaque indice temporel n et chaque indice fréquentiel d'échantillonnage k , un filtre de déréverbération $G_{k,n}$ est construit selon la formule

$$G_{k,n} = \frac{\xi_{k,n}}{1 + \xi_{k,n}} \exp\left(\int_{v_{k,n}}^{\infty} \frac{e^{-t}}{t} dt\right)$$

où $\xi_{k,n}$ est le rapport signal à bruit *a priori*, calculé de la façon suivante

$$\xi_{k,n} = \beta G_{k,n-1}^2 \gamma_{k,n-1} + (1 - \beta) \max\{\gamma_{k,n} - 1, 0\}$$

et où la borne d'intégration $v_{k,n}$ est calculée de la façon suivante

$$v_{k,n} = \gamma_{k,n} \frac{\xi_{k,n}}{1 + \xi_{k,n}}$$

où $\gamma_{k,n}$ est le rapport signal à bruit *a posteriori*, calculé selon la formule

$$\gamma_{k,n} = \frac{|X_{k,n}|^2}{|R_{k,n}|^2}$$

où $R_{k,n}$ est la réverbération tardive lissée calculée de la façon suivante

$$R_{k,n} = \alpha R_{k,n-1} + (1 - \alpha) |X_{k,n}^\ell|$$

où α est une première constante de lissage et β est une seconde constante de lissage. Dans un exemple, la première constante de lissage α vaut 0.77 et la seconde constante de lissage β vaut 0.98.

[0053] En effet, la réverbération estimée est non stationnaire à long terme car le signal sonore émis par la source sonore omnidirectionnelle 100, qui provoque ladite réverbération estimée n'est pas stationnaire à long terme. Des variations trop rapides de la réverbération estimée peuvent introduire des artefacts gênants lors du filtrage. Pour limiter ces effets, un lissage récursif est effectué pour calculer la densité spectrale de puissance de la réverbération tardive.

[0054] Dans une étape 912, pour chaque indice temporel n et chaque indice fréquentiel d'échantillonnage k , les vecteurs d'observations $X_{v,k,n}$ sont filtrés par le filtre de déréverbération $G_{k,n}$ calculé à l'étape 911 afin d'obtenir un module de signal déréverbéré $Y_{k,n}$ calculé de la façon suivante

$$Y_{k,n} = G_{k,n} X_{k,n}.$$

[0055] Le filtre construit à l'étape 911 atténue fortement certains vecteurs d'observations $X_{v,k,n}$ ce qui génère des artefacts nuisibles à la qualité du signal déréverbéré. Pour limiter lesdits artefacts, une borne inférieure est imposée sur l'atténuation du filtre. Ainsi, pour chaque indice fréquentiel d'échantillonnage k et pour chaque indice temporel n , si le filtre de déréverbération $G_{k,n}$ est inférieur ou égal à une valeur minimale du filtre de déréverbération G_{min} , alors ledit filtre de déréverbération $G_{k,n}$ est égal à ladite valeur minimale du filtre de déréverbération G_{min} .

[0056] Dans une étape 913, pour chaque indice fréquentiel d'échantillonnage k et chaque indice temporel n , le module de signal déréverbéré $Y_{k,n}$ et la phase $\angle X_{k,n}$ du signal complexe $X_{k,n}^C$ sont multipliés afin de créer un signal complexe déréverbéré Y^C .

[0057] Dans une étape 914, une transformation fréquence-temps est appliquée par l'unité d'application d'une transformation fréquence-temps 220 au signal complexe déréverbéré $Y_{k,n}^C$ afin d'obtenir un signal temporel déréverbéré $y(t)$ dans le domaine temporel. Dans un exemple, la transformation fréquence-temps est une Transformation de Fourier Inverse à Court Terme.

[0058] Dans une mise en oeuvre, la valeur du nombre de vecteurs d'observation L est égale à 10, la valeur du nombre de longueur d'observation N est égale à 8, la valeur du retard δ est égale à 5, la valeur du paramètre d'intensité maximale λ est égale à 0.5, la valeur du nombre K est égale à 257, la valeur du nombre J est égale à 10, la valeur de la longueur d'une trame M est égale à 512 et la valeur minimale du filtre de déréverbération G_{min} est égale à -12 décibels. Ce choix de paramètres permet l'application du procédé en temps réel.

