



(12) **DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:
31.10.2012 Bulletin 2012/44

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

(21) Numéro de dépôt: **12164777.0**

(22) Date de dépôt: **19.04.2012**

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

(72) Inventeurs:
• **Herve, Michael**
75011 PARIS (FR)
• **Vitte, Guillaume**
75003 PARIS (FR)

(30) Priorité: **26.04.2011 FR 1153572**

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

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

(54) **Combiné audio micro/casque comprenant des moyens de débruitage d'un signal de parole proche, notamment pour un système de téléphonie "mains libres"**

(57) Ce casque comprend : un capteur physiologique (18), apte à être couplé à la joue ou à la tempe du porteur du casque et à capter les vibrations vocales non acoustiques transmises par conduction osseuse interne ; des moyens de filtrage passe-bas (48) du signal capté ; un ensemble de microphones (20, 22) captant les vibrations vocales acoustiques transmises par voie aérienne depuis la bouche du porteur du casque ; des

moyens de filtrage passe-haut (50, 52) et de réduction de bruit (56) des signaux captés par ces microphones ; et des moyens de mixage (54), pour combiner les signaux filtrés et donner en sortie un signal représentatif de la parole émise par le porteur du casque. Le signal du capteur physiologique (18) est également utilisé par des moyens de calcul de la fréquence de coupure (44) des filtres passe-bas et passe-haut et par des moyens de calcul de probabilité d'absence de parole (64, 66).

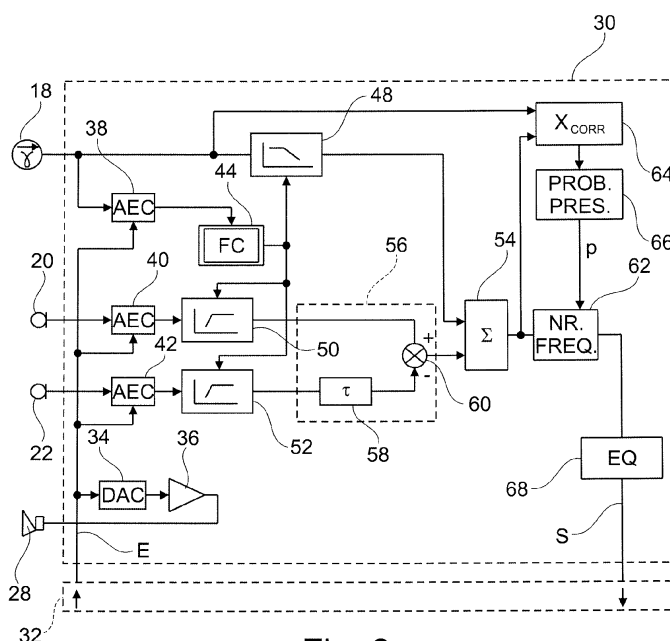


Fig. 2

Description

[0001] L'invention concerne un casque audio du type micro/casque combinés.

[0002] Un tel casque peut notamment être utilisé pour des fonctions de communication telles que des fonctions de téléphonie "mains libres", en complément de l'écoute d'une source audio (musique par exemple) provenant d'un appareil sur lequel est branché le casque.

[0003] Dans les fonctions de communication, l'une des difficultés est d'assurer une intelligibilité suffisante du signal capté par le microphone ("micro"), c'est-à-dire le signal de parole du locuteur proche (le porteur du casque). Le casque peut en effet être utilisé dans un environnement bruyant (métro, rue passante, train, etc.), de sorte que le micro captera non seulement la parole du porteur du casque, mais également les bruits parasites environnants.

[0004] Le porteur peut être protégé de ces bruits par le casque, notamment s'il s'agit d'un modèle à écouteurs fermés isolant l'oreille de l'extérieur, et encore plus si le casque est pourvu d'un "contrôle actif de bruit". En revanche le locuteur distant (celui se trouvant à l'autre bout du canal de communication) souffrira des bruits parasites captés par le micro, venant se superposer et interférer avec le signal de parole du locuteur proche (le porteur du casque).

[0005] En particulier, certains formants de la parole essentiels à la compréhension de la voix sont souvent noyés dans des composantes de bruit couramment rencontrées dans les environnements habituels, composantes qui sont majoritairement concentrées dans les basses fréquences.

[0006] Il a été proposé de recueillir certaines vibrations vocales au moyen d'un capteur physiologique appliqué contre la joue ou la tempe du porteur du casque. En effet, lorsqu'une personne émet un son voisé (c'est-à-dire une composante de parole dont la production s'accompagne d'une vibration des cordes vocales), une vibration se propage depuis les cordes vocales jusqu'au pharynx et à la cavité bucco-nasale, où elle est modulée, amplifiée et articulée. La bouche, le voile du palais, le pharynx, les sinus les fosses nasales servent de caisse de résonance ce son voisé et, leurs parois étant élastiques, elles vibrent à leur tour, et ces vibrations sont transmises par conduction osseuse interne et sont perceptibles au niveau de la joue et de la tempe.

[0007] Ces vibrations vocales au niveau de la joue et de la tempe présentent la caractéristique d'être, par nature, très peu corrompues par le bruit environnant : en effet, en présence de bruits extérieurs, les tissus de la joue et de la tempe ne vibrent quasiment pas et ceci, quelle que soit la composition spectrale du bruit extérieur.

[0008] Par ailleurs, en raison du filtrage engendré par la propagation des vibrations jusqu'à la tempe, le signal recueilli par le capteur physiologique est utilisable uniquement dans les basses fréquences. Mais comme les bruits généralement rencontrés dans un environnement habituel (rue, métro, train ...) sont majoritairement concentrés dans les basses fréquences, le capteur physiologique délivre un signal naturellement dépourvu de composante parasite de bruit (ce qui n'est pas possible avec un micro classique).

[0009] Le JP 2000-261534 A décrit un tel combiné micro-casque comprenant :

- deux écouteurs comportant chacun un transducteur de reproduction sonore d'un signal audio ;
- un capteur physiologique, apte à venir en contact avec la joue ou la tempe du porteur du casque pour y être couplé et capter les vibrations vocales non acoustiques transmises par conduction osseuse interne, ce capteur physiologique délivrant un premier signal de parole ;
- un ensemble microphonique, comprenant au moins un microphone apte à capter les vibrations vocales acoustiques transmises par voie aérienne depuis la bouche du porteur du casque, cet ensemble microphonique délivrant un second signal de parole ; et
- des moyens de mixage, pour combiner le premier signal de parole et le second signal de parole, et donner en sortie un troisième signal de parole représentatif de la parole émise par le porteur du casque.

