(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 684 604**

(51) Int. CI.:

G10L 25/84 (2013.01)
G10L 25/78 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **27.11.2014 PCT/FR2014/053065**

(87) Fecha y número de publicación internacional: **11.06.2015 WO15082807**

(96) Fecha de presentación y número de la solicitud europea: **27.11.2014 E 14814978 (4)**

(97) Fecha y número de publicación de la concesión europea: **23.05.2018 EP 3078027**

(54) Título: **Procedimiento de detección de la voz**

(30) Prioridad:

02.12.2013 FR 1361922

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
03.10.2018

(73) Titular/es:

ADEUNIS RF (100.0%)
283 rue Louis Néel Parc Technologique Pré Roux
38920 Crolles, FR

(72) Inventor/es:

MAOUCHE, KARIM

(74) Agente/Representante:

CURELL AGUILÁ, Mireia

DESCRIPCIÓN

Procedimiento de detección de la voz.

La presente invención se refiere a un procedimiento de detección de la voz que permite detectar la presencia de señales de habla en una señal acústica ruidosa procedente de un micrófono.

Se refiere, más particularmente, a un procedimiento de detección de la voz utilizado en un sistema de comunicación de audio inalámbrico, mono-sensor.

La invención se sitúa en el campo específico de la detección de actividad de la voz, denominado generalmente "VAD" por *Voice Activity Detection*, que consiste en detectar el habla, dicho de otra manera, señales de habla, en una señal acústica procedente de un micrófono.

La invención encuentra una aplicación privilegiada, aunque no limitativa, con un sistema de comunicación de audio inalámbrico multi-usuario, del tipo sistema de comunicación por multiplexado temporal o *full-duplex*, entre diversos terminales de comunicación autónomos, es decir sin conexión a una base de transmisión o a una red, y de utilización sencilla, es decir, sin intervención de un técnico para establecer la comunicación.

Un sistema de comunicación de este tipo, conocido en particular a partir de los documentos WO 10149864 A1, WO 10149875 A1 y EP 1 843 326 A1, se utiliza típicamente en un entorno ruidoso, incluso muy ruidoso, por ejemplo, en el medio marino, dentro del marco de un espectáculo o de un acontecimiento deportivo en interiores o en exteriores, en una obra, etc.

En general, la detección de actividad de la voz consiste en delimitar por medio de criterios cuantificables, los principios y finales de palabras y/o de frases en una señal acústica ruidosa, dicho de otra manera, en un flujo de audio dado. Una detección de este tipo encuentra aplicaciones en campos tales como la codificación del habla, la reducción de ruido o, incluso, el reconocimiento del habla.

La realización de un procedimiento de detección de la voz en la cadena de tratamiento de un sistema de comunicación de audio permite, en particular, no transmitir ninguna señal acústica o de audio durante los periodos de silencio. Por ello, durante estos periodos no se transmitirá el ruido circundante, con el fin de mejorar la reproducción de audio de la comunicación o para reducir el caudal de transmisión. Por ejemplo, en el marco de la codificación del habla, se conoce la utilización de la detección de actividad de la voz para codificar la señal de audio de manera completa solamente cuando el procedimiento "VAD" indica actividad. Por ello, cuando no se produce habla y se está en un periodo de silencio, el caudal de codificación baja significativamente, lo cual, por término medio, en toda la señal, permite lograr unos caudales más reducidos.

Existen, por tanto, muchos procedimientos de detección de actividad de la voz pero estos últimos presentan unos rendimientos mediocres o no funcionan en absoluto en el marco de un entorno ruidoso, incluso muy ruidoso, tal como en un entorno de un encuentro deportivo (en exteriores o en interiores) con árbitros que deben comunicarse por audio y de forma inalámbrica. En efecto, los procedimientos conocidos de detección de actividad de la voz proporcionan resultados erróneos cuando la señal de habla está contaminada con ruido.

Entre los procedimientos conocidos de detección de actividad de la voz, algunos ponen en práctica una detección de la frecuencia fundamental característica de una señal de habla, tal como se da a conocer en particular en el documento FR 2 988 894. En el caso de una señal de habla, denominada señal o sonido sonoro, la señal presenta, en efecto, una frecuencia denominada fundamental, llamada de manera general "pitch", que corresponde a la frecuencia de vibración de las cuerdas vocales de la persona que habla, y que se extiende generalmente entre 70 y 400 Hertz. La evolución de esta frecuencia fundamental determina la melodía del habla y su rango depende del hablante, de sus hábitos aunque, también, de su estado físico y mental.

Así, para lograr la detección de una señal de habla, se sabe que se parte del principio por el cual una señal de habla del tipo mencionado es cuasi periódica y que, por ello, una correlación o una diferencia con la propia señal, aunque desplazada, presentará máximos o mínimos en las proximidades de la frecuencia fundamental y de sus múltiplos.

El documento "YIN, a fundamental frequency estimator for speech and music", de Alain De Cheveigne y Hideki Kawahara, *Journal of the Acoustical Society of America*, vol. 111, n.º 4, págs. 1917 a 1930, abril de 2002, propone y desarrolla un método basado en la diferencia entre la señal y la misma señal desplazada temporalmente.

Diversos métodos descritos a continuación se basan en la detección de la frecuencia fundamental de la señal de habla o *pitch* dentro de una señal acústica x(t) ruidosa.

Un primer método de detección de la frecuencia fundamental utiliza la búsqueda del máximo de la función de autocorrelación $R(\tau)$ definida por la siguiente relación:

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

Sin embargo, este primer método, al utilizar la función de autocorrelación, no ofrece un resultado satisfactorio en cuanto hay presencia de ruido relativamente importante. Además, la función de autocorrelación padece la presencia de máximos que no corresponden a la frecuencia fundamental o con sus múltiplos, sino a submúltiplos de la misma.

Un segundo método de detección de la frecuencia fundamental utiliza la búsqueda del mínimo de la función diferencia $D(\tau)$ definida por la siguiente relación:

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

donde | | es el operador valor absoluto, siendo mínima esta función diferencia en las proximidades de la frecuencia fundamental y de sus múltiplos, y a continuación la comparación de este mínimo con un umbral para deducir la decisión de presencia o no de voz.

Con respecto a la función de autocorrelación $R(\tau)$, la función diferencia $D(\tau)$ tiene la ventaja de ofrecer una carga de cálculo más reducida, consiguiendo así que este segundo método sea más interesante para aplicaciones en tiempo real. No obstante, este segundo método tampoco es completamente satisfactorio en cuanto hay presencia de ruido.

Un tercer método de detección de la frecuencia fundamental utiliza el cálculo, considerando una ventana de tratamiento de longitud H en la que H<N, de la función diferencia al cuadrado $d_t(\tau)$ definida por la relación:

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

A continuación, se prosigue con la búsqueda del mínimo de la función diferencia al cuadrado $d_t(\tau)$, siendo mínima esta función diferencia al cuadrado en las proximidades de la frecuencia fundamental y de sus múltiplos, y, finalmente, la comparación de este mínimo con un umbral para deducir la decisión de presencia o no de voz.

Una mejora conocida de este tercer método consiste en normalizar la función diferencia al cuadrado $d_t(\tau)$ calculando una función diferencia al cuadrado normalizada $d'_t(\tau)$ que responde a la siguiente relación:

$$d'_t(\tau) = \begin{cases} 1, & si\ \tau = 0 \\ si\ no\ \dfrac{d_t(\tau)}{\left(\frac{1}{\tau}\right)\sum_{j=1}^{\tau} d_t(j)} & \end{cases}$$

Este tercer método, aunque presenta una mejor inmunidad al ruido y ofrece, en este escenario, mejores resultados de detección, presenta unos límites en términos de detección de voz, en particular dentro de las zonas de ruido con características de RSB (Relación Señal/Ruido) reducida de un entorno ruidoso.

El estado de la técnica también se puede ilustrar con las enseñanzas de la solicitud de patente FR 2 825 505, que utiliza el tercer método de detección de la frecuencia fundamental citado previamente, para la extracción de esta frecuencia fundamental. En esta solicitud de patente, la función diferencia al cuadrado normalizada $d'_t(\tau)$ se puede comparar con un umbral para determinar esta frecuencia fundamental -pudiendo este umbral ser fijo o pudiendo variar en función del desplazamiento temporal τ- y este método adolece de los inconvenientes antes citados, asociados a este tercer método.

Se conoce también la utilización de un procedimiento de detección de la voz que utiliza la detección de una frecuencia fundamental, a partir del documento “Pitch detection with average magnitude difference function using adaptive threshold algorithm for estimating shimmer and jitter”, de Hae Young Kim *et al.*, *Engineering in Medicine And Biology Society*, 1998, *Proceedings of the 20th Annual International Conference of the IEEE*, vol. 6, 29 de octubre de 1998, páginas 3162 a 6164, XP010320717. En este documento se describe un procedimiento que consiste en buscar el mínimo de una función de autocorrelación, utilizando una comparación con un umbral adaptativo que es función de valores mínimos y máximos de la señal en la trama en curso. Esta adaptación del

umbral es sin embargo muy limitada. En efecto, en una situación de una señal de audio con diferentes valores de la relación señal/ruido pero con la misma amplitud de señal, el umbral sería el mismo para todas las situaciones sin que este último cambie en función del nivel de ruido, lo cual, de este modo, puede provocar cortes en el principio de la frase, incluso no detecciones de la voz, cuando la señal a detectar es una voz, en particular en un contexto en el que el ruido es un ruido de espectadores difuso de tal manera que no se asemeja en absoluto a una señal de habla.

La presente invención tiene como objetivo proponer un procedimiento de detección de la voz que ofrece una detección de las señales de habla contenidas en una señal acústica ruidosa, en particular en entornos ruidosos, incluso muy ruidoso.

Propone, más particularmente, un procedimiento de detección de la voz que está muy adaptado para la comunicación (en particular entre árbitros) en el interior de un estadio en donde el ruido es de nivel relativamente muy alto y es considerablemente no estacionario, con etapas de detección que evitan en particular las detecciones erróneas o falsas (denominadas, en general, “tonches”) debidas a los cánticos de los espectadores, instrumentos de viento, tambores, músicas y silbidos.

