(19) CONFÉDÉRATION SUISSE
INSTITUT FÉDÉRAL DE LA PROPRIÉTÉ INTELLECTUELLE

(11) CH 702 399 A2

(51) Int. Cl.: G06F 3/16 (2006.01)
H04R 1/40 (2006.01)

Demande de brevet pour la Suisse et le Liechtenstein
Traité sur les brevets, du 22 décembre 1978, entre la Suisse et le Liechtenstein

(12) DEMANDE DE BREVET

(21) Numéro de la demande: 01848/09

(22) Date de dépôt: 02.12.2009

(43) Demande publiée: 15.06.2011

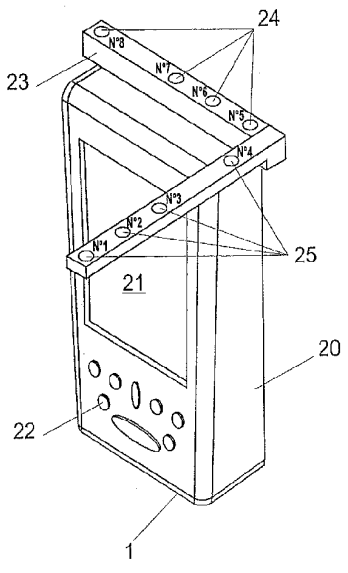
(71) Requérant:
Veovox SA, Chemin des Roches 10
1009 Pully (CH)

(72) Inventeur(s):
Hervé Lissek, 1020 Renens (CH)
Philippe Martin, 1007 Lausanne (CH)
Jorge Carmona, 1009 Pully (CH)
Michel Imhasly, 3984 Fiesch (CH)
Xavier Falourd, 1004 Lausanne (CH)
Patrick Marmaroli, 1024 Ecublens (CH)
Ian Millar, 1521 Curtilles (CH)

(74) Mandataire:
P&TS SA, Av. J.-J. Rousseau 4 P.O. Box 2848
2001 Neuchâtel (CH)

(54) Appareil et procédé pour la saisie et le traitement de la voix.

(57) Appareil portatif (20) de saisie de la voix comprenant: un bras (23) orientable avec un premier réseau linéaire différentiel (25) de microphones comprenant au moins une paire de microphones, la directivité dudit premier réseau étant agencée en sorte à détecter la voix depuis une première direction en fonction de l'orientation dudit bras;
un deuxième réseau linéaire différentiel (24) de microphones comprenant au moins une paire de microphones, la directivité dudit deuxième réseau étant agencée en sorte à détecter le bruit provenant d'une deuxième direction différente de la première direction;
un circuit de réduction de bruit pour fournir un signal vocal avec un bruit réduit, sur la base du signal de sortie dudit premier réseau et du signal de sortie dudit deuxième réseau.



Description

Domaine de l'invention

[0001] La présente invention concerne un appareil et un procédé pour la saisie et le traitement de la voix, en particulier dans des environnements bruyants. L'invention porte entre autres sur un appareil mobile qui peut être utilisé dans des environnements bruyants tels que, à titre d'exemples non limitatifs, dans des restaurants, pour la saisie et le traitement de la voix et pour effectuer de la reconnaissance vocale.

Etat de la technique

[0002] Alors que la fréquence de reconnaissance des algorithmes de reconnaissance vocale s'est récemment améliorée, elle reste faible dans des conditions difficiles, notamment lorsque le rapport du signal au bruit est insuffisant. Pour cette raison, la saisie et la reconnaissance de la voix dans des environnements bruyants reste difficile ou peu fiable.

[0003] Toutefois, il existe un besoin pour des appareils capables d'effectuer de façon fiable de la reconnaissance vocale même dans des environnements très bruyants tels que (à titre d'exemples non limitatifs) dans des bars et des restaurants. Par exemple, il serait utile d'avoir un appareil capable de saisir et de reconnaître la voix d'un serveur dans un restaurant et d'utiliser cet appareil pour recevoir, reconnaître et transmettre des commandes vocales.

[0004] US 7 110 963, dont le contenu est incorporé par référence au sein de la présente demande, divulgue un système de reconnaissance vocale permettant à un serveur dans un restaurant de transmettre des commandes à la cuisine. Une application logicielle de reconnaissance vocale est utilisée pour contrôler le traitement et le flux de données pendant les opérations d'enregistrement des commandes et pour recevoir des informations sur les commandes de la part du serveur en temps réel pendant l'interaction avec le client.

[0005] US-A1-2002/0 007 315, dont le contenu est incorporé par référence au sein de la présente demande, divulgue un autre système à activation vocale de commande dans un établissement de restauration rapide, où les commandes de repas sont introduites dans un enregistreur au point de vente et convertis en messages vocaux pour le préparateur des aliments du restaurant. Un circuit de conversion de voix en texte est utilisé aux points de vente pour introduire les commandes vocales.

[0006] Les solutions présentées ci-dessus sont utiles et permettent une transmission plus rapide et plus naturelle des commandes entre le restaurant et la cuisine. Toutefois, la fiabilité de la reconnaissance vocale dans de nombreux restaurants avec un niveau de bruit élevé ou même moyen n'est pas satisfaisant; le rapport du signal au bruit est insuffisant pour être exécuté de façon fiable par les algorithmes de reconnaissance vocale actuels.

[0007] Il a été constaté que la qualité et la directivité du microphone est d'une importance capitale pour saisir un signal vocal de bonne qualité. Le document US-B2-7 120 477 (Microsoft Corporation) décrit un appareil informatique personnel mobile possédant une antenne avec microphone et reconnaissance vocale. L'antenne comprend un microphone disposé sur son extrémité distale et est adaptée pour être orientée vers un utilisateur, permettant ainsi de réduire la distance avec la bouche de l'utilisateur du microphone tandis que l'utilisateur tient l'appareil dans la paume de sa main. Réduire cette distance permet d'augmenter le rapport du signal au bruit des signaux vocaux fournis par le microphone. Cette solution reste toutefois insuffisante pour des environnements très bruyants.

[0008] Un autre module pour la saisie vocale dans des environnements bruyants est divulgué dans le document EP 694 833. Ce document décrit un premier réseau de microphones à faisceau orientable pour la saisie vocale et un réseau supplémentaire de microphones à faisceau orientable pour la reconnaissance de sources supplémentaires de données audio et de sources de bruit et/ou d'interférence. Le but ici est de repérer le locuteur (source audio) avec un algorithme de triangulation et de contrôler un système d'entraînement mécanique pour focaliser une caméra vidéo sur le locuteur.

[0009] Les deux réseaux de microphones sont bidimensionnels et occupent donc une large surface; il n'est donc pas possible de monter les réseaux sur un faisceau linéaire tout en maintenant une distance suffisante entre les microphones. En outre, le post-traitement des signaux audio fournis par deux réseaux multidimensionnels de microphones est difficile et nécessite un ensemble de circuits ou une puissance de calcul considérable, une consommation d'énergie accrue, et entraîne souvent un filtrage indésirable du signal de sortie.