[0010] Le EP 0 683 621 A2, quant à lui, décrit plus précisément la manière d'intégrer à une même oreillette intra-auriculaire à la fois le capteur physiologique et le microphone extérieur.

[0011] Bien sûr, le signal recueilli par le capteur physiologique n'est pas à proprement parler de la parole puisque la parole n'est pas seulement formée de sons voisés, elle contient des composantes qui ne naissent pas au niveau des cordes vocales : le contenu fréquentiel est par exemple beaucoup plus riche avec le son provenant de la gorge et émis par la bouche. De plus, la conduction osseuse interne et la traversée de la peau a pour effet de filtrer certaines composantes vocales, qui fait que le signal délivré par le capteur physiologique n'est exploitable que dans la partie la plus basse du spectre. C'est pour cela que ce signal est complété par un autre signal, délivré par un capteur microphonique conventionnel, auquel il est combiné.

[0012] Le problème général de l'invention est, dans un tel contexte, de délivrer au locuteur distant un signal vocal représentatif de la parole émise par le locuteur proche, signal qui soit débarrassé des composantes parasites de bruits extérieurs présents dans l'environnement du locuteur proche.

[0013] Un aspect important de ce problème est la nécessité de restituer un signal de parole naturel et intelligible, c'est-à-dire non distordu et dont la plage des fréquences utiles ne soit pas amputée par les traitements de combinaison

des signaux issus de capteurs exploitant des vibrations qui sont de nature différente et transmises par des voies différentes.

[0014] Un autre aspect de l'invention réside dans la possibilité d'utiliser de façon efficace le signal issu du capteur physiologique pour contrôler diverses fonctions de traitement du signal. Ce signal permet en effet d'accéder à de nouvelles informations concernant le contenu de la parole, qui seront ensuite utilisées pour le débruitage ainsi que pour diverses fonctions auxiliaires que l'on exposera plus bas, notamment le calcul d'une fréquence de coupure d'un filtre dynamique.

[0015] Pour résoudre ces problèmes, l'invention propose un combiné micro/casque du type exposé ci-dessus tel qu'enseigné par le JP 2000-261534 A et dans lequel, de façon caractéristique de l'invention :

- le capteur physiologique est incorporé à un coussinet circumaural d'une coque de l'un des écouteurs ;
- l'ensemble microphonique comprend deux microphones placés sur la coque de l'un des écouteurs ;
- les deux microphones sont alignés en un réseau linéaire suivant une direction principale dirigée vers la bouche du porteur du casque ; et
- il est prévu des moyens de réduction de bruit non fréquentielle du second signal de parole, comprenant un combineur apte à appliquer un retard au signal délivré par l'un des microphones et à soustraire ce signal retardé du signal délivré par l'autre microphone, de manière à opérer un débruitage du signal de parole proche émis par le porteur du casque.

[0016] Avantageusement, le combiné micro/casque comprend des moyens de filtrage passe-bas du premier signal de parole avant combinaison par les moyens de mixage, et/ou des moyens de filtrage passe-haut du second signal de parole avant débruitage et combinaison par les moyens de mixage. Ces moyens de filtrage passe-bas et/ou passe-haut comprennent un filtre à fréquence de coupure ajustable, et le casque comprend des moyens de calcul de la fréquence de coupure, opérant en fonction du signal délivré par le capteur physiologique. Les moyens de calcul de la fréquence de coupure peuvent en particulier comprendre des moyens d'analyse du contenu spectral du signal délivré par le capteur physiologique, aptes à déterminer la fréquence de coupure en fonction des niveaux relatifs du rapport signal/bruit évalué dans une pluralité de bandes de fréquences distinctes du signal délivré par le capteur physiologique.

[0017] De préférence, les moyens de débruitage du second signal de parole sont des moyens de réduction de bruit non fréquentielle avec, dans une forme de réalisation particulière de l'invention, l'ensemble microphonique qui comprend deux microphones, et les moyens de réduction de bruit non fréquentielle qui comprennent un combineur apte à appliquer un retard au signal délivré par l'un des microphones et à soustraire ce signal retardé du signal délivré par l'autre microphone.

[0018] Les deux microphones peuvent en particulier être alignés en un réseau linéaire suivant une direction principale dirigée vers la bouche du porteur du casque.

[0019] De préférence également, il est prévu des moyens de débruitage du troisième signal de parole délivré par les moyens de mixage, notamment des moyens de réduction de bruit fréquentielle.

[0020] À cet effet, et selon un aspect original de l'invention, il est prévu des moyens recevant en entrée, et opérant une intercorrélacion entre, le premier et le troisième signal de parole, et délivrant en sortie un signal de probabilité de présence de parole fonction du résultat de l'intercorrélacion. Les moyens de débruitage du troisième signal de parole reçoivent en entrée ce signal de probabilité de présence de parole pour, sélectivement : i) opérer une réduction de bruit différenciée selon les bandes de fréquences en fonction de la valeur du signal de probabilité de présence de parole, et ii) opérer une réduction de bruit maximale sur toutes les bandes de fréquences en l'absence de parole.

[0021] Il peut en outre être prévu des moyens de post-traitement, aptes à opérer une égalisation sélective par bandes de fréquences dans la partie du spectre correspondant au signal recueilli par le capteur physiologique. Ces moyens déterminent un gain d'égalisation pour chacune des bandes de fréquences, ce gain étant calculé à partir des coefficients fréquentiels respectifs des signaux délivrés par le(s) microphones et des signaux délivrés par le capteur physiologique, considérés dans le domaine fréquentiel. Ils opèrent en outre un lissage sur une pluralité des trames successives de signal du gain d'égalisation calculé.

[0022] On va maintenant décrire un exemple de mise en oeuvre du dispositif de l'invention, en référence aux dessins annexés où les mêmes références numériques désignent d'une figure à l'autre des éléments identiques ou fonctionnellement semblables.

La Figure 1 illustre de façon générale le casque de l'invention, posé sur la tête d'un utilisateur.

La Figure 2 est un schéma d'ensemble, sous forme de blocs fonctionnels, expliquant la manière dont est réalisé le traitement du signal permettant de délivrer en sortie un signal débruité représentatif de la parole émise par le porteur du casque.