Con este fin, propone un procedimiento de detección de la voz que permite detectar la presencia de señales de habla en una señal acústica x(t) ruidosa, procedente de un micrófono, y que comprende las etapas sucesivas siguientes:

- una etapa previa de muestreo que comprende una segmentación de la señal acústica x(t) en una señal acústica discreta $\{x_i\}$ compuesta por una secuencia de vectores asociados a tramas i temporales de longitud N, correspondiéndose N con el número de puntos de muestreo, en donde cada vector refleja el contenido acústico de la trama i asociada y está compuesto por N muestras $x_{(i-1)N+1}$, $x_{(i-1)N+2}$, …, x_{iN-1}, x_{iN}, siendo i un entero positivo;

- una etapa de cálculo de una función de detección $FD(\tau)$ basada en el cálculo de una función diferencia $D(\tau)$ que varía en función del desplazamiento τ sobre una ventana de integración de longitud W que comienza en el tiempo t0, con:

$$D(\tau) = \sum_{n=t0}^{t0+W-1} |x(n) - x(n+\tau)| \quad \text{en donde } 0 \leq \tau \leq \max(\tau);$$

  en donde esta etapa de cálculo de la función de detección $FD(\tau)$ consiste en un cálculo de una función de detección discreta $FDi(\tau)$ asociada a las tramas i;

- una etapa de adaptación del umbral dentro de dicho intervalo en curso, en función de valores calculados a partir de la señal acústica x(t) establecidos en dicho intervalo en curso, y en particular valores máximos de dicha señal acústica x(t),

  en donde esta etapa de adaptación del umbral consiste en, para cada trama i, adaptar un umbral Ω_i propio de la trama i en función de valores de referencia calculados a partir de los valores de las muestras de la señal acústica discreta $\{x_i\}$ en dicha trama i;

- una etapa de búsqueda del mínimo de la función de detección $FD(\tau)$ y comparación de este mínimo con un umbral, variando τ dentro de un intervalo de tiempo determinado, que se denomina intervalo en curso, para detectar la presencia o no de una frecuencia fundamental F_0 característica de una señal de habla en dicho intervalo en curso;

  en donde esta etapa de búsqueda del mínimo de la función de detección $FD(\tau)$ y la comparación de este mínimo con un umbral se realizan buscando, en cada trama i, el mínimo rr(i) de la función de detección discreta $FDi(\tau)$ y comparando este mínimo rr(i) con un umbral Ω_i propio de la trama i;

y, en el que, la etapa de adaptación de los umbrales Ω_i para cada trama i comprende las siguientes etapas:

a)- la trama i que comprende N puntos de muestreo se subdivide en T subtramas de longitud L, donde N es un múltiplo de T con el fin de que la longitud L = N/T sea un entero, y de manera que las muestras de la señal acústica discreta $\{x_i\}$ dentro de una subtrama de índice j de la trama i comprendan las siguientes L muestras:

  $x_{(i-1)N+(j-1)L+1}$, $x_{(i-1)N+(j-1)L+2}$, …, $x_{(i-1)N+jL}$, siendo j un entero positivo comprendido entre 1 y T;

b)- se calculan los valores máximos $m_{i,j}$ de la señal acústica discreta $\{x_i\}$ dentro de cada subtrama de índice j de la trama i, con:

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

c)- se calcula por lo menos un valor de referencia $Ref_{i,j}$, $MRef_{i,j}$ propio de la subtrama j de la trama i, calculándose el valor o cada valor de referencia $Ref_{i,j}$, $MRef_{i,j}$, por cada subtrama j, a partir del valor máximo $m_{i,j}$ dentro de la subtrama j de la trama i;

d)- se establece el valor del umbral Ω_i propio de la trama i en función de todos los valores de referencia $Ref_{i,j}$, $MRef_{i,j}$ calculados en las subtramas j de la trama i.

Así, este procedimiento se basa en el principio de un umbral adaptativo, el cual será relativamente bajo durante los periodos de ruido o de silencio y relativamente alto durante los periodos de habla. De este modo, las detecciones falsas se minimizarán y el habla se detectará correctamente con un mínimo de cortes en el principio y el final de las palabras. Con el procedimiento según la invención, para tomar la decisión (voz o ausencia de voz) sobre la trama i completa se consideran los valores máximos $m_{i,j}$ establecidos dentro de las subtramas j.

Según una primera posibilidad, la función de detección $FD(\tau)$ corresponde a la función diferencia $D(\tau)$.

De acuerdo con una segunda posibilidad, la función de detección $FD(\tau)$ corresponde a la función diferencia normalizada $DN(\tau)$ calculada a partir de la función diferencia $D(\tau)$ de la manera siguiente:

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

en donde el cálculo de la función diferencia normalizada $DN(\tau)$ consiste en un cálculo de una función diferencia normalizada discreta $DN_i(\tau)$ asociada a las tramas i, en donde:

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

En una forma de realización particular, la función diferencia discreta $D_i(\tau)$ relativa a la trama i se calcula de la manera siguiente:

- la trama i se subdivide en K subtramas de longitud H, con, por ejemplo, $K = \left\lfloor \frac{N-\max(\tau)}{H} \right\rfloor$, en donde $\lfloor\ \rfloor$ representa el operador de redondeo a la parte entera, de manera que las muestras de la señal acústica discreta $\{x_i\}$ dentro de una subtrama de índice p de la trama i comprenden las H muestras:

  $x_{(i-1)N+(p-1)H+1}, x_{(i-1)N+(p-1)H+2}, \ldots, x_{(i-1)N+pH}$, siendo p un entero positivo comprendido entre 1 y K;

- para cada subtrama de índice p, se calcula la función diferencia $dd_p(\tau)$ siguiente:

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

- se calcula la función diferencia discreta $D_i(\tau)$ relativa a la trama i como la suma de las funciones diferencia $dd_p(\tau)$ de las subtramas de índice p de la trama i, es decir:

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

Además, el procedimiento según la invención destaca por que en la etapa c), se realizan las siguientes subetapas sobre cada trama i:

c1)- se calculan las envolventes suavizadas de los máximos $\overline{m}_{i,j}$ en cada subtrama de índice j de la trama i, con:

$\overline{m}_{i,j} = \lambda\, \overline{m}_{i,j-1} + (1-\lambda) m_{i,j}$, donde λ es un coeficiente predefinido comprendido entre 0 y 1;

c2)- se calculan las señales de variación $\Delta_{i,j}$ en cada subtrama de índice j de la trama i, con:

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

y en donde por lo menos un valor de referencia denominado principal $Ref_{i,j}$ por cada subtrama j se calcula a partir de la señal de variación $\Delta_{i,j}$ en la subtrama j de la trama i.

Así, para tomar la decisión (voz o ausencia de voz) sobre la trama i completa, se consideran las señales de variación $\Delta_{i,j}$ de las envolventes suavizadas establecidas en las subtramas j, fiabilizando la detección del habla (o voz).

Según otra característica, en la etapa c) y a continuación de la subetapa c2), se realizan las siguientes subetapas sobre cada trama i:

c3)- se calculan los máximos de variación $s_{i,j}$ en cada subtrama de índice j de la trama i, en donde $s_{i,j}$ corresponde al máximo de la señal de variación $\Delta_{i,j}$ calculado sobre una ventana deslizante de longitud Lm anterior a dicha subtrama j, siendo variable dicha longitud Lm según que la subtrama j de la trama i corresponda a un periodo de silencio o de presencia de habla.

c4)- se calculan las desviaciones de variación $\delta_{i,j}$ en cada subtrama de índice j de la trama i, con:

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

y en donde, para cada subtrama j de la trama i, se calculan dos valores de referencia principales $Ref_{i,j}$ a partir, respectivamente, de la señal de variación $\Delta_{i,j}$ y de la desviación de variación $\delta_{i,j}$.

Así, se consideran conjuntamente las señales de variación $\Delta_{i,j}$ y las desviaciones de variación $\delta_{i,j}$ establecidas en las subtramas j para elegir el valor del umbral Ω_i adaptativo y, así, tomar la decisión (voz o ausencia de voz) sobre la trama i completa, reforzando la detección del habla. Dicho de otra manera, se estudia el par $(\Delta_{i,j}, \delta_{i,j})$ para determinar el valor del umbral Ω_i adaptativo.

Ventajosamente, en la etapa c) y a continuación de la subetapa c4), se realiza una subetapa c5) de cálculo de las señales de variación normalizadas $\Delta'_{i,j}$ y de las desviaciones de variación normalizadas $\delta'_{i,j}$ en cada subtrama de índice j de la trama i, de la manera siguiente:

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

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

y en donde, para cada subtrama j de una trama i, la señal de variación normalizada $\Delta'_{i,j}$ y la desviación de variación normalizada $\delta'_{i,j}$ constituyen, cada una de ellas, un valor de referencia principal $Ref_{i,j}$ de manera que, en la etapa d), se establece el valor del umbral Ω_i propio de la trama i en función del par $(\Delta'_{i,j}, \delta'_{i,j})$ de las señales de variación normalizadas $\Delta'_{i,j}$ y de las desviaciones de variación normalizadas $\delta'_{i,j}$ en las subtramas j de la trama i.

De esta manera, la variación del umbral Ω_i se puede tratar independientemente de los niveles de las señales $\Delta_{i,j}$ y $\delta_{i,j}$ normalizándolas con el cálculo de las señales normalizadas $\Delta'_{i,j}$ y $\delta'_{i,j}$. Así, los umbrales Ω_i elegidos a partir de estas señales normalizadas $\Delta'_{i,j}$ y $\delta'_{i,j}$ serán independientes del nivel de la señal acústica discreta $\{x_i\}$. Dicho de otra manera, para determinar el valor del umbral Ω_i adaptativo se estudia el par $(\Delta'_{i,j}, \delta'_{i,j})$.

De forma ventajosa, en la etapa d), el valor del umbral Ω_i propio de la trama i se establece dividiendo el espacio definido por el valor del par $(\Delta'_{i,j}, \delta'_{i,j})$, y examinando el valor del par $(\Delta'_{i,j}, \delta'_{i,j})$ sobre una o varias (por ejemplo, entre una y tres) subtramas sucesivas según la zona de valor del par $(\Delta'_{i,j}, \delta'_{i,j})$.

Así, el proceso de cálculo del umbral Ω_i se basa en una partición experimental del espacio definido por el valor del par $(\Delta'_{i,j}, \delta'_{i,j})$. A ello se le añade un mecanismo de decisión que escudriña el valor del par $(\Delta'_{i,j}, \delta'_{i,j})$ sobre una, dos o más subtramas sucesivas según la zona de valor del par. Las condiciones de las pruebas de posicionamiento del valor del par $(\Delta'_{i,j}, \delta'_{i,j})$ dependen en particular de la detección del habla durante la trama precedente, y el mecanismo de escudriñamiento sobre dichas una, dos o más subtramas sucesivas utiliza también una división en particiones experimental.

