

12 DEMANDE DE BREVET D'INVENTION

A1

22 Date de dépôt : 02.12.13.

30 Priorité :

43 Date de mise à la disposition du public de la
demande : 05.06.15 Bulletin 15/23.

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

60 Références à d'autres documents nationaux
apparentés :

○ Demande(s) d'extension :

71 Demandeur(s) : ADEUNIS R F — FR.

72 Inventeur(s) : MAOUCHE KARIM.

73 Titulaire(s) : ADEUNIS R F.

74 Mandataire(s) : CABINET GERMAIN ET MAUREAU
Société civile.

54 PROCÉDE DE DETECTION DE LA VOIX.

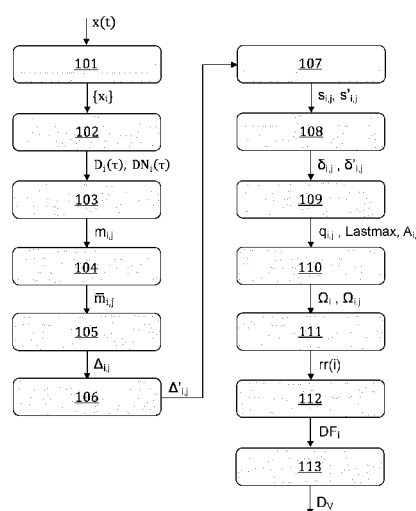
57 Procédé de détection de la voix permettant de détec-
ter la présence de signaux de parole dans un signal acous-
tique $x(t)$ bruité issu d'un microphone, comportant les
étapes successives suivantes :

- calcul d'une fonction de détection $FD(\tau)$ basée sur le
calcul d'une fonction différence $D(\tau)$ variant en fonction du
décalage τ sur une fenêtre d'intégration de longueur W dé-
butant au temps t_0 , avec :

$$D(\tau) = \sum_{n=t_0}^{t_0+W-1} |x(n) - x(n + \tau)| \text{ où } 0 \leq \tau \leq \max(\tau) ;$$

- une étape d'adaptation du seuil dans ledit intervalle
courant, en fonction des valeurs maximales du signal
acoustique $x(t)$ établies dans ledit intervalle courant ;

- recherche du minimum de la fonction de détection
 $FD(\tau)$ et comparaison de ce minimum avec un seuil, pour τ
variant dans un intervalle de temps déterminé dit intervalle
courant pour détecter la présence ou non d'une fréquence
fondamentale F_0 caractéristique d'un signal de parole dans
ledit intervalle courant.



La présente invention se rapporte à un procédé de détection de la voix permettant de détecter la présence de signaux de parole dans un signal acoustique bruité issu d'un microphone.

Elle se rapporte plus particulièrement à un procédé de détection de la voix utilisé dans un système de communication audio sans fil mono-capteur.

L'invention se situe dans le domaine spécifique de la détection d'activité de la voix, généralement appelé « VAD » pour Voice Activity Detection, qui consiste à détecter la parole, autrement dit des signaux de parole, dans un signal acoustique issu d'un microphone.

L'invention trouve une application privilégiée, mais non limitative, avec un système de communication audio sans fils multi-utilisateurs, du type système de communication par multiplexage temporel ou full-duplex, entre plusieurs terminaux de communication autonomes, c'est-à-dire sans raccordement à une base de transmission ou à un réseau, et simple d'utilisation, c'est-à-dire sans intervention d'un technicien pour établir la communication.

Un tel système de communication, notamment connu des documents WO10149864 A1, WO10149875 A1 et EP1843326 A1, est classiquement employé dans un environnement bruité voire très bruité, par exemple en milieu marin, dans le cadre d'un spectacle ou d'un évènement sportif en salle ou en extérieur, sur un chantier, etc.

La détection d'activité de la voix consiste généralement à délimiter au moyen de critères quantifiables, les débuts et fin de mots et/ou de phrases dans un signal acoustique bruité, autrement dit dans un flux audio donné. Une telle détection trouve des applications dans des domaines comme le codage de parole, la réduction de bruit ou encore la reconnaissance de parole.

La mise en œuvre d'un procédé de détection de la voix dans la chaîne de traitement d'un système de communication audio permet en particulier de ne pas transmettre de signal acoustique ou audio durant les périodes de silence. De ce fait, le bruit environnant ne sera pas transmis pendant ces périodes, dans un souci d'améliorer le rendu audio de la communication ou pour réduire le débit de transmission. Par exemple, dans le cadre du codage de parole, il est connu d'employer la détection d'activité de la voix pour coder le signal audio de façon pleine que lorsque le procédé « VAD » indique de l'activité. De ce fait, lorsqu'il n'y a pas de parole et que l'on se trouve dans une période de silence, le débit de codage baisse

significativement, ce qui en moyenne, sur tout le signal, permet d'atteindre des débits plus faibles.

Il existe ainsi beaucoup de procédés de détection d'activité de la voix mais ces derniers présentent des performances médiocres ou ne fonctionnent pas du tout dans le cadre d'un environnement bruité voire très bruité, tel que dans un environnement de rencontre sportive (en extérieur ou en intérieur) avec des arbitres qui doivent communiquer en audio et sans fil. En effet, les procédés connus de détection d'activité de la voix donnent de mauvais résultats lorsque le signal de parole est entaché de bruit.

Parmi les procédés de détection d'activité de la voix connus, certains mettent en œuvre une détection de la fréquence fondamentale caractéristique d'un signal de parole. Dans le cas d'un signal de parole, dit signal ou son voisé, le signal présente en effet une fréquence dite fondamentale, généralement appelée « pitch », qui correspond à la fréquence de vibration des cordes vocales de la personne qui parle, et qui s'étend généralement entre 70 et 400 Hertz. L'évolution de cette fréquence fondamentale détermine la mélodie de la parole et son étendue dépend du locuteur, de ses habitudes mais aussi de son état physique et mental.

Ainsi, pour réaliser la détection d'un signal de parole, il est connu de partir du principe qu'un tel signal de parole est quasi-périodique et que, de ce fait, une corrélation ou une différence avec le signal lui-même mais décalé présentera des maximums ou des minimums au voisinage de la fréquence fondamentale et de ses multiples.

Le document « YIN, a fundamental frequency estimator for speech and music », par Alain De Cheveigne et Hideki Kawahara, Journal of the Acoustical Society of America, Vol. 111, No. 4, pp. 1917-1930, Avril 2002, propose et développe une méthode basée sur la différence entre le signal et le même signal décalé temporellement.

Plusieurs méthodes décrites ci-après sont basées sur la détection de la fréquence fondamentale du signal de parole ou pitch dans un signal acoustique $x(t)$ bruité.

Une première méthode de détection de la fréquence fondamentale met en œuvre la recherche du maximum de la fonction d'auto-corrélation $R(\tau)$ définie par la relation suivante :

$$R(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-\tau} x(n)x(n+\tau) , \quad 0 \leq \tau \leq \max(\tau) .$$

Cette première méthode employant la fonction d'auto-corrélation ne donne cependant pas satisfaction dès lors qu'il y a présence de bruit relativement important. De plus la fonction d'auto-corrélation souffre de la présence de maximums qui ne correspondent pas à la fréquence fondamentale ou à ses multiples, mais à des sous-multiples de celle-ci.

Une deuxième méthode de détection de la fréquence fondamentale met en œuvre la recherche du minimum de la fonction différence $D(\tau)$ définie par la relation suivante :

$$D(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-\tau} |x(n) - x(n+\tau)| , \quad 0 \leq \tau \leq \max(\tau) ,$$

où $| \cdot |$ est l'opérateur valeur absolue, cette fonction différence étant minimale au voisinage de la fréquence fondamentale et de ses multiples, puis la comparaison de ce minimum avec un seuil pour en déduire la décision de présence de voix ou non.

Par rapport à la fonction d'auto-corrélation $R(\tau)$, la fonction différence $D(\tau)$ possède l'avantage d'offrir une charge de calcul plus faible, rendant ainsi cette deuxième méthode plus intéressante pour des applications en temps réel. Cependant, cette deuxième méthode ne donne pas non plus entière satisfaction dès lors qu'il y a présence de bruit.

Une troisième méthode de détection de la fréquence fondamentale met en œuvre le calcul, en considérant une fenêtre de traitement de longueur H où $H < N$, de la fonction différence carrée $d_t(\tau)$ définie par la relation :

$$d_t(\tau) = \sum_{j=t}^{t+H-1} (x_j - x_{j+\tau})^2 ,$$

Puis on poursuit avec la recherche du minimum de la fonction différence carrée $d_t(\tau)$, cette fonction différence carrée étant minimale au voisinage de la fréquence fondamentale et de ses multiples, et enfin la comparaison de ce minimum avec un seuil pour en déduire la décision de présence de voix ou non.

Une amélioration connue de cette troisième méthode consiste à normaliser la fonction différence carrée $d_t(\tau)$ en calculant une fonction différence carrée normalisée $d'_t(\tau)$ répondant à la relation suivante :

$$d'_t(\tau) = \begin{cases} 1, & \text{si } \tau = 0 \\ \frac{d_t(\tau)}{(\frac{1}{\tau}) \sum_{j=1}^{\tau} d_t(j)} & \text{sinon} \end{cases}$$

Bien que présentant une meilleure immunité au bruit et donnant, dans ce cadre, de meilleurs résultats de détection, cette troisième méthode présente des limites en termes de détection de voix, en particulier dans des zones de bruit à faible RSB (Rapport Signal sur Bruit) caractéristiques d'un
5 environnement très bruyé.

L'état de la technique peut également être illustré par l'enseignement de la demande de brevet FR 2 825 505 qui met en œuvre la troisième méthode de détection de la fréquence fondamentale précitée, pour l'extraction de cette fréquence fondamentale. Dans cette demande de brevet,
10 la fonction différence carrée normalisée $d'_t(\tau)$ peut être comparée à un seuil pour déterminer cette fréquence fondamentale - ce seuil pouvant être fixe ou varier en fonction du décalage temporel τ - et cette méthode présente les inconvénients précités associés à cette troisième méthode.

La présente invention a pour but de proposer un procédé de
15 détection de la voix qui offre une détection des signaux de parole contenus dans un signal acoustique bruyé, en particulier dans des environnements bruyés voire très bruyés.

Elle propose plus particulièrement un procédé de détection de la voix qui est très adapté pour la communication (notamment entre arbitres) à
20 l'intérieur d'un stade où le bruit est relativement très fort en niveau et est fortement non stationnaire, avec des étapes de détection qui évitent surtout les mauvaises ou fausses détections (généralement appelées « tonches ») dues aux chants des spectateurs, tambours, musiques et sifflets.

A cet effet, elle propose un procédé de détection de la voix
25 permettant de détecter la présence de signaux de parole dans un signal acoustique $x(t)$ bruyé issu d'un microphone, comportant les étapes successives suivantes :

- calcul d'une fonction de détection $FD(\tau)$ basée sur le calcul d'une fonction différence $D(\tau)$ variant en fonction du décalage τ sur une fenêtre d'intégration
30 de longueur W débutant au temps t_0 , avec :

$$D(\tau) = \sum_{n=t_0}^{t_0+W-1} |x(n) - x(n + \tau)| \quad \text{où } 0 \leq \tau \leq \max(\tau) \quad ;$$

- recherche du minimum de la fonction de détection $FD(\tau)$ et comparaison de ce minimum avec un seuil, pour τ variant dans un intervalle de temps déterminé dit intervalle courant pour détecter la présence ou non d'une fréquence fondamentale F_0 caractéristique d'un signal de parole dans ledit intervalle courant ;

ledit procédé étant remarquable en qu'il comporte, avant l'étape de recherche et de comparaison, une étape d'adaptation du seuil dans ledit intervalle courant, en fonction de valeurs calculées à partir du signal acoustique $x(t)$ établies dans au moins un intervalle de temps précédent ledit intervalle courant, et notamment des valeurs maximales dudit signal acoustique $x(t)$.

Ainsi, ce procédé est basé sur le principe d'un seuil adaptatif, qui sera relativement bas pendant les périodes de bruit ou de silence et relativement haut pendant les périodes de parole. De ce fait, les fausses détections seront minimisées et la parole sera détectée correctement avec un minimum de coupures en début et fin de mots.

Selon une première possibilité, la fonction de détection $FD(\tau)$ correspond à la fonction différence $D(\tau)$.