[0010] Un but de la présente invention est donc de développer un appareil portatif amélioré qui soit capable d'effectuer la saisie et le traitement de la voix et de générer un signal vocal avec un rapport du signal au bruit suffisant pour des applications de reconnaissance vocale fiables.

[0011] Un autre but de l'invention est de développer un appareil avec microphone qui soit capable d'améliorer la détection de la voix de l'utilisateur tout en minimisant le bruit de fond et les locuteurs potentiellement parasitiques dans des conditions diffuses.

[0012] Les performances du dispositif doivent couvrir au moins la bande passante vocale moyenne mais doivent également s'étendre de façon à améliorer le processus de reconnaissance vocale, à savoir [300Hz-6kHz].

[0013] Un autre but est de développer un appareil qui soit capable d'extraire des informations vocales utiles (telle une commande ou un ordre dans un restaurant) hors du bruit de fond qui peut être plus ou moins diffus (aucun angle d'incidence

privilégié), plus ou moins intense (en termes de niveaux de pression acoustique) et posséder différentes caractéristiques spectrales (musique amplifiée, voix individuelles, bruit «cocktail party», etc.).

[0014] Un autre but est de développer un appareil amélioré qui permette de détecter la voix émanant de la bouche du locuteur et le bruit provenant d'autres directions et que l'utilisateur puisse tenir dans la paume de sa main.

Bref résumé de l'invention

[0015] Selon un aspect de l'invention, un appareil portatif de saisie de la voix comprend:
un bras orientable adapté pour être orienté en direction de la bouche d'un utilisateur, ledit bras comprenant un premier réseau linéaire différentiel de microphones, la directivité dudit premier réseau étant agencée en sorte à détecter la voix émanant de la bouche dudit utilisateur;
un deuxième réseau linéaire différentiel de microphones, la directivité dudit deuxième réseau étant agencée en sorte à détecter le bruit provenant d'une direction différente de la bouche de l'utilisateur;
un circuit de réduction de bruit pour fournir un signal vocal avec un bruit réduit, sur la base du signal de sortie dudit premier réseau et du signal de sortie dudit deuxième réseau.

[0016] Dans une forme d'exécution préférentielle, le premier réseau différentiel est utilisé pour saisir le bruit de fond depuis une direction arrière.

[0017] Des réseaux différentiels de microphones sont connus en tant que tels et sont décrits par exemple dans Elko, G.W. «Superdirectional Microphone Arrays», dans J. Benesty et S. Gay (éds), «Acoustic Signal Processing for Télécommunication», pp.181-236, Kluwer Académie Publishers, 2000. La plupart des réseaux de microphones sont relativement encombrants et peu adaptés à des appareils portatifs.

[0018] La présente invention concerne un arrangement spécifique de réseaux linéaires qui permet de saisir le son selon différentes directions. L'utilisateur peut orienter le bras vers sa bouche et s'assurer que la première direction est adaptée pour saisir la voix de l'utilisateur tandis que la deuxième direction saisit essentiellement le bruit de fond. Le circuit de réduction de bruit peut ensuite améliorer le signal vocal en supprimant le bruit de fond, en utilisant par exemple des techniques de cohérence.

[0019] Dans une forme d'exécution, le premier réseau de microphones effectue la saisie vocale dans une première direction avant et saisit le bruit de fond depuis une direction arrière, tandis que le deuxième réseau de microphones saisit le bruit de fond et d'autres voix depuis la droite et la gauche.

[0020] D'autres formes d'exécution peuvent utiliser un nombre de réseaux de microphones plus grand que deux et/ou des réseaux complexes en sorte à offrir un meilleur contrôle de la directivité de l'appareil. Les microphones sont de préférence, à titre d'exemple non limitatif, des microphones électrets.

[0021] Dans une forme d'exécution, le bras est en forme de L et comprend un réseau linéaire de microphones sur chacune des deux branches. D'autres dispositions, y compris des microphones avec une pluralité de branches non perpendiculaires, des microphones disposés en forme de U avec trois réseaux de microphones ou des dispositions avec des paires de microphones sur différentes branches d'un arbre commun, peuvent également être utilisées dans le cadre de l'invention.

[0022] Selon un autre aspect éventuellement indépendant de l'invention, le signal vocal de sortie du microphone est post-traité par un filtre de post-traitement comprenant une pluralité de couches de traitement de signal, permettant d'extraire la voix hors du bruit, de réduire le bruit résiduel et d'estimer la cohérence du signal résultant avec la détection vocale originale.

[0023] Selon un autre aspect de l'invention, un détecteur d'activité vocale automatique supplémentaire permet d'améliorer davantage le signal en supprimant les segments temporels pendant lesquels aucune activité vocale n'est détectée.

[0024] Les techniques de post-traitement de signaux vocaux sont connues en tant que telles et décrites par exemple par Kim KM, Choi YK, Park KS, «A new approach for rustle noise cancelling in pen-type voice recorder», IEEE Transactions on Consumer Electronics, Vol. 49(4), pp. 1118-1124, nov. 2003. Un autre exemple de procédé de post-traitement est décrit par O. Yilmaz et S. Rickard, «Blind Séparation of Speech Mixtures via Time-Frequency Masking», IEEE Transactions on Signal Processing, Vol. 52(7), pp. 1830-1847, juillet 2004. La combinaison spécifique des procédés décrits et revendiqués s'est avérée, au moyen de tests, être particulièrement efficace dans le but susmentionné et efficace pour supprimer le bruit d'un signal vocal saisi avec le microphone spécifique décrit et revendiqué au sein de cette demande, tout en évitant des composants matériels et logiciels requis par des dispositions plus complexes.

[0025] Un avantage clé du dispositif divulgué au sein de la description et dans les revendications est la capacité d'ajuster la directivité afin d'effectuer la saisie et la reconnaissance vocale à une distance confortable; le locuteur peut parler à une distance confortable (plus grande que 10 cm, de préférence plus grande que 15 cm même dans des conditions bruyantes telles que dans un restaurant) au travers de l'appareil portatif sans devoir approcher sa bouche très près de l'appareil avec microphone.

[0026] Dans une forme d'exécution préférentielle de l'invention, le réseau de microphones est suffisamment réduit pour garantir l'ergonomie et la portabilité du système et n'excède pas les dimensions des assistants numériques personnels ou ordinateurs de poche conventionnels (approximativement 150 mm x 70 mm, en tout cas plus petit que 180 mm x 100 mm).

Brève description des figures

[0027] La présente invention sera mieux comprise au moyen de la description de quelques formes d'exécution illustrées par les figures, dans lesquelles:

la fig. 1	illustre de façon schématique un système pour la saisie et la transmission de commandes vocales dans un restaurant;
la fig. 2	illustre de façon schématique un sous-système de microphones;
la fig. 3	est un diagramme démontrant l'influence de u sur le schéma de directivité d'un sous-système de premier degré de microphones;
la fig. 4	est un diagramme démontrant à quel point la sensibilité d'un réseau différentiel de premier degré dépend de l'angle et de la fréquence;
la fig. 5	illustre de façon schématique un réseau différentiel de deuxième degré;
la fig. 6	illustre un exemple d'appareil comprenant une installation différentielle bidimensionnelle de microphones (à gauche: réseau à rayonnement transversal; à droite: réseau orientable à rayonnement longitudinal).
la fig. 7	illustre de façon schématique la disposition des microphones de l'appareil de la fig. 5.
la fig. 8	est un organigramme du procédé de post-traitement appliqué au signal vocal pour le rehaussement de la voix.