Según una característica, en la subetapa c3), la longitud Lm de la ventana deslizante responde a las siguientes ecuaciones:

- Lm = L0 si la subtrama j de la trama i corresponde a un periodo de silencio;
- Lm = L1 si la subtrama j de la trama i corresponde a un periodo de presencia de habla;

con L1 < L0, y en particular con L1=k1·L y L0=k0·L, siendo L la longitud de las subtramas de índice j y siendo k0, k1 enteros positivos.

Según otra característica, en la subetapa c3), para cada cálculo del máximo de variación $s_{i,j}$ en la subtrama j de la trama i, la ventana deslizante de longitud Lm está retardada Mm tramas de longitud N con respecto a dicha subtrama j.

Según otra característica, se realizan los siguientes perfeccionamientos:

- en la subetapa c3), se calculan también los máximos de variación normalizados $s'_{i,j}$ en cada subtrama de índice j de la trama i, en donde $s'_{i,j}$ corresponde al máximo de la señal de variación normalizado $\Delta'_{i,j}$ calculado sobre una ventana deslizante de longitud Lm anterior a dicha subtrama j, en donde:

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

y en donde cada máximo de variación normalizado $s'_{i,j}$ se calcula según un método de minimización que comprende las siguientes etapas iterativas:

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

- si rem(i, Lm) = 0, donde rem es el operador resto de la división entera de dos enteros, entonces:

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

con $s'_{0,1} = 0$ y $\tilde{s}'_{0,1} = 0$; y

- en la etapa c4), se calculan las desviaciones de variación normalizadas $\delta'_{i,j}$ en cada subtrama de índice j de la trama i, de la manera siguiente:

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

De manera ventajosa, en la etapa c), se realiza una subetapa c6) en la que se calculan los máximos del $q_{i,j}$ máximo en cada subtrama de índice j de la trama i, en donde $q_{i,j}$ corresponde al máximo del valor máximo $m_{i,j}$ calculado sobre una ventana deslizante de longitud fija Lq anterior a dicha subtrama j, en donde la ventana deslizante de longitud Lq está retardada Mq tramas de longitud N con respecto a dicha subtrama j, y en donde otro valor de referencia denominado secundario $MRef_{i,j}$ por cada subtrama j corresponde a dicho máximo del $q_{i,j}$ máximo dentro de la subtrama j de la trama i.

Así, para evitar provechosamente las detecciones falsas, resulta ventajoso tener en cuenta también esta señal $q_{i,j}$ (valor de referencia secundario $MRef_{i,j} = q_{i,j}$) que se calcula de una manera similar al cálculo de la señal $s_{i,j}$ citada previamente, pero que actúa sobre los valores máximos $m_{i,j}$ en lugar de actuar sobre las señales de variación $\Delta_{i,j}$ o sobre las señales de variación normalizadas $\Delta'_{i,j}$.

En un modo de realización particular, en la etapa d), el umbral Ω_i propio de la trama i se segmenta en varios subumbrales $\Omega_{i,j}$ propios de cada subtrama j de la trama i, y el valor de cada subumbral $\Omega_{i,j}$ se establece por lo menos en función del valor o valores de referencia $Ref_{i,j}$, $MRef_{i,j}$ calculados en la subtrama j de la trama i correspondiente.

Así, se tiene $\Omega_i = \{\Omega_{i,1}\ ;\ \Omega_{i,2}\ ;\ ...;\ \Omega_{i,T}\}$, que refleja la segmentación del umbral Ω_i en varios subumbrales $\Omega_{i,j}$ propios de las subtramas j, aportando una resolución suplementaria en el establecimiento del umbral Ω_i adaptativo.

De manera ventajosa, en la etapa d), se establece el valor de cada umbral $\Omega_{i,j}$ propio de la subtrama j de la trama i comparando los valores del par ($\Delta'_{i,j}$, $\delta'_{i,j}$) con varios pares de umbrales fijos, seleccionándose el valor de cada umbral $\Omega_{i,j}$ entre varios valores fijos en función de las comparaciones del par ($\Delta'_{i,j}$, $\delta'_{i,j}$) con dichos pares de umbrales fijos.

Estos pares de umbrales fijos se determinan, por ejemplo, experimentalmente mediante una repartición del espacio de los valores ($\Delta'_{i,j}$, $\delta'_{i,j}$) en zonas de decisiones.

De manera complementaria, se establece el valor de cada umbral $\Omega_{i,j}$ propio de la subtrama j de la trama i también llevando a cabo una comparación del par ($\Delta'_{i,j}$, $\delta'_{i,j}$) en una o varias subtramas sucesivas según la zona inicial del par ($\Delta'_{i,j}$, $\delta'_{i,j}$).

Las condiciones de las pruebas de posicionamiento del valor del par ($\Delta'_{i,j}$, $\delta'_{i,j}$) dependen de la detección del habla durante la trama precedente, y el mecanismo de comparación en la subtrama o subtramas sucesivas utiliza también una división en particiones experimental.

Evidentemente, también es previsible establecer el valor de cada umbral $\Omega_{i,j}$ propio de la subtrama j de la trama i comparando:

- los valores del par ($\Delta'_{i,j}$, $\delta'_{i,j}$) (los valores de referencia principales $Ref_{i,j}$) con varios pares de umbrales fijos;
- los valores de $q_{i,j}$ (el valor de referencia secundario $MRef_{i,j}$) con otros diversos umbrales fijos.

Así, el mecanismo de decisión basado en la comparación del par ($\Delta'_{i,j}$, $\delta'_{i,j}$) con pares de umbrales fijos, se completa mediante otro mecanismo de decisión basado en la comparación de $q_{i,j}$ con otros umbrales fijos.

Ventajosamente, en la etapa d), se realiza un proceso denominado de decisión, que comprende las siguientes subetapas, para cada trama i:

- para cada subtrama j de la trama i, se establece un índice de decisión $DEC_i(j)$ que ocupa o bien un estado "1" de detección de una señal de habla o bien un estado "0" de no detección de una señal de habla;

- se establece una decisión temporal VAD(i) basada en la comparación de los índices de decisión $DEC_i(j)$ con operadores "O" lógicos, de manera que la decisión temporal VAD(i) ocupa un estado "1" de detección de una señal de habla si por lo menos uno de dichos índices de decisión $DEC_i(j)$ ocupa este estado "1" de detección de una señal de habla.

Así, para evitar las detecciones tardías (cortes de palabras en el principio de la detección), la decisión final (voz o ausencia de voz) se toma a continuación de este proceso de decisión basándose en la decisión temporal VAD(i) que, a su vez, se toma sobre la trama i completa, con la aplicación de un operador "O" lógico sobre las decisiones tomadas en las subtramas j, y, preferentemente, en subtramas j sucesivas con un horizonte corto y finito a partir del principio de la trama i.

Durante este proceso de decisión, se pueden realizar las siguientes subetapas, para cada trama i:

- se memoriza un valor máximo de umbral Lastmax que corresponde al valor variable de un umbral de comparación para la amplitud de la señal acústica discreta $\{x_i\}$ por debajo del cual se considera que la señal acústica no comprende ninguna señal de habla, determinándose este valor variable durante la última trama de índice k que precede a dicha trama i y en la que la decisión temporal VAD(k) ocupaba un estado "1" de detección de una señal de habla;

- se memoriza un valor máximo medio $A_{i,j}$ que corresponde al valor máximo medio de la señal acústica discreta $\{x_i\}$ en la subtrama j de la trama i, calculado de la manera siguiente:

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

  en donde $a_{i,j}$ corresponde al máximo de la señal acústica discreta $\{x_i\}$ contenido en una trama k formada por la subtrama j de la trama i y por lo menos por una o varias subtramas sucesivas que preceden a dicha subtrama j; y

  es un coeficiente predefinido comprendido entre 0 y 1, con $\theta < \lambda$

- se establece el valor de cada subumbral $\Omega_{i,j}$ en función de la comparación entre dicho valor máximo de umbral Lastmax y valores máximos medios $A_{i,j}$ y $A_{i,j-1}$ considerados sobre dos subtramas j y j-1 sucesivas.

En muchos casos, las falsas detecciones llegan con una amplitud inferior a la de la señal de habla (al estar situado el micrófono al lado de la boca de la persona que se está comunicando). Así, este proceso de decisión pretende eliminar todavía más las detecciones erróneas memorizando el valor máximo de umbral Lastmax de la señal de habla actualizado de nuevo en el último periodo de activación y los valores máximos medios $A_{i,j}$ y $A_{i,j-1}$ que corresponden al valor máximo medio de la señal acústica discreta $\{x_i\}$ en las subtramas j y j-1 de la trama i.

Teniendo en cuenta estos valores (Lastmax, $A_{i,j}$, y $A_{i,j-1}$), se vuelve a añadir una condición en el nivel del establecimiento del umbral Ω_i adaptativo.

Es importante que el valor de θ se seleccione de manera que sea inferior al coeficiente λ para ralentizar las fluctuaciones de $A_{i,j}$.

En el proceso de decisión mencionado anteriormente, se actualiza de nuevo el valor máximo de umbral Lastmax cada vez que el procedimiento ha considerado que una subtrama p de una trama k contiene una señal de habla, poniendo en práctica el proceso siguiente:

- la detección de una señal de habla en la subtrama p de la trama k sucede a un periodo de ausencia de habla, y, en este caso, Lastmax adopta el valor actualizado [α ($A_{k,p}$ + LastMax)], en donde α es un coeficiente predefinido, comprendido entre 0 y 1, y, por ejemplo, comprendido entre 0,2 y 0,7;

- la detección de una señal de habla en la subtrama p de la trama k sucede a un periodo de presencia de habla, y, en este caso, Lastmax adopta el valor actualizado $A_{k,p}$ si $A_{k,p}$ > Lastmax.

Así, la actualización del valor Lastmax se realiza únicamente durante los periodos de activación del procedimiento (dicho de otra manera, los periodos de detección de la voz). En una situación de detección de habla, el valor Lastmax valdrá $A_{k,p}$ > cuando se tenga $A_{k,p}$ > LastMax. Sin embargo, es importante que esta actualización se realice de la manera siguiente durante la activación de la primera subtrama p que sucede a una zona de silencio: el valor Lastmax valdrá [α ($A_{k,p}$ + LastMax)].