Selon une seconde possibilité, la fonction de détection $FD(\tau)$ correspond à la fonction différence normalisée $DN(\tau)$ calculée à partir de la fonction différence $D(\tau)$ comme suit :

$$DN(\tau) = 1 \text{ si } \tau = 0 ,$$

$$DN(\tau) = \frac{D(\tau)}{(1/\tau) \sum_{j=1}^{\tau} D(j)} \text{ si } \tau \neq 0 .$$

Il est bien entendu avantageux de réaliser le procédé sur un signal acoustique échantillonné, autrement dit le procédé intègre une étape préalable d'échantillonnage comprenant un découpage du signal acoustique $x(t)$ en un signal acoustique discret $\{x_i\}$ composé d'une suite de vecteurs associés à des trames i temporelles de longueur N , N correspondant au nombre de points d'échantillonnage, où chaque vecteur traduit le contenu acoustique de la trame i associée et est composé des N échantillons $x_{(i-1)N+1}, x_{(i-1)N+2}, \dots, x_{iN-1}, x_{iN}$, i entier positif, de sorte que :

- le calcul de la fonction de détection $FD(\tau)$ consiste en un calcul d'une fonction de détection discrète $FD_i(\tau)$ associée aux trames i ;

- l'adaptation du seuil consiste à, pour chaque trame i , adapter un seuil Ω_i propre à la trame i en fonction de valeurs de référence calculées à partir des valeurs des échantillons du signal acoustique discret $\{x_i\}$ dans ladite trame i ;
 - la recherche du minimum de la fonction de détection $FD(\tau)$ et la
- 5 comparaison de ce minimum avec un seuil sont réalisées en recherchant, sur chaque trame i , le minimum $rr(i)$ de la fonction de détection discrète $FDi(\tau)$ et en comparant ce minimum $rr(i)$ avec le seuil Ω_i propre à la trame i .

Dans une réalisation particulière, la fonction différence discrète $D_i(\tau)$ relative à la trame i est calculée comme suit :

- 10 - on subdivise la trame i en K sous-trames de longueur H , avec par exemple $K = \left\lfloor \frac{N - \max(\tau)}{H} \right\rfloor$ où $\lfloor \cdot \rfloor$ représente l'opérateur d'arrondi en partie entière, de sorte que les échantillons du signal acoustique discret $\{x_i\}$ dans une sous-trame d'indice p de la trame i comprennent les H échantillons :

$X_{(i-1)N+(p-1)H+1}, X_{(i-1)N+(p-1)H+2}, \dots, X_{(i-1)N+pH}$, p entier positif compris entre 1 et K ;

- 15 - pour chaque sous-trame d'indice p , on calcule la fonction différence $dd_p(\tau)$ suivante :

$$dd_p(\tau) = \sum_{j=(i-1)N+(p-1)H+1}^{(i-1)N+pH} |x_j - x_{j+\tau}| ,$$

- on calcule la fonction différence discrète $D_i(\tau)$ relative à la trame i comme la somme des fonctions différence $dd_p(\tau)$ des sous-trames d'indice p de la trame
- 20 i , soit :

$$D_i(\tau) = \sum_{p=1}^K dd_p(\tau) .$$

Dans le cas de la seconde possibilité citée ci-dessus, le calcul de la fonction différence normalisée $DN(\tau)$ consiste en un calcul d'une fonction différence normalisée discrète $DN_i(\tau)$ associée aux trames i , où :

- 25 $DN_i(\tau) = 1$ si $\tau = 0$,

$$DN_i(\tau) = \frac{D_i(\tau)}{(1/\tau) \sum_{j=1}^{\tau} D_i(j)} \text{ si } \tau \neq 0 .$$

De manière avantageuse, l'étape d'adaptation des seuils Ω_i pour chaque trame i comporte les étapes suivantes :

- a)- on subdivise la trame i comprenant N points d'échantillonnage en T sous-trames de longueur L , où N est un multiple de T afin que la longueur $L=N/T$ soit
- 30 entière, et de sorte que les échantillons du signal acoustique discret $\{x_i\}$ dans

une sous-trame d'indice j de la trame i comprennent les L échantillons suivants :

$X_{(i-1)N+(j-1)L+1}, X_{(i-1)N+(j-1)L+2}, \dots, X_{(i-1)N+jL}$, j entier positif compris entre 1 et T ;

b)- on calcule les valeurs maximales $m_{i,j}$ du signal acoustique discret $\{x_i\}$ dans chaque sous-trame d'indice j de la trame i , avec :

$$m_{i,j} = \max \{X_{(i-1)N+(j-1)L+1}, X_{(i-1)N+(j-1)L+2}, \dots, X_{(i-1)N+jL}\};$$

c)- on calcule au moins une valeur de référence $Ref_{i,j}$, $MRef_{i,j}$ propre à la sous-trame j de la trame i , la ou chaque valeur de référence $Ref_{i,j}$, $MRef_{i,j}$ par sous-trame j étant calculée à partir de la valeur maximale $m_{i,j}$ dans la sous-trame j de la trame i ;

d)- on établit la valeur du seuil Ω_i propre à la trame i en fonction de toutes les valeurs de référence $Ref_{i,j}$, $MRef_{i,j}$ calculées dans les sous-frames j de la trame i .

Ainsi, on considère les valeurs maximales $m_{i,j}$ établies dans les sous-frames j pour prendre la décision (voix ou absence de voix) sur la trame i entière.

Selon une caractéristique, lors de l'étape c), on réalise les sous-étapes suivantes sur chaque trame i :

c1)-on calcule les enveloppes lissées des maximum $\bar{m}_{i,j}$ dans chaque sous-trame d'indice j de la trame i , avec :

$$\bar{m}_{i,j} = \lambda \bar{m}_{i,j-1} + (1 - \lambda)m_{i,j}, \text{ où } \lambda \text{ est un coefficient prédéfini compris entre 0 et 1 ;}$$

c2)-on calcule les signaux de variation $\Delta_{i,j}$ dans chaque sous-trame d'indice j de la trame i , avec :

$$\Delta_{i,j} = m_{i,j} - \bar{m}_{i,j} = \lambda (m_{i,j} - \bar{m}_{i,j-1});$$

et où au moins une valeur de référence dite principale $Ref_{i,j}$ par sous-trame j est calculée à partir du signal de variation $\Delta_{i,j}$ dans la sous-trame j de la trame i .

Ainsi, on considère les signaux de variation $\Delta_{i,j}$ établies dans les sous-frames j pour prendre la décision (voix ou absence de voix) sur la trame i entière, fiabilisant la détection de la parole (ou voix).

Selon une autre caractéristique, lors de l'étape c) et à la suite de la sous-étape c2), on réalise les sous-étapes suivantes sur chaque trame i :

c3)-on calcule les maximums de variation $s_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où $s_{i,j}$ correspond au maximum du signal de variation $\Delta_{i,j}$ calculé sur une fenêtre glissante de longueur L_m antérieure à ladite sous-trame j , ladite longueur L_m étant variable selon que la sous-trame j de la trame i correspond à
 5 une période de silence ou de présence de parole ;

c4)-on calcule les écarts de variation $\delta_{i,j}$ dans chaque sous-trame d'indice j de la trame i , avec :

$$\delta_{i,j} = \Delta_{i,j} - s_{i,j} ;$$

et où on calcule, pour chaque sous-trame j de la trame i , deux valeurs de
 10 référence principales $Ref_{i,j}$ à partir respectivement du signal de variation $\Delta_{i,j}$ et de l'écart de variation $\delta_{i,j}$.

Ainsi, on considère conjointement les signaux de variation $\Delta_{i,j}$ et les écarts de variation $\delta_{i,j}$ établies dans les sous-frames j pour choisir la valeur du seuil Ω_i adaptatif et ainsi prendre la décision (voix ou absence de voix) sur la
 15 trame i entière, renforçant la détection de la parole. Autrement dit, on étudie le couple $(\Delta_{i,j} ; \delta_{i,j})$ pour déterminer la valeur du seuil Ω_i adaptatif.

Avantageusement, lors de l'étape c) et à la suite de la sous-étape c4), on réalise une sous-étape c5) de calcul des signaux de variation normalisés $\Delta'_{i,j}$ et des écarts de variation normalisés $\delta'_{i,j}$ dans chaque sous-
 20 trame d'indice j de la trame i , comme suit :

$$\Delta'_{i,j} = \frac{\Delta_{i,j}}{\bar{m}_{i,j}} = \frac{m_{i,j} - \bar{m}_{i,j}}{\bar{m}_{i,j}} ;$$

$$\delta'_{i,j} = \frac{\delta_{i,j}}{\bar{m}_{i,j}} = \frac{m_{i,j} - \bar{m}_{i,j} - s_{i,j}}{\bar{m}_{i,j}} ;$$

et où, pour chaque sous-trame j d'une trame i , le signal de variation normalisé $\Delta'_{i,j}$ et l'écart de variation normalisé $\delta'_{i,j}$ constituent chacun une valeur de
 25 référence principale $Ref_{i,j}$ de sorte que, lors de l'étape d), on établit la valeur du seuil Ω_i propre à la trame i en fonction du couple $(\Delta'_{i,j}, \delta'_{i,j})$ des signaux de variation normalisés $\Delta'_{i,j}$ et des écarts de variation normalisés $\delta'_{i,j}$ dans les sous-frames j de la trame i .

De cette façon, on peut traiter la variation du seuil Ω_i
 30 indépendamment des niveaux des signaux $\Delta_{i,j}$ et $\delta_{i,j}$ en les normalisant avec le calcul des signaux normalisés $\Delta'_{i,j}$ et $\delta'_{i,j}$. Ainsi, les seuils Ω_i choisis à partir de ces signaux normalisés $\Delta'_{i,j}$ et $\delta'_{i,j}$ seront indépendants du niveau du signal

acoustique discret $\{x_i\}$. Autrement dit, on étudie le couple $(\Delta'_{i,j}; \delta'_{i,j})$ pour déterminer la valeur du seuil Ω_i adaptatif.

De manière avantageuse, lors de l'étape d), la valeur du seuil Ω_i propre à la trame i est établie en partitionnant l'espace défini par la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$, et en examinant la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$ sur une ou plusieurs (par exemple entre une et trois) sous-trames successives selon la zone de valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$.

Ainsi, la procédure de calcul du seuil Ω_i se base sur une partition expérimentale de l'espace défini par la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$. A cela s'adjoint un mécanisme de décision qui scrute la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$ sur une, deux ou plus sous-trames successives selon la zone de valeur du couple. Les conditions de tests de positionnement de la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$ dépendent surtout de la détection de parole lors de la trame précédente et le mécanisme de scrutation sur les une, deux ou plus sous-trames successives utilise aussi un partitionnement expérimental.

Selon une caractéristique, lors de la sous-étape c3), la longueur L_m de la fenêtre glissante répond aux équations suivantes :

- $L_m = L_0$ si la sous-trame j de la trame i correspond à une période de silence ;
- $L_m = L_1$ si la sous-trame j de la trame i correspond à une période de présence de parole ;

avec $L_1 < L_0$, et notamment avec $L_1 = k_1.L$ et $L_0 = k_0.L$, L étant la longueur des sous-trame d'indice j et k_0, k_1 étant des entiers positifs.

Selon une autre caractéristique, lors de la sous-étape c3), pour chaque calcul du maximum de variation $s_{i,j}$ dans la sous-trame j de la trame i , la fenêtre glissante de longueur L_m est en retard de M_m trames de longueur N vis-à-vis de ladite sous-trame j .