Description détaillée de formes d'exécution préférentielles

[0028] La description qui suit est donnée en mettant l'accent sur la forme d'exécution comprenant un ordinateur de poche pour enregistrer les commandes vocales dans un restaurant. Toutefois, le dispositif de l'invention peut être également utilisé avec d'autres équipements, y compris et sans limitations un ordinateur fixe, des ordinateurs portables, des stations de travail, d'autres appareils mobiles tels que des téléphones portables et autres appareils ainsi que pour des applications autres que pour les restaurants et bars (dans l'industrie hôtelière, les hôpitaux, l'industrie du divertissement, les magasins d'alimentation, les laboratoires etc.).

[0029] Un exemple d'environnement dans lequel le procédé et l'appareil peuvent être utilisés est illustré à la fig. 1. Dans cet exemple, un serveur 2 dans un bar ou un restaurant prend une commande de clients 3 assis à une table. Le serveur du restaurant répète chaque commande et les énonce dans le microphone de son appareil mobile 1. Dans cette forme d'exécution, le signal vocal enregistré est post-traité localement, par exemple par le processeur de l'appareil mobile 1 ou de préférence par des moyens de traitement dédiés, afin d'améliorer le rapport du signal au bruit. Ce post-traitement peut également être effectué dans une autre forme d'exécution par un ordinateur ou un serveur à distance, bien que cela puisse entraîner un retard. Le signal vocal traité est ensuite transmis à travers l'air à un point d'accès 7 en utilisant un protocole de communication sans fil standard tel que 802.11, Bluetooth, etc. Le point d'accès 7 appartient à un réseau local 8 (LAN) auquel sont connectés divers autres appareils tels qu'un ordinateur personnel 5, un serveur 6 etc. Le signal vocal reçu depuis le point d'accès 7 est converti en commandes texte par le serveur 6 qui exécute un algorithme de reconnaissance vocale. L'algorithme de reconnaissance vocale pourrait être exécuté par l'appareil mobile si celui-ci possède une puissance de traitement suffisante; cela peut toutefois rendre plus difficile une mise à jour des modèles de voix et de langage (tels que la liste des commandes à reconnaître et la grammaire associée).

[0030] Dans une forme d'exécution préférentielle, la reconnaissance vocale dépend du locuteur et utilise des profils dépendants du locuteur et stockés dans une base de données 60. Une grammaire est également stockée dans la base de données 60 afin de limiter le nombre de mots ou d'expressions à reconnaître et pour définir certaines règles caractérisant le texte prononcé par le serveur du restaurant. Cette grammaire est avantageusement mise à jour chaque fois que des nouveaux articles sont proposés aux clients 3, par exemple chaque fois que le menu du restaurant est modifié.

[0031] Pour cette application, l'algorithme de reconnaissance vocale est avantageusement basé sur un classifieur statistique, par exemple un réseau de neurones artificiels, en combinaison avec un classifieur basé sur profiles. Des tests ont révélé que cet arrangement offre un taux de reconnaissance amélioré et une introduction de nouveaux mots ou de nouvelles expressions dans la grammaire facilitée. La grammaire peut inclure des unités de reconnaissance de profiles de différentes tailles (syntagme, phrase, mot, phonème). Une grammaire dépendante de l'utilisateur peut également être utilisée.

[0032] La grammaire et/ou le classifieur sont de préférence adaptatifs et des unités de reconnaissance de profiles appris sont incorporées dans les données vocales d'entrée. Cela permet un apprentissage en ligne de nouveaux mots ou de nouveaux profiles. Un feedback d'utilisateur peut être utilisé, par exemple dans l'appareil de l'utilisateur, pour entrer ou choisir le texte équivalent d'un profile nouvellement appris.

[0033] En outre, la grammaire est avantageusement organisée en catégories et sous-catégories séparées; cela accroît la qualité de la reconnaissance vocale puisque le système connaît la catégorie du prochain profile escompté. Cela facilite également l'introduction manuelle de nouveaux profiles. Par exemple, une catégorie de profiles peut correspondre à la carte des vins et une autre catégorie au menu de desserts.

[0034] Le texte reconnu par le système de reconnaissance vocale dans le serveur 6 est transmis par le biais du réseau local 8 et au travers du canal sans fil en retour à l'appareil 1 du serveur de restaurant et affiché en temps réel. Dans un autre environnement, la reconnaissance pourrait s'effectuer directement sur l'appareil du serveur de restaurant. Le serveur peut vérifier si la reconnaissance est correcte et confirmer ou corriger la commande reconnue par le serveur et affichée par l'appareil. Ce feedback de l'utilisateur peut être utilisé pour adapter le profile dépendant du locuteur, la grammaire et/ou pour ajouter des nouvelles unités de reconnaissance.

[0035] Lorsque le niveau de confiance atteint par l'algorithme de reconnaissance vocale se situe en dessous d'un niveau prédéfini ou lorsqu'il existe différentes options possibles qui sont très proches l'une de l'autre, un menu avec une liste à choix multiple des données vocales d'entrée les plus probables est affichée au serveur du restaurant qui peut choisir la commande visée dans ce menu en utilisant par exemple un écran tactile, un stylet ou tout autre moyen d'entrée approprié y compris la voix. Le serveur de restaurant peut également sélectionner d'autres options, par exemple pour préciser le nombre d'articles commandés (nombre ou volume), le type (par exemple le millésime d'un vin, les préférences du client concernant la cuisson, etc.) en fonction de l'article commandé ou si la commande originale n'est pas suffisamment précise.

[0036] Une fois validé par le serveur du restaurant, ce texte ainsi que la réponse du serveur de restaurant aux options de menu sont également affichés sur un ordinateur personnel 5 ou imprimés et lus par les employés du restaurant afin de préparer et fournir la commande requise. Dans une autre forme d'exécution, ce texte est prononcé en cuisine. La liste des articles commandés peut être stockée dans une base de données du serveur 6 qui peut être utilisée ultérieurement pour préparer la facture pour le client. Dans une variante, le signal vocal enregistré est post-traité par un ordinateur ou un serveur.

[0037] Dans une alternative, la reconnaissance vocale est effectuée localement, dans l'appareil 1 de l'utilisateur. Cela nécessite toutefois des appareils 1 avec une puissance de traitement accrue et une synchronisation des modèles dépendants du locuteur plus difficile si un utilisateur utilise plusieurs appareils différents.