Este mecanismo de actualización del valor máximo de umbral Lastmax permite que el procedimiento detecte la voz del usuario incluso si este último ha reducido la intensidad de su voz (dicho de otra manera, habla menos fuerte) con respecto a la última vez en la que el procedimiento ha detectado que él había hablado.

Dicho de otra manera, para mejorar todavía más la eliminación de las falsas detecciones, se realiza un tratamiento sutil en el que el valor máximo de umbral Lastmax es variable y se compara con los valores máximos medios $A_{i,j}$ y $A_{i,j-1}$ de la señal acústica discreta.

Efectivamente, con el procedimiento se podrían captar voces lejanas, ya que dichas voces presentan frecuencias fundamentales susceptibles de ser detectadas, igual que la voz del usuario. Para garantizar que las voces lejanas, que pueden ser molestas en varios casos prácticos, no sean tenidas en cuenta por el procedimiento, se considera un tratamiento en el transcurso del cual el valor máximo medio de la señal (sobre dos tramas sucesivas), en este caso $A_{i,j}$ y $A_{i,j-1}$, se compara con Lastmax que constituye un umbral variable según la amplitud de la voz del usuario medida en la última activación. Así, el valor del umbral Ω_i se fija a un valor mínimo muy bajo, cuando la señal esté por debajo del umbral.

Esta condición para establecer el valor del umbral Ω_i en función del valor máximo de umbral Lastmax se basa ventajosamente en la comparación entre:

- el valor máximo de umbral Lastmax; y

- los valores [Kp·$A_{i,j}$] y [Kp· $A_{i,j-1}$], en donde Kp es un coeficiente fijo de ponderación comprendido entre 1 y 2.

De esta manera, el valor máximo de umbral Lastmax se compara con los valores máximos medios de la señal acústica discreta $\{x_i\}$ en las subtramas j y j-1 ($A_{i,j}$ y $A_{i,j-1}$) ponderados con un coeficiente de ponderación Kp comprendido entre 1 y 2, para reforzar la detección. Esta comparación se realiza únicamente cuando la trama precedente no ha dado lugar a una detección de voz.

De manera ventajosa, el procedimiento comprende además una fase denominada de bloqueo, que comprende una etapa de conmutación de un estado de no detección de una señal de habla a un estado de detección de una señal de habla después de haber detectado la presencia de una señal de habla sobre N_p tramas i temporales sucesivas.

Así, el procedimiento pone en práctica una etapa del tipo *hangover* configurada de tal manera que la transición de una situación sin voz a una situación con presencia de voz se realiza únicamente después de Np tramas sucesivas con presencia de voz.

Asimismo, el procedimiento consta además de una fase denominada de bloqueo que comprende una etapa de conmutación de un estado de detección de una señal de habla a un estado de no detección de una señal de habla después de no haber detectado ninguna presencia de una señal sonora sobre N_A tramas i temporales sucesivas.

Así, el procedimiento pone en práctica una etapa del tipo *hangover* configurada de tal manera que la transición de una situación con presencia de voz a una situación sin voz se realiza únicamente después de N_A tramas sucesivas sin voz.

Sin estas etapas de conmutación, el procedimiento corre el riesgo de cortar ocasionalmente la señal acústica durante las frases o incluso en mitad de las palabras pronunciadas. Para remediar esto, estas etapas de conmutación ponen en práctica una etapa de bloqueo o de *hangover* sobre una serie dada de tramas.

Según una posibilidad de la invención, el procedimiento comprende una etapa de interrupción de la fase de bloqueo en zonas de decisión que intervienen al final de palabras y en una situación sin ruido, detectándose dichas zonas de decisión al analizar el mínimo rr(i) de la función de detección discreta $FDi(\tau)$.

Así, la fase de bloqueo se interrumpe al final de una frase o palabra durante una detección particular en el espacio de decisión. Esta interrupción sobreviene únicamente en una situación ruidosa inexistente o reducida. Por ello, el procedimiento prevé el aislamiento de una zona de decisión particular que sobreviene únicamente al final de palabras y en una situación sin ruido. Para reforzar la decisión de detección de esta zona, el procedimiento utiliza también el mínimo rr(i) de la función de detección discreta $FDi(\tau)$, en donde la función de detección discreta $FDi(\tau)$ corresponde o bien a la función de diferencia discreta $D_i(\tau)$ o bien a la función de diferencia normalizada discreta $DN_i(\tau)$. Por ello, la voz se cortará más rápidamente al final del habla, confiriendo así al sistema una mejor calidad de audio.

La invención tiene también como objetivo un programa de ordenador que comprende instrucciones de código aptas para controlar la ejecución de las etapas del procedimiento de detección de la voz tal como se ha definido anteriormente cuando el mismo es ejecutado por un procesador.

La invención tiene también como objetivo un soporte de grabación de datos de grabación en el que se almacena un programa de ordenador según se ha definido anteriormente en la presente.

La invención tiene como objetivo adicional poner a disposición un programa de ordenador según se ha definido anteriormente en la presente en una red de telecomunicación con vistas a su descarga.

Otras características y ventajas de la presente invención se pondrán de manifiesto al leer la descripción detallada que se ofrece posteriormente, de un ejemplo de puesta en práctica, no limitativo, y realizada en referencia a las figuras adjuntas en las que:

- la figura 1 es un esquema sinóptico del procedimiento de acuerdo con la invención;
- la figura 2 es una vista esquemática de un bucle de limitación puesto en práctica por una etapa de bloqueo de decisión denominada etapa del tipo *hangover*;
- la figura 3 ilustra el resultado de un procedimiento de detección de la voz que utiliza un umbral fijo con, en la parte superior, una representación de la curva del mínimo rr(i) de la función de detección y de la línea de umbral fijo Ω_{fijo} y, en la parte inferior, una representación de la señal acústica discreta $\{x_i\}$ y de la señal de salida DF_i;
- la figura 4 ilustra el resultado de un procedimiento de detección de la voz de acuerdo con la invención utilizando un umbral adaptativo con, en la parte superior, una representación de la curva del mínimo rr(i) de la función de detección y de la línea de umbral adaptativo Ωi y, en la parte inferior, una representación de la señal acústica discreta $\{x_i\}$ y de la señal de salida DF_i.

La descripción del procedimiento de detección de la voz se realiza en referencia a la figura 1 que ilustra esquemáticamente la sucesión de las diferentes etapas necesarias para la detección de la presencia de señales de habla (o de voz) en una señal acústica ruidosa x(t) procedente de un micrófono único que está funcionando en un medio ruidoso.

El procedimiento comienza por una etapa 101 previa de muestreo que comprende una segmentación de la señal acústica x(t) en una señal acústica discreta $\{x_i\}$ compuesta por una secuencia de vectores asociados a tramas i temporales de longitud N, correspondiéndose N con el número de puntos de muestreo, en donde cada vector refleja el contenido acústico de la trama i asociada y está compuesto por N muestras $x_{(i-1)N+1}$, $x_{(i-1)N+2}$,…,x_{iN-1}, x_{iN}, siendo i un entero positivo:

A título de ejemplo, la señal acústica ruidosa x(t) se segmenta en tramas de 240 o 256 muestras, lo cual, a una frecuencia de muestreo F_e de 8 kHz, corresponde a unas tramas temporales de 30 o 32 milisegundos.

El procedimiento prosigue con una etapa 102 de cálculo de una función diferencia discreta $D_i(\tau)$ relativa a la trama i que se calcula de la manera siguiente:

- cada trama i se subdivide en K subtramas de longitud H, con la siguiente relación:

  $K = \left\lfloor \frac{N-\max(\tau)}{H} \right\rfloor$ en donde $\lfloor\ \rfloor$ representa el operador de redondeo a la parte entera,

  de manera que las muestras de la señal acústica discreta $\{x_i\}$ dentro de una subtrama de índice p de la trama i comprenden las H muestras siguientes:

  $x_{(i-1)N+(p-1)H+1}$, $x_{(i-1)N+(p-1)H+2}$, ..., $x_{(i-1)N+pH}$, siendo p un entero positivo comprendido entre 1 y K; a continuación

- para cada subtrama de índice p, se calcula la función diferencia $dd_p(\tau)$ siguiente:

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

- se calcula la función diferencia discreta $D_i(\tau)$ relativa a la trama i como la suma de las funciones diferencia $dd_p(\tau)$ de las subtramas de índice p de la trama i, es decir:

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

Es también posible que la etapa 102 comprenda asimismo el cálculo de una función diferencia normalizada discreta $DN_i(\tau)$ a partir de la función diferencia discreta $D_i(\tau)$, de la manera siguiente:

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

El procedimiento prosigue con una etapa 103 en la que, para cada trama i:

- la trama i que comprende N puntos de muestreo se subdivide en T subtramas de longitud L, donde N es un múltiplo de T con el fin de que la longitud L=N/T sea un entero, y de manera que las muestras de la señal acústica discreta $\{x_i\}$ dentro de una subtrama de índice j de la trama i comprendan las siguientes L muestras:

  $x_{(i-1)N+(j-1)L+1}$, $x_{(i-1)N+(j-1)L+2}$, ..., $x_{(i-1)N+jL}$, siendo j un entero positivo comprendido entre 1 y T;

b)- se calculan los valores máximos $m_{i,j}$ de la señal acústica discreta $\{x_i\}$ dentro de cada subtrama de índice j de la trama i, con:

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

A título de ejemplo, cada trama i de longitud 240 (es decir, N=240) se subdivide en cuatro subtramas j de longitudes 60 (es decir, T=4, y L=60).

A continuación, en una etapa 104, se calculan las envolventes suavizadas de los máximos $\overline{m}_{i,j}$ en cada subtrama de índice j de la trama i, definidos por:

$\overline{m}_{i,j} = \lambda \overline{m}_{i,j-1} + (1-\lambda) m_{i,j}$, donde λ es un coeficiente predefinido comprendido entre 0 y 1.

A continuación, en una etapa 105, se calculan las señales de variación $\Delta_{i,j}$ en cada subtrama de índice j de la trama i, definidas por:

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

A continuación, en una etapa 106, se calculan las señales de variación normalizadas $\Delta'_{i,j}$ definidas por:

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

A continuación, en una etapa 107, se calculan los máximos de variación $s_{i,j}$ en cada subtrama de índice j de la trama i, en donde $s_{i,j}$ corresponde al máximo de la señal de variación $\Delta_{i,j}$ calculada sobre una ventana deslizante de longitud Lm anterior a dicha subtrama j. Durante esta etapa 106, la longitud Lm es variable según que la subtrama j de la trama i corresponda a un periodo de silencio o de presencia de habla, con:

- Lm = L0 si la subtrama j de la trama i corresponde a un periodo de silencio;
- Lm = L1 si la subtrama j de la trama i corresponde a un periodo de presencia de habla;

con L1 < L0. A título de ejemplo, L1=k1·L y L0=k0·L, siendo L, a título recordatorio, la longitud de las subtramas de índice j y siendo k0, k1 enteros positivos con k1<k0. Además, la ventana deslizante de longitud Lm está retrasada Mm tramas de longitud N con respecto a dicha subtrama j.