Selon une autre caractéristique, on réalise les perfectionnements suivants :

- lors de la sous-étape c3), on calcule également les maximums de variation normalisés $s'_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où $s'_{i,j}$ correspond au maximum du signal de variation normalisé $\Delta'_{i,j}$ calculé sur une fenêtre glissante de longueur L_m antérieure à ladite sous-trame j , où :

$$s'_{i,j} = \frac{s_{i,j}}{m_{i,j}} ;$$

et où chaque maximum de variation normalisé $s'_{i,j}$ est calculée selon une méthode de minimisation comprenant les étapes itératives suivantes :

- calcul de $s'_{i,j} = \max \{s'_{i,j-1} ; \Delta'_{i-M_m,j}\}$ et $\tilde{s}'_{i,j} = \max \{s'_{i,j-1} ; \Delta'_{i-M_m,j}\}$
- 5 - si $\text{rem}(i, L_m) = 0$, où rem est l'opérateur reste de la division entière de deux entiers, alors :

$$s'_{i,j} = \max \{\tilde{s}'_{i,j-1} ; \Delta'_{i-M_m,j}\},$$

$$\tilde{s}'_{i,j} = \Delta'_{i-M_m,j}$$

avec $s'_{0,1} = 0$ et $\tilde{s}'_{0,1} = 0$; et

- 10 - lors de l'étape c4), on calcule les écarts de variation normalisés $\delta'_{i,j}$ dans chaque sous-trame d'indice j de la trame i , comme suit :

$$\delta'_{i,j} = \Delta'_{i,j} - s'_{i,j} .$$

- De manière avantageuse, lors de l'étape c), on réalise une sous-étape c6) dans laquelle on calcule les maximums de maximale $q_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où $q_{i,j}$ correspond au maximum de la valeur maximale $m_{i,j}$ calculée sur une fenêtre glissante de longueur fixe L_q antérieure à ladite sous-trame j , où la fenêtre glissante de longueur L_q est en retard de M_q trames de longueur N vis-à-vis de ladite sous-trame j , et où une autre valeur de référence dite secondaire $M\text{Ref}_{i,j}$ par sous-trame j correspond audit
- 15
 - 20

- maximum de maximale $q_{i,j}$ dans la sous-trame j de la trame i .
- Ainsi, pour éviter d'avantage les fausses détections, il est avantageux de prendre en compte également ce signal $q_{i,j}$ (valeur de référence secondaire $M\text{Ref}_{i,j} = q_{i,j}$) qui est calculé d'une façon similaire au calcul du signal $s_{i,j}$ précité, mais qui opère sur les valeurs maximales $m_{i,j}$ au lieu d'opérer sur
- 25

- les signaux de variation $\Delta_{i,j}$ ou sur les signaux de variation normalisés $\Delta'_{i,j}$.
- Dans un mode de réalisation particulier, lors de l'étape d), le seuil Ω_i propre à la trame i est découpé en plusieurs sous-seuils $\Omega_{i,j}$ propres à chaque sous-trame j de la trame i , et la valeur de chaque sous-seuil $\Omega_{i,j}$ est au moins établie en fonction de la ou des valeurs de référence $\text{Ref}_{i,j}$, $M\text{Ref}_{i,j}$
- 30

calculées dans la sous-trame j de la trame i correspondante.

Ainsi, on a $\Omega_i = \{\Omega_{i,1} ; \Omega_{i,2} ; \dots ; \Omega_{i,T}\}$, traduisant le découpage du seuil Ω_i en plusieurs sous-seuils $\Omega_{i,j}$ propres aux sous-frames j , apportant une finesse supplémentaire dans l'établissement du seuil Ω_i adaptatif.

De manière avantageuse, lors de l'étape d), on établit la valeur de chaque seuil $\Omega_{i,j}$ propre à la sous-trame j de la trame i en comparant les valeurs du couple $(\Delta'_{i,j}, \delta'_{i,j})$ avec plusieurs couples de seuils fixes, la valeur de chaque seuil $\Omega_{i,j}$ étant sélectionnée parmi plusieurs valeurs fixes en fonction des comparaisons du couple $(\Delta'_{i,j}, \delta'_{i,j})$ avec lesdits couples de seuils fixes.

Ces couples de seuils fixes sont par exemple déterminés expérimentalement par une répartition de l'espace des valeurs $(\Delta'_{i,j}, \delta'_{i,j})$ en zones de décisions.

De manière complémentaire, on établit la valeur de chaque seuil $\Omega_{i,j}$ propre à la sous-trame j de trame i également en réalisant une comparaison du couple $(\Delta'_{i,j}, \delta'_{i,j})$ sur une ou plusieurs sous-frames successives selon la zone initiale du couple $(\Delta'_{i,j}, \delta'_{i,j})$.

Les conditions de tests de positionnement de la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$ dépendent de la détection de parole lors de la trame précédente et le mécanisme de comparaison sur la ou les sous-frames successives utilise aussi un partitionnement expérimental.

Bien entendu, il est également envisageable que l'on établisse la valeur de chaque seuil $\Omega_{i,j}$ propre à la sous-trame j de trame i en comparant :

- les valeurs du couple $(\Delta'_{i,j}, \delta'_{i,j})$ (les valeurs de référence principales $\text{Ref}_{i,j}$) avec plusieurs couples de seuils fixes ;
- les valeurs de $q_{i,j}$ (la valeur de référence secondaire $\text{MRef}_{i,j}$) avec plusieurs autres seuils fixes.

Ainsi, le mécanisme de décision basé sur la comparaison du couple $(\Delta'_{i,j}, \delta'_{i,j})$ avec des couples de seuils fixes, est complété par une autre mécanisme de décision basée sur la comparaison de $q_{i,j}$ avec des autres seuils fixes.

Avantageusement, lors de l'étape d), on réalise une procédure dite de décision comprenant les sous-étapes suivantes, pour chaque trame i :

- pour chaque sous-trame j de la trame i, on établit un indice de décision $\text{DEC}_i(j)$ qui occupe soit un état « 1 » de détection d'un signal de parole soit un état « 0 » de détection d'un signal de parole ;
- on établit une décision temporaire $\text{VAD}(i)$ basée sur la comparaison des indices de décision $\text{DEC}_i(j)$ avec des opérateurs « OU » logiques, de sorte que la décision temporaire $\text{VAD}(i)$ occupe un état « 1 » de détection d'un signal de

parole si au moins l'un desdits indices de décision $DEC_i(j)$ occupe cet état « 1 » de détection d'un signal de parole.

Ainsi, pour éviter les détections tardives (coupures de mots en début de détection), la décision finale (voix ou absence de voix) est prise à la suite de cette procédure de décision en s'appuyant sur la décision temporaire $VAD(i)$ qui est elle-même prise sur la trame i entière, en mettant en œuvre un opérateur « OU » logique sur les décisions prises dans les sous-trames j , et de préférence dans des sous-trames j successives sur un horizon court et fini à partir du début de la trame i .

Lors de cette procédure de décision, on peut réaliser les sous-étapes suivantes, pour chaque trame i :

- on mémorise une valeur maximale de seuil $Lastmax$ qui correspond à la valeur variable d'un seuil de comparaison pour l'amplitude du signal acoustique discret $\{x_i\}$ en-dessous duquel on considère que le signal acoustique ne comprend pas de signal de parole, cette valeur variable étant déterminé durant la dernière trame d'indice k qui précède ladite trame i et dans laquelle la décision temporaire $VAD(k)$ occupait un état « 1 » de détection d'un signal de parole ;
- on mémorise une valeur maximale moyenne $A_{i,j}$ qui correspond à la valeur maximale moyenne du signal acoustique discret $\{x_i\}$ dans la sous-trame j de la trame i calculée comme suit :

$$A_{i,j} = \theta A_{i,j-1} + (1 - \theta)a_{i,j}$$

où $a_{i,j}$ correspond au maximum du signal acoustique discret $\{x_i\}$ contenu dans une trame k formée par la sous-trame j de la trame i et par au moins une ou plusieurs sous-trames successives qui précèdent ladite sous-trame j ; et

θ est un coefficient prédéfini compris entre 0 et 1 avec $\theta < \lambda$

- on établit la valeur de chaque sous-seuil $\Omega_{i,j}$ en fonction de la comparaison entre ladite valeur maximale de seuil $Lastmax$ et des valeurs maximales moyennes $A_{i,j}$ et $A_{i,j-1}$ considérées sur deux sous-trames j et $j-1$ successives.

Dans beaucoup de cas, les fausses détections arrivent avec une amplitude inférieure à celle du signal de parole (le microphone étant situé à côté de la bouche de la personne qui communique). Ainsi, cette procédure de décision vise à éliminer encore plus les mauvaises détections en mémorisant la

valeur maximale de seuil Lastmax du signal de parole réactualisé dans la dernière période d'activation et les valeurs maximales moyennes $A_{i,j}$ et $A_{i,j-1}$ qui correspondent à la valeur maximale moyenne du signal acoustique discret $\{x_i\}$ dans les sous-frames j et $j-1$ de la trame i . En prenant en compte ces valeurs
 5 (Lastmax, $A_{i,j}$ et $A_{i,j-1}$), on rajoute une condition au niveau de l'établissement du seuil Ω_i adaptatif.

Il est important que la valeur de θ soit choisie comme étant inférieure au coefficient λ pour ralentir les fluctuations de $A_{i,j}$.

Lors de la procédure de décision mentionnée ci-dessus, on
 10 réactualise la valeur maximale de seuil Lastmax à chaque fois que le procédé a considéré qu'une sous-trame p d'une trame k contient un signal de parole, en mettant en œuvre la procédure suivante :

- la détection d'un signal de parole dans la sous-trame p de la trame k fait suite à une période d'absence de parole, et dans ce cas Lastmax prend la
 15 valeur actualisée $[\alpha (A_{k,p} + \text{LastMax})]$, où α est un coefficient prédéfini compris entre 0 et 1, et par exemple compris entre 0,2 et 0,7 ;
- la détection d'un signal de parole dans la sous-trame p de la trame k fait suite à une période de présence de parole, et dans ce cas Lastmax prend la valeur actualisée $A_{k,p}$ si $A_{k,p} > \text{Lastmax}$.

20 La mise à jour de la valeur Lastmax se fait ainsi uniquement durant les périodes d'activation du procédé (autrement dit les périodes de détection de la voix). Dans une situation de détection de parole, la valeur Lastmax vaudra $A_{k,p}$ lorsqu'on aura $A_{k,p} > \text{LastMax}$. Cependant, il est important que cette mise à jour soit faite comme suit lors de l'activation de la première sous-trame p qui
 25 suit une zone de silence : la valeur Lastmax vaudra $[\alpha (A_{k,p} + \text{LastMax})]$.

Ce mécanisme d'actualisation de la valeur maximale de seuil Lastmax permet au procédé de détecter la voix de l'utilisateur même si celui-ci a réduit l'intensité de sa voix (autrement dit parle moins fort) par rapport à la dernière fois où le procédé a détecté qu'il avait parlé.

30 Autrement dit, pour améliorer encore plus l'élimination des fausses détections, on réalise un traitement fin dans lequel la valeur maximale de seuil Lastmax est variable et est comparée aux valeurs maximales moyennes $A_{i,j}$ et $A_{i,j-1}$ du signal acoustique discret.

En effet, des voix lointaines pourraient être captées avec le
 35 procédé, car de telles voix présentent des fréquences fondamentales

susceptibles d'être détectées, tout comme la voix de l'utilisateur. Pour faire en sorte que les voix lointaines, qui peuvent être gênantes dans plusieurs cas d'utilisation, ne soient pas prises en compte par le procédé, on considère un traitement au cours duquel on compare la valeur maximale moyenne du signal (sur deux trames successives), en l'occurrence $A_{i,j}$ et $A_{i,j-1}$, avec Lastmax qui constitue un seuil variable selon l'amplitude de la voix de l'utilisateur mesurée à la dernière activation. Ainsi, on fixe la valeur du seuil Ω_i à une valeur minimale très basse, lorsque le signal sera en dessous du seuil.

Cette condition pour établir la valeur du seuil Ω_i en fonction de la valeur maximale de seuil Lastmax est avantageusement basée sur la comparaison entre :

- la valeur maximale de seuil Lastmax ; et
- les valeurs $[Kp.A_{i,j}]$ et $[Kp.A_{i,j-1}]$, où Kp est un coefficient fixe de pondération compris entre 1 et 2.

De cette manière, on compare la valeur maximale de seuil Lastmax avec les valeurs maximales moyennes du signal acoustique discret $\{x_i\}$ dans les sous-trame j et j-1 ($A_{i,j}$ et $A_{i,j-1}$) pondérés d'un coefficient de pondération Kp compris entre 1 et 2, pour renforcer la détection. Cette comparaison se fait uniquement quand la trame précédente n'a pas donné lieu à une détection de voix.

De manière avantageuse, le procédé comporte en outre une phase dite de blocage comprenant une étape de basculement d'un état de non détection d'un signal de parole à un état de détection d'un signal de parole après avoir détecté la présence d'un signal de parole sur N_P trames i temporelles successives.

Ainsi, le procédé met en œuvre une étape du type hangover configurée de telle façon que la transition d'une situation sans voix à une situation avec présence de voix ne se fait qu'au bout de N_P trames successives avec présence de voix.

