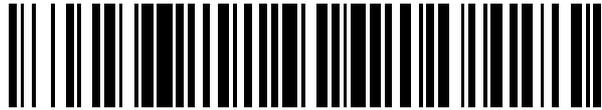


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 489 472**

51 Int. Cl.:

G10L 25/78 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.12.2010 E 10861147 (6)**

97 Fecha y número de publicación de la concesión europea: **21.05.2014 EP 2619753**

54 Título: **Método y aparato para una detección adaptativa de la actividad vocal en una señal de audio de entrada**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
02.09.2014

73 Titular/es:

**HUAWEI TECHNOLOGIES CO., LTD. (100.0%)
Huawei Administration Building Bantian
Longgang District
Shenzhen, Guangdong 518129, CN**

72 Inventor/es:

WANG, ZHE

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 489 472 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Método y aparato para una detección adaptativa de la actividad vocal en una señal de audio de entrada

5 CAMPO DE LA INVENCION

La presente invención se refiere a un método y un aparato para una detección adaptativa de una actividad vocal en una señal de audio de entrada constituida por tramas y en particular, a un método y aparato de detección de actividad vocal que utiliza parámetros de la relación de señal a ruido segmentada de sub-banda procesada de forma no lineal.

ANTECEDENTES DE LA INVENCION

La detección de actividad vocal (VAD) suele ser una técnica que se proporciona para detectar una actividad vocal en una señal. Los detectores de actividad vocal son ampliamente utilizados en el sector de las telecomunicaciones. La función de un detector de actividad vocal es detectar, en los canales de comunicación, la presencia o ausencia de señales activas tales como voz o música. El detector de actividad vocal puede estar contenido dentro de una red de comunicaciones en donde la red puede decidir comprimir un ancho de banda de transmisión en periodos en donde las señales activas estén ausentes o realizar otro procesamiento en función de una decisión de detección de actividad vocal que indica si existe, o no, una señal activa. Un detector de actividad vocal puede comprar un parámetro característico o un conjunto de parámetros característicos extraídos desde la señal de entrada con los correspondientes valores de umbral y determinar si la entrada comprende una señal activa, o no, sobre la base del resultado de la comparación. La función de un detector de actividad vocal se determina, en una gran medida, por la elección de los parámetros característicos utilizados. Son numerosos los parámetros característicos propuestos para la detección de actividad vocal tales como los parámetros basados en la energía, parámetros basados en la envolvente espectral, parámetros basados en la entropía o parámetros basados en estadísticas de orden superior. En general, los parámetros basados en la energía proporcionan un buen rendimiento de la detección de actividad vocal. En los últimos años, parámetros basados en la relación señal a ruido SNR de sub-bandas, como una clase de parámetros basados en la energía, han sido ampliamente utilizados en el sector de las telecomunicaciones. En los detectores de actividad vocal basados en la relación de señal a ruido SNR de sub-bandas, la relación SNR para cada sub-banda de frecuencia de una trama de entrada se detecta y los valores de SNRs de todas las sub-bandas se añaden para proporcionar una SNR segmentada. Este parámetro de relación SNR segmentada, SSNR, puede compararse por un valor de umbral para tomar una decisión de detección de actividad vocal VADD. El umbral utilizado suele ser una variable que es adaptativa para una relación SNR a largo plazo de la señal de entrada o un nivel de ruido de fondo.

En una Recomendación de ITU-T recientemente completada G.720.1, el parámetro de la relación SSNR convencional ha sido mejorado aplicando un procesamiento no lineal para obtener una relación SSNR modificada. La relación SNR segmentada modificada calculada se compara también con un valor de umbral que se determina a partir de una tabla de valores de umbral en función de la relación SNR a largo plazo de la señal de entrada, de la variación del ruido de fondo y del punto operativo de detección de actividad vocal, en donde el punto operativo VAD define la solución de compromiso de la decisión de VAD entre detección activa y detección inactiva, a modo de ejemplo, un punto operativo de calidad preferida hará que la decisión VAD sea a favor de la detección de señales activas y viceversa.

Aunque el parámetro de SNR segmentada modificada, utilizado por la recomendación G.720.1, aumenta el rendimiento de la detección de actividad vocal, el rendimiento de VAD no es estacionario y todavía necesita mejora el entorno de fondo de baja relación SNR. Los detectores de actividad vocal convencionales están diseñados para equilibrar sus rendimientos en varias condiciones de ruido de fondo. En consecuencia, los detectores de actividad vocal convencionales tienen un rendimiento que es sub-óptimo para condiciones específicas y en particular, en un entorno de fondo de baja relación SNR y de carácter no estacionario.

En consecuencia, un objetivo de la presente invención es dar a conocer un método y un aparato para detectar una actividad vocal en una señal de audio de entrada que tenga un alto rendimiento de VAD.

El documento EP 2159788 A1 da a conocer un método y dispositivo de detección de actividad vocal (VAD), de modo que el valor de umbral de VAD pueda ser adaptativo con la variación del ruido de fondo. El dispositivo de detección de actividad vocal VAD incluye: una unidad de análisis de fondo, adaptada para: analizar las características de ruido de fondo de una señal actual en función de un resultado de determinación de VAD de entrada, para obtener parámetros relacionados con la variación del ruido de fondo y para proporcionar estos parámetros; una unidad de ajuste del valor de umbral de VAD, adaptada para: obtener un sesgo del umbral de VAD en función de los parámetros proporcionados por la unidad de análisis de fondo y proporcionar el sesgo del valor de umbral de VAD y una unidad de determinación de VAD adaptada para: modificar un umbral de VAD a modificarse en función del sesgo del rendimiento de VAD proporcionado por la unidad de ajuste del valor de umbral de VAD, para determinar el ruido de fondo utilizando el umbral VAD modificado y proporcionar un resultado de la determinación de VAD.

SUMARIO DE LA INVENCION

La invención da a conocer, según un primer aspecto de la idea inventiva, un método para la detección adaptativa de una actividad vocal en una señal de audio de entrada constituida por tramas que comprende las etapas de:

(a) determinar una característica de ruido de la señal de entrada sobre la base de al menos en una trama de entrada recibida de la señal de audio de entrada,

(e) la derivación de un parámetro de VAD (vp) adaptado para, o seleccionado en función de, dicha característica de ruido de la señal de audio de entrada y

(f) la comparación del parámetro VAD derivado con un valor de umbral para proporcionar una decisión de detección de actividad vocal, según la reivindicación 1.

Las formas de puesta en práctica del primer aspecto de la idea inventiva pueden utilizar parámetros basados en la energía, parámetros basados en la envolvente espectral, parámetros basados en la entropía o parámetros basados en estadísticas de orden superior como parámetros de VAD.