Durante esta etapa 106, se calculan también los máximos de variación normalizados $s'_{i,j}$ en cada subtrama de índice j de la trama i, en donde:

$$s'_{i,j} = \frac{s_{i,j}}{\overline{m}_{i,j}}.$$

Se puede prever el cálculo de los máximos de variación normalizados $s'_{i,j}$ según un método de minimización que comprende las siguientes etapas iterativas:

- cálculo de $s'_{i,j} = \max\{s'_{i,j-1}; \Delta'_{i-M_m,j}\}$ y $\tilde{s}'_{i,j} = \max\{s'_{i,j-1}; \Delta'_{i-M_m,j}\}$
- si rem(i, Lm) = 0, donde rem es el operador resto de la división entera de dos enteros, entonces:

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

- fin del si

con $s'_{0,1} = 0$ y $\tilde{s}'_{0,1} = 0$.

A continuación, en una etapa 108, se calculan las desviaciones de variación $\delta_{i,j}$ en cada subtrama de índice j de la trama i, definidas por:

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

En esta misma etapa 108, se calculan las desviaciones de variación normalizadas $\delta'_{i,j}$ en cada subtrama de índice j de la trama i, definidas por:

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

A continuación, en una etapa 109, se calculan los máximos del $q_{i,j}$ máximo en cada subtrama de índice j de la trama i, en donde $q_{i,j}$ corresponde al máximo del valor máximo $m_{i,j}$ calculado sobre una ventana deslizante de longitud fija Lq anterior a dicha subtrama j, en donde la ventana deslizante de longitud Lq está retardada Mq tramas de longitud N con respecto a dicha subtrama j. Ventajosamente, Lq > L0, y en especial Lq=kq·L, siendo kq un entero positivo y kq > k0. Además, se tiene Mq > Mm.

Durante esta etapa 109, se puede prever el cálculo de los máximos del $q_{i,j}$ máximo según un método de minimización que comprende las siguientes etapas iterativas:

- cálculo de $q_{i,j} = \max\left\{q_{i,j-1}; m_{i-M_q,j}\right\}$ y $\tilde{q}_{i,j} = \max\left\{q_{i,j-1}; m_{i-M_q,j}\right\}$
- si rem(i, Lq) = 0, en donde rem es el operador resto de la división entera de dos enteros, entonces:

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

- fin del si

con $q_{0,1} = 0$ y $q_{0,1} = 0$.

A continuación, en una etapa 110, se establecen los valores de umbrales Ω_i propios de cada trama i, entre varios valores fijos Ωa, Ωb, Ωc, etc. De forma más precisa, se establecen los valores de los subumbrales $\Omega_{i,j}$ propios de cada subtrama j de la trama i, segmentándose el umbral Ω_i en varios subumbrales $\Omega_{i,j}$. A título de ejemplo, cada umbral Ω_i o subumbral $\Omega_{i,j}$ adopta un valor fijo escogido entre seis valores fijos Ωa, Ωb, Ωc, Ωd, Ωe, Ωf, estando comprendidos estos valores fijos, por ejemplo, entre 0,05 y 1, y, en especial, entre 0,1 y 0,7.

Cada umbral Ω_i o subumbral $\Omega_{i,j}$ se fija a un valor fijo Ωa, Ωb, Ωc, Ωd, Ωe, Ωf, mediante la puesta en práctica de dos análisis:

- primer análisis: la comparación de los valores del par ($\Delta'_{i,j}$, $\delta'_{i,j}$) en la subtrama de índice j de la trama i con varios pares de umbrales fijos;

- segundo análisis: la comparación de los máximos del máximo $q_{i,j}$ en la subtrama de índice j de la trama i con umbrales fijos.

A continuación de estos dos análisis, un proceso denominado de decisión aportará la decisión final sobre la presencia de la voz en la trama i. Este proceso de decisión comprende las siguientes subetapas, para cada trama i:

- para cada subtrama j de la trama i, se establece un índice de decisión $DEC_i(j)$ que ocupa o bien un estado "1" de detección de una señal de habla o bien un estado "0" de no detección de una señal de habla;

- se establece una decisión temporal VAD(i) basada en la comparación de los índices de decisión $DEC_i(j)$ con operadores "O" lógicos, de manera que la decisión temporal VAD(i) ocupa un estado "1" de detección de una señal de habla si por lo menos uno de dichos índices de decisión $DEC_i(j)$ ocupa este estado "1" de detección de una señal de habla, dicho de otra manera se tiene la siguiente relación:

$$VAD(i) = DEC_i(1) + DEC_i(2) + ... + DEC_i(T), \text{ en donde "+" es el operador "O".}$$

Así, en función de las comparaciones realizadas durante el primer y el segundo análisis, y en función del estado de la decisión temporal VAD(i), el umbral Ω_i se fija a uno de los valores fijos Ωa, Ωb, Ωc, Ωd, Ωe, Ωf y se deduce la decisión final comparando el mínimo rr(i) con el umbral Ω_i fijado a uno de sus valores fijos (consúltese la descripción más adelante).

En muchos casos, las falsas detecciones (o *tonches*) llegan con una amplitud inferior a la de la señal de habla, al estar situado el micrófono al lado de la boca del usuario. Teniendo en cuenta este hecho, es previsible eliminar todavía más las falsas detecciones memorizando el valor máximo de umbral Lastmax deducido a partir de la señal de habla en el último periodo de activación del "VAD" y añadiendo una condición en el procedimiento basada en este valor máximo de umbral Lastmax.

Así, en la etapa 109 descrita anteriormente, se añade la memorización del valor máximo de umbral Lastmax que corresponde al valor variable (o actualizado) de un umbral de comparación para la amplitud de la señal acústica discreta $\{x_i\}$ por debajo del cual se considera que la señal acústica no comprende ninguna señal de habla, determinándose este valor variable durante la última trama de índice k que precede a dicha trama i y en la cual la decisión temporal VAD(k) ocupaba un estado "1" de detección de una señal de habla.

En esta etapa 109, se memoriza también un valor máximo medio $A_{i,j}$ que corresponde al valor máximo medio de la señal acústica discreta $\{x_i\}$ en la subtrama j de la trama i, calculado de la manera siguiente:

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

en donde $a_{i,j}$ corresponde al máximo de la señal acústica discreta $\{x_i\}$ contenido en la trama teórica k formada por la subtrama j de la trama i y por lo menos por una o más subtramas sucesivas que preceden a dicha subtrama j; y

θ es un coeficiente predefinido comprendido entre 0 y 1, con $\theta < \lambda$.

En esta etapa 109, se actualiza de nuevo el valor máximo de umbral Lastmax cada vez que el procedimiento ha considerado que una subtrama p de una trama k contiene una señal de habla, poniendo en práctica el proceso siguiente:

- la detección de una señal de habla en la subtrama p de la trama k sucede a un periodo de ausencia de habla, y, en este caso, Lastmax adopta el valor actualizado [α ($A_{k,p}$ + LastMax)], en donde α es un coeficiente predefinido, comprendido entre 0 y 1, y, por ejemplo, comprendido entre 0,2 y 0,7;

- la detección de una señal de habla en la subtrama p de la trama k sucede a un periodo de presencia de habla, y, en este caso, Lastmax adopta el valor actualizado $A_{k,p}$ si $A_{k,p}$ > Lastmax.

A continuación, en la etapa 110 descrita anteriormente, se añade una condición basada en el valor máximo de umbral Lastmax para fijar el umbral Ω_i.

Para cada trama i, esta condición se basa en la comparación entre:

- el valor máximo de umbral Lastmax; y

- los valores [Kp·$A_{i,j}$] y [Kp· $A_{i,j-1}$], en donde Kp es un coeficiente fijo de ponderación comprendido entre 1 y 2.

Se puede prever también la disminución del valor máximo de umbral Lastmax después de un periodo de temporización dado (por ejemplo, fijado entre varios segundos y varias decenas de segundos) entre la trama i y la última trama de índice k citada previamente, con el fin de evitar la no detección del habla si el usuario/hablante disminuye la amplitud de su voz de forma significativa.

A continuación, en una etapa 111, se calcula, para cada trama en curso i, el mínimo rr(i) de una función de detección discreta FDi(τ), en donde la función de detección discreta FDi(τ) corresponde o bien a la función diferencia discreta $D_i(\tau)$ o bien a la función de diferencia normalizada discreta $DN_i(\tau)$.

Finalmente, en una última etapa 112, se compara, para cada trama en curso i, este mínimo rr(i) con el umbral Ω_i propio de la trama i, para detectar la presencia o no de una señal de habla (o señal sonora), con:

- si rr(i) ≤ Ω_i, entonces la trama i se considera que presenta una señal de habla y el procedimiento entrega una señal de salida DF_i que adopta el valor “1” (dicho de otra manera, la decisión final para la trama i es “presencia de voz en la trama i”);

- si rr(i) > Ω_i, entonces se considera que la trama i no presenta ninguna señal de habla y el procedimiento entrega una señal de salida DF_i que adopta el valor “0” (dicho de otra manera, la decisión final para la trama i es “ausencia de voz en la trama i”).

En referencia a las figuras 1 y 2, se puede prever la aportación de un perfeccionamiento al procedimiento, introduciendo una etapa suplementaria 113 de bloqueo de decisión (o etapa de *hangover*), para evitar los cortes de sonido en una frase y durante la pronunciación de las palabras, teniendo como objetivo esta etapa 113 de bloqueo de decisión reforzar la decisión de presencia/ausencia de voz mediante la puesta en práctica de las dos etapas siguientes:

- conmutación de un estado de no detección de una señal de habla a un estado de detección de una señal de habla después de haber detectado la presencia de una señal de habla sobre N_p tramas i temporales sucesivas;

- conmutación de un estado de detección de una señal de habla a un estado de no detección de una señal de habla después de no haber detectado ninguna presencia de una señal sonora sobre N_A tramas i temporales sucesivas.