[0059] Le procédé de suppression de la réverbération tardive d'un signal sonore selon l'invention est rapide et présente une complexité réduite. Ledit procédé est donc utilisable en temps réel. En outre, ce procédé n'introduit pas d'artefacts et est robuste au bruit de fond. De plus, ledit procédé réduit le bruit de fond et est compatible avec des procédés de réduction de bruit.

[0060] Le procédé de suppression de la réverbération tardive d'un signal sonore selon l'invention nécessite un seul microphone pour traiter avec précision la réverbération.

Revendications

1. Procédé de suppression de la réverbération tardive d'un signal sonore **caractérisé en ce qu'il** comporte les étapes suivantes :

- captation (900) d'un signal d'entrée (x) formé par la superposition de plusieurs versions retardées et atténuées du signal sonore,
- application (901) d'une transformation temps-fréquence au signal d'entrée (x) afin d'obtenir une transformée temps-fréquence complexe (X^C) du signal d'entrée (x),
- création d'un module sous échantillonné (\tilde{X}) en fréquence à partir du module de la transformée temps-fréquence complexe (X^C) du signal d'entrée (x),
- création (905) d'une pluralité de vecteurs d'observation sous échantillonnés à partir dudit module sous échantillonné (\tilde{X}) en fréquence,

EP 3 025 342 B1

- construction (906) d'une pluralité de dictionnaires d'analyse (D^a) à partir de la pluralité de vecteurs d'observation sous échantillonnés,
- calcul (907) d'une pluralité de vecteurs de prédiction (α) à partir de la pluralité de vecteurs d'observation sous échantillonnés et de la pluralité de dictionnaires d'analyse (D^a), en minimisant, pour chaque vecteur de prédiction (α), l'expression $\|\tilde{X}v - D^a\alpha\|_2$, qui est la norme euclidienne de la différence entre le vecteur d'observation sous échantillonné associé audit vecteur de prédiction (α) et du dictionnaire d'analyse (D^a) associé audit vecteur de prédiction (α) multiplié par ledit vecteur de prédiction (α), en tenant compte de la contrainte $\|\alpha\|_1 \leq \lambda$, selon laquelle la norme 1 dudit vecteur de prédiction (α) est inférieure ou égale à un paramètre d'intensité maximale de la réverbération tardive (λ),
- création (908) d'une pluralité de vecteurs d'observation à partir du module de la transformée temps-fréquence complexe (X^C) du signal d'entrée (x),
- construction (909) d'une pluralité de dictionnaires de synthèse (D^s) à partir de la concaténation de la pluralité de vecteurs d'observations,
- estimation (910) d'un spectre de réverbération tardive (X^ℓ) à partir de la multiplication de la pluralité de dictionnaires de synthèse (D^s) avec la pluralité de vecteurs de prédiction (α),
- filtrage (912) de la pluralité de vecteurs d'observations afin d'éliminer le spectre de réverbération tardive (X^ℓ) et d'obtenir un module de signal déréverbéré (Y).

2. Procédé selon la revendication 1, **caractérisé en ce que** la valeur du paramètre d'intensité maximale de la réverbération tardive (λ) est comprise entre 0 et 1.

3. Procédé selon l'une des revendications 1 ou 2, **caractérisé en ce qu'il** comporte en outre l'étape suivante :

- création (913) d'un signal complexe déréverbéré (Y^C) à partir du module de signal déréverbéré (Y) et de la phase ($\angle X$) de la transformée temps-fréquence complexe (X^C) du signal d'entrée (x).

4. Procédé selon la revendication 3, **caractérisé en ce qu'il** comporte en outre l'étape suivante :

- application (914) d'une transformation fréquence-temps au signal complexe déréverbéré (Y^C) afin d'obtenir un signal temporel déréverbéré (y).