[0038] Un exemple d'appareil 1 selon l'invention est illustré à la fig. 6. Il est avantageusement réalisé autour d'un assistant numérique personnel (PDA), un mini-ordinateur portatif (netbook) ou autre appareil similaire. Il comprend:
un boîtier adapté pour transporter et manipuler l'appareil dans la paume de l'utilisateur;
un affichage 21 pour afficher à l'utilisateur 2 le texte reconnu et tout autre texte ou des images;
des moyens haptiques 22 tels que clavier, clavier numérique, bouton/touche électronique, molette cliquable etc.
une interface de communication (non représentée), par exemple une interface WLAN et/ou Bluetooth;
des moyens de traitement (non représentés) tels qu'un microprocesseur avec une mémoire RAM et ROM appropriée, pour le traitement audio du signal audio saisi par le biais du microphone et pour exécuter d'autres programmes et fonctions;
un bras 23 orientable en forme de L comprenant plusieurs réseaux linéaires de microphones 24, 25 avec un espace différent entre les microphones de chaque réseau. L'utilisation d'une pluralité de réseaux de microphones offre une détection d'activité vocale améliorée et un contrôle de la directivité à large bande. Le bras est relié au boîtier au travers d'un lien rotatif afin de diriger de façon précise la jambe la plus longue en direction de la bouche du locuteur.

[0039] Le bras 23 est avantageusement un accessoire qui est adapté pour être ultérieurement installé et monté de façon semi permanente sur un appareil mobile existant. Un ensemble de circuits électroniques, tels que convertisseur analogique-numérique, retardeur, additionneur, etc. et/ou des processeurs de traitement numérique du signal (DSPs) peuvent être associés de façon opérationnelle avec ce bras pour le traitement des signaux audio de sortie par les réseaux de microphones. Cet accessoire (bras détachable avec ensemble de circuits optionnels) peut être vendu séparément de l'appareil mobile et installé ultérieurement sur un appareil mobile existant afin de le transformer en un appareil selon l'invention. L'installation peut également comprendre l'installation de pilotes de périphérique et de logiciels d'application appropriés dans l'appareil mobile pour accéder aux signaux depuis l'accessoire, post-traiter ces signaux, les envoyer au serveur à distance ou à l'appareil mobile et afficher le feedback depuis le serveur. La connexion électrique entre le bras et l'appareil utilise de préférence un connecteur existant de l'appareil portable, par exemple un USB, un RS-232 ou un connecteur propriétaire, ou une connexion sans fil.

[0040] Dans une autre forme d'exécution, non représentée, le bras avec les réseaux de microphones et l'ensemble de circuits électroniques associés est relié à un appareil mobile existant par le biais d'une interface sans fil, par exemple une interface Bluetooth ou Zigbee. Dans ce cas, le bras peut être détaché de l'appareil mobile et manipulé séparément. Il est également possible de séparer le bras en plusieurs parties et d'utiliser l'une des jambes comme stylet tenu vers la bouche et relié (sans fil ou par fil) aux autres parties et/ou à l'appareil mobile. En outre, le bras, ou chaque élément du bras, peut être un composant entièrement passif qui comprend uniquement des microphones ou un composant «intelligent» possédant un microprocesseur, un réseau prédiffusé programmable (FPGA, field programmable gâte array) ou un processeur audio. Les différents éléments peuvent être reliés mutuellement et reliés à l'appareil mobile et/ou à un module récepteur de l'appareil mobile au travers d'une interface par fil ou sans fil. En outre, le microphone ou les parties du microphone et/ou l'appareil mobile peuvent être reliés à distance depuis un module de contrôle à distance pour commander l'amplification,

la réduction de bruit, la directivité etc. Dans une forme d'exécution, le système comprend des moyens de traitement de signal qui sont distribués entre le bras, ou différentes parties du bras, et l'appareil mobile.

[0041] Un exemple de réseau linéaire de microphones 24 est illustré à la fig. 2. Ce réseau simple comprend deux microphones 240, 241 espacés d'une distance d. Le signal de sortie d'un microphone est ajouté au moyen d'un élément additionneur 243 au signal de sortie différé de l'autre microphone à une distance d, le retard appliqué par l'élément retardeur 242 étant indiqué par $\#_e$. Ce réseau forme un système formateur de faisceau; un choix approprié du retard $\#_e$ améliore le rapport du signal au bruit et améliore la sensibilité aux signaux audio en provenance de la direction du réseau linéaire.

[0042] Si l'on considère un signal acoustique entrant avec un angle d'incidence # (par rapport à l'axe du sous-système) et en supposant un signal harmonique de fréquence f [Hz] (ou pulsation #=2.#.f), le retard acoustique entre les deux microphones est $\#_d$ =d/c [s] (où c'est la vitesse du son dans l'air) et la tension de sortie résultante U [V] du sous-système dépend de l'angle d'incidence #[rad]:

$$\underline{U} = \underline{U}_1 - \underline{U}_2 e^{(-j\omega\tau_e)} = \underline{M}_1 \underline{p}_1 \left(1 - e^{-j\omega(\tau_e + \tau_d \cos\theta)}\right) \cong \underline{M}_1 \underline{p}_1 j\omega(\tau_e + \tau_d \cos\theta) \quad (1)$$

où M_1 [V/Pa] est la sensibilité du premier microphone, g_1 [Pa] est la pression acoustique d'une onde plane au niveau du premier microphone, $\#_e$ [s] est le retard appliqué au deuxième microphone et $\#_d$ est le temps de propagation du premier au deuxième microphone. Avec # = #e+ $\#_d$ et μ= $\#_d$/#, l'on obtient finalement la sensibilité M du sous-système:

$$\underline{M} = \frac{\underline{U}}{\underline{p}} \cong \underline{M}_1 j\omega[(1-\mu) + \mu\cos\theta] \quad (2)$$

qui est la caractéristique d'un microphone directif du premier ordre.

[0043] De cette équation, il ressort que la réponse en fréquence correspond à un filtre passe-haut avec une pente de +6dB/octave. Cela signifie que la sensibilité décroît dans la bande baisse-fréquence (LF). Cela peut constituer un désavantage dans la mesure où cela entraîne un rapport du signal au bruit plus bas dans le cas d'un champ diffus.

[0044] En posant μ=0.5, l'on obtient une directivité cardioïde du réseau de microphones et en posant μ=1, un microphone bidirectionnel. La fig. 3 illustre les schémas de directivité caractéristiques pour différentes valeurs de μ.

[0045] La directivité dépend hautement de la fréquence, comme illustré à la fig. 4. Pour assurer un schéma de directivité constant sur l'entière bande passante de fréquence, différentes paires de réseaux avec différentes distances entre les paires et différentes limites de fréquence sont combinées dans les réseaux de microphones 24, 25.

[0046] Le bras à microphones de l'invention utilise alors plusieurs paires de microphones qui sont disposés le long du même axe pour obtenir un réseau plus directif (dans l'axe du réseau). Chaque réseau est ainsi monodimensionnel et comprend une pluralité de paires toutes disposées sur une rangée.