En el primer aspecto de la idea inventiva de la presente invención, la invención da a conocer un método para la detección adaptativa de una actividad vocal en una señal de audio de entrada constituida por tramas, que comprende las etapas de:

(a) la determinación de una característica de ruido de la señal de entrada sobre la base de al menos en una trama de entrada recibida de la señal de audio de entrada,

(b) la división de dicha trama de entrada recibida de dicha señal de audio en varias sub-bandas,

(c) la obtención de una relación de señal a ruido SNR para cada sub-banda de dicha trama de entrada,

(d) el cálculo de un parámetro específico de sub-banda sobre la base de dicha relación SNR de la sub-banda respectiva para cada sub-banda utilizando una función adaptativa,

en donde al menos un parámetro de la función adaptativa se selecciona en función de dicha característica de ruido de la señal de audio de entrada,

(e) la derivación de una relación SNR segmentada modificada como dicho parámetro de VAD añadiendo los parámetros específicos de sub-bandas y

(f) la comparación de la relación SNR segmentada modificada derivada con un valor de umbral para proporcionar una decisión de VAD.

La invención según el primer aspecto de la idea inventiva da a conocer una solución de más eficiencia y de mejor calidad de VAD. La eficiencia de VAD es la capacidad para detectar la característica de ruido, tal como ruido de fondo, mientras la calidad de VAD se refiere a la capacidad para detectar señales activas, tales como de voz o de música en la señal de audio de entrada.

En una puesta en práctica posible del primer aspecto de la idea inventiva, la característica de ruido determinada de dicha señal de audio de entrada estación de trabajo formada por la relación de señal a ruido SNR a largo plazo de la señal de audio de entrada.

En otra puesta en práctica posible del primer aspecto de la idea inventiva, la característica de ruido determinada de dicha señal de audio de entrada se forma por una variación del ruido de fondo de la señal de audio de entrada.

En otra puesta en práctica posible del primer aspecto de la idea inventiva, la característica de ruido determinada de dicha señal de audio de entrada está constituida por una combinación de las relaciones SNR a largo plazo y de la variación del ruido de fondo de la señal de audio de entrada.

En el primer aspecto de la idea inventiva de la presente invención, la función adaptativa utilizada para calcular un parámetro específico de sub-banda está formada por una función no lineal.

En una posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la idea inventiva de la presente invención, se obtiene una relación SNR para cada sub-banda de la trama de entrada obteniendo una energía de señal para cada sub-banda, p.e., una energía de señal para cada sub-banda de la trama de entrada.

En otra posible puesta en práctica del método para la detección adaptativa de una actividad vocal por una señal de

audio de entrada en función del primer aspecto de la presente invención, la relación SNR para cada sub-banda de dicha trama de entrada se obtiene estimando una energía de ruido de fondo para cada sub-banda.

5 En otra posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la relación SNR para cada sub-banda de dicha trama de entrada se obtiene calculando la relación SNR para cada sub-banda dependiendo de la energía de la señal y de la energía del ruido de fondo de la respectiva sub-banda.

10 En otra posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la energía de señal para cada sub-banda de dicha trama de entrada es la energía de señal atenuada que forma una media ponderada entre la trama de entrada y al menos una trama anterior.

15 En la puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada en función del primer aspecto de la presente invención, la relación SSNR modificada se calcula sumando los parámetros específicos de sub-banda como sigue:

$$mssnr = \sum_{i=1}^N sbasp(i)$$

20 en donde N es el número de sub-bandas de frecuencias en las que se divide dicha trama de entrada,

en donde sbasp(i) es un parámetro específico de sub-banda calculado sobre la base de la relación SNR de sub-banda para cada sub-banda que utiliza una función adaptativa.

25 En la puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la relación SNR segmentada modificada se calcula como sigue.

$$mssnr = \sum_{i=1}^N (f(snr(i)) + \alpha)^\beta$$

30 en donde snr(i) es una relación SNR de la i-ésima sub-banda de la trama de entrada,

N es el número de sub-bandas de frecuencia en las que se divide dicha trama de entrada,

35 $(f(snr(i)) + \alpha)^\beta$ es la función adaptativa (AF) utilizada para calcular el parámetro específico de sub-banda sbasp(i) y

α , β son dos variables configurables de la función adaptativa (AF).

40 En una posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la primera variable α de la función adaptativa (AF) depende de una relación SNR a largo plazo (lsnr) de la señal de audio de entrada como sigue:

$$\alpha = g(i, lsnr)$$

45 en donde g es una función lineal o no lineal y

en donde la segunda variable β de la función adaptativa (AF) depende de la relación SNR a largo plazo (lsnr) y φ :

$$\beta = h(lsnr, \varphi)$$

50 en donde h es una función no lineal y

$$\varphi = f(snr(i)) + \alpha.$$

55 En una puesta en práctica adicional del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la primera variable α de la función adaptativa ((AF)) se calcula por:

$$\alpha = g(i, \text{lsnr}) = a(i) \cdot \text{lsnr} + b(i)$$

en donde a (i), b (i) son números reales que dependen de un índice de sub-banda i y

5 la segunda variable β de la función adaptativa ((AF)) se calcula por:

$$\beta = h(\text{lsnr}, \varphi) = \begin{cases} \beta_1 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \\ \beta_2 & \varphi \geq d \text{ y } e_1 < \text{lsnr} \leq e_2 \\ \beta_3 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \\ \beta_4 & \text{de no ser así} \end{cases}$$

10 en donde $\beta_1 < \beta_2 < \beta_3$ y β_4 y d y $e_1 < e_2$ son números enteros o flotantes y en donde lsnr es la relación SNR a largo plazo de la señal de audio de entrada.

En una posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la relación SNR segmentada modificada derivada (mssnr) compara con un valor de umbral (thr) que se establece para:

15

$$\text{thr} = \begin{cases} k_1 & \text{lsnr} > e_2 \\ k_2 & e_1 < \text{lsnr} \leq e_2 \\ k_3 & \text{lsnr} \leq e_1 \end{cases}$$

en donde $k_1 > k_2 > k_3$ y $e_1 < e_2$ son números enteros o flotantes y en donde la decisión de detección de actividad vocal (VADD) se genera como sigue:

20

$$VADD = \begin{cases} 1 & \text{mssnr} > \text{thr} \\ 0 & \text{mssnr} \leq \text{thr} \end{cases}$$

en donde VADD = 1 representa una trama activa con la actividad vocal estando presente y

25 VADD =0 representa una trama pasiva con la actividad vocal estando ausente.

En una posible puesta en práctica del método para la detección adaptativa de una señal de audio de entrada de actividad vocal, según el primer aspecto de la presente invención, la primera variable alfa de la función adaptativa ((AF)) se calcula por:

30

$$\alpha = g(i, \text{lsnr}, \varepsilon) = a(i) \cdot \text{lsnr} + b(i) + c(\varepsilon)$$

en donde a (i), b (i) son números reales que dependen de un índice de sub-banda i y

35 c (ε) es un número real que depende de la fluctuación estimada del ruido de fondo de la señal de audio de entrada y

en donde la segunda variable β de la función adaptativa ((AF)) se calcula por:

$$\beta = h(\text{lsnr}, \varphi, \varepsilon) = \begin{cases} \beta_1 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \text{ y } \varepsilon \leq p \\ \beta_2 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \text{ y } \varepsilon > p \\ \beta_3 & \varphi \geq d \text{ y } e_1 < \text{lsnr} < e_2 \text{ y } \varepsilon \leq p \\ \beta_4 & \varphi \geq d \text{ y } e_1 < \text{lsnr} < e_2 \text{ y } \varepsilon > p \\ \beta_5 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \text{ y } \varepsilon \leq p \\ \beta_6 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \text{ y } \varepsilon > p \\ \beta_7 & \varphi < d \end{cases}$$

40

en donde $\varphi = f(\text{snr}(i)) + \alpha$ y ε es la fluctuación estimada del ruido de fondo y

d y $e_1 < e_2$ son números enteros o flotantes.