De même, le procédé comporte en outre une phase dite de blocage comprenant une étape de basculement d'un état de détection d'un signal de parole à un état de non détection d'un signal de parole après avoir détecté aucune présence d'un signal voisé sur N_A trames i temporelles successives.

Ainsi, le procédé met en œuvre une étape du type hangover configurée de telle façon que la transition d'une situation avec présence de voix

à une situation sans voix ne se fait qu'au bout de N_A trames successives sans voix.

Sans ces étapes de basculement, le procédé risque de couper ponctuellement le signal acoustique pendant les phrases ou même au milieu
5 des mots prononcés. Pour remédier à cela, ces étapes de basculement mettent en œuvre une étape de blocage ou de hangover sur une série donnée de trames.

Selon une possibilité de l'invention, le procédé comprend une étape d'interruption de la phase de blocage dans des zones de décision intervenant
10 en fin de mots et dans une situation non-bruitée, lesdites zones de décision étant détectées en analysant le minimum $rr(i)$ de la fonction de détection discrète $FDi(\tau)$.

Ainsi, la phase de blocage est interrompue à la fin d'une phrase ou mot lors d'une détection particulière dans l'espace de décision. Cette
15 interruption ne survient que dans une situation non ou peu bruitée. A ce titre, le procédé prévoit d'isoler une zone de décision particulière qui ne survient qu'en fin de mots et dans une situation non-bruitée. Pour renforcer la décision de détection de cette zone, le procédé utilise aussi le minimum $rr(i)$ de la fonction de détection discrète $FDi(\tau)$, où la fonction de détection discrète $FDi(\tau)$
20 correspond soit à la fonction différence discrète $D_i(\tau)$ soit à la fonction différence normalisée discrète $DN_i(\tau)$. De ce fait, la voix sera coupée plus rapidement en fin de parole, conférant ainsi au système une meilleure qualité audio.

L'invention a également pour objet un programme d'ordinateur
25 comprenant des instructions de code aptes à commander l'exécution des étapes du procédé de détection de la voix tel que défini ci-dessus lorsqu'il est exécuté par un processeur.

L'invention a en outre pour objet un support d'enregistrement de données d'enregistrement sur lequel est stocké un programme d'ordinateur tel
30 que défini ci-dessus.

L'invention a pour autre objet une mise à disposition d'un programme d'ordinateur tel que défini ci-dessus sur un réseau de télécommunication en vue de son téléchargement.

D'autres caractéristiques et avantages de la présente invention
35 apparaîtront à la lecture de la description détaillée ci-après, d'un exemple de

mise en œuvre non limitatif, faite en référence aux figures annexées dans lesquelles :

- la figure 1 est un schéma synoptique du procédé conforme à l'invention ;
- 5 - la figure 2 est une vue schématique d'une boucle de limitation mis en œuvre par une étape de blocage de décision dite étape du type hangover ;
- la figure 3 illustre le résultat d'un procédé de détection de la voix utilisant un seuil fixe avec, en haut, une représentation de la courbe du minimum $rr(i)$ de la fonction de détection et de la ligne de seuil fixe Ω_{fix} et, en bas, une représentation du signal acoustique discret $\{x_i\}$ et du signal de sortie DF_i ;
- 10 - la figure 4 illustre le résultat d'un procédé de détection de la voix conforme à l'invention en utilisant un seuil adaptatif avec, en haut, une représentation de la courbe du minimum $rr(i)$ de la fonction de détection et de la ligne de seuil adaptatif Ω_i et, en bas, une représentation du signal acoustique discret $\{x_i\}$ et du signal de sortie DF_i .

La description du procédé de détection de la voix est faite en référence à la figure 1 qui illustre schématiquement la succession des différentes étapes nécessaires à la détection de la présence de signaux de parole (ou voix) dans un signal acoustique bruité $x(t)$ issu d'un microphone unique opérant dans un milieu bruité.

Le procédé commence par une étape 101 préalable d'échantillonnage comprenant un découpage du signal acoustique $x(t)$ en un signal acoustique discret $\{x_i\}$ composé d'une suite de vecteurs associés à des trames i temporelles de longueur N , N correspondant au nombre de points d'échantillonnage, où chaque vecteur traduit le contenu acoustique de la trame i associée et est composé des N échantillons $x_{(i-1)N+1}, x_{(i-1)N+2}, \dots, x_{iN-1}, x_{iN}$, i entier positif:

30 A titre d'exemple, le signal acoustique bruité $x(t)$ est découpé en trames de 240 ou 256 échantillons, ce qui à une fréquence d'échantillonnage F_e de 8 kHz correspond à des trames temporelles de 30 ou 32 millisecondes.

Le procédé se poursuit avec une étape 102 de calcul d'une fonction différence discrète $D_i(\tau)$ relative à la trame i est calculée comme suit :

- on subdivise chaque trame i en K sous-trames de longueur H , avec la relation suivante :

$$K = \left\lfloor \frac{N - \max(\tau)}{H} \right\rfloor \text{ où } \lfloor \cdot \rfloor \text{ représente l'opérateur d'arrondi en partie entière,}$$

de sorte que les échantillons du signal acoustique discret $\{x_i\}$ dans une sous-trame d'indice p de la trame i comprennent les H échantillons suivants :

$x_{(i-1)N+(p-1)H+1}, x_{(i-1)N+(p-1)H+2}, \dots, x_{(i-1)N+pH}$, p entier positif compris entre 1 et K ; puis

- pour chaque sous-trame d'indice p , on calcule la fonction différence $dd_p(\tau)$ suivante :

$$10 \quad dd_p(\tau) = \sum_{j=(i-1)N+(p-1)H+1}^{(i-1)N+pH} |x_j - x_{j+\tau}| ,$$

- on calcule la fonction différence discrète $D_i(\tau)$ relative à la trame i comme la somme des fonctions différence $dd_p(\tau)$ des sous-trames d'indice p de la trame i , soit :

$$D_i(\tau) = \sum_{p=1}^K dd_p(\tau) .$$

15 Il est également possible que l'étape 102 comprennent également le calcul d'une fonction différence normalisée discrète $DN_i(\tau)$ à partir de la fonction différence discrète $D_i(\tau)$, comme suit :

$$DN_i(\tau) = 1 \text{ si } \tau = 0 ,$$

$$DN_i(\tau) = \frac{D_i(\tau)}{(1/\tau) \sum_{j=1}^{\tau} D_i(j)} \text{ si } \tau \neq 0 .$$

20 Le procédé se poursuit avec une étape 103 dans laquelle, pour chaque trame i :

- on subdivise la trame i comprenant N points d'échantillonnage en T sous-trames de longueur L , où N est un multiple de T afin que la longueur $L=N/T$ soit entière, et de sorte que les échantillons du signal acoustique discret $\{x_i\}$ dans une sous-trame d'indice j de la trame i comprennent les L échantillons suivants :

$$x_{(i-1)N+(j-1)L+1}, x_{(i-1)N+(j-1)L+2}, \dots, x_{(i-1)N+jL} , j \text{ entier positif compris entre 1 et } T ;$$

b)- on calcule les valeurs maximales $m_{i,j}$ du signal acoustique discret $\{x_i\}$ dans chaque sous-trame d'indice j de la trame i , avec :

$$30 \quad m_{i,j} = \max \{x_{(i-1)N+(j-1)L+1}, x_{(i-1)N+(j-1)L+2}, \dots, x_{(i-1)N+jL}\};$$

A titre d'exemple, on subdivise chaque trame i de longueur 240 (soit $N=240$) en quatre sous-trame j de longueurs 60 (soit $T=4$, et $L=60$).

Ensuite, dans une étape 104, on calcule les enveloppes lissées des maximum $\bar{m}_{i,j}$ dans chaque sous-trame d'indice j de la trame i , définis par :

$$5 \quad \bar{m}_{i,j} = \lambda \bar{m}_{i,j-1} + (1 - \lambda)m_{i,j} , \text{ où } \lambda \text{ est un coefficient prédéfini compris entre 0 et 1.}$$

Puis, dans une étape 105, on calcule les signaux de variation $\Delta_{i,j}$ dans chaque sous-trame d'indice j de la trame i , définis par :

$$\Delta_{i,j} = m_{i,j} - \bar{m}_{i,j} = \lambda (m_{i,j} - \bar{m}_{i,j-1}).$$

10 Ensuite, dans une étape 106, on calcule les signaux de variation normalisés $\Delta'_{i,j}$ définis par :

$$\Delta'_{i,j} = \frac{\Delta_{i,j}}{\bar{m}_{i,j}} = \frac{m_{i,j} - \bar{m}_{i,j}}{\bar{m}_{i,j}} .$$

Ensuite, dans une étape 107, on calcule les maximums de variation $s_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où $s_{i,j}$ correspond au
15 maximum du signal de variation $\Delta_{i,j}$ calculé sur une fenêtre glissante de longueur L_m antérieure à ladite sous-trame j . Lors de cette étape 106, la longueur L_m est variable selon que la sous-trame j de la trame i correspond à une période de silence ou de présence de parole, avec :

- $L_m = L_0$ si la sous-trame j de la trame i correspond à une période de
20 silence ;
- $L_m = L_1$ si la sous-trame j de la trame i correspond à une période de présence de parole ;

avec $L_1 < L_0$. A titre d'exemple, $L_1=k_1.L$ et $L_0=k_0.L$, L étant pour rappel la longueur des sous-trame d'indice j et k_0, k_1 étant des entiers positifs avec
25 $k_1 < k_0$. En outre, la fenêtre glissante de longueur L_m est en retard de M_m trames de longueur N vis-à-vis de ladite sous-trame j .

Durant cette étape 106, on calcule également les maximums de variation normalisés $s'_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où :

$$30 \quad s'_{i,j} = \frac{s_{i,j}}{\bar{m}_{i,j}}.$$

Il est envisageable de calculer les maximums de variation normalisés $s'_{i,j}$ selon une méthode de minimisation comprenant les étapes itératives suivantes :

- calcul de $s'_{i,j} = \max \{s'_{i,j-1} ; \Delta'_{i-M_m,j}\}$ et $\tilde{s}'_{i,j} = \max \{s'_{i,j-1} ; \Delta'_{i-M_m,j}\}$
- 5 - si $\text{rem}(i, L_m) = 0$, où rem est l'opérateur reste de la division entière de deux entiers, alors :

$$s'_{i,j} = \max \{\tilde{s}'_{i,j-1} ; \Delta'_{i-M_m,j}\},$$

$$\tilde{s}'_{i,j} = \Delta'_{i-M_m,j}$$

- fin

- 10 avec $s'_{0,1} = 0$ et $\tilde{s}'_{0,1} = 0$.

Ensuite, dans une étape 108, on calcule les écarts de variation $\delta_{i,j}$ dans chaque sous-trame d'indice j de la trame i , définis par :

$$\delta_{i,j} = \Delta_{i,j} - s_{i,j}.$$

- Dans cette même étape 108, on calcule les écarts de variation normalisés $\delta'_{i,j}$ dans chaque sous-trame d'indice j de la trame i , définis par :
- 15

$$\delta'_{i,j} = \frac{\delta_{i,j}}{\bar{m}_{i,j}} = \frac{m_{i,j} - \bar{m}_{i,j} - s_{i,j}}{\bar{m}_{i,j}}.$$

- Ensuite, dans une étape 109, on calcule les maximums de maximale $q_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où $q_{i,j}$ correspond au maximum de la valeur maximale $m_{i,j}$ calculée sur une fenêtre glissante de longueur fixe L_q antérieure à ladite sous-trame j , où la fenêtre glissante de longueur L_q est en retard de M_q trames de longueur N vis-à-vis de ladite sous-trame j . Avantagusement, $L_q > L_0$, et notamment $L_q = k_q \cdot L$ avec k_q un entier positif et $k_q > k_0$. De plus, on a $M_q > M_m$.
- 20

- Durant cette étape 109, il est envisageable de calculer les maximums de maximale $q_{i,j}$ selon une méthode de minimisation comprenant les étapes itératives suivantes :
- 25

- calcul de $q_{i,j} = \max \{q_{i,j-1} ; m_{i-M_q,j}\}$ et $\tilde{q}_{i,j} = \max \{q_{i,j-1} ; m_{i-M_q,j}\}$
- si $\text{rem}(i, L_q) = 0$, où rem est l'opérateur reste de la division entière de deux entiers, alors :

- 30 $q_{i,j} = \max \{\tilde{q}_{i,j-1} ; m_{i-M_q,j}\},$

$$\tilde{q}_{i,j} = m_{i-M_m,j}$$

- finsi

avec $q_{0,1} = 0$ et $q_{0,1} = 0$.