Así, esta etapa 113 de bloqueo permite entregar en la salida una señal de decisión de la detección de la voz D_v que adopta el valor “1” correspondiente a una decisión de la detección de la voz y el valor “0” correspondiente a una decisión de la no detección de la voz, en donde:

- la señal de decisión de la detección de la voz D_v conmuta de un estado “1” a un estado “0” si y solo si la señal de salida DF_i adopta el valor “0” en N_A tramas i temporales sucesivas; y

- la señal de decisión de la detección de la voz D_v conmuta de un estado “0” a un estado “1” si y solo si la señal de salida DF_i adopta el valor “1” en N_P tramas i temporales sucesivas.

En referencia a la figura 2, si se supone que se parte de un estado “D_V=1”, se conmuta a un estado “D_V=0” si la señal de salida DF_i adopta el valor “0” en N_A tramas sucesivas, si no, el estado permanece en “D_V=1” (representando Ni el número de la trama en el inicio de la serie). Asimismo, si se supone que se parte de un estado “D_V=0”, se conmuta a un estado “D_V=1” si la señal de salida DF_i adopta el valor “1” en N_P tramas sucesivas, si no, el estado permanece en “D_V=0”.

La decisión final se aplica a las primeras H muestras de la trama tratada. Preferentemente, N_A es superior a N_P, con, por ejemplo, N_A=100 y N_P=3, ya que es mejor correr el riesgo de detectar silencio antes de cortar una conversación.

La descripción trata a continuación sobre dos resultados de detección de voz obtenidos con un procedimiento típico que utiliza un umbral fijo (Figura 3) y con el procedimiento de acuerdo con la invención que utiliza un umbral adaptativo (Figura 4).

En las figuras 3 y 4 (parte inferior), se observa que los dos procedimientos actúan sobre la misma señal acústica discreta $\{x_i\}$, con la amplitud en las ordenadas y las muestras en la abscisa. Esta señal acústica discreta $\{x_i\}$ presenta una sola zona de presencia de habla “PAR”, y numerosas zonas de presencia de ruidos parásitos tales como música, tambores, gritos de un gentío y silbidos. Esta señal acústica discreta $\{x_i\}$ refleja un entorno representativo de una comunicación entre personas (tales como árbitros) en el interior de un estado o de un gimnasio en donde el ruido es relativamente muy fuerte en cuanto a nivel y es notablemente no estacionario.

En las figuras 3 y 4 (parte superior), se observa que los dos procedimientos aprovechan la misma función rr(i) correspondiente, a título recordatorio, al mínimo de la función de detección discreta FDi[τ] seleccionada.

En la figura 3 (en la parte superior), la función mínima rr(i) se compara con un umbral fijo Ωfijo seleccionado de manera óptima para garantizar la detección de la voz. En la figura 3 (parte inferior), se observa la forma de la señal de salida DF_i que ocupa un estado “1” si rr(i) $\leq$ Ωfijo y un estado “0” si rr(i) > Ωfijo.

En la figura 4 (parte superior), la función mínima rr(i) se compara con un umbral adaptativo Ω_i calculado según las etapas descritas anteriormente en referencia a la figura 1. En la figura 4 (parte inferior), se observa la forma de la señal de salida DF_i que ocupa un estado “1” si rr(i) $\leq \Omega_i$ y un estado “0” si rr(i) > Ω_i.

En la figura 3 se observa que el procedimiento de acuerdo con la invención permite una detección de la voz en la zona de presencia de habla “PAR” con la señal de salida DF_i que ocupa un estado “1”, y que esta misma señal de salida DF_i ocupa varias veces un estado “1” en las otras zonas en las que el habla, sin embargo, está ausente, lo cual corresponde a unas falsas detecciones no deseadas con el procedimiento clásico.

Por el contrario, en la figura 4 se observa que el procedimiento de acuerdo con la invención permite una detección óptima de la voz en la zona de presencia de habla “PAR” con la señal de salida DF_i que ocupa un estado “1”, y que esta misma señal de salida DF_i ocupa un estado “0” en las otras zonas en las que el habla está ausente. Así, el procedimiento de acuerdo con la invención garantiza una detección de la voz con una fuerte reducción del número de falsas detecciones.

REIVINDICACIONES

1. Procedimiento de detección de la voz que permite detectar la presencia de señales de habla en una señal acústica x(t) ruidosa procedente de un micrófono, que comprende las etapas sucesivas siguientes:

- una etapa previa de muestreo que comprende una segmentación de la señal acústica x(t) en una señal acústica discreta $\{x_i\}$ compuesta por una secuencia de vectores asociados a unas tramas i temporales de longitud N, correspondiendo N al número de puntos de muestreo, en donde cada vector traduce el contenido acústico de la trama i asociada y está compuesto por N muestras $x_{(i-1)N+1}$, $x_{(i-1)N+2}$,...,x_{iN-1}, x_{iN}, siendo i un entero positivo;

- una etapa de cálculo de una función de detección $FD(\tau)$ basada en el cálculo de una función diferencia $D(\tau)$ que varía en función del desplazamiento τ sobre una ventana de integración de longitud W que comienza en el tiempo t0, con:

$$D(\tau) = \sum_{n=t0}^{t0+W-1} |x(n) - x(n+\tau)| \text{ en donde } 0 \leq \tau \leq \max(\tau);$$

en donde esta etapa de cálculo de la función de detección $FD(\tau)$ consiste en un cálculo de una función de detección discreta $FDi(\tau)$ asociada a las tramas i;

- una etapa de búsqueda del mínimo de la función de detección $FD(\tau)$ y comparación de este mínimo con un umbral, variando τ dentro de un intervalo de tiempo determinado, denominado intervalo en curso, para detectar la presencia o no de una frecuencia fundamental F_0 característica de una señal de habla en dicho intervalo en curso, en donde esta etapa de búsqueda del mínimo de la función de detección $FD(\tau)$ y la comparación de este mínimo con un umbral se realizan buscando, en cada trama i, el mínimo rr(i) de la función de detección discreta $FDi(\tau)$;

estando dicho procedimiento caracterizado por que comprende:

- una etapa de adaptación del umbral en dicho intervalo en curso, en función de valores calculados a partir de la señal acústica x(t) establecidos en dicho intervalo en curso,

en el que esta etapa de adaptación del umbral consiste en, para cada trama i, adaptar un umbral Ω_i propio de la trama i en función de valores de referencia calculados a partir de los valores de las muestras de la señal acústica discreta $\{x_i\}$ en dicha trama i;

en el que dicha etapa de búsqueda del mínimo de la función de detección $FD(\tau)$ y la comparación de este mínimo con un umbral se realizan comparando, en cada trama i, el mínimo rr(i) de la función de detección discreta $FDi(\tau)$ con un umbral Ω_i propio de la trama i;

y, en el que, la etapa de adaptación de los umbrales Ω_i para cada trama i comprende las etapas siguientes:

a)- se subdivide la trama i que comprende N puntos de muestreo en T subtramas de longitud L, donde N es un múltiplo de T con el fin de que la longitud L=N/T sea un entero, y de manera que las muestras de la señal acústica discreta $\{x_i\}$ dentro de una subtrama de índice j de la trama i comprendan las L muestras siguientes:

$x_{(i-1)N+(j-1)L+1}$, $x_{(i-1)N+(j-1)L+2}$, ..., $x_{(i-1)N+jL}$, siendo j un entero positivo comprendido entre 1 y T;

b)- se calculan los valores máximos $m_{i,j}$ de la señal acústica discreta $\{x_i\}$ en cada subtrama de índice j de la trama i, con:

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

c)- se calcula por lo menos un valor de referencia $Ref_{i,j}$, $MRef_{i,j}$ propio de la subtrama j de la trama i, siendo el valor o cada valor de referencia $Ref_{i,j}$, $MRef_{i,j}$, por cada subtrama j calculado a partir del valor máximo $m_{i,j}$ en la subtrama j de la trama i;

d)- se establece el valor del umbral Ω_i propio de la trama i en función de todos los valores de referencia $Ref_{i,j}$, $MRef_{i,j}$ calculados en las subtramas j de la trama i;

y en el que, en la etapa c), se realizan las siguientes subetapas sobre cada trama i:

c1)- se calculan las envolventes suavizadas de los máximos $\overline{m}_{i,j}$ en cada subtrama de índice j de la trama i, con:

$$\overline{m}_{i,j} = \lambda\, \overline{m}_{i,j-1} + (1-\lambda) m_{i,j}$$, en donde λ es un coeficiente predefinido comprendido entre 0 y 1;

c2)- se calculan las señales de variación $\Delta_{i,j}$ en cada subtrama de índice j de la trama i, con:

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

y en el que se calcula por lo menos un valor de referencia denominado principal $Ref_{i,j}$ por cada subtrama j a partir de la señal de variación $\Delta_{i,j}$ en la subtrama j de la trama i.

2. Procedimiento según la reivindicación 1, en el que, en la etapa c) y a continuación de la subetapa c2), se realizan las siguientes subetapas sobre cada trama i:

c3)- se calculan los máximos de variación $s_{i,j}$ en cada subtrama de índice j de la trama i, en donde $s_{i,j}$ corresponde al máximo de la señal de variación $\Delta_{i,j}$ calculado sobre una ventana deslizante de longitud Lm anterior a dicha subtrama j, siendo dicha longitud Lm variable según que la subtrama j de la trama i corresponda a un periodo de silencio o de presencia de habla;

c4)- se calculan las desviaciones de variación $\delta_{i,j}$ en cada subtrama de índice j de la trama i, con:

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

y en el que, para cada subtrama j de la trama i, se calculan dos valores de referencia principales $Ref_{i,j}$ a partir respectivamente de la señal de variación $\Delta_{i,j}$ y de la desviación de variación $\delta_{i,j}$.

3. Procedimiento según la reivindicación 2, en el que, en la etapa c) y a continuación de la subetapa c4), se realiza una subetapa c5) de cálculo de las señales de variación normalizadas $\Delta'_{i,j}$ y de las desviaciones de variación normalizadas $\delta'_{i,j}$ en cada subtrama de índice j de la trama i, de la manera siguiente:

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

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

y en el que, para cada subtrama j de una trama i, la señal de variación normalizada $\Delta'_{i,j}$ y la desviación de variación normalizada $\delta'_{i,j}$ constituyen, cada una de ellas, un valor de referencia principal $Ref_{i,j}$ de manera que, en la etapa d), se establece el valor del umbral Ω_i propio de la trama i en función del par $(\Delta'_{i,j}, \delta'_{i,j})$ de las señales de variación normalizadas $\Delta'_{i,j}$ y de las desviaciones de variación normalizadas $\delta'_{i,j}$ en las subtramas j de la trama i.