- 5 En una posible puesta en práctica de un método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la relación SNR segmentada modificada derivada (mssnr) se compara con un valor de umbral (thr) que se establece para:

$$\text{thr} = \begin{cases} q_1 + r_1 \cdot \text{Min} \left[\frac{\text{lsnr} - v_1}{W_1}, 1 \right] & \text{lsnr} > e_2 \\ q_2 + r_2 \cdot \text{Min} \left[\frac{\text{lsnr} - v_2}{W_2} \right] & e_1 < \text{lsnr} \leq e_2 \\ q_3 + r_3 \cdot \text{Max} \left[\text{Min} \frac{\text{lsnr} - v_3}{W_3}, 1 \right] & \text{lsnr} \leq e_1 \end{cases}$$

10

en donde q_1, q_2, q_3 y r_1, r_2, r_3 y $e_1 < e_2$ y

v_1, v_2, v_3 y w_1, w_2, w_3 son números enteros o flotantes,

15

en donde la decisión de detección de actividad vocal (VADD) se genera como sigue:

$$\text{VADD} = \begin{cases} 1 & \text{mssnr} > \text{thr} \\ 0 & \text{mssnr} \leq \text{thr} \end{cases}$$

20

en donde VADD = 1 representa una trama activa con la actividad vocal estando presente y

VADD = 0 representa una trama pasiva con la actividad vocal estando ausente.

La invención da a conocer, según un segundo aspecto de la invención, un aparato de VAD para detectar una actividad vocal en una señal de audio de entrada que consiste en tramas,

25

en donde dicho aparato de VAD comprende:

30

una unidad de cálculo de parámetros VAD basados en la relación SNR que calcula una SNR (snr) para cada sub-banda de una trama de entrada aplicada y un parámetro específico de sub-banda (sbsp) basado en dicha respectiva SNR de sub-banda (snr) para cada sub-banda que utiliza una función adaptativa (AF) y deriva una relación SNR segmentada modificada (mssnr) añadiendo los parámetros específicos de la sub-banda y

35

una unidad de generación de decisión de VAD que genera una decisión de VAD (VADD) comparando la relación SNR segmentada modificada (mssnr) con un valor de umbral, según se especifica en la reivindicación 11.

En la puesta en práctica del aparato de VAD según el segundo aspecto de la presente invención, el aparato comprende una unidad de determinación de característica de ruido que determina una característica de ruido (nc) de la señal de entrada sobre la base de al menos en una trama de entrada recibida de la señal de audio de entrada.

40

En una posible puesta en práctica del aparato de VAD según el segundo aspecto de la presente invención, la unidad de determinación de característica de ruido comprende una unidad de estimación de SNR a largo plazo que calcula una relación SNR a largo plazo de dicha señal de audio de entrada.

45

En otra posible puesta en práctica del aparato de VAD según el segundo aspecto de la presente invención, la unidad de determinación de característica de ruido comprende una unidad de estimación de variación de ruido de fondo que calcula una estacionalidad o fluctuación del ruido de fondo de la señal de audio de entrada.

50

En otra posible puesta en práctica del aparato de VAD según el segundo aspecto de la presente invención, la unidad de determinación de la característica de ruido comprende una unidad de estimación de SNR a largo plazo que calcula una relación SNR a largo plazo de la señal de audio de entrada y una unidad de estimación de variación de ruido de fondo que calcula una estacionalidad o fluctuación del ruido de fondo de dicha señal de audio de entrada.

En la puesta en práctica del aparato de VAD según el segundo aspecto de la presente invención, la función

adaptativa (AF) se selecciona en función de al menos una característica de ruido (nc) determinada por dicha unidad de determinación de característica de ruido.

5 La invención da a conocer, además, un dispositivo de procesamiento de señal de audio según un tercer aspecto de la presente invención en donde dicho dispositivo de procesamiento de señal de audio comprende una unidad de procesamiento de señal de audio para procesar una señal de entrada dependiendo de una decisión de VAD (VADD) proporcionada por el aparato de VAD según el segundo aspecto de la presente invención tal como se estipula en la reivindicación 14.

10 BREVE DESCRIPCIÓN DE LOS DIBUJOS

A continuación, posibles puestas en práctica de diferentes aspectos de la presente invención se describen haciendo referencia a las Figuras adjuntas con más detalle.

15 La Figura 1 ilustra un diagrama de flujo para representar una posible puesta en práctica de un método para la detección adaptativa de una actividad vocal en una señal de audio de entrada según un primer aspecto de la presente invención;

20 La Figura 2 ilustra un diagrama de flujo de un aparato de VAD para detectar una actividad vocal en una señal de audio de entrada según un segundo aspecto de la presente invención,

La Figura 3 ilustra un diagrama de flujo de un dispositivo de procesamiento de señales de audio según un tercer aspecto de la presente invención.

25 DESCRIPCIÓN DETALLADA DE LAS FORMAS DE REALIZACIÓN

La Figura 1 representa un diagrama de flujo de una posible puesta en práctica de un método para la detección adaptativa de una actividad vocal en una señal de audio de entrada según un primer aspecto de la presente invención. En una primera etapa S1 de la puesta en práctica, a modo de ejemplo, del primer aspecto de la presente invención, una característica de ruido nc de la señal de audio de entrada se determina sobre la base de al menos una trama de entrada recibida de la señal de audio de entrada. La señal de audio de entrada comprende tramas de señales. En una posible puesta en práctica, la señal de entrada está segmentada en tramas de longitud predeterminada de p.e., 20 ms e introducidas trama por trama. En otras formas de realización, la longitud de las tramas de entrada puede ser variable. La característica de ruido nc de la señal de audio de entrada, que se determina en la etapa S1, puede ser una relación SNR a largo plazo, Isnr, calculada por una unidad de estimación de SNR a largo plazo. En otra posible forma de realización, la característica de ruido nc determinada en la etapa S1 se forma por una variación del ruido de fondo calculada por una unidad de estimación de variación del ruido de fondo, que calcula un carácter estacionario o fluctuación ε del ruido de fondo bn de la señal de audio de entrada. Además, es posible que la característica de ruido nc determinada en la etapa S1 esté constituida por una relación SNR a largo plazo, Isnr, y una variación del ruido de fondo.

En una etapa S2 posterior, la trama de entrada recibida de la señal de audio de entrada se divide en varias sub-bandas de frecuencia.