La Figure 3 est une représentation spectrale amplitude/fréquence illustrant le calcul d'intercorrélacion servant à évaluer une probabilité de présence de parole.

La Figure 4 est une représentation spectrale amplitude/fréquence illustrant le traitement final d'égalisation automa-

tique opéré après la réduction de bruit.

[0023] Sur la Figure 1, la référence 10 désigne de façon générale le casque selon l'invention, qui comporte deux oreillettes 12 réunies par un arceau. Chacune des oreillettes est de préférence constituée d'une coque fermée 12, logeant un transducteur de reproduction sonore, appliquée autour de l'oreille de l'utilisateur avec interposition d'un coussinet isolant 16 isolant l'oreille de l'extérieur.

[0024] Ce casque est pourvu d'un capteur physiologique 18 permettant de recueillir les vibrations produites par un signal voisé émis par le porteur du casque, et qui peuvent être captées au niveau de la joue ou de la tempe. Le capteur 18 est de préférence un accéléromètre intégré dans le coussinet 16 de manière à venir s'appliquer contre la joue ou la tempe de l'utilisateur avec un couplage le plus étroit possible. Le capteur physiologique peut notamment être placé sur la face intérieure de la peau du coussinet de sorte que, une fois le casque mis en place, le capteur physiologique soit appliqué contre la joue ou la tempe de l'utilisateur sous l'effet d'une légère pression résultant de l'écrasement du matériau du coussinet, avec seulement interposition de la peau du coussinet.

[0025] Le casque comporte également un réseau ou antenne de microphones, par exemple deux micros omnidirectionnels 20, 22, placés sur la coque de l'écouteur 12. Ces deux micros avant 20 et arrière 22 sont des micros omnidirectionnels disposés l'un par rapport à l'autre de manière que leur direction d'alignement 24 soit approximativement dirigée vers la bouche 26 du porteur du casque.

[0026] La Figure 2 est un schéma par blocs montrant les différents blocs et fonctions mis en oeuvre par le procédé de l'invention ainsi que leurs interactions.

[0027] Le procédé de l'invention est mis en oeuvre par des moyens logiciels, qu'il est possible de décomposer et schématiser par un certain nombre de blocs 30 à 64 illustrés Figure 2. Ces traitements sont mis en oeuvre sous forme d'algorithmes appropriés exécutés par un microcontrôleur ou un processeur numérique de signal. Bien que, pour la clarté de l'exposé, ces divers traitements soient présentés sous forme de blocs distincts, ils mettent en oeuvre des éléments communs et correspondent en pratique à une pluralité de fonctions globalement exécutées par un même logiciel.

[0028] On retrouve sur cette figure le capteur physiologique 18 et les deux micros omnidirectionnels avant 20 et arrière 22. La référence 28 désigne par ailleurs le transducteur de reproduction sonore placé à l'intérieur de la coque de l'écouteur. Ces divers éléments délivrent des signaux qui font l'objet d'un traitement par le bloc référencé 30, qui peut être couplé à une interface 32 aux circuits de communication (circuits téléphoniques) et reçoit en entrée E le son destiné à être reproduit par le transducteur 28 (parole du locuteur distant pendant une communication téléphonique, source musicale hors des périodes de communication téléphonique), et délivre sur la sortie S un signal représentatif de la parole du locuteur proche, c'est-à-dire du porteur du casque.

[0029] Le signal à reproduire appliqué sur l'entrée E est un signal numérique converti en analogique par le convertisseur 34, puis amplifié par l'amplificateur 36 pour reproduction par le transducteur 28.

[0030] On va maintenant décrire la manière dont est produit le signal débruité représentatif de la parole du locuteur proche, à partir des signaux respectifs recueillis par le capteur physiologique 18 et les micros 20 et 22.

[0031] Le signal recueilli par le capteur physiologique 18 est un signal comprenant principalement des composantes dans la région inférieure du spectre sonore (typiquement 0-1500 Hz). Comme on l'a expliqué plus haut, ce signal est naturellement non bruité.

[0032] Les signaux recueillis par les micros 20, 22 seront utilisés principalement pour le haut du spectre (au-dessus de 1500 Hz), mais ces signaux sont fortement bruités et il sera indispensable d'opérer un traitement de débruitage fort pour en éliminer les composantes de bruit parasites, dont le niveau peut être tel, dans certains environnements, qu'elles occultent complètement le signal de parole capté par ces micros 20, 22.

[0033] La première étape du traitement est un traitement anti-écho, appliqué aux signaux du capteur physiologique et des micros.

[0034] En effet, le son reproduit par le transducteur 28 est capté par le capteur physiologique 18 et les micros 20, 22, générant un écho qui perturbe le fonctionnement du système, et qui doit donc être éliminé en amont.

[0035] Ce traitement anti-écho est mis en oeuvre par les blocs 38, 40 et 42, chacun de ces blocs recevant sur une première entrée le signal émis par le capteur 18, 20 ou bien 22 et sur une seconde entrée le signal reproduit par le transducteur 28 (signal générateur d'écho), et délivre en sortie, pour traitement ultérieur, un signal dont l'écho a été éliminé.

[0036] Le traitement anti-écho est par exemple réalisé par un traitement à algorithme adaptatif tel que celui décrit dans le FR 2 792 146 A1 (Parrot SA), auquel on pourra se reporter pour plus de détails. Il s'agit d'une technique d'annulation d'écho ou AEC consistant à définir de façon dynamique un filtre de compensation modélisant le couplage acoustique entre le transducteur 28 et le capteur physiologique 18 (ou le micro 20, ou le micro 22, respectivement) par une transformation linéaire entre le signal reproduit par le transducteur 28 (c'est-à-dire le signal E appliqué en entrée des blocs 38, 40 ou 42) et l'écho capté par le capteur physiologique 18 (ou le micro 20 ou 22). Cette transformation définit un filtre adaptatif qui est appliqué au signal incident reproduit, et le résultat de ce filtrage est soustrait du signal

recueilli par le capteur physiologique 18 (ou le micro 20 ou 22), ce qui a pour effet d'annuler la majeure partie de l'écho acoustique.

[0037] Cette modélisation repose sur la recherche d'une corrélation entre le signal reproduit par le transducteur 28 et le signal recueilli par le capteur physiologique 18 (ou le micro 20 ou 22), c'est-à-dire sur une estimation de la réponse impulsienne du couplage constituée par le corps de l'écouteur 12 supportant ces divers éléments.