4. Procedimiento según la reivindicación 3, en el que, en la etapa d), el valor del umbral Ω_i propio de la trama i se establece dividiendo el espacio definido por el valor del par $(\Delta'_{i,j}, \delta'_{i,j})$, y examinando el valor del par $(\Delta'_{i,j}, \delta'_{i,j})$ sobre una o varias subtramas sucesivas según la zona de valor del par $(\Delta'_{i,j}, \delta'_{i,j})$.

5. Procedimiento según cualquiera de las reivindicaciones 2 a 4, en el que, en la subetapa c3), la longitud Lm de la ventana deslizante responde a las ecuaciones siguientes:

- Lm = L0 si la subtrama j de la trama i corresponde a un periodo de silencio;
- Lm = L1 si la subtrama j de la trama i corresponde a un periodo de presencia de habla;

con L1 < L0, y en particular con L1=k1·L y L0=k0·L, siendo L la longitud de las subtramas de índice j y siendo k0, k1 unos enteros positivos.

6. Procedimiento según la reivindicación 2, en el que, en la subetapa c3), para cada cálculo del máximo de variación $s_{i,j}$ en la subtrama j de la trama i, la ventana deslizante de longitud Lm está retardada Mm tramas de longitud N con respecto a dicha subtrama j.

7. Procedimiento según las reivindicaciones 3 y 6, en el que, en la subetapa c3), se calculan también los máximos de variación normalizados $s'_{i,j}$ en cada subtrama de índice j de la trama i, en donde $s'_{i,j}$ corresponde al máximo de la señal de variación normalizado $\Delta'_{i,j}$ calculado sobre una ventana deslizante de longitud Lm anterior a dicha subtrama j, en donde:

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

y en donde cada máximo de variación normalizado s'$_{i,j}$ se calcula según un método de minimización que comprende las etapas iterativas siguientes:

- cálculo de $s'_{i,j} = \max\{s'_{i,j-1}; \Delta'_{i-M_m,j}\}$ y $\tilde{s}'_{i,j} = \max\{s'_{i,j-1}; \Delta'_{i-M_m,j}\}$
- si rem(i, Lm) = 0, donde rem es el operador resto de la división entera de dos enteros, entonces:

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

con $s'_{0,1} = 0$ y $\tilde{s}'_{0,1} = 0$;

y en el que, en la etapa c4), se calculan las desviaciones de variación normalizadas δ'$_{i,j}$ en cada subtrama de índice j de la trama i, de la manera siguiente:

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

8. Procedimiento según cualquiera de las reivindicaciones 1 a 7, en el que, en la etapa c), se realiza una subetapa c6) en la que se calculan los máximos del $q_{i,j}$ máximo en cada subtrama de índice j de la trama i, en donde $q_{i,j}$ corresponde al máximo del valor máximo $m_{i,j}$ calculado sobre una ventana deslizante de longitud fija Lq anterior a dicha subtrama j, en donde la ventana deslizante de longitud Lq está retardada Mq tramas de longitud N con respecto a dicha subtrama j, y en donde otro valor de referencia denominado secundario $MRef_{i,j}$ por cada subtrama j corresponde a dicho máximo del $q_{i,j}$ máximo en la subtrama j de la trama i.

9. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que, en la etapa d), el umbral Ω_i propio de la trama i se segmenta en varios subumbrales $\Omega_{i,j}$ propios de cada subtrama j de la trama i, y el valor de cada subumbral $\Omega_{i,j}$ se establece por lo menos en función del o de los valores de referencia $Ref_{i,j}$, $MRef_{i,j}$ calculados en la subtrama j de la trama i correspondiente.

10. Procedimiento según las reivindicaciones 3 y 9, en el que, en la etapa d), se establece el valor de cada umbral $\Omega_{i,j}$ propio de la subtrama j de la trama i comparando los valores del par (Δ'$_{i,j}$, δ'$_{i,j}$) con varios pares de umbrales fijos, siendo el valor de cada umbral $\Omega_{i,j}$ seleccionado de entre varios valores fijos en función de las comparaciones del par (Δ'$_{i,j}$, δ'$_{i,j}$) con dichos pares de umbrales fijos.

11. Procedimiento según cualquiera de las reivindicaciones 1 a 10, en el que, en la etapa d), se realiza un proceso denominado de decisión, que comprende las siguientes subetapas, para cada trama i:

- para cada subtrama j de la trama i, se establece un índice de decisión $DEC_i(j)$ que ocupa o bien un estado "1" de detección de una señal de habla, o bien un estado "0" de no detección de una señal de habla;

- se establece una decisión temporal VAD(i) basada en la comparación de los índices de decisión $DEC_i(j)$ con unos operadores "O" lógicos, de manera que la decisión temporal VAD(i) ocupa un estado "1" de detección de una señal de habla si por lo menos uno de dichos índices de decisión $DEC_i(j)$ ocupa este estado "1" de detección de una señal de habla.

12. Procedimiento según las reivindicaciones 9 y 11, en el que, en el proceso de decisión, se realizan las siguientes subetapas, para cada trama i:

- se memoriza un valor máximo de umbral Lastmax que corresponde al valor variable de un umbral de comparación para la amplitud de la señal acústica discreta $\{x_i\}$ por debajo del cual se considera que la señal acústica no comprende ninguna señal de habla, siendo este valor variable determinado durante la última trama de índice k que precede a dicha trama i y en la que la decisión temporal VAD(k) ocupaba un estado "1" de detección de una señal de habla;

- se memoriza un valor máximo medio $A_{i,j}$ que corresponde al valor máximo medio de la señal acústica discreta $\{x_i\}$ en la subtrama j de la trama i, calculado de la manera siguiente:

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

en donde $a_{i,j}$ corresponde al máximo de la señal acústica discreta $\{x_i\}$ contenido en una trama formada por la subtrama j de la trama i y por lo menos una o varias subtramas sucesivas que preceden a dicha subtrama j; y

θ es un coeficiente predefinido comprendido entre 0 y 1, con $\theta < \lambda$;

- se establece el valor de cada subumbral $\Omega_{i,j}$ en función de la comparación entre dicho valor máximo de umbral Lastmax y unos valores máximos medios $A_{i,j}$ y $A_{i,j-1}$ considerados sobre dos subtramas j y j-1 sucesivas.

13. Procedimiento según la reivindicación 12, en el que, en el proceso de decisión, se actualiza de nuevo el valor máximo de umbral Lastmax cada vez que el procedimiento ha considerado que una subtrama p de una trama k contiene una señal de habla, poniendo en práctica el siguiente proceso:

- la detección de una señal de habla en la subtrama p de la trama k sucede a un periodo de ausencia de habla, y, en este caso, Lastmax adopta el valor actualizado $[\alpha(A_{k,p} + LastMax)]$, en donde α es un coeficiente predefinido comprendido entre 0 y 1, y, por ejemplo, comprendido entre 0,2 y 0,7;

- la detección de una señal de habla en la subtrama p de la trama k sucede a un periodo de presencia de habla, y, en este caso, Lastmax adopta el valor actualizado $A_{k,p}$ si $A_{k,p} > Lastmax$.

14. Procedimiento según las reivindicaciones 12 o 13, en el que se establece el valor del umbral Ω_i en función de dicho valor máximo Lastmax basándose en la comparación entre:

- el valor máximo de umbral Lastmax; y

- los valores $[Kp.A_{i,j}]$ y $[Kp.A_{i,j-1}]$, en donde Kp es un coeficiente fijo de ponderación comprendido entre 1 y 2.

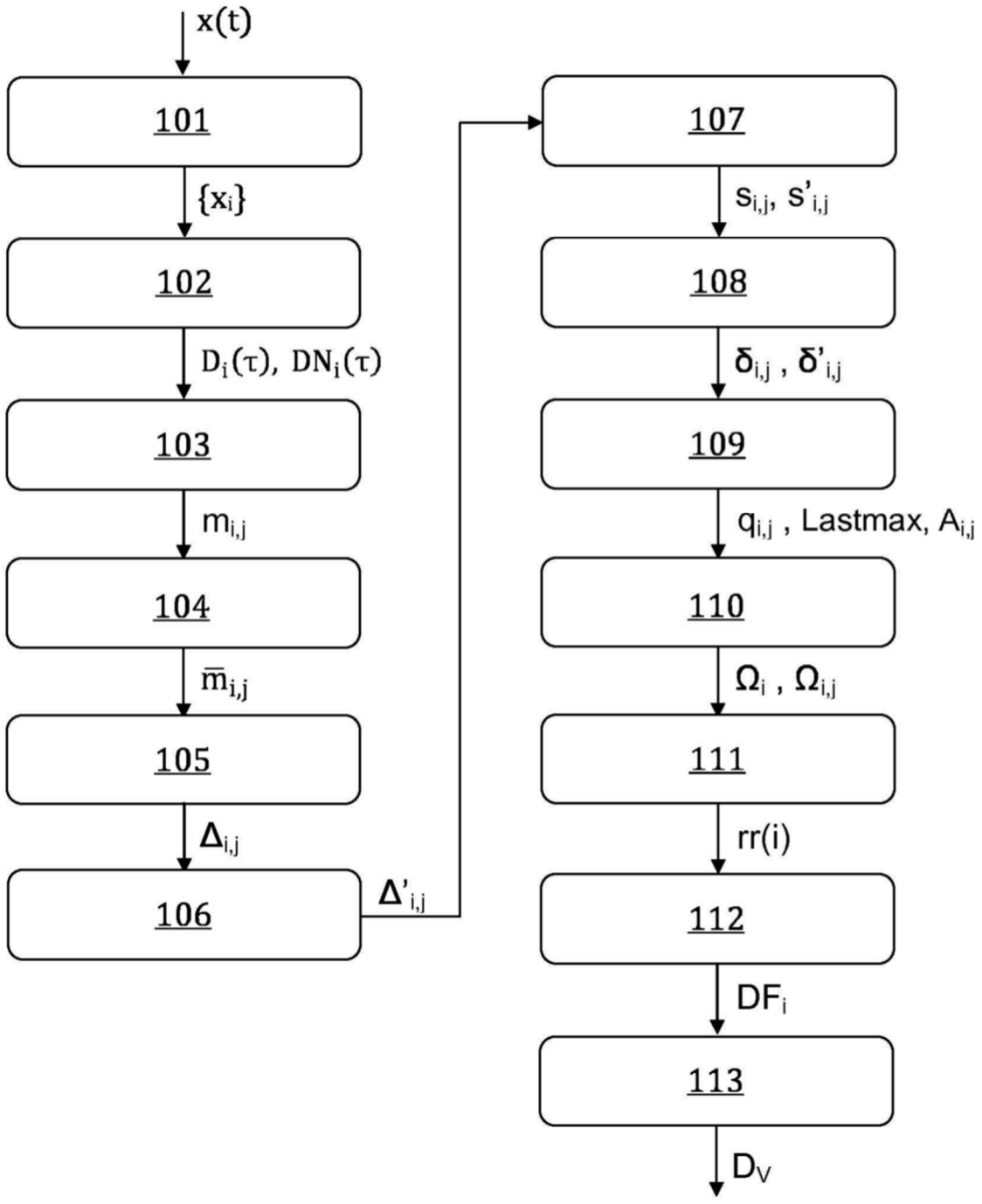
x(t)
101
{xi}
102
Di(τ), DNi(τ)
103
mi,j
104
m̄i,j
105
Δi,j
106
Δ'i,j
107
si,j, s'i,j
108
δi,j , δ'i,j
109
qi,j , Lastmax, Ai,j
110
Ωi , Ωi,j
111
rr(i)
112
DFi
113
DV