45 En una etapa adicional S3, se calcula un parámetro específico de sub-banda sbsp sobre la base de la relación SNR de la sub-banda para cada sub-banda que utiliza una función adaptativa AF. En una posible forma de realización, se obtiene un espectro de potencia para cada trama de entrada por intermedio de una Transformada de Fourier Rápida (FFT) y el espectro de potencia obtenido se divide en un número predeterminado de sub-bandas con anchura no lineal. Las energías para cada sub-banda se calculan en donde la energía para cada sub-banda de la trama de entrada puede formarse, en una posible forma de realización, por una energía atenuada que se forma mediante una media ponderada de las energías para la misma sub-banda entre la trama de entrada y al menos una trama anterior. En una posible puesta en práctica del primer aspecto de la presente invención, las relaciones SNRs de sub-bandas (snr) pueden calcularse como las relaciones SNRs logarítmicas modificadas de la sub-banda de frecuencia:

55
$$snr(i) = \log_{10} \left(\frac{E(i)}{E_n(i)} \right)$$

en donde E(i) de la energía de la i-ésima sub-banda de la trama de entrada y En(i) es la energía de la i-ésima sub-banda de la estimación del ruido de fondo. La estimación del ruido de fondo puede calcularse por la unidad de estimación del ruido de fondo, en donde la energía de cada sub-banda de la estimación de ruido de fondo se calcula mediante el desplazamiento-promediación de las energías de cada sub-banda entre las tramas de ruido de fondo detectadas. Lo que antecede puede expresarse como:

$$E_n(i) = \lambda \cdot E_n(i) + (1 - \lambda) \cdot E(i)$$

en donde $E(i)$ es la energía de la i -ésima sub-banda de la trama detectada como ruido de fondo, λ es un, así denominado, "factor de olvido" normalmente comprendido entre 0.9 y 0.99.

5 Después de haber obtenido una relación SNR (snr) para cada sub-banda de dicha trama de entrada en la etapa S3, se calcula un parámetro específico de sub-banda ($sbsp$) en la etapa S4 sobre la base de la respectiva relación SNR (snr) de la respectiva sub-banda utilizando una función adaptativa (AF). En una posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, al menos un parámetro de la función adaptativa (AF) se selecciona en función de la característica de ruido determinada de la
10 señal de audio de entrada. La característica de ruido nc , determinada en la etapa S1, puede comprender una relación SNR a largo plazo y/o una variación de ruido de fondo de la señal de audio de entrada. La función adaptativa AF es una función no lineal.

15 En una posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, en la etapa S5 se deriva una relación SNR segmentada modificada ($mssnr$) sumando los parámetros específicos de sub-banda ($sbsp$) como sigue:

$$mssnr = \sum_{i=1}^N sbsp(i)$$

20 en donde N es el número de sub-bandas de frecuencia en las que se divide dicha trama de entrada y

en donde $sbsp(i)$ es un parámetro específico de sub-banda calculado sobre la base de la relación SNR de sub-banda para cada sub-banda que utiliza la función adaptativa (AF). En una posible puesta en práctica del primer aspecto de la presente invención, la relación SNR segmentada modificada ($mssnr$) se calcula como sigue:

25

$$mssnr = \sum_{i=1}^N (f(snr(i)) + \alpha)^\beta$$

en donde $snr(i)$ es la relación SNR de la i -ésima sub-banda de la trama de entrada;

30 N es el número de sub-banda de frecuencia en la que se divide dicha trama de entrada y

AF = $(f(snr(i)) + \alpha)^\beta$ es la función adaptativa utilizada para calcular el parámetro específico de sub-banda $sbsp(i)$,

35 en donde α , β son dos variables configurables de la función adaptativa (AF).

En una posible puesta en práctica del primer aspecto de la presente invención, la primera variable α de la función adaptativa (AF) depende de una relación SNR a largo plazo ($lsnr$) de la señal de audio de entrada como sigue:

$$40 \quad \alpha = g(i, lsnr)$$

en donde g es una función lineal o no lineal y

en donde la segunda variable β de la función adaptativa (AF) depende de la SNR a largo plazo ($lsnr$) y de un valor φ :

$$45 \quad \beta = h(lsnr, \varphi)$$

en donde h es una función no lineal y

$$50 \quad \varphi = f(snr(i)) + \alpha.$$

En una posible puesta en práctica del método según el primer aspecto de la presente invención la primera variable α de la función adaptativa (AF) se calcula por:

$$55 \quad \alpha = g(i, lsnr) = a(i) \cdot lsnr + b(i)$$

en donde a (i), b (i) son números reales que dependen de un índice de sub-banda i y

la segunda variable β de la función adaptativa (AF) se calcula por:

$$\beta = h(lsnr, \varphi) = \begin{cases} \beta_1 & \varphi \geq d \text{ y } lsnr > e_2 \\ \beta_2 & \varphi \geq d \text{ y } e_1 < lsnr \leq e_2 \\ \beta_3 & \varphi \geq d \text{ y } lsnr \leq e_1 \\ \beta_4 & \text{de no ser así} \end{cases}$$

5

en donde $\beta_1 < \beta_2 < \beta_3$ y β_4 y d así como $e_1 < e_2$ son números enteros o flotantes y en donde lsnr es la SNR a largo plazo de la señal de audio de entrada.

10 En una posible puesta en práctica específica $\beta_1 = 4$, $\beta_2 = 10$, $\beta_3 = 15$ y $\beta_4 = 9$. En esta puesta en práctica específica d se establece en 1 y $e_1 = 8$ y $e_2 = 18$.

15 La relación SNR segmentada modificada (msnr) se deriva en la etapa S5 sumando los parámetros específicos de sub-banda (sbsp). En una etapa S6 posterior de la puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada según se ilustra en la Figura 1, la relación SNR segmentada modificada derivada (mssnr) se compara con un valor de umbral thr para proporcionar una decisión de VAD (VADD).

20 En una posible puesta en práctica, la relación SNR segmentada modificada derivada (mssnr) se compara un valor de umbral thr que se establece a:

20

$$thr = \begin{cases} k_1 & lsnr > e_2 \\ k_2 & e_1 < lsnr \leq e_2 \\ k_3 & lsnr \leq e_1 \end{cases}$$

en donde $k_1 > k_2 > k_3$ y $e_1 < e_2$ son números enteros o flotantes y en donde la decisión de VAD (VADD) se genera como sigue:

25

$$VADD = \begin{cases} 1 & mssnr > thr \\ 0 & mssnr \leq thr \end{cases}$$

en donde VADD = 1 representa una trama activa con la actividad vocal estando presente

30 y VADD = 0 representa una trama pasiva con la actividad vocal estando ausente.

En una posible puesta en práctica específica, $k_1 = 135$, $k_2 = 35$, $k_3 = 10$ y e_1 se establece en 8, mientras que e_2 se establece en 18.

35 En otra posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, la primera variable α de la función adaptativa (AF) se calcula por:

$$\alpha = g(i, lsnr, \varepsilon) = a(i) \cdot lsnr + b(i) + c(\varepsilon)$$

40 en donde a (i), b (i) son números reales que dependen de un índice de sub-banda i y

c (ε) es un número real que depende de la fluctuación estimada del ruido de fondo b_n de la señal de audio de entrada y

45 en donde la segunda variable β de la función adaptativa (AF) se calcula por:

$$\beta = h(\text{lsnr}, \varphi, \varepsilon) = \begin{cases} \beta_1 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \text{ y } \varepsilon \leq p \\ \beta_2 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \text{ y } \varepsilon > p \\ \beta_3 & \varphi \geq d \text{ y } e_1 < \text{lsnr} < e_2 \text{ y } \varepsilon \leq p \\ \beta_4 & \varphi \geq d \text{ y } e_1 < \text{lsnr} < e_2 \text{ y } \varepsilon > p \\ \beta_5 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \text{ y } \varepsilon \leq p \\ \beta_6 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \text{ y } \varepsilon > p \\ \beta_7 & \varphi < d \end{cases}$$

en donde $\varphi = f(\text{snr}(i)) + \alpha$ y ε es la fluctuación estimada del ruido de fondo b_n y

5 d y $e_1 < e_2$ y p son números enteros o flotantes.