Ensuite, dans une étape 110, on établit les valeurs de seuils Ω_i propres à chaque trame i , parmi plusieurs valeurs fixes Ω_a , Ω_b , Ω_c , etc. De manière plus fine, on établit les valeurs des sous-seuils $\Omega_{i,j}$ propres à chaque sous-trame j de la trame i , le seuil Ω_i étant découpée en plusieurs sous-seuils $\Omega_{i,j}$. A titre d'exemple, chaque seuil Ω_i ou sous-seuil $\Omega_{i,j}$ prend une valeur fixe choisie parmi six valeurs fixes Ω_a , Ω_b , Ω_c , Ω_d , Ω_e , Ω_f , ces valeurs fixes étant par exemple comprises entre 0,05 et 1, et notamment entre 0,1 et 0,7.

Chaque seuil Ω_i ou sous-seuil $\Omega_{i,j}$ est fixé à une valeur fixe Ω_a , Ω_b , Ω_c , Ω_d , Ω_e , Ω_f par la mise en œuvre de deux analyses :

- première analyse : la comparaison des valeurs du couple $(\Delta'_{i,j}, \delta'_{i,j})$ dans la sous-trame d'indice j de la trame i avec plusieurs couples de seuils fixes ;
- deuxième analyse : la comparaison des maximums de maximale $q_{i,j}$ dans la sous-trame d'indice j de la trame i avec des seuils fixes.

A la suite de ces deux analyses, une procédure dite de décision donnera la décision finale sur la présence de la voix dans la trame i . Cette procédure de décision comprend les sous-étapes suivantes, pour chaque trame i :

- pour chaque sous-trame j de la trame i , on établit un indice de décision $DEC_i(j)$ qui occupe soit un état « 1 » de détection d'un signal de parole soit un état « 0 » de détection d'un signal de parole ;
- on établit une décision temporaire $VAD(i)$ basée sur la comparaison des indices de décision $DEC_i(j)$ avec des opérateurs « OU » logiques, de sorte que la décision temporaire $VAD(i)$ occupe un état « 1 » de détection d'un signal de parole si au moins l'un desdits indices de décision $DEC_i(j)$ occupe cet état « 1 » de détection d'un signal de parole, autrement dit on a la relation suivante :

$VAD(i) = DEC_i(1) + DEC_i(2) + \dots + DEC_i(T)$, où "+" est l'opérateur « OU ».

Ainsi, en fonction des comparaisons faites lors des première et deuxième analyses, et en fonction de l'état de la décision temporaire $VAD(i)$, on fixe le seuil Ω_i à l'une des valeurs fixes Ω_a , Ω_b , Ω_c , Ω_d , Ω_e , Ω_f et on en déduit la décision finale en comparant le minimum $rr(i)$ avec le seuil Ω_i fixé à l'une des ses valeurs fixes (voir description ci-après).

Dans beaucoup de cas, les fausses détections (ou tonches) arrivent avec une amplitude inférieure à celle du signal de parole, le

microphone étant situé à côté de la bouche de l'utilisateur. En tenant compte de ce fait, il est envisageable d'éliminer encore plus les fausses détections en mémorisant la valeur maximale de seuil Lastmax déduit du signal de parole dans la dernière période d'activation du « VAD » et en ajoutant une condition

5 dans le procédé basée sur cette valeur maximale de seuil Lastmax.

Ainsi, dans l'étape 109 décrite précédemment, on rajoute la mémorisation de la valeur maximale de seuil Lastmax qui correspond à la valeur variable (ou actualisé) d'un seuil de comparaison pour l'amplitude du signal acoustique discret $\{x_i\}$ en-dessous duquel on considère que le signal

10 acoustique ne comprend pas de signal de parole, cette valeur variable étant déterminé durant la dernière trame d'indice k qui précède ladite trame i et dans laquelle la décision temporaire VAD(k) occupait un état « 1 » de détection d'un signal de parole.

Dans cette étape 109, on mémorise également une valeur

15 maximale moyenne $A_{i,j}$ qui correspond à la valeur maximale moyenne du signal acoustique discret $\{x_i\}$ dans la sous-trame j de la trame i calculée comme suit :

$$A_{i,j} = \theta A_{i,j-1} + (1 - \theta)a_{i,j}$$

où $a_{i,j}$ correspond au maximum du signal acoustique discret $\{x_i\}$ contenu dans la trame théorique k formée par la sous-trame j de la trame i et par au moins une ou plus sous-frames successives qui précèdent ladite sous-trame j ; et

20 θ est un coefficient prédéfini compris entre 0 et 1 avec $\theta < \lambda$.

Dans cette étape 109, on réactualise la valeur maximale de seuil Lastmax à chaque fois que le procédé a considéré qu'une sous-trame p d'une trame k contient un signal de parole, en mettant en œuvre la procédure suivante :

- 25 - la détection d'un signal de parole dans la sous-trame p de la trame k fait suite à une période d'absence de parole, et dans ce cas Lastmax prend la valeur actualisée $[\alpha (A_{k,p} + \text{LastMax})]$, où α est un coefficient prédéfini compris entre 0 et 1, et par exemple compris entre 0,2 et 0,7 ;
- la détection d'un signal de parole dans la sous-trame p de la trame k fait
- 30 suite à une période de présence de parole, et dans ce cas Lastmax prend la valeur actualisée $A_{k,p}$ si $A_{k,p} > \text{Lastmax}$.

Puis, dans l'étape 110 décrite précédemment, on rajoute une condition basée sur la valeur maximale de seuil Lastmax pour fixer le seuil Ω_i .

Pour chaque trame i , cette condition est basée sur la comparaison entre :

- la valeur maximale de seuil Lastmax, et
- les valeurs $[Kp.A_{i,j}]$ et $[Kp.A_{i,j-1}]$, où Kp est un coefficient fixe de pondération compris entre 1 et 2.

Il est également envisageable d'abaisser la valeur maximale de seuil Lastmax après une période de temporisation donnée (par exemple fixée entre quelques secondes et quelques dizaines de secondes) entre la trame i et la dernière trame d'indice k précitée, afin d'éviter la non détection de la parole si l'utilisateur/locuteur abaisse l'amplitude de sa voix de manière significative.

Ensuite, dans une étape 111, on calcule, pour chaque trame courante i , le minimum $rr(i)$ d'une fonction de détection discrète $FDi(\tau)$, où la fonction de détection discrète $FDi(\tau)$ correspond soit à la fonction différence discrète $D_i(\tau)$ soit à la fonction différence normalisée discrète $DN_i(\tau)$.

Enfin, dans une dernière étape 112, on compare, pour chaque trame courante i , ce minimum $rr(i)$ avec le seuil Ω_i propre à la trame i , pour détecter la présence ou non d'un signal de parole (oui signal voisé), avec :

- si $rr(i) \leq \Omega_i$, alors la trame i est considérée comme présentant un signal de parole et le procédé délivre un signal de sortie DF_i prenant la valeur « 1 » (autrement dit, la décision finale pour la trame i est « présence de voix dans la trame i ») ;
- si $rr(i) > \Omega_i$, alors la trame i est considérée comme ne présentant pas de signal de parole et le procédé délivre un signal de sortie DF_i prenant la valeur « 0 » (autrement dit, la décision finale pour la trame i est « absence de voix dans la trame i »).

En référence aux figures 1 et 2, il est envisageable d'apporter un perfectionnement au procédé, en introduisant une étape supplémentaire 113 de blocage de décision (ou étape de hangover), pour éviter les coupures de son dans une phrase et pendant la prononciation des mots, cette étape 113 de blocage de décision visant à renforcer la décision de présence/absence de voix par la mise en œuvre des deux étapes suivantes :

- basculement d'un état de non détection d'un signal de parole à un état de détection d'un signal de parole après avoir détecté la présence d'un signal de parole sur N_p trames i temporelles successives ;

- basculement d'un état de détection d'un signal de parole à un état de non détection d'un signal de parole après avoir détecté aucune présence d'un signal voisé sur N_A trames i temporelles successives.

Ainsi, cette étape 113 de blocage permet de délivrer en sortie un
 5 signal de décision de la détection de la voix D_V qui prend la valeur « 1 » correspondant à une décision de la détection de la voix et la valeur « 0 » correspondant à une décision de la non détection de la voix, où :

- le signal de décision de la détection de la voix D_V bascule d'un état « 1 » à un état « 0 » si et seulement si le signal de sortie DF_i prend la valeur « 0 » sur N_A
 10 trames i temporelles successives ; et
- le signal de décision de la détection de la voix D_V bascule d'un état « 0 » à un état « 1 » si et seulement si le signal de sortie DF_i prend la valeur « 1 » sur N_P trames i temporelles successives.

En référence à la figure 2, si on suppose que l'on parte d'un état
 15 « $D_V=1$ », on bascule vers un état « $D_V=0$ » si le signal de sortie DF_i prend la valeur « 0 » sur N_A trames successives, sinon l'état reste à « $D_V=1$ » (N_i représentant le numéro de la trame au début de la série). De même, si on suppose que l'on parte d'un état « $D_V=0$ », on bascule vers un état « $D_V=1$ » si le signal de sortie DF_i prend la valeur « 1 » sur N_P trames successives, sinon
 20 l'état reste à « $D_V=0$ ».

La décision finale s'applique aux premiers H échantillons de la trame traitée. De préférence N_A est supérieur à N_P , avec par exemple $N_A=100$ et $N_P=3$, car il vaut mieux risquer de détecter du silence plutôt que de couper une conversation.

25 La suite de la description porte sur deux résultats de détection de voix obtenus avec un procédé classique utilisant un seuil fixe (Figure 3) et avec le procédé conforme à l'invention utilisant un seuil adaptatif (Figure 4).

Sur les figures 3 et 4 (en bas), on note que les deux procédés travaillent sur le même signal acoustique discret $\{x_i\}$, avec en ordonnées
 30 l'amplitude et en abscisse les échantillons. Ce signal acoustique discret $\{x_i\}$ présente une seule zone de présence de parole « PAR », et de nombreuses zones de présence de bruits parasites tels que musique, tambours, cris de foule et sifflets. Ce signal acoustique discret $\{x_i\}$ traduit un environnement représentatif d'une communication entre personnes (comme des arbitres) à
 35 l'intérieur d'un stade ou d'un gymnase où le bruit est relativement très fort en niveau et est fortement non stationnaire.

Sur les figures 3 et 4 (en haut), on note que les deux procédés exploitent la même fonction $rr(i)$ correspondant pour rappel au minimum de la fonction de détection discrète $FDi(\tau)$ sélectionnée.

Sur la figure 3 (en haut), la fonction minimum $rr(i)$ est comparée à un seuil fixe Ω_{fix} sélectionné de manière optimale pour assurer la détection de la voix. Sur la figure 3 (en bas), on note la forme du signal de sortie DF_i qui occupe un état « 1 » si $rr(i) \leq \Omega_{fix}$ et un état « 0 » si $rr(i) > \Omega_{fix}$.

Sur la figure 4 (en haut), la fonction minimum $rr(i)$ est comparée à un seuil adaptatif Ω_i calculé selon les étapes décrites précédemment en référence à la figure 1. Sur la figure 4 (en bas), on note la forme du signal de sortie DF_i qui occupe un état « 1 » si $rr(i) \leq \Omega_i$ et un état « 0 » si $rr(i) > \Omega_i$.

On note sur la figure 3 que le procédé conforme à l'invention permet une détection de la voix dans la zone de présence de parole « PAR » avec le signal de sortie DF_i qui occupe un état « 1 », et que ce même signal de sortie DF_i occupe plusieurs fois un état « 1 » dans les autres zones où la parole est pourtant absente, ce qui correspond par des fausses détections non désirées avec le procédé classique.

Par contre, on note sur la figure 4 que le procédé conforme à l'invention permet une détection optimale de la voix dans la zone de présence de parole « PAR » avec le signal de sortie DF_i qui occupe un état « 1 », et que ce même signal de sortie DF_i occupe un état « 0 » dans les autres zones où la parole est absente. Ainsi, le procédé conforme à l'invention assure une détection de la voix avec une forte réduction du nombre de fausses détections.