[0047] En combinant deux réseaux différentiels du premier ordre et après avoir introduit un retard temporel supplémentaire, un réseau général différentiel de microphones du deuxième ordre peut être élaboré. La sensibilité globale d'un tel système peut être calculée en multipliant les sensibilités des sous-systèmes concernés, résultant en une directivité améliorée avec deux sous-systèmes en cascade plutôt qu'avec un seul, mais avec le désavantage d'un comportement d'un filtre passe-haut du deuxième ordre. En choisissant les dimensions de chaque sous-système, des bandes passantes de fréquence plus larges peuvent être couvertes avec des directivités et sensibilités constantes, formant ainsi des réseaux différentiels.

[0048] Un réseau différentiel est décrit par son ordre, c'est-à-dire par le nombre d'»étapes» de retards, comme décrit à la fig. 5 pour un réseau 24 du deuxième ordre. Dans cet exemple, le réseau comprend N=3 microphones disposés en quatre paires: {1;2}, {2;3}, {3;1}, {3;2}. Les distances dj entre les microphones successifs au sein des paires sont variables.

[0049] Le signal analogique u1(t), ui(t), ..., uN(t) à la sortie de chaque microphone 240, 241, 244 est converti en un signal numérique par des convertisseurs analogiques-numériques 245_1, 245_2, 245_3. Pour chaque paire, une première étape de traitement 246 effectue ensuite la différentiation numérique entre un signal et le signal retardé de l'autre microphone de la paire. Une deuxième étape de traitement 247 effectue ensuite la différentiation entre les données de sortie d'un élément additionneur 243 et les données de sortie retardées d'un autre élément additionneur de la première étape. Le premier signal numérique fourni par cette deuxième étape forme un signal de faisceau avant 248 tandis que l'autre signal numérique fourni par cette deuxième étape forme un signal de faisceau arrière 248.

[0050] Théoriquement, il est possible de combiner autant de paires que souhaité, mais en pratique il est difficile d'aller au-delà d'un réseau de deuxième degré. Cela est dû principalement au fait qu'un réseau différentiel est un réseau différentiateur (filtre passe-haut) du même ordre que l'ordre du réseau, ce qui signifie que les basses fréquences sont hautement atténuées et que le rapport du signal au bruit est dégradé. Il y a ainsi un compromis à faire concernant les dimensions de chaque réseau, la bande passante de fréquence qui présente un intérêt et le nombre de canaux disponibles pour le traitement du signal.

[0051] Le bras à microphones de l'appareil 1 est disposé pour détecter le son non seulement depuis la direction utile (la direction de la bouche), mais également depuis au moins une autre direction, correspondant au bruit. Une meilleure

connaissance du bruit émanant de différentes directions permet d'extraire le signal utile et de rejeter le signal de bruit, en utilisant des techniques de cohérence, et d'améliorer l'efficacité du post-filtrage ultérieur.

[0052] Dans une forme d'exécution, le bras à microphones 23 de la présente invention comprend un réseau de microphones bidimensionnel (plutôt qu'un réseau à une dimension comme décrit jusqu'à présent). Ce réseau bidimensionnel est constitué de deux réseaux monodimensionnels, comme représenté sur la fig. 7. Un premier réseau 24 est disposé sur la première et plus longue jambe du bras 23 en forme de L, tandis que le deuxième réseau est disposé sur l'autre jambe, plus courte, du même bras. Ce deuxième réseau transversal de microphones est utilisé pour améliorer la suppression du bruit d'interférence.

[0053] Comme mentionné, ce bras en forme de L est orientable, par rotation autour de l'axe de l'une des deux jambes (ici, la plus courte), en sorte que l'utilisateur peut ajuster la position afin qu'elle soit optimale (devant la bouche). Lorsque le bras 23 est orienté correctement, la jambe la plus longue (dans cet exemple) détecte le signal utile avant depuis la direction de la bouche du locuteur ainsi que le bruit depuis l'arrière. La deuxième jambe (ici la plus courte, sans que cela soit nécessairement le cas) détecte le bruit diffus à partir des directions gauche et droite.

[0054] Dans l'arrangement illustré, l'orientation de la deuxième jambe reste essentiellement inchangée lorsque le bras est pivoté; il n'y a qu'un degré de liberté pour orienter la première jambe en direction de la bouche de l'utilisateur.

[0055] Dans une forme d'exécution préférentielle, les deux jambes sont perpendiculaires l'une à l'autre; d'autres dispositions sont toutefois possibles.

[0056] Chaque jambe est équipée d'au moins un réseau différentiel linéaire de microphones.

[0057] Dans une autre forme d'exécution, le microphone est en forme de U et comprend deux jambes reliées par une troisième jambe, de préférence mais à titre d'exemple non limitatif perpendiculaire aux deux premières jambes.

[0058] Le dispositif de l'invention peut en outre utiliser des microphones ou réseaux de microphones supplémentaires y compris des microphones non orientables sur le boîtier de l'appareil ou des réseaux de microphones supplémentaires pour saisir le bruit de fond depuis différentes directions.

[0059] En outre, des microphones de différentes jambes peuvent être couplés par paires pour offrir une détection supplémentaire du bruit diffus le long d'autres directions.

[0060] Les différents signaux fournis par les différents réseaux sur le microphone sont ensuite post-traités afin de fournir un signal vocal avec un meilleur rapport du signal au bruit et capable de servir de donnée d'entrée pour un logiciel de reconnaissance vocale. La fig. 8 est un organigramme illustrant divers filtres et procédés utilisés pour améliorer la sensibilité de détection de la voix.

[0061] Dans une première étape, les procédés formateurs de faisceaux (comme décrits ci-dessus) sont appliqués pour réduire le bruit et contrôler la directivité en calculant les différences entre les signaux fournis par différents microphones ou sous-systèmes de microphones.

[0062] Le bruit est en outre réduit davantage en utilisant un filtre Wiener. A cette fin, une estimation des caractéristiques spectrales d'un temps de 50 ms de bruit est effectuée (avant que la voix n'active le processus) et soustraite du reste du signal.

[0063] L'étape de post-filtrage entraîne une comparaison, dans le domaine fréquentiel, des quatre signaux fournis par le réseau de microphones (avant, arrière, gauche, droit), calculés par l'étape de formation de faisceau et débruités par la phase de réduction de bruit en utilisant un filtre adaptatif basé sur un algorithme DUET modifié (DUET, Degenerate Unmixing Estimation Technique). Pour chaque canal du formateur de faisceau, ces filtres adaptatifs permettent de diminuer l'influence du bruit dans le canal avant, par soustraction spectrale des signaux des trois autres canaux qui détectent essentiellement le bruit.

[0064] La quatrième étape implique un calcul de cohérence qui est effectué entre le signal avant fourni par le formateur de faisceau et le résultat du post-filtrage, afin de filtrer les signaux résiduels qui ne proviennent pas du locuteur. Deux signaux sont cohérents si l'un est une version proportionnellement à échelle et retardée de l'autre.

[0065] Finalement, l'appareil comprend également un détecteur d'activité vocale pour détecter lorsque le locuteur est en train de parler. La détection vocale est de préférence effectuée par analyse de la puissance du signal. Lorsqu'il n'y a pas de voix, le signal résiduel est supprimé afin d'éliminer tout bruit entre périodes de locution.