En una puesta en práctica específica, los parámetros se establecen como sigue:

10 $\beta_1 = 3, \beta_2 = 4, \beta_3 = 7, \beta_4 = 10, \beta_5 = 8, \beta_6 = 15, \beta_7 = 15$ y

$d = 1$ y $e_1 = 8$ y $e_2 = 18$ y $p = 40$.

En una puesta en práctica del método de detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, la relación SNR segmentada modificada derivada (mssnr) se compara con un valor de umbral thr que se establece para:

$$thr = \begin{cases} q_1 + r_1 \cdot \text{Min} \left[\frac{\text{lsnr} - v_1}{W_1}, 1 \right] & \text{lsnr} > e_2 \\ q_2 + r_2 \cdot \text{Min} \left[\frac{\text{lsnr} - v_2}{W_2}, 1 \right] & e_1 < \text{lsnr} \leq e_2 \\ q_3 + r_3 \cdot \text{Max} \left[\text{Min} \left[\frac{\text{lsnr} - v_3}{W_3}, 1 \right], 1 \right] & \text{lsnr} \leq e_1 \end{cases}$$

en donde q_1, q_2, q_3 y r_1, r_2, r_3 y $e_1 < e_2$ y

20 v_1, v_2, v_3 y w_1, w_2, w_3 son números enteros o flotantes.

En una forma de realización específica del primer aspecto de la presente invención, $q_1 = 20, q_2 = 30, q_3 = 9$ y $r_1 = 30, r_2 = 10$ y $r_3 = 2$. Además, $v_1 = 18, v_2 = 8$ y $v_3 = 5$ y $w_1 = 8, w_2 = 10$ y $w_3 = 3$. Además, los parámetros e_1, e_2 se establecen a $e_1 = 8$ y $e_2 = 18$.

En consecuencia, en una posible forma de realización, no solamente una estimación del ruido de fondo y una estimación de la relación SNR a largo plazo se realizan, sino que además se realiza una estimación de variación del ruido de fondo para determinar una fluctuación del ruido de fondo ε del ruido de fondo de la señal de audio de entrada.

Dos factores, α, β de la función adaptativa (AF) ajustan una solución de compromiso de la potencia discriminadora del parámetro de SNR segmentada modificada. Diferentes soluciones de compromiso significan que la detección es más probable para la detección activa o inactiva para las tramas recibidas. En general, cuanto más alta es la relación SNR a largo plazo (lsnr) de la señal de audio de entrada tanto más favorable es ajustar la relación SNR segmentada modificada (mssnr) para detección activa por medio de ajustes de los correspondientes coeficientes α, β de la función adaptativa (AF).

La decisión de VAD, realizada en la etapa S6, puede pasar a través de un procedimiento de activación difícil. Un procedimiento de secuencia de corte fuerza las decisiones de VAD para varias tramas que sean activas inmediatamente después de la decisión de VAD obtenida en la etapa S6 cambie desde activa a inactiva.

En una posible puesta en práctica del método para la detección adaptativa de una actividad vocal en una señal de audio de entrada, según el primer aspecto de la presente invención, el ruido de fondo de la señal de audio de

entrada se analiza y se genera un número que representa la magnitud de la estacionalidad o fluctuación del ruido de fondo, que se indica por ε . Esta fluctuación ε del ruido de fondo bn puede calcularse, a modo de ejemplo, por:

$$\varepsilon = \omega \cdot \varepsilon + (1 - \omega) \cdot ssnr_n$$

5 en donde ω es el denominado factor de olvido que varía normalmente entre 0.9 y 0.99 y $ssnr_n$ es la suma de $snr(i)$ a través de todas las sub-bandas de la trama detectada como una trama de fondo multiplicada por un factor de p.e., 10.

10 La Figura 2 ilustra un diagrama de flujo de un aparato de VAD 1 según un segundo aspecto de la presente invención. El aparato de VAD 1 comprende una unidad de cálculo de parámetros VAD sobre la base de SNR 2 que recibe una señal de audio de entrada aplicada a una entrada 3 del aparato de VAD 1. La unidad de cálculo de parámetros 2 de VAD sobre la base de SNR calcula una relación SNR para cada sub-banda de una trama de entrada aplicada de la señal de audio de entrada y un parámetro específico de sub-banda (sbsp) basado en dicha respectiva relación SNR de sub-banda (snr) para cada sub-banda que utiliza una función adaptativa (AF) y deriva una relación SNR segmentada modificada (mssnr) sumando los parámetros específicos de sub-banda (sbsp). La relación SNR segmentada modificada derivada (mssnr) se aplica por la unidad de cálculo 2 de parámetros de VAD sobre la base de SNR para una unidad de generación de decisiones de VAD 4 del aparato de VAD 1. La unidad de generación de decisiones de VAD 4 genera una decisión de VAD (VADD) comparando la relación SNR segmentada modificada (mssnr) con un valor de umbral (thr). La decisión de VAD generada (VADD) se proporciona por el aparato de VAD 1 en una salida 5.

25 En una posible puesta en práctica del aparato de VAD 1, según el segundo aspecto de la presente invención, el aparato de detección de VAD 1 comprende, además, una unidad de determinación de la característica de ruido 6 según se ilustra en la Figura 2. La unidad de determinación de característica de ruido 6 determina una característica de ruido (nc) de la señal de entrada sobre la base de al menos una trama de entrada recibida de la señal de audio de entrada aplicada a la entrada 3 del aparato de VAD 1. En una puesta en práctica alternativa, la característica de ruido (nc) se aplica a la unidad de cálculo de parámetros de VAD sobre la base SNR 2 desde una entidad de determinación de la característica de ruido externa. En una posible puesta en práctica del aparato de VAD 1, según el segundo aspecto del aplicable, la unidad de determinación de la característica de ruido 6, según se ilustra en la Figura 2, puede comprender una unidad de estimación de la relación SNR a largo plazo que calcula una SNR a largo plazo (lsnr) de la señal de audio de entrada. En otra posible puesta en práctica de la invención, la unidad de determinación de la característica de ruido 6 puede comprender también una unidad de estimación de la variación del ruido de fondo que calcula una forma estacionaria o fluctuación ε del ruido de fondo bn de la señal de audio de entrada. En consecuencia, la característica de ruido (nc) proporcionada por la unidad de determinación de la característica de ruido 6 puede comprender una relación SNR a largo plazo (lsnr) de la señal de audio de entrada y/o una estacionalidad o fluctuación (ε) del ruido de fondo de la señal de audio de entrada. En una posible puesta en práctica, una función adaptativa (AF) utilizada por la unidad de cálculo 2 de parámetros de VAD sobre la base de SNR se selecciona en función de al menos una característica de ruido nc determinada por dicha unidad de determinación de la característica de ruido 6.