[0038] Le traitement est notamment opéré par un algorithme de type APA (*Affine Projection Algorithm*) adaptatif, qui assure une convergence rapide, bien adaptée à des applications de type "mains libres" avec un débit vocal intermittent et un niveau pouvant rapidement varier.

[0039] Avantagusement, l'algorithme itératif est exécuté avec un pas variable, comme décrit dans le FR 2 792 146 A1 précité. Avec cette technique, le pas μ varie de façon continue en fonction des niveaux d'énergie du signal capté par le micro, avant et après filtrage. Ce pas est accru lorsque l'énergie du signal capté est dominée par l'énergie de l'écho, et, inversement, réduit lorsque l'énergie du signal capté est dominée par l'énergie du bruit de fond et/ou de la parole du locuteur distant.

[0040] Le signal recueilli par le capteur physiologique 18 après le traitement anti-écho par le bloc 38 sera utilisé comme signal d'entrée d'un bloc 44 de calcul d'une fréquence de coupure FC.

[0041] L'étape suivante consiste à opérer un filtrage des signaux, avec un filtre passe-bas 48 pour le signal du capteur physiologique 18 et avec un filtre passe-haut 50, 52 pour les signaux recueillis par les micros 20, 22, respectivement.

[0042] Ces filtres 48, 50 et 52 sont de préférence des filtres numériques du type à réponse impulsienne infinie IIR (filtres récurrents), qui présentent une transition relativement abrupte entre la bande passante et la bande rejetée.

[0043] Avantagusement, ces filtres sont des filtres adaptatifs dont la fréquence de coupure est variable et déterminée dynamiquement par le bloc 44.

[0044] Ceci permet d'adapter le filtrage aux conditions particulières d'utilisation du casque : voix plus ou moins haute du porteur lorsqu'il parle, couplage plus ou moins étroit entre le capteur physiologique 18 et la joue ou la tempe du porteur, etc. La fréquence de coupure FC, qui est de préférence la même pour le filtre passe-bas 48 et les filtres passe-haut 50 et 52, est déterminée à partir du signal du capteur physiologique 18 après le traitement anti-écho 38. Pour cela, un algorithme calcule le rapport signal/bruit pour plusieurs bandes de fréquences situées dans une plage comprise entre par exemple 0 et 2500 Hz (le niveau de bruit étant donné par un calcul de l'énergie dans une bande de fréquences plus haute, par exemple entre 3000 et 4000 Hz, car l'on sait que dans cette zone le signal ne peut être que du bruit, du fait des propriétés du composant constituant le capteur physiologique 18). La fréquence de coupure choisie correspondra à la fréquence maximale pour laquelle le rapport signal/bruit dépasse un seuil prédéterminé, par exemple 10 dB.

[0045] L'étape suivante consiste à opérer au moyen du bloc 54 un mixage pour reconstruire le spectre complet avec, d'une part, la région inférieure du spectre donnée par le signal filtré du capteur physiologique 18 et, d'autre part, le haut du spectre donné par le signal filtré des micros 20 et 22 après passage dans un combineur-déphaseur 56 permettant d'opérer un débruitage dans cette partie du spectre. Cette reconstruction est opérée par sommation des deux signaux, qui sont appliqués en synchronisme au bloc de mixage 54 de manière à éviter toute déformation.

[0046] On va maintenant décrire plus précisément la manière dont est opérée la réduction du bruit par le combineur-déphaseur 56.

[0047] Le signal que l'on souhaite débruiter (c'est-à-dire le signal du locuteur proche situé dans la partie haute du spectre, typiquement les composantes de fréquence supérieure à 1500 Hz) est issu des deux micros 20, 22 disposés à quelques centimètres l'un de l'autre sur la coque 14 de l'un des écouteurs du casque. Comme on l'a indiqué, ces deux micros sont disposés l'un par rapport à l'autre de manière que la direction 24 qu'ils définissent soit approximativement orientée dans la direction de la bouche 26 du porteur du casque. De ce fait, un signal de parole émis depuis la bouche atteindra le micro avant 20 puis le micro arrière 22 avec un retard, et donc un déphasage, sensiblement constant, tandis que les bruits ambiants seront captés sans déphasage par les deux micros 20 et 22 (qui sont des micros omnidirectionnels), compte tenu de l'éloignement des sources de bruits parasites par rapport aux deux micros 20 et 22.

[0048] La réduction de bruit sur les signaux captés par les micros 20 et 22 n'est pas opérée dans le domaine fréquentiel (comme cela est souvent le cas), mais dans le domaine temporel, au moyen du combineur-déphaseur 56 qui comprend un déphaseur 58 appliquant un retard τ au signal du micro arrière 22 et un combineur 60 permettant de soustraire ce signal retardé au signal issu du micro avant 20.

[0049] On constitue ainsi un réseau différentiel de micros du premier ordre, équivalent à un micro virtuel unique dont la directivité pourra être ajustée en fonction de la valeur de τ , avec $0 \leq \tau \leq \tau_A$ (τ_A étant la valeur correspondant au déphasage naturel entre les deux micros 20 et 22, égale à la distance entre les deux micros divisée par la vitesse du son, soit un retard d'environ 30 μ s pour un espacement de 1 cm). Une valeur $\tau = \tau_A$ donnera un diagramme de directivité cardioïde, une valeur $\tau = \tau_A/3$ un diagramme hypercardioïde, et une valeur $\tau = 0$ un diagramme dipolaire. On peut obtenir par un choix approprié de ce paramètre une atténuation d'environ 6 dB sur des bruits diffus environnants. Pour plus de détails sur cette technique, on pourra par exemple se référer à :

[0050] [1] M. Buck et M. Rößler, First Order Differential Microphones Arrays for Automotive Applications, Proceedings of the 7th International Workshop on Acoustic echo and Noise control (IWAENC), Darmstadt, 10-13 Sept 2001.

On va maintenant décrire les traitements opérés sur le signal global (haut et bas du spectre) délivré en sortie des moyens de mixage 54.

[0051] Ce signal est soumis par le bloc 62 à une réduction de bruit fréquentielle.

[0052] De préférence, cette réduction de bruit fréquentielle est opérée de façon différente en présence ou en l'absence de parole, en évaluant une probabilité p d'absence de parole à partir du signal recueilli par le capteur physiologique 18.

[0053] Avantagusement, cette probabilité d'absence de parole est dérivée de l'information donnée par le capteur physiologique.