5. Procédé selon l'une des revendications 1 à 4, **caractérisé en ce qu'il** comporte en outre une étape de construction d'un filtre de déréverbération selon le modèle

$$G = \frac{\xi}{1 + \xi} \exp\left(\int_v^\infty \frac{e^{-t}}{t} dt\right),$$

où ξ est le rapport signal à bruit *a priori*, et où la borne d'intégration v est calculée selon le modèle

$$v = \gamma \frac{\xi}{1 + \xi}$$

où γ est le rapport signal à bruit *a posteriori*.

6. Dispositif de suppression de la réverbération tardive d'un signal sonore **caractérisé en ce qu'il** comporte des moyens pour :

- capter un signal d'entrée (x) formé par la superposition de plusieurs versions retardées et atténuées du signal sonore,
- appliquer une transformation temps-fréquence au signal d'entrée (x) afin d'obtenir une transformée temps-fréquence complexe (X^C) du signal d'entrée (x),
- créer un module sous échantillonné (\tilde{X}) en fréquence à partir du module de la transformée temps-fréquence complexe (X^C) du signal d'entrée (x),
- créer d'une pluralité de vecteurs d'observation sous échantillonnés à partir dudit module sous échantillonné (\tilde{X}) en fréquence,

- construire une pluralité de dictionnaires d'analyse (D^a) à partir de la pluralité de vecteurs d'observation sous échantillonnés,
- calculer une pluralité de vecteurs de prédiction (α) à partir de la pluralité de vecteurs d'observation sous échantillonnés et de la pluralité de dictionnaires d'analyse (D^a), en minimisant, pour chaque vecteur de prédiction (α), l'expression $\|\tilde{X}_v - D^a \alpha\|_2$, qui est la norme euclidienne de la différence entre le vecteur d'observation sous échantillonné associé audit vecteur de prédiction (α) et du dictionnaire d'analyse (D^a) associé audit vecteur de prédiction (α) multiplié par ledit vecteur de prédiction (α), en tenant compte de la contrainte $\|\alpha\|_1 \leq \lambda$, selon laquelle la norme 1 dudit vecteur de prédiction (α) est inférieure ou égale à un paramètre d'intensité maximale de la réverbération tardive (λ),
- créer une pluralité de vecteurs d'observation à partir du module de la transformée temps-fréquence complexe (X^c) du signal d'entrée (x),
- construire une pluralité de dictionnaires de synthèse (D^s) à partir de la concaténation de la pluralité de vecteurs d'observations,
- estimer un spectre de réverbération tardive (X^l) à partir de la multiplication de la pluralité de dictionnaires de synthèse (D^s) avec la pluralité de vecteurs de prédiction (α),
- filtrer la pluralité de vecteurs d'observations afin d'éliminer le spectre de réverbération tardive (X^l) et d'obtenir un module de signal déréverbéré (Y).

20 Patentansprüche

1. Verfahren zur Unterdrückung des späten Nachhalls eines akustischen Signals, **dadurch gekennzeichnet, dass** es die folgenden Schritte umfasst:

- Erfassung (900) eines Eingangssignals (x), das durch die Überlagerung mehrerer verzögerter und gedämpfter Versionen des akustischen Signals gebildet ist,
- Anwendung (901) einer Zeit-Frequenz-Transformation am Eingangssignal (x), um eine komplexe Zeit-Frequenz-Transformierte (X^c) des Eingangssignals (x) zu erhalten,
- Erzeugung eines hinsichtlich der Frequenz unterabgetasteten Moduls (X) aus dem Modul der komplexen Zeit-Frequenz-Transformierten (X^c) des Eingangssignals (x),
- Erzeugung (905) einer Vielzahl von unterabgetasteten Beobachtungsvektoren aus dem hinsichtlich der Frequenz unterabgetasteten Modul (X),
- Erstellung (906) einer Vielzahl von Analysewörterbüchern (D^a) aus der Vielzahl von unterabgetasteten Beobachtungsvektoren,
- Berechnung (907) einer Vielzahl von Prädiktionsvektoren (α) aus der Vielzahl von unterabgetasteten Beobachtungsvektoren und der Vielzahl von Analysewörterbüchern (D^a), wobei für jeden Prädiktionsvektor (α) der Ausdruck $\|\tilde{X}_v - D^a \alpha\|_2$ minimiert wird, der die euklidische Norm der Differenz zwischen dem unterabgetasteten Beobachtungsvektor, der dem Prädiktionsvektor (α) zugeordnet ist, und dem Analysewörterbuch (D^a), das dem Prädiktionsvektor (α) zugeordnet ist, multipliziert mit dem Prädiktionsvektor (α), unter Berücksichtigung der Anforderung $\|\alpha\|_1 \leq \lambda$, nach der die Norm 1 des Prädiktionsvektors (α) kleiner oder gleich einem Parameter maximaler Intensität des späten Nachhalls (λ) ist,
- Erzeugung (908) einer Vielzahl von Beobachtungsvektoren aus dem Modul der komplexen Zeit-Frequenz-Transformierten (X^c) des Eingangssignals (x),
- Erstellung (909) einer Vielzahl von Synthesewörterbüchern (D^s) aus der Verkettung der Vielzahl von Beobachtungsvektoren,
- Schätzung (910) eines späten Nachhallspektrums (X^l) aus der Multiplikation der Vielzahl von Synthesewörterbüchern (D^s) mit der Vielzahl von Prädiktionsvektoren (α),
- Filterung (912) der Vielzahl von Beobachtungsvektoren, um das späte Nachhallspektrum (X^l) zu eliminieren und ein Signalmodul ohne Nachhall (Y) zu erhalten.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** der Wert des maximalen Intensitätsparameters des späten Nachhalls (λ) zwischen 0 und 1 beträgt.

3. Verfahren nach einem der Ansprüche 1 oder 2, **dadurch gekennzeichnet, dass** es ferner den folgenden Schritt umfasst:

- Erzeugung (913) eines komplexen Signals ohne Nachhall (Y^c) aus dem Signalmodul ohne Nachhall (Y) und der Phase ($\angle X$) der komplexen Zeit-Frequenz-Transformierten (X^c) des Eingangssignals (x).

4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet, dass** es ferner den folgenden Schritt umfasst:

- Anwendung (914) einer Zeit-Frequenz-Transformierten am komplexen Signal ohne Nachhall (Y^c), um ein Zeitsignal ohne Nachhall (y) zu erhalten.

5. Verfahren nach einem der Ansprüche 1 bis 4, **dadurch gekennzeichnet, dass** es ferner den Schritt der Erstellung eines Nachhallbeseitigungsfilters nach folgendem Modell umfasst:

$$G = \frac{\xi}{1+\xi} \exp\left(\int_v^\infty \frac{e^{-t}}{t} dt\right),$$

wobei ζ das Signal-Rausch-Verhältnis a priori ist, und wobei die Integrationsklemme v nach folgendem Modell berechnet wird

$$v = \gamma \frac{\xi}{1+\xi}$$

wobei γ das Signal-Rausch-Verhältnis a posteriori ist.

6. Vorrichtung zur Unterdrückung des späten Nachhalls eines akustischen Signals, **dadurch gekennzeichnet, dass** sie Mittel umfasst, um:

- ein Eingangssignal (x) zu erfassen, das durch die Überlagerung mehrerer verzögerter und gedämpfter Versionen des akustischen Signals gebildet ist,
- eine Zeit-Frequenz-Transformation am Eingangssignal (x) anzuwenden, um eine komplexe Zeit-Frequenz-Transformierte (X^c) des Eingangssignals (x) zu erhalten,
- ein hinsichtlich der Frequenz unterabgetastetes Modul (X) aus dem Modul der komplexen Zeit-Frequenz-Transformierten (X^c) des Eingangssignals (x) zu erzeugen,
- eine Vielzahl von unterabgetasteten Beobachtungsvektoren aus dem hinsichtlich der Frequenz unterabgetasteten Modul (\tilde{X}) zu erzeugen,
- eine Vielzahl von Analysewörterbüchern (D^a) aus der Vielzahl von unterabgetasteten Beobachtungsvektoren zu erstellen,
- eine Vielzahl von Prädiktionsvektoren (α) aus der Vielzahl von unterabgetasteten Beobachtungsvektoren und der Vielzahl von Analysewörterbüchern (D^a), wobei für jeden Prädiktionsvektor (α) der Ausdruck $\|\tilde{X}v - D^a\alpha\|_2$ minimiert wird, der die euklidische Norm der Differenz zwischen dem unterabgetasteten Beobachtungsvektor, der dem Prädiktionsvektor (α) zugeordnet ist, und dem Analysewörterbuch (D^a), das dem Prädiktionsvektor (α) zugeordnet ist, multipliziert mit dem Prädiktionsvektor (α), unter Berücksichtigung der Anforderung $\|\alpha\|_1 \leq \lambda$, nach der die Norm 1 des Prädiktionsvektors (α) kleiner oder gleich einem Parameter maximaler Intensität des späten Nachhalls (λ) ist,
- eine Vielzahl von Beobachtungsvektoren aus dem Modul der komplexen Zeit-Frequenz-Transformierten (X^c) des Eingangssignals (x) zu erzeugen,
- eine Vielzahl von Synthesewörterbüchern (D^s) aus der Verkettung der Vielzahl von Beobachtungsvektoren zu erstellen,
- ein spätes Nachhallspektrum (X^l) aus der Multiplikation der Vielzahl von Synthesewörterbüchern (D^s) mit der Vielzahl von Prädiktionsvektoren (α) zu schätzen,
- die Vielzahl von Beobachtungsvektoren zu filtern, um das späte Nachhallspektrum (X^l) zu eliminieren und ein Signalmodul ohne Nachhall (Y) zu erhalten.

Claims

1. Method for suppressing the late reverberation of a sound signal, **characterized in that** it includes the following steps:

- capturing (900) an input signal (x) formed by the superimposition of a multiplicity of delayed and attenuated

versions of the sound signal,

- applying (901) a time-frequency transformation to the input signal (x) in order to obtain a complex time-frequency transform (X^C) of the input signal (x),
- generating a frequency-subsampled modulus (\tilde{X}) from the modulus of the complex time-frequency transform (X^C) of the input signal (x),
- generating (905) a plurality of subsampled observation vectors from said frequency-subsampled modulus (\tilde{X}),
- constructing (906) a plurality of analysis dictionaries (D^a) from the plurality of subsampled observation vectors,
- calculating (907) a plurality of prediction vectors (α) from the plurality of subsampled observation vectors and from the plurality of analysis dictionaries (D^a), by minimizing, for each prediction vector (α), the expression $\|\tilde{X}v - D^a\alpha\|_2$, which is the Euclidean norm of the difference between the subsampled observation vector associated with said prediction vector (α) and of the analysis dictionary (D^a) associated with said prediction vector (α) multiplied by said prediction vector (α), while taking into consideration the constraint $\|\alpha\|_1 \leq \lambda$, according to which the norm 1 of said prediction vector (α) is less than or equal to a maximum intensity parameter of the late reverberation (λ),
- generating (908) a plurality of observation vectors from the modulus of the complex time-frequency transform (X^C) of the input signal (x),
- constructing (909) a plurality of synthesis dictionaries (D^s) from the concatenation of the plurality of observation vectors,
- estimating (910) a late reverberation spectrum (X^l) from the multiplication of the plurality of synthesis dictionaries (D^s) by the plurality of prediction vectors (α),
- filtering (912) the plurality of observation vectors in order to eliminate the late reverberation spectrum (X^l) and to obtain a dereverberated signal modulus (Y).

2. Method according to Claim 1, **characterized in that** the value of the maximum intensity parameter of the late reverberation (A) is between 0 and 1.

3. Method according to either of Claims 1 and 2, **characterized in that** it furthermore includes the following step:

- generating (913) a dereverberated complex signal (Y^C) from the dereverberated signal modulus (Y) and from the phase ($\angle X$) of the complex time-frequency transform (X^C) of the input signal (x).

4. Method according to Claim 3, **characterized in that** it furthermore includes the following step:

- applying (914) a frequency-time transformation to the dereverberated complex signal (Y^C) in order to obtain a dereverberated time signal (y).

5. Method according to one of Claims 1 to 4, **characterized in that** it furthermore includes a step of constructing a dereverberation filter in accordance with the model

$$G = \frac{\xi}{1 + \xi} \exp\left(\int_v^\infty \frac{e^{-t}}{t} dt\right),$$

where ξ is the *a priori* signal-to-noise ratio, and where the bound of integration v is calculated in accordance with the model

$$v = \gamma \frac{\xi}{1 + \xi}$$

where γ is the *a posteriori* signal-to-noise ratio.