40 La Figura 3 ilustra un diagrama de flujo de un dispositivo de procesamiento de señales de audio 7 según un tercer aspecto de la presente invención, que comprende un aparato de VAD 1 que proporciona una decisión de VAD (VADD) para una unidad de procesamiento de señal de audio 8 dentro del dispositivo de procesamiento de señales de audio 7. El procesamiento de señales de audio de una señal de audio de entrada se realiza por la unidad de procesamiento de señales de audio 8 dependiendo de la decisión de VAD recibida (VADD) generada por el aparato de VAD 1 según el primer aspecto de la presente invención. La unidad de procesamiento de señales de audio 8 puede realizar, a modo de ejemplo, una codificación de la señal de audio de entrada sobre la base de la decisión de VAD (VADD). El dispositivo de procesamiento de señales de audio 7 puede formar parte del dispositivo de comunicación vocal, tal como un teléfono móvil. Además, el dispositivo de procesamiento de señal de audio 7 puede proporcionarse dentro de un sistema de comunicación vocal, tal como un sistema de conferencias de audio, un sistema de cancelación de señales de eco, un sistema de reducción de ruido vocal, un sistema de reconocimiento de voz o un sistema de codificación de voz. La decisión de VAD (VADD) generada por el aparato de VAD 1 puede controlar, en una posible puesta en práctica, un módulo DTX de transmisión discontinua de una entidad, a modo de ejemplo, una entidad en un sistema de radio celular, a modo de ejemplo, un sistema GSM o LTE o CDMA. El aparato de VAD 1 puede ampliar la capacidad del sistema de un sistema tal como un sistema de radio celular reduciendo las interferencias entre canales. Además, el consumo de energía de un dispositivo digital portátil, de un sistema de radio celular, se puede reducir en gran medida.

60

REIVINDICACIONES

1. Un método para la detección adaptativa de una actividad vocal en una señal de audio de entrada constituida por tramas, cuyo método comprende:

5 la determinación de una característica de ruido (nc) de la señal de audio de entrada sobre la base de al menos una trama de entrada recibida de la señal de audio de entrada;

10 la derivación de un parámetro VAD (vp) adaptado a la característica de ruido (nc) de la señal de audio de entrada y la comparación del parámetro VAD (vp) derivado con un valor de umbral (thr) para proporcionar una decisión de detección de actividad vocal (VADD); en donde el procesamiento de derivación de un parámetro VAD (vp) adaptado a la característica de ruido (nc) de la señal de audio de entrada comprende:

15 la división de la trama de entrada recibida de dicha señal de audio en varias sub-bandas; la obtención de una relación de señal a ruido (snr) para cada una de las varias sub-bandas de dicha trama de entrada recibida;

20 el cálculo de un parámetro específico de sub-banda (sbsp) sobre la base de la relación de señal a ruido (snr) de la sub-banda respectiva para cada una de las múltiples sub-bandas que utilizan una función adaptativa (AF), en donde al menos un parámetro de la función adaptativa (AF) se selecciona en función de la característica de ruido (nc) de la señal de audio de entrada y

25 la derivación de una relación de señal a ruido segmentada modificada (mssnr) en tanto como dicho parámetro de VAD (vp) añadiendo los parámetros específicos de sub-bandas (sbsp) calculados;

y en donde la relación de señal a ruido segmentada modificada (mssnr) se calcula como sigue:

$$mssnr = \sum_{i=1}^N (f(snr(i)) + \alpha)^\beta$$

30 en donde snr(i) es la relación de señal a ruido de la i-ésima sub-banda de la trama de entrada, N es el número de sub-banda de frecuencia en las que se divide dicha trama de entrada, $(f(snr(i)) + \alpha)^\beta$ es la función adaptativa (AF) utilizada para calcular el parámetro específico de sub-banda (sbsp(i)) y α, β son dos variables configurables de la función adaptativa (AF).

40 2. El método según la reivindicación 1, en donde la característica de ruido (nc) de la señal de audio de entrada es una relación de señal a ruido a largo plazo (lsnr) o una variación del ruido de fondo (ϵ) o bien, una relación de señal a ruido a largo plazo (lsnr) y una variación del ruido de fondo.

45 3. El método según la reivindicación 1, en donde la función adaptativa (AF) es una función no lineal.

50 4. El método según una de las reivindicaciones precedentes 1 o 3, en donde dicha relación de señal a ruido (snr) para cada una de las múltiples sub-bandas de dicha trama de entrada se obtiene mediante:

- 55 - obtención de una energía de señal para cada una de las múltiples sub-bandas;
- la estimación de una energía de ruido de fondo (bn) para cada una de las múltiples sub-bandas y
- el cálculo de la relación de señal a ruido (snr) para cada una de las sub-bandas en función de la energía de la señal y de la energía del ruido de fondo de la respectiva sub-banda.

60 5. El método según la reivindicación 4, en donde la energía de señal para cada sub-banda de dicha trama de entrada es una energía de señal atenuada que forma una media ponderada entre la trama de entrada y al menos una trama anterior.

6. El método según la reivindicación 1,

en donde la primera variable α de la función adaptativa (AF) depende de una relación de señal a ruido a largo plazo (lsnr) de la señal de audio de entrada como sigue:

5
$$\alpha = g(i, \text{lsnr})$$

en donde g es una función lineal o no lineal y

10 en donde la segunda variable β de la función adaptativa (AF) depende de la relación de señal a ruido a largo plazo (lsnr) y de φ :

$$\beta = h(\text{lsnr}, \varphi)$$

15 en donde h es una función no lineal y

$$\varphi = f(\text{snr}(i)) + \alpha$$

7. El método según la reivindicación 6,

20 en donde la primera variable α de la función adaptativa (AF) se calcula mediante:

$$\alpha = g(i, \text{lsnr}) = a(i) \cdot \text{lsnr} + b(i)$$

25 en donde a (i), b (i) son números reales que dependen de un índice de sub-banda i y la segunda variable β de la función adaptativa (AF) se calcula por:

$$\beta = h(\text{lsnr}, \varphi) = \begin{cases} \beta_1 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \\ \beta_2 & \varphi \geq d \text{ y } e_1 < \text{lsnr} \leq e_2 \\ \beta_3 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \\ \beta_4 & \text{de no ser así} \end{cases}$$

30 En donde $\beta_1 < \beta_2 < \beta_3$ y β_4 y d y $e_1 < e_2$ son números enteros o flotantes y en donde lsnr es la relación de señal a ruido a largo plazo (lsnr).

8. El método según la reivindicación 7,

35 en donde la relación de señal a ruido segmentada modificada derivada (mssnr) se compara con un valor de umbral (thr) que se establece para:

$$\text{thr} = \begin{cases} k_1 & \text{lsnr} > e_2 \\ k_2 & e_1 < \text{lsnr} \leq e_2 \\ k_3 & \text{lsnr} \leq e_1 \end{cases}$$

40 en donde $k_1 > k_2 > k_3$ y $e_1 < e_2$ son números enteros o flotantes y en donde la decisión de detección de actividad vocal (VADD) se genera como sigue:

$$VADD = \begin{cases} 1 & \text{mssnr} > \text{thr} \\ 0 & \text{mssnr} \leq \text{thr} \end{cases}$$

45 en donde VADD = 1 representa una trama activa con la actividad vocal estando presente y VADD = 0 representa una trama pasiva con la actividad vocal estando ausente.