[0066] Cet appareil peut être utilisé par exemple pour des applications pour prendre des commandes vocales et des applications de reconnaissance vocale dans des restaurants, bars, discothèques, hôtels, hôpitaux, dans l'industrie du divertissement, magasins d'alimentation etc.

Revendications

1. Appareil portatif de saisie de la voix (1) comprenant:

un bras orientable (23) adapté pour être orienté en direction de la bouche d'un utilisateur, ledit bras comprenant un premier réseau linéaire différentiel (25) de microphones, la directivité dudit premier réseau étant agencée en sorte à améliorer la détection de la voix émanant de la bouche dudit utilisateur;
un deuxième réseau linéaire différentiel (24) de microphones, la directivité dudit deuxième réseau étant agencée en sorte à améliorer la détection du bruit provenant d'une direction différente de la bouche de l'utilisateur;
un circuit de réduction de bruit pour fournir un signal vocal avec un bruit réduit, sur la base du signal de sortie dudit premier réseau et du signal de sortie dudit deuxième réseau.

2. L'appareil de la revendication 1, dans lequel le circuit de réduction de bruit est basé sur des techniques de cohérence pour supprimer le bruit du signal de sortie dudit premier réseau.
3. L'appareil de l'une des revendications 1 ou 2, dans lequel ledit bras comprend une première jambe avec ledit premier réseau linéaire (25) et une deuxième jambe avec ledit deuxième réseau linéaire (24), ladite première jambe et ladite deuxième jambe possédant différentes orientations.
4. L'appareil de la revendication 3, comprenant une connexion rotative pour pivoter ledit bras autour de l'axe de l'une desdites jambes, en sorte que l'utilisateur (2) puisse orienter une jambe vers sa bouche.
5. L'appareil de l'une des revendications 3 ou 4, ledit bras étant en forme de L, le premier réseau linéaire (25) étant disposé sur une première jambe et le deuxième réseau linéaire (24) sur une deuxième jambe dudit bras en forme de L, dans lequel ledit bras peut être pivoté autour d'un axe parallèle à ladite jambe.
6. L'appareil de l'une des revendications 3 ou 4, ledit bras étant en forme de U et comprenant trois réseaux de microphones.
7. L'appareil de l'une des revendications 1 à 6, ledit bras comprenant une pluralité de jambes, au moins un réseau de microphones comprenant un microphone sur deux jambes différentes.
8. L'appareil de l'une des revendications 1 à 7, construit autour d'un assistant numérique personnel avec ledit bras étant un accessoire externe détachable monté sur ledit assistant numérique personnel.
9. L'appareil de l'une des revendications 1 à 7, construit autour d'un assistant numérique personnel avec ledit bras étant connecté sans fil audit assistant numérique personnel.
10. L'appareil de l'une des revendications 1 à 9, comprenant en outre: des moyens de traitement de données;
un écran d'affichage (21); une interface de communication sans fil; un filtre Wiener pour la réduction de bruit; un détecteur d'activité vocale.
11. L'appareil de l'une des revendications 1 à 10, connecté de manière opérationnelle à un module logiciel de reconnaissance vocale dépendant de l'utilisateur.
12. L'appareil de la revendication 11, ledit module de reconnaissance vocale dépendant de l'utilisateur comprenant une grammaire et un dictionnaire adapté pour des applications de commande et de contrôle et/ou pour prendre des commandes dans des restaurants.
13. L'appareil de l'une des revendications 1 à 12, la directivité dudit microphone étant adapté pour la saisie vocale à une distance de la bouche plus grande que 15 cm dans des conditions bruyantes.
14. Un procédé pour saisir la voix, comprenant:
la saisie du signal vocal avec un premier réseau linéaire différentiel (25) de microphones montés sur une première jambe d'un bras rotatif (23) d'un appareil portatif (1), ledit bras étant dirigé vers la bouche du locuteur;
la saisie du bruit à partir d'au moins une direction différente de la direction dudit signal utile, en utilisant un deuxième réseau différentiel linéaire (24) de microphones monté sur une deuxième jambe dudit bras, ladite première et deuxième jambe possédant différentes directions;
la réduction du bruit à partir dudit signal vocal, utilisant les données de sortie dudit deuxième réseau.