6. Device for suppressing the late reverberation of a sound signal, **characterized in that** it includes means for:

- capturing an input signal (x) formed by the superimposition of a multiplicity of delayed and attenuated versions of the sound signal,

EP 3 025 342 B1

- applying a time-frequency transformation to the input signal (x) in order to obtain a complex time-frequency transform (X^C) of the input signal (x),
- generating a frequency-subsampled modulus (\tilde{X}) from the modulus of the complex time-frequency transform (X^C) of the input signal (x),
- generating a plurality of subsampled observation vectors from said frequency-subsampled modulus (\tilde{X}),
- constructing a plurality of analysis dictionaries (D^a) from the plurality of subsampled observation vectors,
- calculating a plurality of prediction vectors (α) from the plurality of subsampled observation vectors and from the plurality of analysis dictionaries (D^a), by minimizing, for each prediction vector (α), the expression $\|\tilde{X}v - D^a\alpha\|_2$, which is the Euclidean norm of the difference between the subsampled observation vector associated with said prediction vector (α) and of the analysis dictionary (D^a) associated with said prediction vector (α) multiplied by said prediction vector (α), while taking into consideration the constraint $\|\alpha\|_1 \leq \lambda$, according to which the norm 1 of said prediction vector (α) is less than or equal to a maximum intensity parameter of the late reverberation (λ),
- generating a plurality of observation vectors from the modulus of the complex time-frequency transform (X^C) of the input signal (x),
- constructing a plurality of synthesis dictionaries (D^s) from the concatenation of the plurality of observation vectors,
- estimating a late reverberation spectrum (X^l) from the multiplication of the plurality of synthesis dictionaries (D^s) by the plurality of prediction vectors (α),
- filtering the plurality of observation vectors in order to eliminate the late reverberation spectrum (X^l) and to obtain a dereverberated signal modulus (Y).