9. El método según la reivindicación 6,

en donde la primera variable α de la función adaptativa (AF) se calcula mediante

$$\alpha = g(i, \text{lsnr}, \varepsilon) = a(i) \cdot \text{lsnr} + b(i) + c(\varepsilon)$$

en donde a (i), b (i) son números reales que dependen de un índice de sub-banda i y

c (ε) es un número real que depende de la fluctuación estimada del ruido de fondo de la señal de audio de entrada y

en donde la segunda variable β de la función adaptativa (AF) se calcula mediante:

$$\beta = h(\text{lsnr}, \varphi, \varepsilon) = \begin{cases} \beta_1 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \text{ y } \varepsilon \leq p \\ \beta_2 & \varphi \geq d \text{ y } \text{lsnr} > e_2 \text{ y } \varepsilon > p \\ \beta_3 & \varphi \geq d \text{ y } e_1 < \text{lsnr} < e_2 \text{ y } \varepsilon \leq p \\ \beta_4 & \varphi \geq d \text{ y } e_1 < \text{lsnr} < e_2 \text{ y } \varepsilon > p \\ \beta_5 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \text{ y } \varepsilon \leq p \\ \beta_6 & \varphi \geq d \text{ y } \text{lsnr} \leq e_1 \text{ y } \varepsilon > p \\ \beta_7 & \varphi < d \end{cases}$$

en donde φ = f(snr(i)) + α y ε es la fluctuación estimada del ruido de fondo y d y e₁ < e₂ y p son números enteros o flotantes.

10. El método según la reivindicación 9,

en donde la relación de señal de ruido segmentada modificada derivada (mssnr) se compara con un valor de umbral (thr) que se establece para:

$$\text{thr} = \begin{cases} q_1 + r_1 \cdot \text{Min} \left[\frac{\text{lsnr} - v_1}{W_1}, 1 \right] & \text{lsnr} > e_2 \\ q_2 + r_2 \cdot \text{Min} \left[\frac{\text{lsnr} - v_2}{W_2}, 1 \right] & e_1 < \text{lsnr} \leq e_2 \\ q_3 + r_3 \cdot \text{Max} \left[\text{Min} \frac{\text{lsnr} - v_3}{W_3}, 1 \right] & \text{lsnr} \leq e_1 \end{cases}$$

en donde q₁, q₂, q₃ y r₁, r₂, r₃ y e₁ < e₂ y

v₁, v₂, v₃ y w₁, w₂, w₃ son números enteros o flotantes,

en donde la decisión de detección de actividad vocal (VADD) se genera como sigue:

$$VADD = \begin{cases} 1 & \text{mssnr} > \text{thr} \\ 0 & \text{mssnr} \leq \text{thr} \end{cases}$$

en donde VADD = 1 representa una trama activa con la actividad vocal estando presente

y VADD = 0 representa una trama pasiva con la actividad vocal estando ausente.

11. Un aparato de detección de actividad vocal para detectar una actividad vocal en una señal de audio de entrada constituida por tramas,

en donde el aparato de detección de actividad vocal (VAD) comprende:

una unidad de cálculo de parámetro VAD sobre la base de la relación de señal a ruido, destinada a calcular una relación de señal a ruido (snr) para cada sub-banda de una trama de entrada aplicada y de un parámetro específico

de sub-banda (sbsp) basado en la respetiva relación de señal a ruido (snr) para cada sub-banda que utiliza una función adaptativa (AF) y para derivar en tanto como parámetro VAD (vp) una relación de señal a ruido segmentada modificada (mmsnr) añadiendo los parámetros específicos de sub-bandas (sbsp) y

5 una unidad generadora de decisión de detección de actividad vocal (VAD) destinada a generar una decisión de detección de actividad vocal (VADD) comparando la relación de señal a ruido segmentada modificada (mssnr) con un valor de umbral;

10 en donde dicho aparato comprende una unidad de determinación de característica de ruido destinada a determinar una característica de ruido (nc) de la señal de entrada sobre la base de al menos una trama de entrada recibida de la señal de audio de entrada;

15 en donde al menos un parámetro de la función adaptativa (AF) se selecciona en función de la característica de ruido (nc) de la señal de audio de entrada y

en donde la relación de señal a ruido segmentada modificada (mssnr) se calcula como sigue:

$$mssnr = \sum_{i=1}^N (f(snr(i)) + \alpha)^\beta$$

20 en donde snr(i) es la relación de señal a ruido de la i-ésima sub-banda de la trama de entrada,

N es el número de sub-bandas de frecuencias, en las que se divide dicha trama de entrada,

25 $(f(snr(i)) + \alpha)^\beta$ es la función adaptativa (AF) utilizada para calcular un parámetro específico de sub-banda sbsp(i) y α , β son dos variables configurables de la función adaptativa (AF).

12. El aparato de detección de actividad vocal según la reivindicación 11,

30 en donde dicha unidad de determinación de característica de ruido comprende una unidad de estimación de la relación de señal a ruido a largo plazo que calcula una relación de señal a ruido a largo plazo (lsnr) de dicha señal de audio de entrada.

13. El aparato de detección de actividad vocal según la reivindicación 11,

35 en donde dicha unidad de determinación de característica de ruido comprende una unidad de estimación de variación de ruido de fondo que calcula una forma estacionaria o una fluctuación (ϵ) del ruido de fondo de dicha señal de audio de entrada.

40 14. Un dispositivo de procesamiento de señal de audio que comprende un aparato de detección de actividad vocal según una de las reivindicaciones 11 a 13 precedentes y una unidad de procesamiento de señal de audio destinada a procesar una señal de entrada de audio en función de una decisión de detección de actividad vocal (VADD) proporcionada por el aparato de detección de actividad vocal.