Bien entendu l'exemple de mise en œuvre évoqué ci-dessus ne présente aucun caractère limitatif et d'autres améliorations et détails peuvent être apportés au procédé selon l'invention, sans pour autant sortir du cadre de l'invention où d'autres algorithmes de calcul de la fonction de détection $FD(\tau)$ peuvent par exemple être utilisés.

REVENDEICATIONS

1. Procédé de détection de la voix permettant de détecter la présence de signaux de parole dans un signal acoustique $x(t)$ bruité issu d'un microphone, comportant les étapes successives suivantes :

- calcul d'une fonction de détection $FD(\tau)$ basée sur le calcul d'une fonction différence $D(\tau)$ variant en fonction du décalage τ sur une fenêtre d'intégration de longueur W débutant au temps t_0 , avec :

$$D(\tau) = \sum_{n=t_0}^{t_0+W-1} |x(n) - x(n + \tau)| \quad \text{où } 0 \leq \tau \leq \max(\tau) ;$$

10 - recherche du minimum de la fonction de détection $FD(\tau)$ et comparaison de ce minimum avec un seuil, pour τ variant dans un intervalle de temps déterminé dit intervalle courant pour détecter la présence ou non d'une fréquence fondamentale F_0 caractéristique d'un signal de parole dans ledit intervalle courant ;

15 ledit procédé étant caractérisé en ce qu'il comporte, avant l'étape de recherche et de comparaison, une étape d'adaptation du seuil dans ledit intervalle courant, en fonction de valeurs calculées à partir du signal acoustique $x(t)$ établies dans ledit intervalle courant, et notamment des valeurs maximales dudit signal acoustique $x(t)$.

20

2. Procédé de détection selon la revendication 1, dans lequel la fonction de détection $FD(\tau)$ correspond à la fonction différence $D(\tau)$.

25 3. Procédé de détection selon la revendication 1, dans lequel la fonction de détection $FD(\tau)$ correspond à la fonction différence normalisée $DN(\tau)$ calculée à partir de la fonction différence $D(\tau)$ comme suit :

$$DN(\tau) = 1 \text{ si } \tau = 0 ,$$

$$DN(\tau) = \frac{D(\tau)}{(1/\tau) \sum_{j=1}^{\tau} D(j)} \text{ si } \tau \neq 0 .$$

30 4. Procédé selon l'une quelconque des revendications précédentes, comprenant une étape préalable d'échantillonnage comprenant un découpage du signal acoustique $x(t)$ en un signal acoustique discret $\{x_i\}$ composé d'une suite de vecteurs associés à des trames i temporelles de longueur N , N

correspondant au nombre de points d'échantillonnage, où chaque vecteur traduit le contenu acoustique de la trame i associée et est composé des N échantillons $x_{(i-1)N+1}, x_{(i-1)N+2}, \dots, x_{iN-1}, x_{iN}$, i entier positif, de sorte que :

- le calcul de la fonction de détection $FD(\tau)$ consiste en un calcul d'une
5 fonction de détection discrète $FDi(\tau)$ associée aux trames i ;
- l'adaptation du seuil consiste à, pour chaque trame i , adapter un seuil Ω_i propre à la trame i en fonction de valeurs de référence calculées à partir des valeurs des échantillons du signal acoustique discret $\{x_i\}$ dans ladite trame i ;
- la recherche du minimum de la fonction de détection $FD(\tau)$ et la
10 comparaison de ce minimum avec un seuil sont réalisées en recherchant, sur chaque trame i , le minimum $rr(i)$ de la fonction de détection discrète $FDi(\tau)$ et en comparant ce minimum $rr(i)$ avec le seuil Ω_i propre à la trame i .

5. Procédé selon la revendication 4, dans lequel la fonction différence
15 discrète $D_i(\tau)$ relative à la trame i est calculée comme suit :

- on subdivise la trame i en K sous-trames de longueur H , avec par exemple
 $K = \left\lfloor \frac{N - \max(\tau)}{H} \right\rfloor$ où $\lfloor \cdot \rfloor$ représente l'opérateur d'arrondi en partie entière, de
sorte que les échantillons du signal acoustique discret $\{x_i\}$ dans une sous-trame
d'indice p de la trame i comprennent les H échantillons :
20 $x_{(i-1)N+(p-1)H+1}, x_{(i-1)N+(p-1)H+2}, \dots, x_{(i-1)N+pH}$, p entier positif compris entre 1 et K ;
- pour chaque sous-trame d'indice p , on calcule la fonction différence $dd_p(\tau)$
suivante :

$$dd_p(\tau) = \sum_{j=(i-1)N+(p-1)H+1}^{(i-1)N+pH} |x_j - x_{j+\tau}| ,$$

- on calcule la fonction différence discrète $D_i(\tau)$ relative à la trame i comme la
25 somme des fonctions différence $dd_p(\tau)$ des sous-trames d'indice p de la trame
 i , soit :

$$D_i(\tau) = \sum_{p=1}^K dd_p(\tau) .$$

6. Procédé selon les revendications 3 et 5, dans le calcul de la fonction
30 différence normalisée $DN(\tau)$ consiste en un calcul d'une fonction différence
normalisée discrète $DNi(\tau)$ associée aux trames i , où :

$$DN_i(\tau) = 1 \text{ si } \tau = 0 ,$$

$$DN_i(\tau) = \frac{D_i(\tau)}{(1/\tau) \sum_{j=1}^{\tau} D_i(j)} \text{ si } \tau \neq 0 .$$

7. Procédé selon l'une quelconque des revendications 4 à 6, dans lequel
5 l'étape d'adaptation des seuils Ω_i pour chaque trame i comporte les étapes suivantes :

a)- on subdivise la trame i comprenant N points d'échantillonnage en T sous-
trames de longueur L , où N est un multiple de T afin que la longueur $L=N/T$ soit
entière, et de sorte que les échantillons du signal acoustique discret $\{x_i\}$ dans
10 une sous-trame d'indice j de la trame i comprennent les L échantillons
suivants :

$$X_{(i-1)N+(j-1)L+1}, X_{(i-1)N+(j-1)L+2}, \dots, X_{(i-1)N+jL}, j \text{ entier positif compris entre } 1 \text{ et } T ;$$

b)- on calcule les valeurs maximales $m_{i,j}$ du signal acoustique discret $\{x_i\}$ dans
chaque sous-trame d'indice j de la trame i , avec :

$$15 \quad m_{i,j} = \max \{X_{(i-1)N+(j-1)L+1}, X_{(i-1)N+(j-1)L+2}, \dots, X_{(i-1)N+jL}\};$$

c)- on calcule au moins une valeur de référence $Ref_{i,j}$, $MRef_{i,j}$ propre à la sous-
trame j de la trame i , la ou chaque valeur de référence $Ref_{i,j}$, $MRef_{i,j}$ par sous-
trame j étant calculée à partir de la valeur maximale $m_{i,j}$ dans la sous-trame j de
la trame i ;

20 d)- on établit la valeur du seuil Ω_i propre à la trame i en fonction de toutes les
valeurs de référence $Ref_{i,j}$, $MRef_{i,j}$ calculées dans les sous-trames j de la trame
 i .

8. Procédé selon la revendication 7, dans lequel, lors de l'étape c), on
25 réalise les sous-étapes suivantes sur chaque trame i :

c1)-on calcule les enveloppes lissées des maximum $\bar{m}_{i,j}$ dans chaque sous-
trame d'indice j de la trame i , avec :

$$\bar{m}_{i,j} = \lambda \bar{m}_{i,j-1} + (1 - \lambda)m_{i,j} , \text{ où } \lambda \text{ est un coefficient prédéfini compris entre } 0 \text{ et } 1 ;$$

30 c2)-on calcule les signaux de variation $\Delta_{i,j}$ dans chaque sous-trame d'indice j
de la trame i , avec :

$$\Delta_{i,j} = m_{i,j} - \bar{m}_{i,j} = \lambda (m_{i,j} - \bar{m}_{i,j-1}) ;$$

et où au moins une valeur de référence dite principale $Ref_{i,j}$ par sous-trame j est calculée à partir du signal de variation $\Delta_{i,j}$ dans la sous-trame j de la trame i .

9. Procédé selon la revendication 8, dans lequel, lors de l'étape c) et à la suite de la sous-étape c2), on réalise les sous-étapes suivantes sur chaque trame i :

c3)-on calcule les maximums de variation $s_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où $s_{i,j}$ correspond au maximum du signal de variation $\Delta_{i,j}$ calculé sur une fenêtre glissante de longueur L_m antérieure à ladite sous-trame j , ladite longueur L_m étant variable selon que la sous-trame j de la trame i correspond à une période de silence ou de présence de parole ;

c4)-on calcule les écarts de variation $\delta_{i,j}$ dans chaque sous-trame d'indice j de la trame i , avec :

$$\delta_{i,j} = \Delta_{i,j} - s_{i,j} ;$$

et où on calcule, pour chaque sous-trame j de la trame i , deux valeurs de référence principales $Ref_{i,j}$ à partir respectivement du signal de variation $\Delta_{i,j}$ et de l'écart de variation $\delta_{i,j}$.

10. Procédé selon la revendication 9, dans lequel, lors de l'étape c) et à la suite de la sous-étape c4), on réalise une sous-étape c5) de calcul des signaux de variation normalisés $\Delta'_{i,j}$ et des écarts de variation normalisés $\delta'_{i,j}$ dans chaque sous-trame d'indice j de la trame i , comme suit :

$$\Delta'_{i,j} = \frac{\Delta_{i,j}}{\overline{m}_{i,j}} = \frac{m_{i,j} - \overline{m}_{i,j}}{\overline{m}_{i,j}} ;$$

$$\delta'_{i,j} = \frac{\delta_{i,j}}{\overline{m}_{i,j}} = \frac{m_{i,j} - \overline{m}_{i,j} - s_{i,j}}{\overline{m}_{i,j}} ;$$

et où, pour chaque sous-trame j d'une trame i , le signal de variation normalisé $\Delta'_{i,j}$ et l'écart de variation normalisé $\delta'_{i,j}$ constituent chacun une valeur de référence principale $Ref_{i,j}$ de sorte que, lors de l'étape d), on établit la valeur du seuil Ω_i propre à la trame i en fonction du couple $(\Delta'_{i,j}, \delta'_{i,j})$ des signaux de variation normalisés $\Delta'_{i,j}$ et des écarts de variation normalisés $\delta'_{i,j}$ dans les sous-frames j de la trame i .

11. Procédé selon la revendication 10, dans laquelle, lors de l'étape d), la valeur du seuil Ω_i propre à la trame i est établie en partitionnant l'espace défini par la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$, et en examinant la valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$ sur une ou plusieurs sous-trames successives selon la zone de valeur du couple $(\Delta'_{i,j}, \delta'_{i,j})$.

12. Procédé selon l'une quelconque des revendications 9 à 11, dans lequel, lors de la sous-étape c3), la longueur L_m de la fenêtre glissante répond aux équations suivantes :

- $L_m = L_0$ si la sous-trame j de la trame i correspond à une période de silence ;
- $L_m = L_1$ si la sous-trame j de la trame i correspond à une période de présence de parole ;

avec $L_1 < L_0$, et notamment avec $L_1 = k_1.L$ et $L_0 = k_0.L$, L étant la longueur des sous-trame d'indice j et k_0, k_1 étant des entiers positifs.

13. Procédé selon la revendication 10, dans lequel, lors de la sous-étape c3), pour chaque calcul du maximum de variation $s_{i,j}$ dans la sous-trame j de la trame i , la fenêtre glissante de longueur L_m est en retard de M_m trames de longueur N vis-à-vis de ladite sous-trame j .