[0054] En effet, comme on l'a indiqué plus haut, le signal délivré par ce capteur présente un très bon rapport signal/bruit jusqu'à la fréquence de coupure FC déterminée par le bloc 44. Mais au-delà de cette fréquence de coupure le rapport signal/bruit reste encore bon, et souvent meilleur que celui des micros 20 et 22. L'information du capteur est exploitée en calculant (bloc 64) l'intercorrélation fréquentielle entre le signal combiné délivré par le bloc de mixage 54 et le signal non filtré du capteur physiologique, avant filtrage par le filtre passe-bas 48.

[0055] Ainsi, pour chaque fréquence f comprise par exemple entre FC et 4000 Hz, et pour chaque trame n , le calcul suivant est réalisé par le bloc 64 :

$$InterCorrelation(n, f) = \alpha_{InterCorr} \cdot InterCorrelation(n - 1, f) + (1 - \alpha_{InterCorr}) \cdot \overline{Smix(f)} \cdot \overline{Sacc(f)}$$

[0056] $Smix(f)$ et $sacc(f)$ étant les représentations vectorielles (complexes) fréquentielles, pour la trame n , respectivement du signal combiné délivré par le bloc de mixage 54, et du signal du capteur physiologique 18.

[0057] Pour évaluer une probabilité d'absence de parole, l'algorithme recherche les fréquences pour lesquelles il n'y a que du bruit (situation d'absence de parole) : sur le spectrogramme du signal délivré par le bloc de mixage 54 certaines harmoniques sont noyées dans le bruit, alors qu'elles ressortent plus sur le signal du capteur physiologique.

[0058] Le calcul d'intercorrélation par la formule décrite ci-dessus produit un résultat dont la figure 3 montre un exemple, dans le domaine fréquentiel.

[0059] Les pics $P_1, P_2, P_3, P_4, \dots$ de ce calcul d'intercorrélation indiquent une forte corrélation entre le signal combiné délivré par le bloc de mixage 54, et le signal du capteur physiologique 18, et l'émergence de ces fréquences corrélées indique la présence probable de parole pour ces fréquences.

[0060] Pour obtenir une probabilité d'absence de parole (bloc 66), on considère la valeur complémentaire :

$$AbsProba(n, f) = 1 - InterCorrelation(n, f) / coefficient_normalisation$$

[0061] La valeur *coefficient_normalisation* permet de régler la répartition des probabilités en fonction de la valeur de l'intercorrélation, et obtenir des valeurs entre 0 et 1.

[0062] La probabilité p d'absence de parole ainsi obtenue est appliquée au bloc 62 qui opère sur le signal délivré par le bloc de mixage 54 une réduction de bruit fréquentielle de façon sélective par rapport à un seuil donné de probabilité d'absence de parole :

- en l'absence probable de parole, la réduction de bruit est appliquée sur toutes les bandes de fréquences, c'est-à-dire que le gain maximal de réduction est appliqué de la même façon sur toutes les composantes du signal (puisque dans ce cas celui-ci ne contient vraisemblablement pas de composante utile) ;
- en revanche, en présence probable de parole, la réduction de bruit est une réduction de bruit fréquentielle appliquée sélectivement selon les différentes bandes de fréquences en fonction de la valeur p de la probabilité de présence de parole, selon un schéma classique, par exemple comparable à celui décrit dans le WO 2007/099222 A1 (Parrot).

[0063] Le système que l'on vient de décrire permet d'obtenir d'excellentes performances globales, typiquement de l'ordre de 30 à 40 dB de réduction de bruit sur le signal de parole du locuteur proche. Grâce à l'élimination de tous les bruits parasites, notamment les plus gênants (train, métro, etc.) qui sont concentrés dans les basses fréquences, cela donne l'impression au locuteur distant (celui avec lequel le porteur du casque est en communication) que son interlocuteur (le porteur du casque) se trouve dans une pièce silencieuse.

[0064] Enfin, il est avantageux d'appliquer au signal une égalisation finale (bloc 68), notamment sur le bas du spectre.

[0065] En effet, le contenu basse fréquence recueilli au niveau de la joue ou de la tempe par le capteur physiologique 18 est différent du contenu basse fréquence du son émis par la bouche de l'utilisateur, tel qu'il serait capté par un micro situé à quelques centimètres de la bouche, ou même par l'oreille d'un interlocuteur. L'utilisation du capteur physiologique

et le filtrage que l'on a décrit plus haut permet certes d'obtenir un signal très bon en termes de rapport signal/bruit, mais qui peut présenter pour l'interlocuteur qui l'entend un timbre un peu sourd et peu naturel.

[0066] Pour pallier cette difficulté, il est avantageux d'opérer une égalisation du signal de sortie avec des gains ajustés sélectivement sur différentes bandes de fréquences dans la région du spectre correspondant au signal recueilli par le capteur physiologique. L'égalisation peut être réalisée de manière automatique, à partir du signal délivré par les micros 20, 22, avant filtrage.

[0067] La Figure 4 montre un exemple, dans le domaine fréquentiel (donc après transformée de Fourier) du signal ACC produit par le capteur physiologique 18, par rapport à un signal microphonique MIC qui serait capté à quelques centimètres de la bouche.

[0068] De manière à optimiser le rendu du signal recueilli par le capteur physiologique, des gains différenciés $G_1, G_2, G_3, G_4, \dots$ sont appliqués à différentes bandes de fréquences de la partie du spectre située dans les basses fréquences.

[0069] Ces gains sont évalués par comparaison des signaux captés, dans une bande de fréquences commune, à la fois par le capteur physiologique 18 et par les micros 20 et/ou 22.

[0070] Plus précisément, l'algorithme calcule les transformées de Fourier respectives de ces deux signaux, donnant une série de coefficients fréquentiels (exprimés en dB) $NormPhysioFreq_dB(i)$ et $NormMicFreq_dB(i)$, correspondant respectivement à la norme du $i^{\text{ème}}$ coefficient de Fourier du signal du capteur physiologique et à la norme du $i^{\text{ème}}$ coefficient Fourier du signal microphonique.

[0071] Pour chaque coefficient fréquentiel de rang i , si la différence :

$$DifferenceFreq_dB(i) = NormPhysioFreq_dB(i) - NormMicFreq_dB(i).$$

est positive, le gain qui sera appliqué sera inférieur à l'unité (négatif en dB) ; réciproquement si la différence est négative le gain à appliquer sera supérieur à l'unité (positif en dB).