45

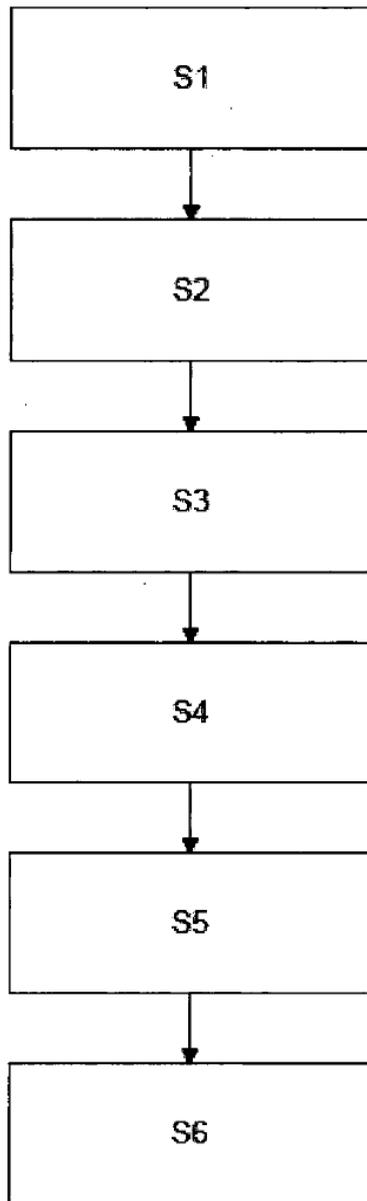


Fig. 1

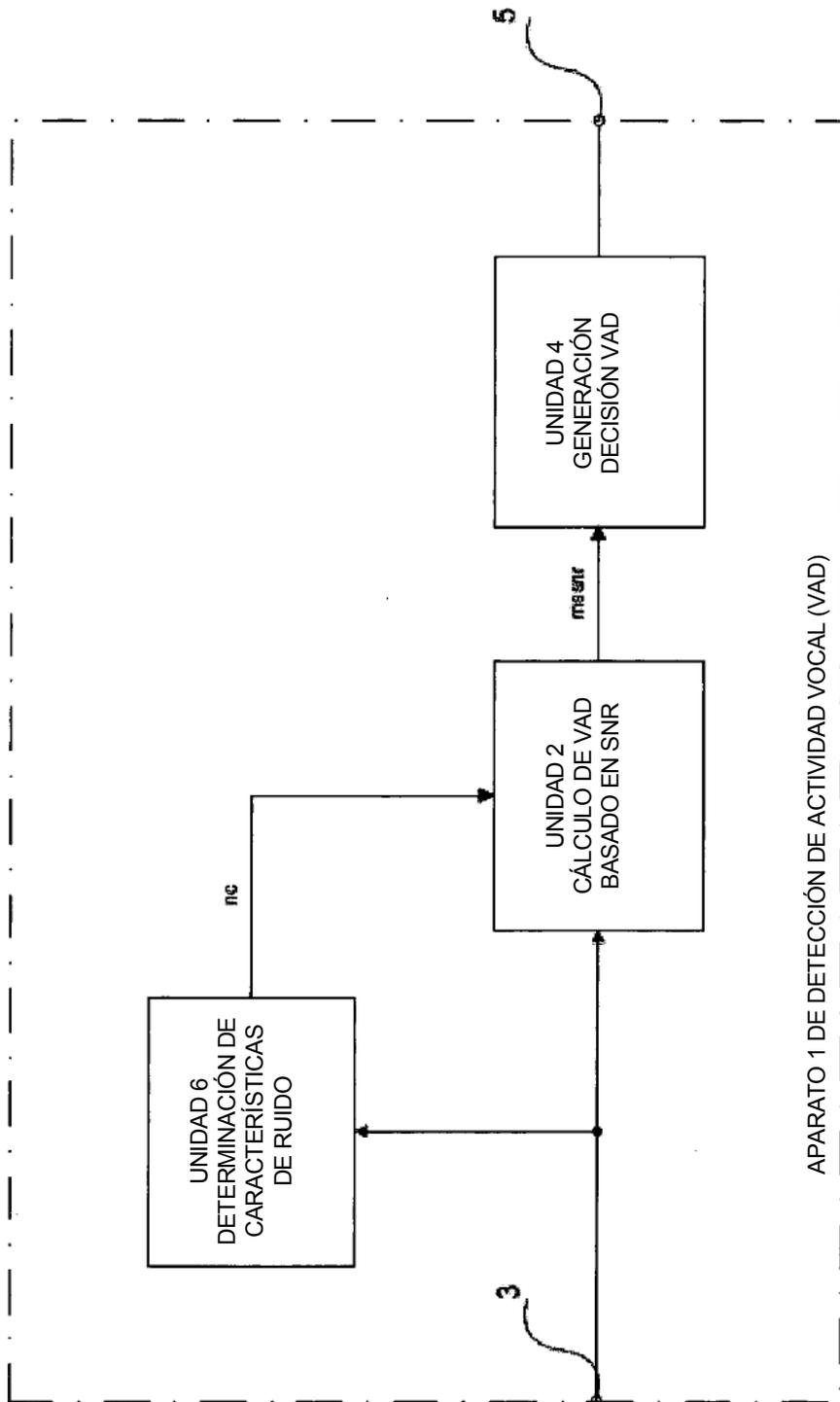


Fig. 2

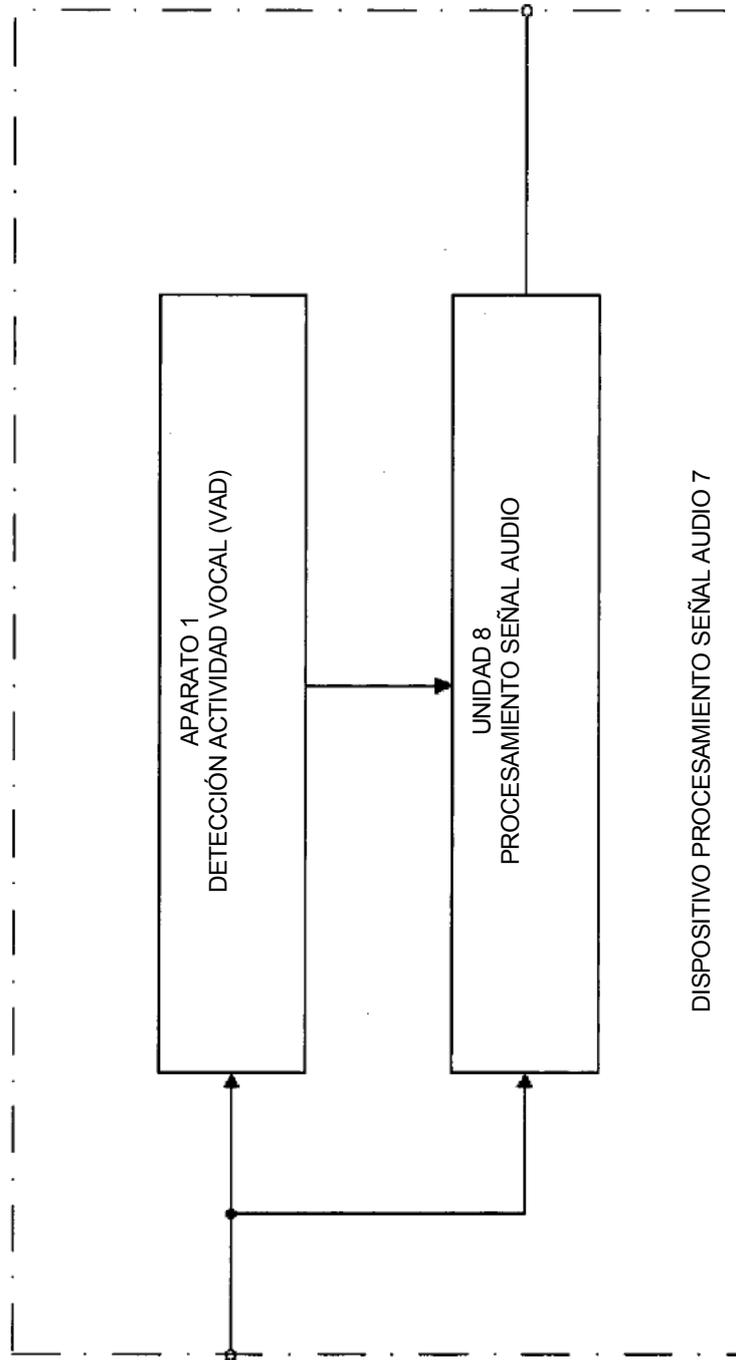


Fig. 3