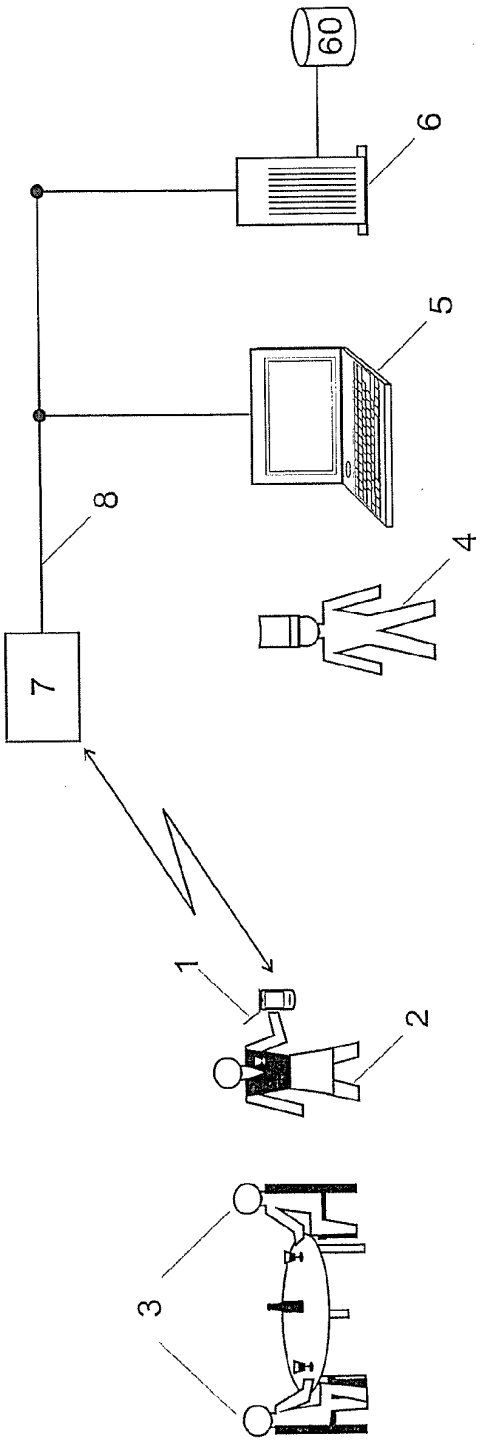


Fig. 1

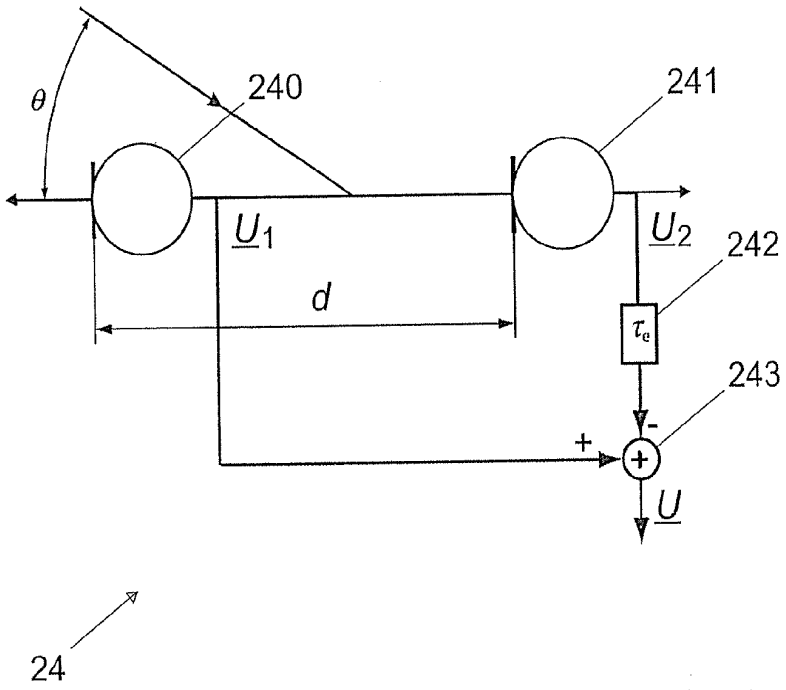


Fig. 2

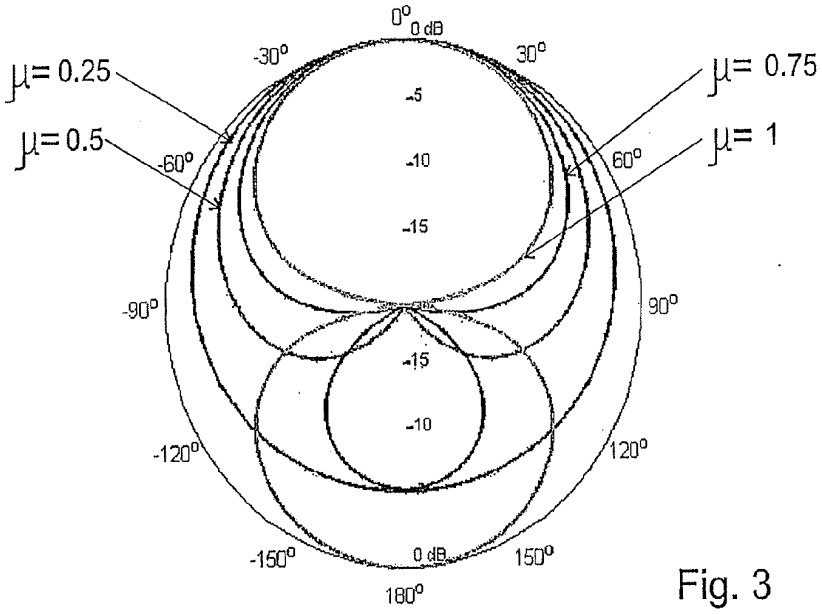

μ= 0.25
μ= 0.5
μ= 0.75
μ= 1
0° 0 dB
-30°
30°
-60°
60°
-90°
90°
-120°
120°
-150°
150°
180°
-5
-10
-15
-15
-10
0 dB

Fig. 3

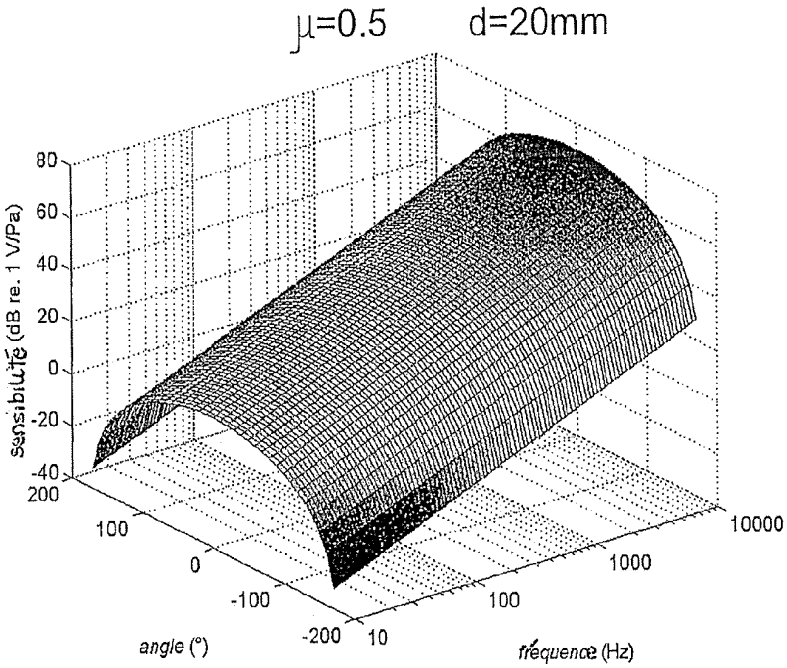
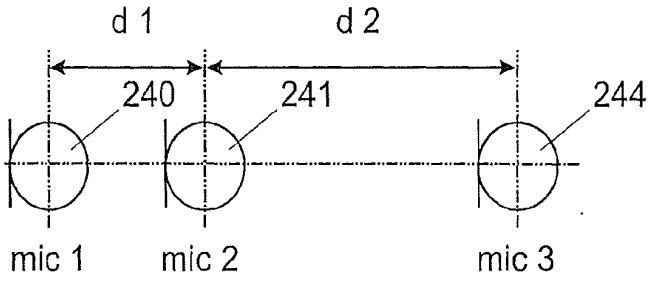

μ=0.5 d=20mm
sensibilité (dB re. 1 V/Pa)
80
60
40
20
0
-20
-40
200
100
0
-100
-200
angle (°)
10
100
1000
10000
fréquence (Hz)