[0072] Si le gain était appliqué tel quel, les différences n'étant pas exactement constantes d'une trame à une autre, notamment lorsqu'il ne s'agit pas de sons voisés, il y aurait des variations importantes d'égalisation dans le timbre. Pour éviter ces variations, l'algorithme opère un lissage de la différence, qui permet d'affiner l'égalisation :

$$Gain_dB(i) = \lambda \cdot Gain_dB(i) - (1-\lambda) DifferenceFreq_dB(i).$$

[0073] Plus le coefficient λ sera proche de 1, moins l'information de la trame courante sera prise en compte pour le calcul du gain du $i^{\text{ème}}$ coefficient. Inversement, plus le coefficient λ sera proche de 0, plus l'information instantanée sera prise en compte. En pratique, pour que le lissage soit efficace, on prendra une valeur λ proche de 1, par exemple $\lambda = 0,99$. Le gain appliqué sur chaque bande de fréquences du signal issu du capteur physiologique donnera, pour la $i^{\text{ème}}$ fréquence modifiée :

$$NormPhysioFreq_dB_corrigée(i) = NormPhysioFreq_dB(i) + Gain_dB(i)$$

[0074] C'est cette norme qui sera utilisée par l'algorithme d'égalisation.

[0075] L'application de gains différenciés permet de rendre plus naturel le signal de parole dans le bas du spectre. Une étude subjective a montré que, dans un environnement silencieux et lorsqu'une telle égalisation est appliquée, la différence entre un signal microphonique de référence et le signal produit par le capteur physiologique dans le bas du spectre est pratiquement imperceptible.

Revendications

1. Un casque audio (10) du type combiné micro/casque, comprenant :

- deux écouteurs (12) comportant chacun un transducteur (28) de reproduction sonore d'un signal audio ;
- un capteur physiologique (18), apte à venir en contact avec la joue ou la tempe du porteur du casque pour y être couplé et capter les vibrations vocales non acoustiques transmises par conduction osseuse interne, ce

capteur physiologique délivrant un premier signal de parole ;

- un ensemble microphonique, comprenant au moins un microphone (20, 22) apte à capter les vibrations vocales acoustiques transmises par voie aérienne depuis la bouche du porteur du casque, cet ensemble microphonique délivrant un second signal de parole ; et

- des moyens de mixage (54), pour combiner le premier signal de parole et le second signal de parole, et donner en sortie un troisième signal de parole représentatif de la parole émise par le porteur du casque, **caractérisé en ce que** :

- le capteur physiologique (18) est incorporé à un coussinet circumaural (16) d'une coque (14) de l'un des écouteurs (12) ;

- l'ensemble microphonique comprend deux microphones (20, 22) placés sur la coque (14) de l'un des écouteurs (12) ;

- les deux microphones (20, 22) sont alignés en un réseau linéaire suivant une direction principale (24) dirigée vers la bouche (26) du porteur du casque ; et

- il est prévu des moyens (56) de réduction de bruit non fréquentielle du second signal de parole, comprenant un combineur apte à appliquer un retard au signal délivré par l'un des microphones et à soustraire ce signal retardé du signal délivré par l'autre microphone, de manière à opérer un débruitage du signal de parole proche émis par le porteur du casque.

2. Le casque audio de la revendication 1, comprenant en outre :

- des moyens (48) de filtrage passe-bas du premier signal de parole avant combinaison par les moyens de mixage, et/ou des moyens (50, 52) de filtrage passe-haut du second signal de parole avant débruitage et combinaison par les moyens de mixage, ces moyens de filtrage passe-bas et/ou passe-haut (48, 50, 52) comprenant un filtre à fréquence de coupure ajustable ; et

- des moyens (44) de calcul de la fréquence de coupure opérant en fonction du signal délivré par le capteur physiologique.

3. Le casque audio de la revendication 2, dans lequel les moyens (44) de calcul de la fréquence de coupure comprennent des moyens d'analyse du contenu spectral du signal délivré par le capteur physiologique, aptes à déterminer la fréquence de coupure en fonction des niveaux relatifs du rapport signal/bruit évalué dans une pluralité de bandes de fréquences distinctes du signal délivré par le capteur physiologique.

4. Le casque audio de la revendication 1, comprenant en outre :

- des moyens (62) de débruitage du troisième signal de parole délivré par les moyens de mixage, opérant par réduction de bruit fréquentielle.

5. Le casque audio de la revendication 4, comprenant en outre des moyens recevant en entrée, et opérant une intercorrélation entre, ledit premier et ledit troisième signal de parole, et délivrant en sortie un signal de probabilité de présence de parole fonction du résultat de ladite intercorrélation.

6. Le casque audio de la revendication 5, dans lequel les moyens (62) de débruitage du troisième signal de parole reçoivent en entrée ledit signal de probabilité de présence de parole et sont aptes à, sélectivement :

i) opérer une réduction de bruit différenciée selon les bandes de fréquences en fonction de la valeur dudit signal de probabilité de présence de parole, et

ii) opérer une réduction de bruit maximale sur toutes les bandes de fréquences en l'absence de parole.

7. Le casque audio de la revendication 1, comprenant en outre :

- des moyens (64) de post-traitement, aptes à opérer une égalisation sélective par bandes de fréquences dans la partie du spectre correspondant au signal recueilli par le capteur physiologique.

8. Le casque audio de la revendication 7, dans lequel les moyens de post-traitement sont aptes à déterminer un gain d'égalisation pour chacune desdites bandes de fréquences, ce gain étant calculé à partir des coefficients fréquentiels respectifs des signaux délivrés par le(s) microphones et des signaux délivrés par le capteur physiologique, considérés dans le domaine fréquentiel.

EP 2 518 724 A1

9. Le casque audio de la revendication 8, dans lequel les moyens de post-traitement sont en outre aptes à opérer un lissage sur une pluralité des trames successives de signal dudit gain d'égalisation calculé.