FIG.1

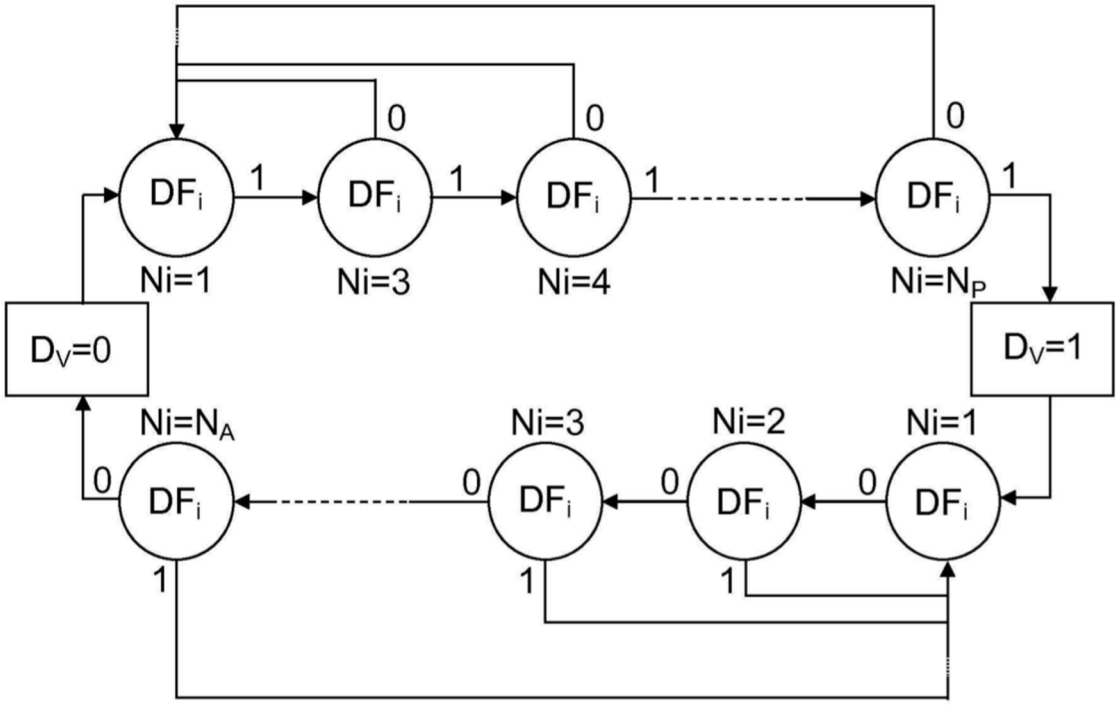
DFi
Ni=1
Ni=3
Ni=4
Ni=NP
DV=0
DV=1
Ni=NA
Ni=2
0
1

FIG.2

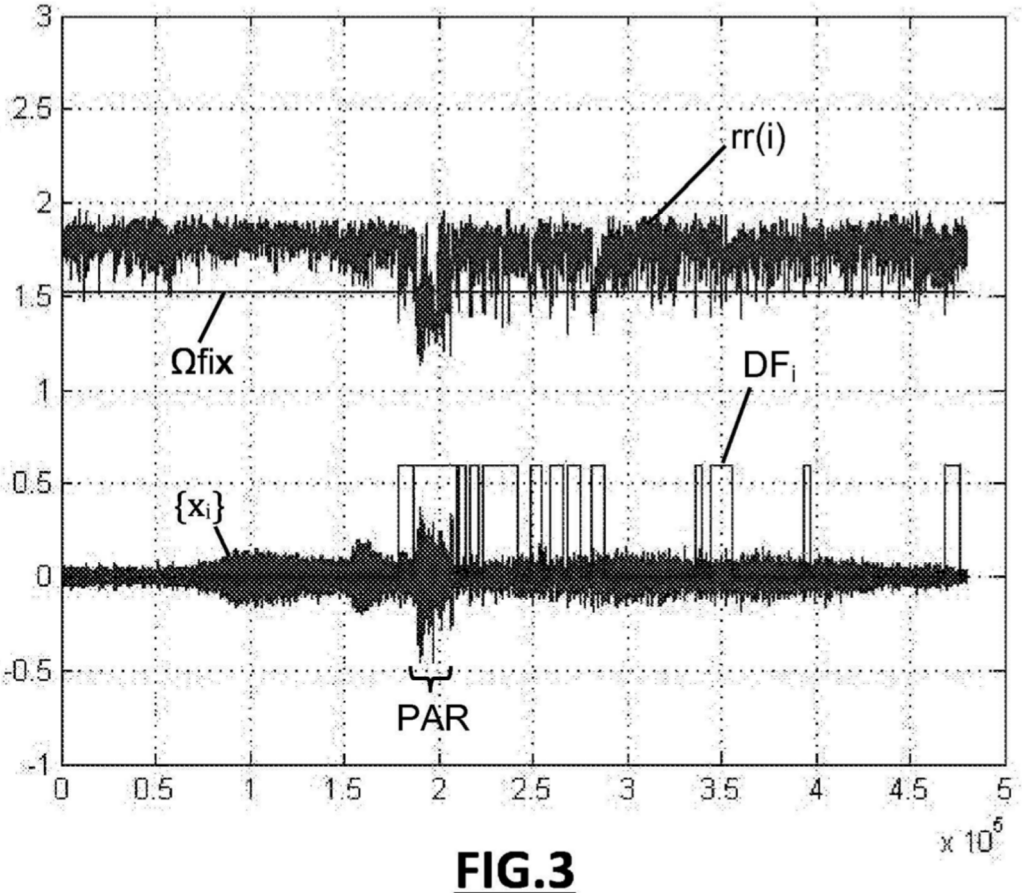
rr(i)
Ωfix
DFi
{xi}
PAR
x 10^5

FIG.3

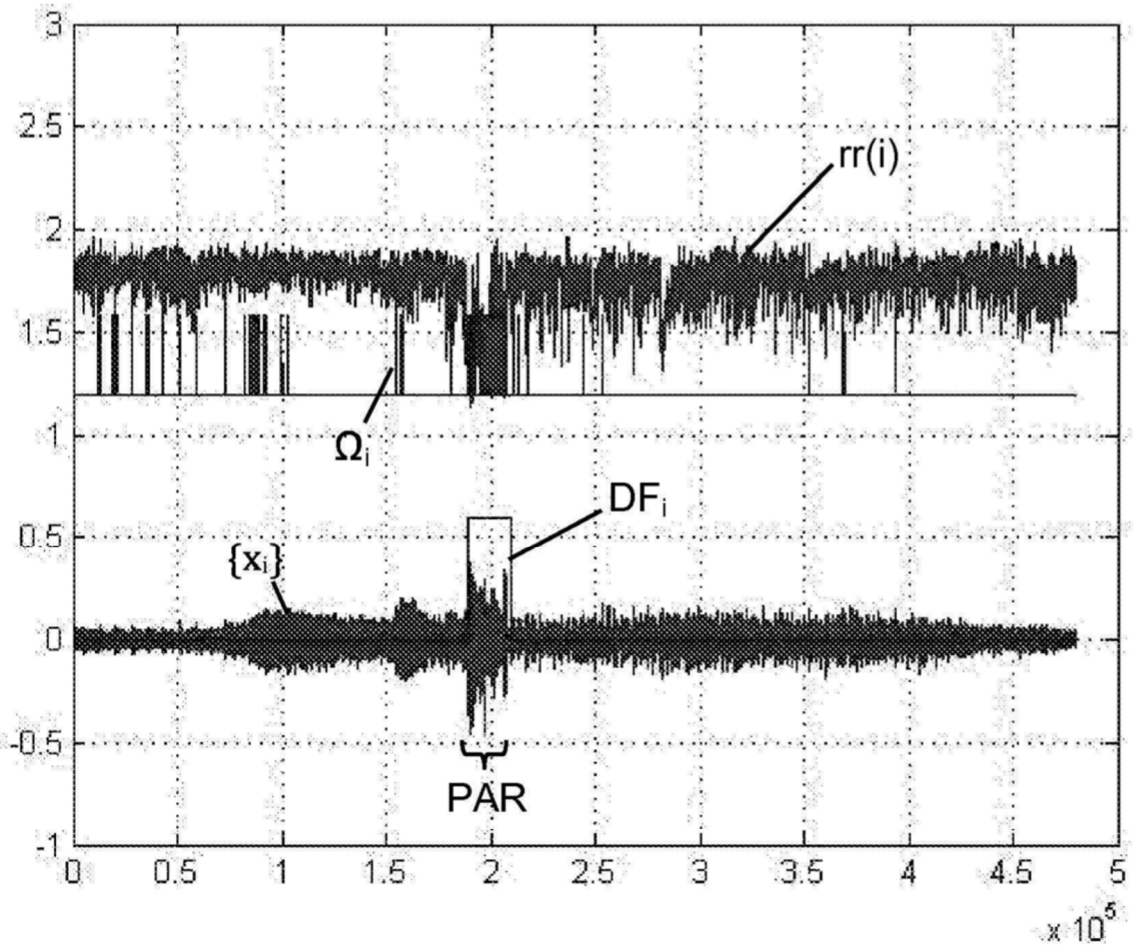

FIG.4