Fig. 4


d 1
d 2
240
241
244
mic 1
mic 2
mic 3

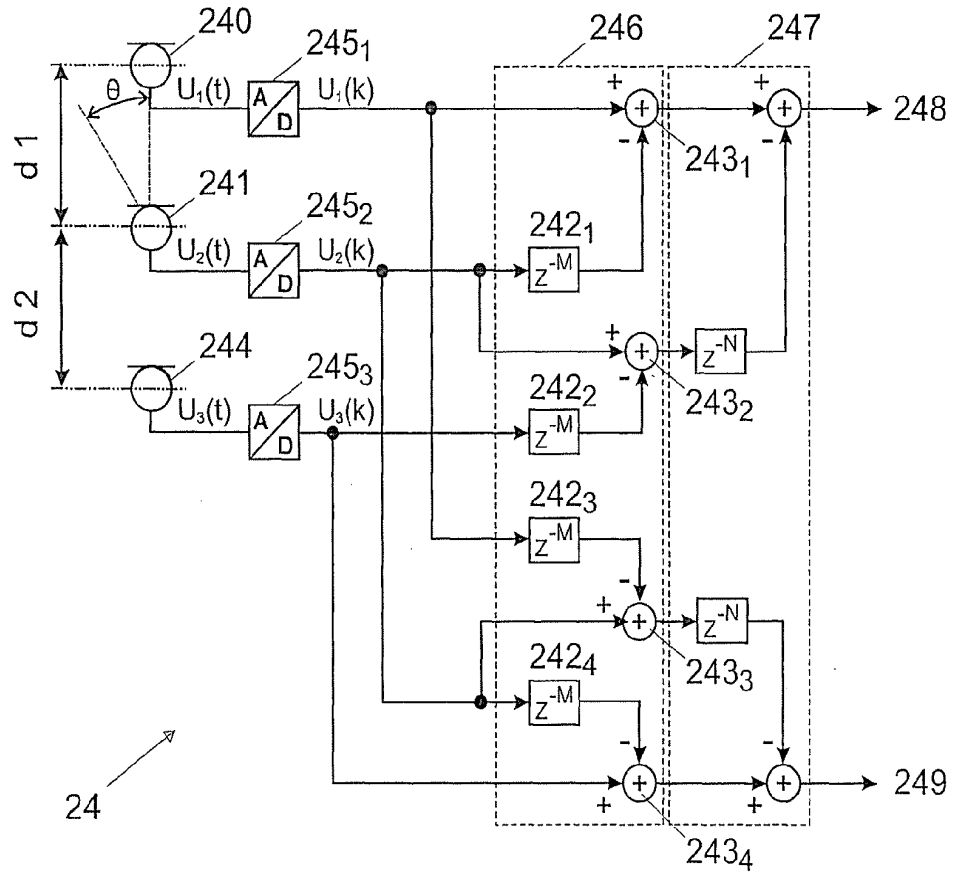

240
245₁
246
247
θ
U₁(t)
U₁(k)
248
d 1
2431
241
245₂
2421
U₂(t)
U₂(k)
d 2
244
245₃
2422
U₃(t)
U₃(k)
2432
2423
2433
2424
24
249
2434

Fig. 5

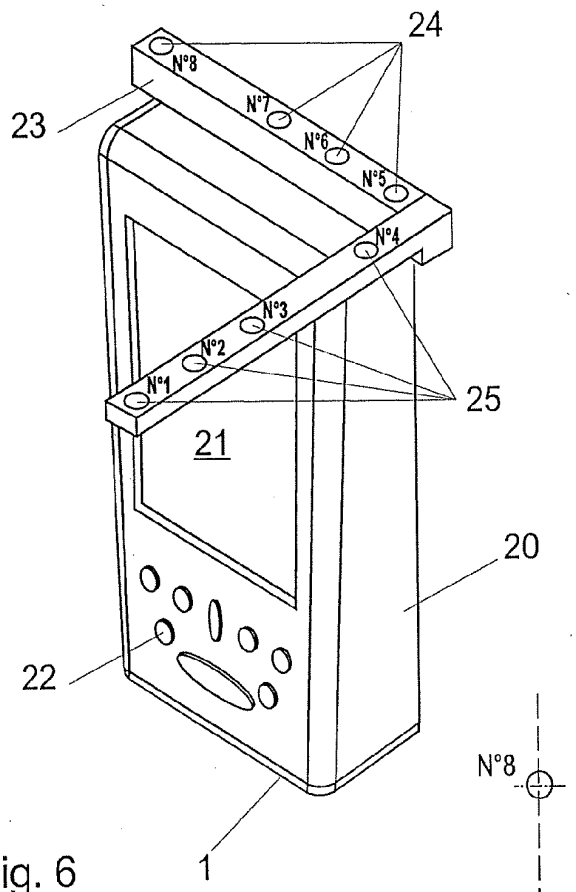


Fig. 6

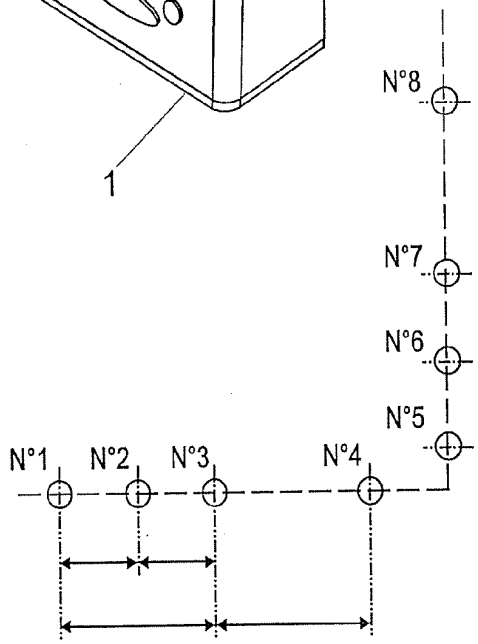


Fig. 7

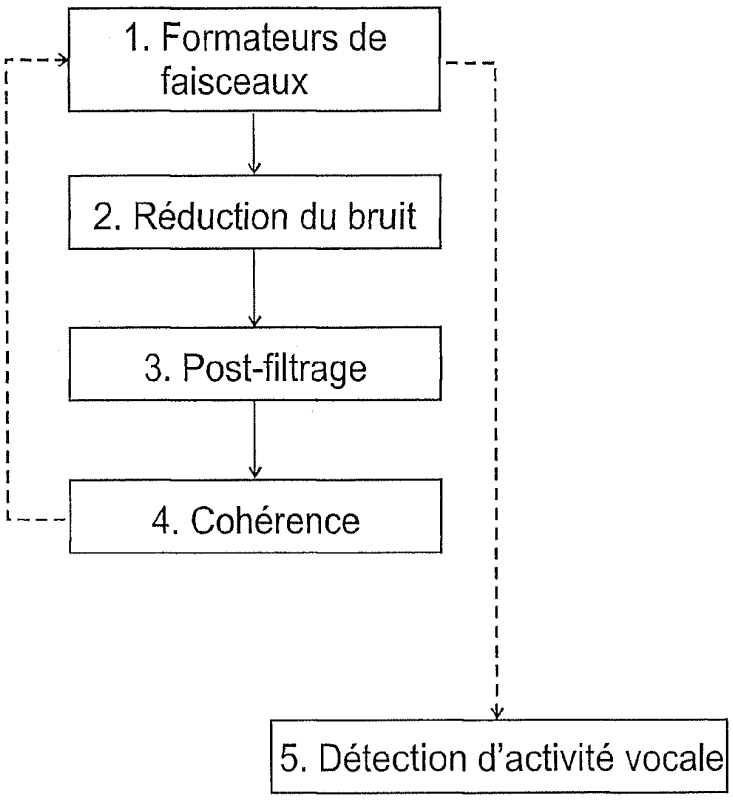

1. Formateurs de faisceaux
2. Réduction du bruit
3. Post-filtrage
4. Cohérence
5. Détection d'activité vocale

Fig. 8