5

10

15

20

25

30

35

40

45

50

55

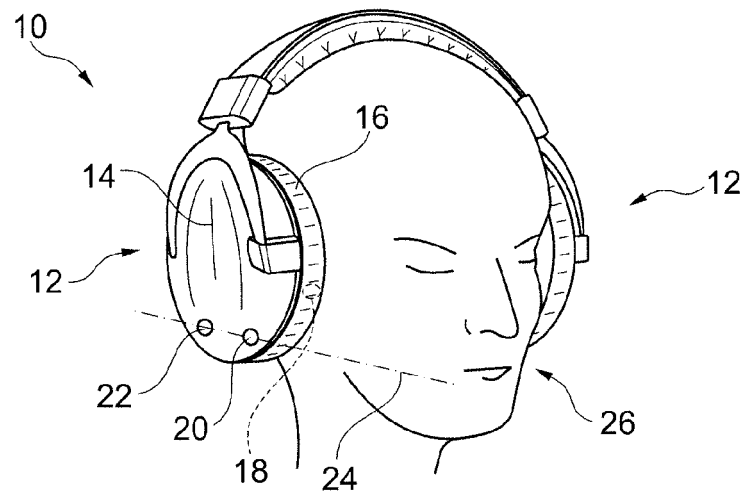


Fig. 1

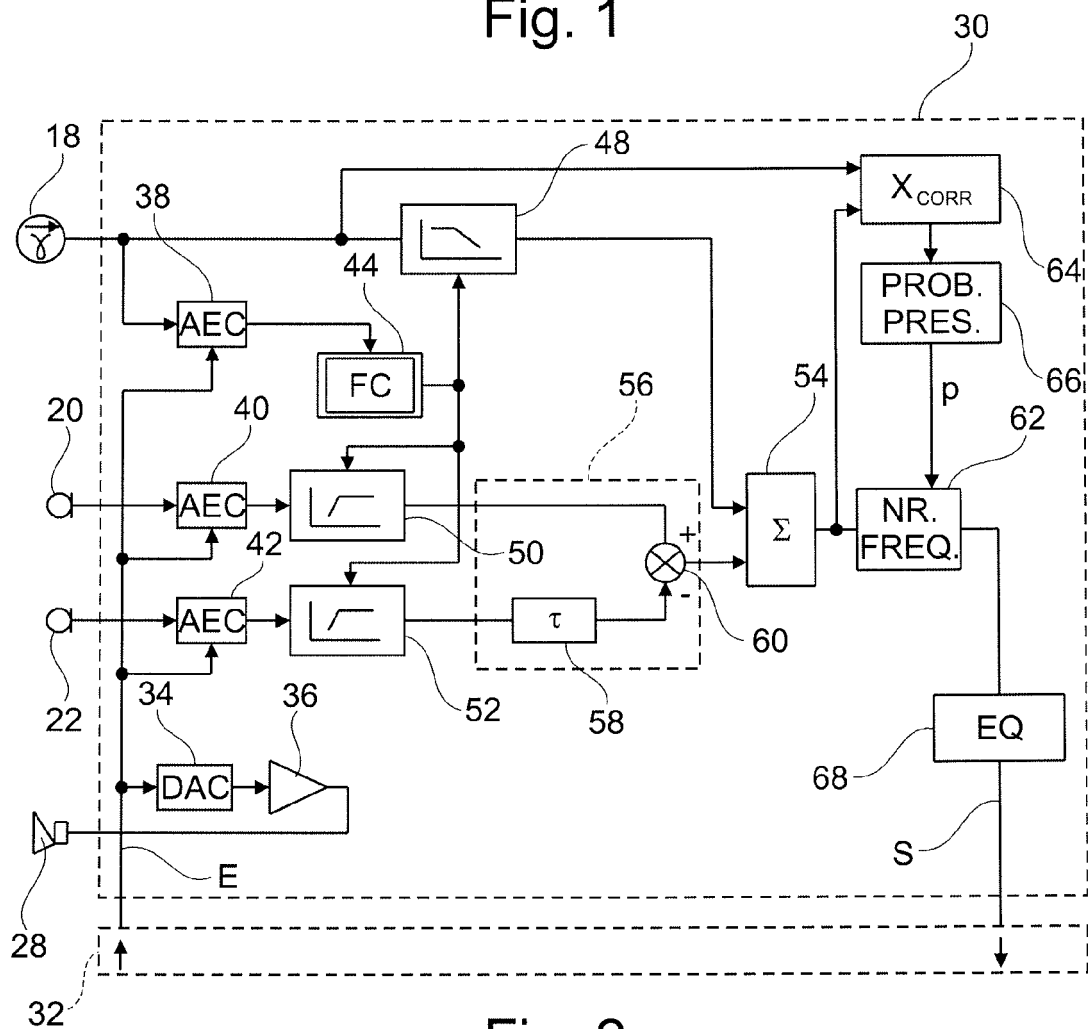


Fig. 2

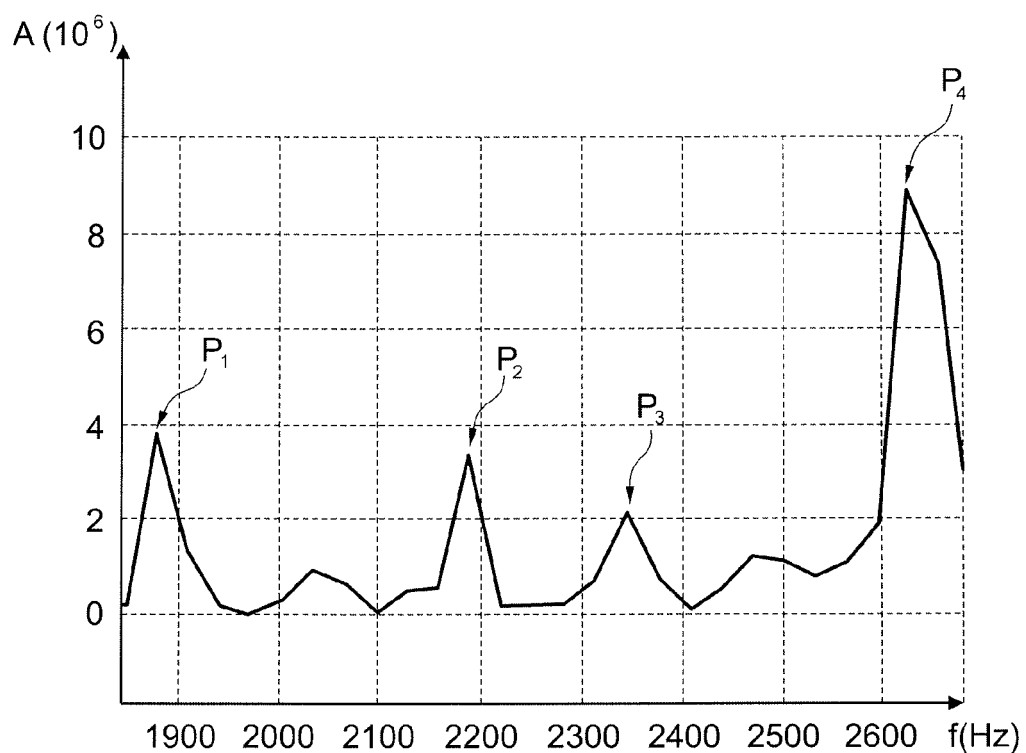


Fig. 3

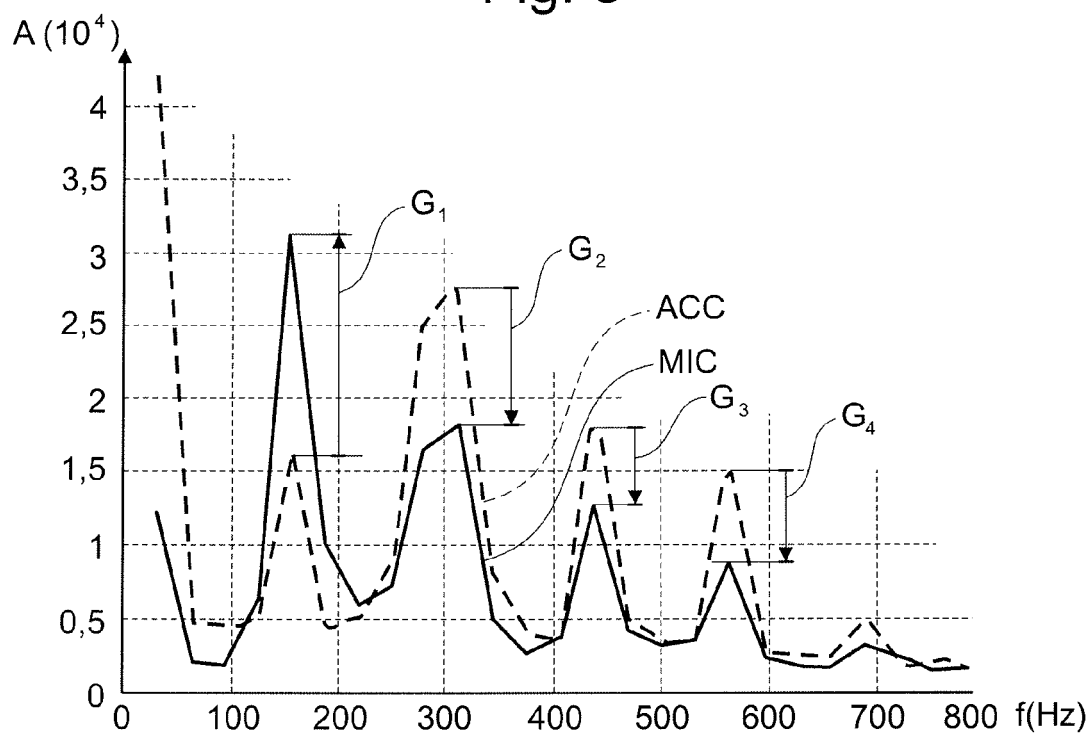


Fig. 4



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande

EP 12 16 4777

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
Y	EP 0 683 621 A2 (NIPPON TELEGRAPH & TELEPHONE [JP]) 22 novembre 1995 (1995-11-22)	1,2,4-9	INV. G10L21/02 H04R3/00
A	* figures 1,2 * * page 6, ligne 47 - ligne 52 * * page 7, ligne 45-53 *	3	
Y	US 2007/088544 A1 (ACERO ALEJANDRO [US] ET AL) 19 avril 2007 (2007-04-19)	1,2,4-9	
A	* figure 3 * * alinéas [0026] - [0031] *	3	
Y	M. BUCK, M. RÖSSLER: "First Order Differential Microphones Arrays for Automotive Applications", PROCEEDINGS OF THE 7TH INTERNATIONAL WORKSHOP ON ACOUSTIC ECHO AND NOISE CONTROL (IWAENC), 10 septembre 2001 (2001-09-10), XP002680249,	1,2,4-9	
A	* le document en entier *	3	DOMAINES TECHNIQUES RECHERCHES (IPC)
Y	JP 2000 261534 A (NIPPON TELEGRAPH & TELEPHONE) 22 septembre 2000 (2000-09-22)	2	G10L H04R
A	* figures 1,5-7 * * alinéas [0011] - [0014], [0052] *		
A	JP 8 214391 A (IWATSU ELECTRIC CO LTD; NIPPON TELEGRAPH & TELEPHONE) 20 août 1996 (1996-08-20)	1-9	
A	* figure 1 * * alinéa [0011] *		
A	WO 00/21194 A1 (RESOUND CORP [US]) 13 avril 2000 (2000-04-13)	1-9	
A	* figures 10,11 *		