5

10

15

20

25

30

35

40

45

50

55

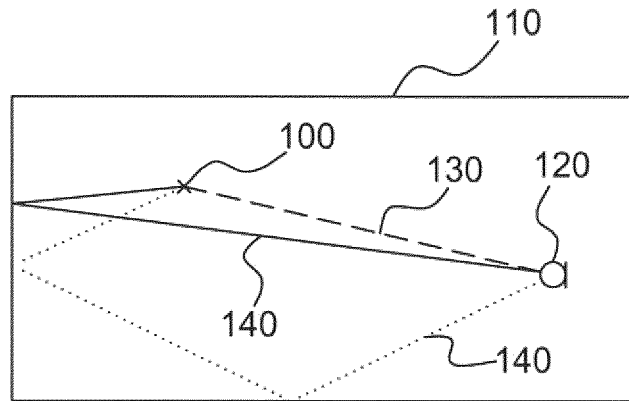


Fig.1
-Etat de la Technique-

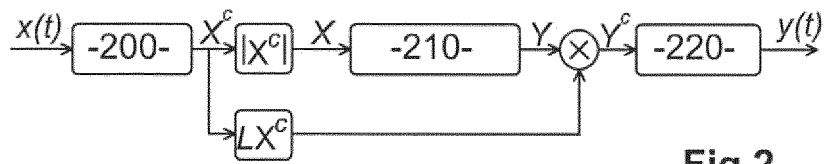


Fig.2

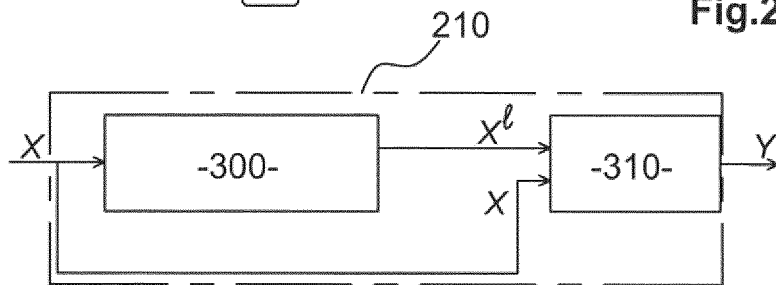


Fig.3

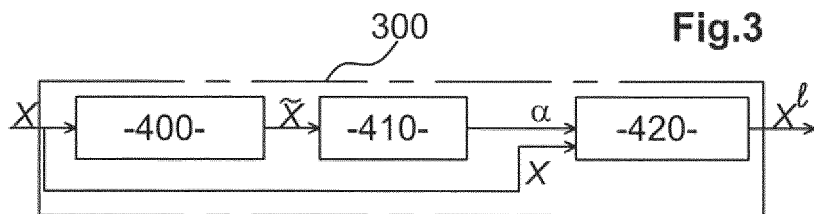
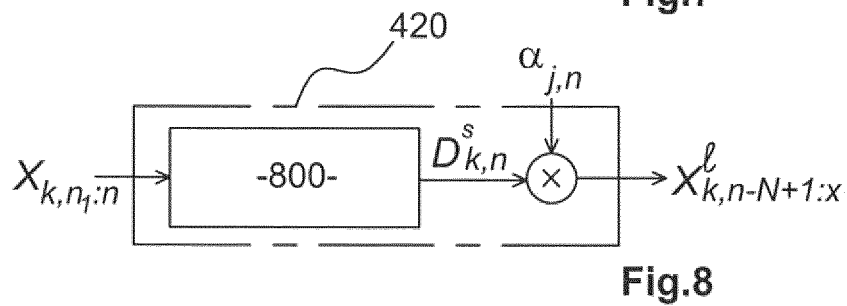
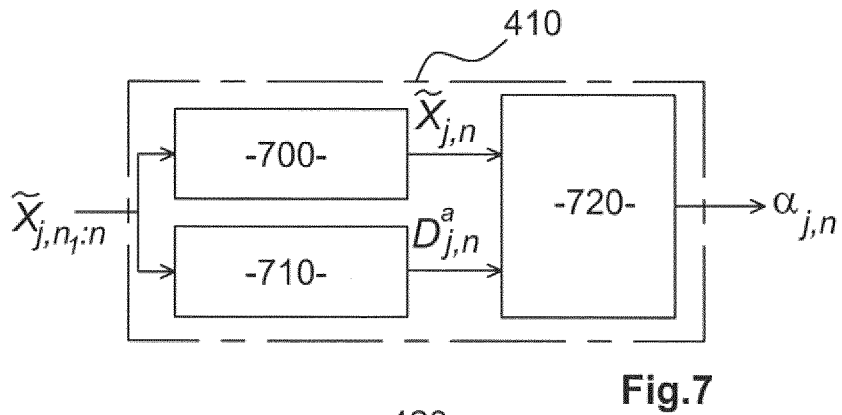
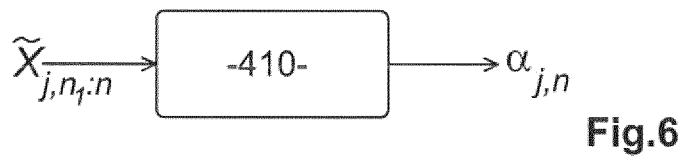
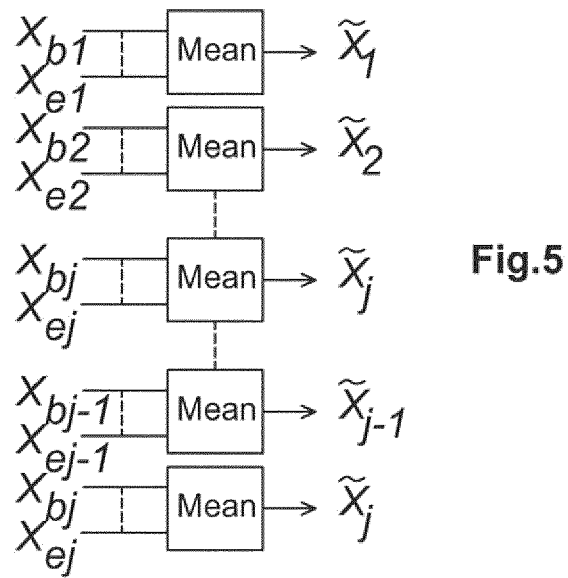