14. Procédé selon les revendications 10 et 13, dans lequel, lors de la sous-étape c3), on calcule également les maximums de variation normalisés $s'_{i,j}$ dans chaque sous-trame d'indice j de la trame i , où $s'_{i,j}$ correspond au maximum du signal de variation normalisé $\Delta'_{i,j}$ calculé sur une fenêtre glissante de longueur L_m antérieure à ladite sous-trame j , où :

$$s'_{i,j} = \frac{s_{i,j}}{m_{i,j}} ;$$

et où chaque maximum de variation normalisé $s'_{i,j}$ est calculée selon une méthode de minimisation comprenant les étapes itératives suivantes :

- calcul de $s'_{i,j} = \max \{s'_{i,j-1} ; \Delta'_{i-M_m,j}\}$ et $\tilde{s}'_{i,j} = \max \{s'_{i,j-1} ; \Delta'_{i-M_m,j}\}$
- si $\text{rem}(i, L_m) = 0$, où rem est l'opérateur reste de la division entière de deux entiers, alors :

$$s'_{i,j} = \max \{\tilde{s}'_{i,j-1} ; \Delta'_{i-M_m,j}\},$$

$$\tilde{s}'_{i,j} = \Delta'_{i-M_m,j}$$

avec $s'_{0,1} = 0$ et $\tilde{s}'_{0,1} = 0$;

et dans lequel, lors de l'étape c4), on calcule les écarts de variation normalisés $\delta'_{i,j}$ dans chaque sous-trame d'indice j de la trame i, comme suit :

$$5 \quad \delta'_{i,j} = \Delta'_{i,j} - s'_{i,j}.$$

15. Procédé selon l'une quelconque des revendications 8 à 14, dans lequel, lors de l'étape c), on réalise une sous-étape c6) dans laquelle on calcule les maximums de maximale $q_{i,j}$ dans chaque sous-trame d'indice j de la trame i, où $q_{i,j}$ correspond au maximum de la valeur maximale $m_{i,j}$ calculée sur une fenêtre glissante de longueur fixe L_q antérieure à ladite sous-trame j, où la fenêtre glissante de longueur L_q est en retard de M_q trames de longueur N vis-à-vis de ladite sous-trame j, et où une autre valeur de référence dite secondaire $MRef_{i,j}$ par sous-trame j correspond audit maximum de maximale $q_{i,j}$ dans la sous-trame j de la trame i.

16. Procédé selon l'une quelconque des revendications 8 à 15, dans lequel, lors de l'étape d), le seuil Ω_i propre à la trame i est découpée en plusieurs sous-seuils $\Omega_{i,j}$ propres à chaque sous-trame j de la trame i, et la valeur de chaque sous-seuil $\Omega_{i,j}$ est au moins établie en fonction de la ou des valeurs de référence $Ref_{i,j}$, $MRef_{i,j}$ calculées dans la sous-trame j de la trame i correspondante.

17. Procédé selon les revendications 10 et 16, dans lequel, lors de l'étape d), on établit la valeur de chaque seuil $\Omega_{i,j}$ propre à la sous-trame j de trame i en comparant les valeurs du couple $(\Delta'_{i,j}, \delta'_{i,j})$ avec plusieurs couples de seuils fixes, la valeur de chaque seuil $\Omega_{i,j}$ étant sélectionnée parmi plusieurs valeurs fixes en fonction des comparaisons du couple $(\Delta'_{i,j}, \delta'_{i,j})$ avec lesdits couples de seuils fixes.

30

18. Procédé selon l'une quelconque des revendications 8 à 17, dans lequel, lors de l'étape d), on réalise une procédure dite de décision comprenant les sous-étapes suivantes, pour chaque trame i :

- pour chaque sous-trame j de la trame i , on établit un indice de décision $DEC_i(j)$ qui occupe soit un état « 1 » de détection d'un signal de parole soit un état « 0 » de détection d'un signal de parole ;
 - on établit une décision temporaire $VAD(i)$ basée sur la comparaison des indices de décision $DEC_i(j)$ avec des opérateurs « OU » logiques, de sorte que la décision temporaire $VAD(i)$ occupe un état « 1 » de détection d'un signal de parole si au moins l'un desdits indices de décision $DEC_i(j)$ occupe cet état « 1 » de détection d'un signal de parole.
19. Procédé selon les revendications 16 et 18, dans lequel, lors de la procédure de décision, on réalise les sous-étapes suivantes, pour chaque trame i :
- on mémorise une valeur maximale de seuil $Lastmax$ qui correspond à la valeur variable d'un seuil de comparaison pour l'amplitude du signal acoustique discret $\{x_i\}$ en-dessous duquel on considère que le signal acoustique ne comprend pas de signal de parole, cette valeur variable étant déterminé durant la dernière trame d'indice k qui précède ladite trame i et dans laquelle la décision temporaire $VAD(k)$ occupait un état « 1 » de détection d'un signal de parole ;
 - on mémorise une valeur maximale moyenne $A_{i,j}$ qui correspond à la valeur maximale moyenne du signal acoustique discret $\{x_i\}$ dans la sous-trame j de la trame i calculée comme suit :

$$A_{i,j} = \theta A_{i,j-1} + (1 - \theta)a_{i,j}$$
 où $a_{i,j}$ correspond au maximum du signal acoustique discret $\{x_i\}$ contenu dans une trame formée par la sous-trame j de la trame i et par au moins une ou plusieurs sous-frames successives qui précèdent ladite sous-trame j ; et
 θ est un coefficient prédéfini compris entre 0 et 1 avec $\theta < \lambda$;
 - on établit la valeur de chaque sous-seuil $\Omega_{i,j}$ en fonction de la comparaison entre ladite valeur maximale de seuil $Lastmax$ et des valeurs maximales moyennes $A_{i,j}$ et $A_{i,j-1}$ considérées sur deux sous-frames j et $j-1$ successives.
20. Procédé selon la revendication 19, dans lequel, lors de la procédure de décision, on réactualise la valeur maximale de seuil $Lastmax$ à chaque fois

que le procédé a considéré qu'une sous-trame p d'une trame k contient un signal de parole, en mettant en œuvre la procédure suivante :

- la détection d'un signal de parole dans la sous-trame p de la trame k fait suite à une période d'absence de parole, et dans ce cas Lastmax prend la
5 valeur actualisée $[\alpha (A_{k,p} + \text{LastMax})]$, où α est un coefficient prédéfini compris entre 0 et 1, et par exemple compris entre 0,2 et 0,7 ;
- la détection d'un signal de parole dans la sous-trame p de la trame k fait suite à une période de présence de parole, et dans ce cas Lastmax prend la valeur actualisée $A_{k,p}$ si $A_{k,p} > \text{Lastmax}$.

10

21. Procédé selon les revendications 19 ou 20, dans lequel, on établit la valeur du seuil Ω_i en fonction de ladite valeur maximale Lastmax en se basant sur la comparaison entre :

- la valeur maximale de seuil Lastmax ; et
- 15 - les valeurs $[Kp.A_{i,j}]$ et $[Kp.A_{i,j-1}]$, où Kp est un coefficient fixe de pondération compris entre 1 et 2.

22. Procédé selon l'une quelconque des revendications 4 à 21, comportant en outre une phase dite de blocage comprenant une étape de
20 basculement d'un état de non détection d'un signal de parole à un état de détection d'un signal de parole après avoir détecté la présence d'un signal de parole sur N_p trames i temporelles successives.

23. Procédé selon l'une quelconque des revendications 4 à 22, comportant en outre une phase dite de blocage comprenant une étape de
25 basculement d'un état de détection d'un signal de parole à un état de non détection d'un signal de parole après avoir détecté aucune présence d'un signal voisé sur N_A trames i temporelles successives.

30 24. Procédé selon l'une quelconque des revendications 22 et 23, comportant en outre une étape d'interruption de la phase de blocage dans des zones de décision intervenant en fin de mots et dans une situation non-bruitée, lesdites zones de décision étant détectées en analysant le minimum $rr(i)$ de la fonction de détection discrète $FDi(\tau)$.

35

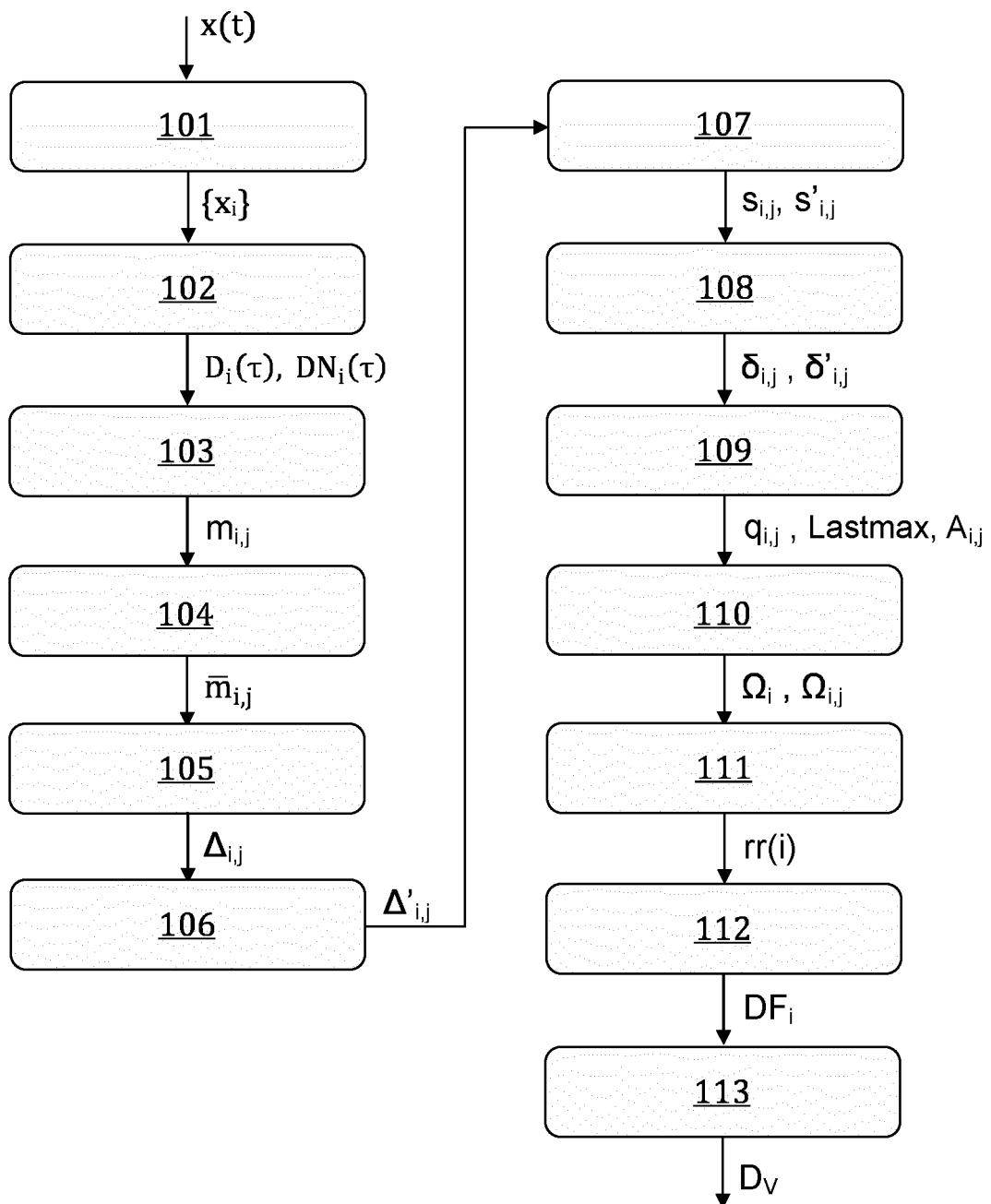
25. Programme d'ordinateur, caractérisé en ce qu'il comprend des instructions de code aptes à commander l'exécution des étapes du procédé de détection de la voix selon l'une quelconque des revendications précédentes lorsqu'il est exécuté par un processeur.