2 Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche Munich		Date d'achèvement de la recherche 19 juillet 2012	Examineur Chétry, Nicolas
CATEGORIE DES DOCUMENTS CITES X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

EPO FORM 1503 03.82 (P04C02)

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

EP 12 16 4777

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche européenne visé ci-dessus.
Lesdits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

19-07-2012

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP 0683621 A2	22-11-1995	CA 2149563 A1	19-11-1995
		DE 69525987 D1	02-05-2002
		DE 69525987 T2	19-09-2002
		DE 69527731 D1	12-09-2002
		DE 69527731 T2	03-04-2003
		DE 69531413 D1	04-09-2003
		DE 69531413 T2	15-04-2004
		EP 0683621 A2	22-11-1995
		EP 0984660 A2	08-03-2000
		EP 0984661 A2	08-03-2000
		US 5933506 A	03-08-1999
US 2007088544 A1	19-04-2007	AUCUN	
JP 2000261534 A	22-09-2000	AUCUN	
JP 8214391 A	20-08-1996	JP 2835009 B2	14-12-1998
		JP 8214391 A	20-08-1996
WO 0021194 A1	13-04-2000	AU 6291599 A	26-04-2000
		WO 0021194 A1	13-04-2000

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

- JP 2000261534 A [0009] [0015]
- EP 0683621 A2 [0010]
- FR 2792146 A1 [0036] [0039]
- WO 2007099222 A1, Parrot [0062]

Littérature non-brevet citée dans la description

- **M. BUCK ; M. RÖBLER.** First Order Differential Microphones Arrays for Automotive Applications. *Proceedings of the 7th International Workshop on Acoustic echo and Noise control (IWAENC, 10 Septembre 2001* [0050]