Fig.4



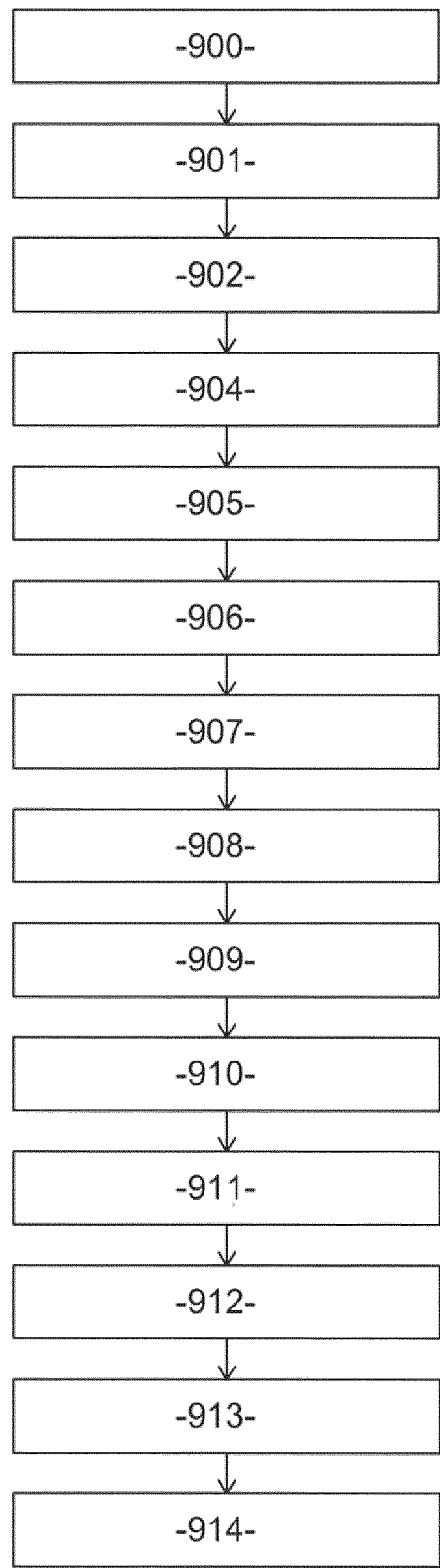


Fig.9

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.

Littérature non-brevet citée dans la description

- Speech dereverberation via maximum-kurtosis sub-band adaptive filtering, Proc. **BWGILLESPIE ; H S MALVAR ; D A F FLORÊNCIO**,. International Conference on Acoustics, Speech, and Signal Processing. IEEE, 2001, vol. 6, 3701-3704 **[0006]**
- **MWU ; DLWANG**. A two-stage algorithm for one-microphone reverberant speech enhancement, Audio, Speech, and Language Processing. *IEEE Transactions on*, 2006, vol. 14 (3), 774-784 **[0006]**
- **SAEED MOSAYYEBPOUR ; ABOLGHASEM SAYYADIYAN ; MOHSEN ZAREIAN ; ALI SHAHBAZI**. Single Channel Inverse Filtering of Room Impulse Response. *Maximizing Skewness of LP Residual* **[0006]**
- **D BEES ; M BLOSTEIN ; P KABAL**. Reverberant speech enhancement using cepstral processing, ICASSP '91. *Proceedings of the Acoustics, Speech, and Signal Processing*, 1991 **[0010]**
- E.A.P Habets, Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement. PhD thesis. Technische Universiteit Eindhoven, 2007 **[0013]**
- Speech Enhancement, Reverberant Environments. **T. YOSHIOKA**. PhD thesis. 2010 **[0013]**
- Speech Enhancement in Reverberant Environments. **T. YOSHIOKA**. PhD thesis,. 2010 **[0017]**
- Robust speech dereverberation based on nonnegativity and sparse nature of speech spectrograms, Proceedings of the 2009. **H KAMEOKA ; T NAKATANI ; T YOSHIOKA**. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '09. IEEE Computer Society, 2009, 45-48 **[0019]**