5

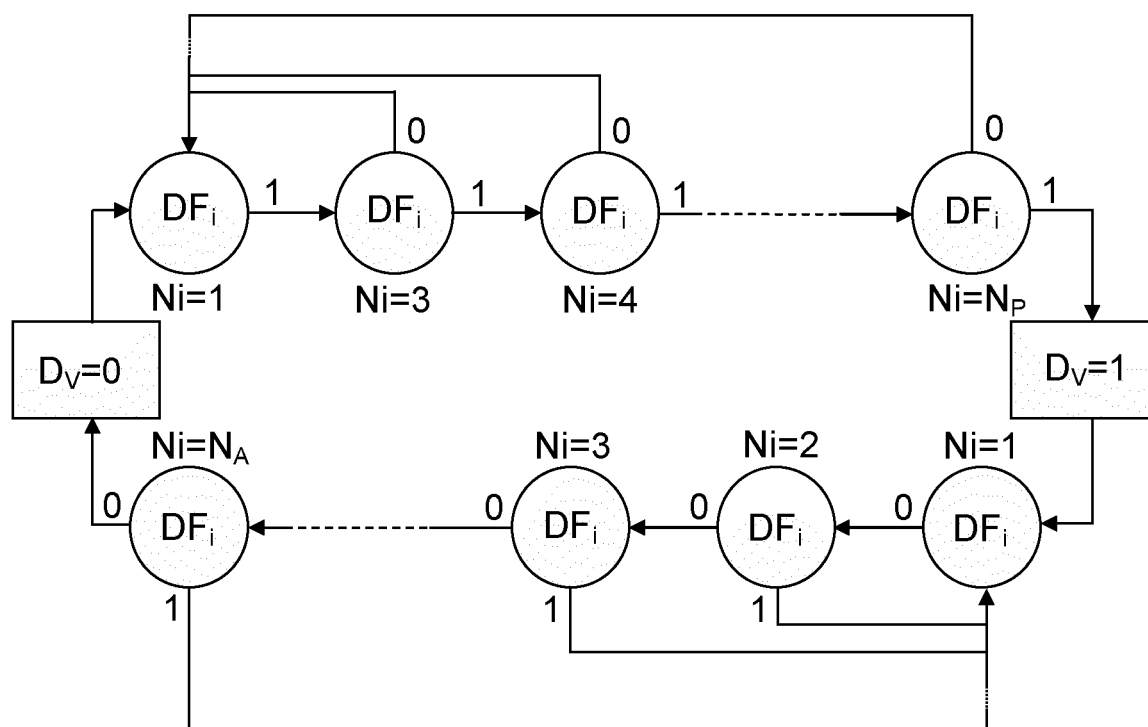
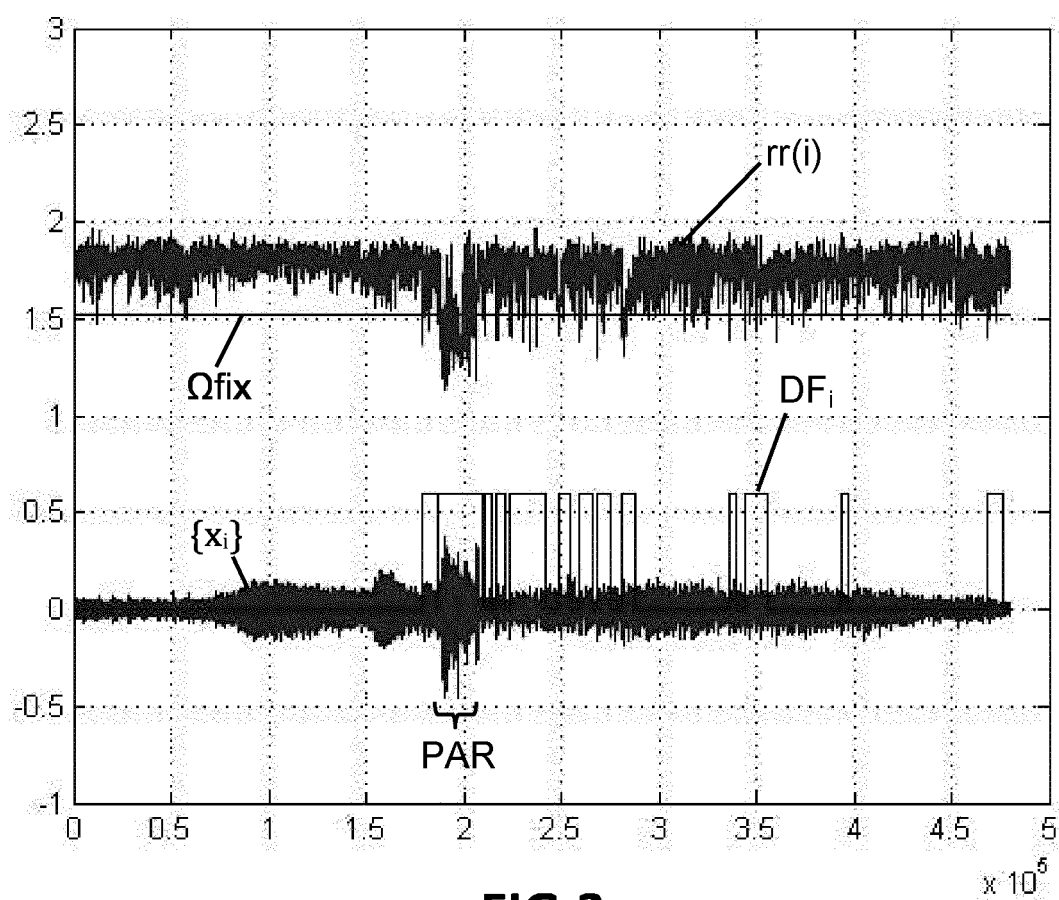
26. Support d'enregistrement de données sur lequel est stocké le programme d'ordinateur selon la revendication 25.

27. Mise à disposition d'un programme selon la revendication 25 sur un
10 réseau de télécommunication en vue de son téléchargement.

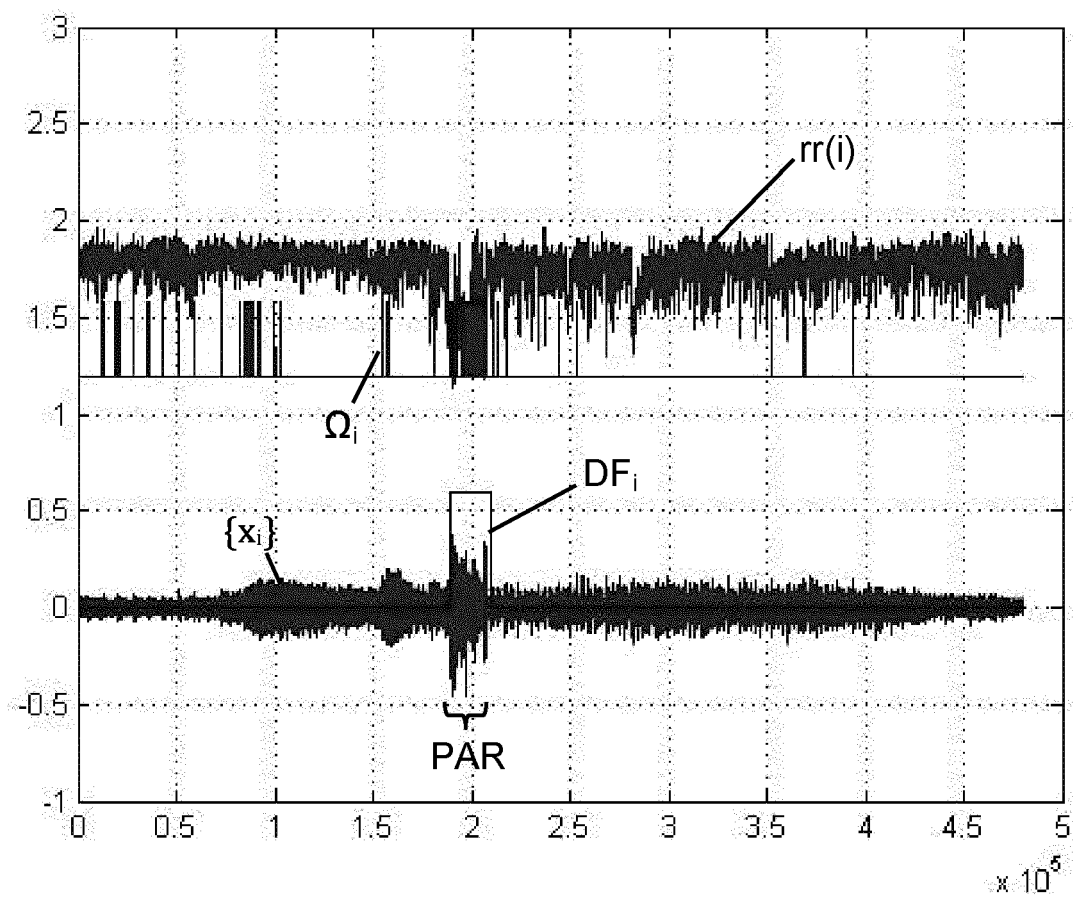
1/3

**FIG.1**

2/3

**FIG.2****FIG.3**

3/3

**FIG.4**



RAPPORT DE RECHERCHE PRÉLIMINAIRE

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

N° d'enregistrement
national

FA 791476
FR 1361922

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
Y	FR 2 988 894 A1 (ADEUNIS R F [FR]) 4 octobre 2013 (2013-10-04)	1-7, 22-27	G10L25/78
A	* page 1, ligne 4 - ligne 16 * * page 2, ligne 26 - page 5, ligne 23 * * page 6, ligne 14 - page 7, ligne 26 * * page 10, ligne 13 - page 11, ligne 10 * -----	8-21	
Y	HAE YOUNG KIM ET AL: "Pitch detection with average magnitude difference function using adaptive threshold algorithm for estimating shimmer and jitter", ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 1998. PROCEEDINGS OF THE 20TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, IEEE - PISCATAWAY, NJ, US, vol. 6, 29 octobre 1998 (1998-10-29), pages 3162-3164, XP010320717, ISBN: 978-0-7803-5164-6	1-7, 22-27	
A	* page 3162, alinéa II. - page 3163 * -----	8-21	
A	BERISHA V ET AL: "Real-Time Implementation of a Distributed Voice Activity Detector", SENSOR ARRAY AND MULTICHANNEL PROCESSING, 2006. FOURTH IEEE WORKSHOP ON, IEEE, PISCATAWAY, NJ, USA, 12 juillet 2006 (2006-07-12), pages 659-662, XP031331558, ISBN: 978-1-4244-0308-0 * page 660, alinéa II. - page 661; figure 2 *	1-27	DOMAINES TECHNIQUES RECHERCHÉS (IPC) G10L
A	US 2009/076814 A1 (LEE SUNG JOO [KR]) 19 mars 2009 (2009-03-19) * page 1, alinéa [0007] - alinéa [0011] * * page 5, alinéa [0054] - alinéa [0056]; figure 4 * -----	1-27	
Date d'achèvement de la recherche		Examineur	
8 juillet 2014		Virette, David	
CATÉGORIE DES DOCUMENTS CITÉS X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant			

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 1361922 FA 791476

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

Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **08-07-2014**

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

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
FR 2988894	A1	04-10-2013	AUCUN

US 2009076814	A1	19-03-2009	KR 20090030063 A 24-03-2009
		US 2009076814 A1	19-03-2009



RAPPORT DE RECHERCHE PRÉLIMINAIRE

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

N° d'enregistrement
national

FA 791476
FR 1361922

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
Y	FR 2 988 894 A1 (ADEUNIS R F [FR]) 4 octobre 2013 (2013-10-04)	1-7, 22-27	G10L25/78
A	* page 1, ligne 4 - ligne 16 * * page 2, ligne 26 - page 5, ligne 23 * * page 6, ligne 14 - page 7, ligne 26 * * page 10, ligne 13 - page 11, ligne 10 * -----	8-21	
Y	HAE YOUNG KIM ET AL: "Pitch detection with average magnitude difference function using adaptive threshold algorithm for estimating shimmer and jitter", ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 1998. PROCEEDINGS OF THE 20TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, IEEE - PISCATAWAY, NJ, US, vol. 6, 29 octobre 1998 (1998-10-29), pages 3162-3164, XP010320717, ISBN: 978-0-7803-5164-6	1-7, 22-27	
A	* page 3162, alinéa II. - page 3163 * -----	8-21	
A	BERISHA V ET AL: "Real-Time Implementation of a Distributed Voice Activity Detector", SENSOR ARRAY AND MULTICHANNEL PROCESSING, 2006. FOURTH IEEE WORKSHOP ON, IEEE, PISCATAWAY, NJ, USA, 12 juillet 2006 (2006-07-12), pages 659-662, XP031331558, ISBN: 978-1-4244-0308-0 * page 660, alinéa II. - page 661; figure 2 *	1-27	DOMAINES TECHNIQUES RECHERCHÉS (IPC) G10L
A	US 2009/076814 A1 (LEE SUNG JOO [KR]) 19 mars 2009 (2009-03-19) * page 1, alinéa [0007] - alinéa [0011] * * page 5, alinéa [0054] - alinéa [0056]; figure 4 * -----	1-27	
Date d'achèvement de la recherche		Examineur	
8 juillet 2014		Virette, David	
CATÉGORIE DES DOCUMENTS CITÉS			
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 1361922 FA 791476

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

Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **08-07-2014**

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

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
FR 2988894	A1	04-10-2013	AUCUN	

US 2009076814	A1	19-03-2009	KR 20090030063 A	24-03-2009
			US 2009076814 A1	19-03-2009
