

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 740 173**

51 Int. Cl.:

G10L 25/78 (2013.01)

G10L 25/93 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.12.2010** **E 17174901 (3)**

97 Fecha y número de publicación de la concesión europea: **01.05.2019** **EP 3252771**

54 Título: **Un método y un aparato para realizar una detección de actividad de voz**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.02.2020

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 2 740 173 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

Un método y un aparato para realizar una detección de actividad de voz

5 ANTECEDENTES

La invención se refiere a un método y un aparato para realizar una detección de actividad de voz y en particular, a un aparato de detección de actividad de voz que tiene al menos dos estados de funcionamiento diferentes, utilizando parámetros de relación de señal a ruido segmental de sub-banda procesada no linealmente.

10 La detección de actividad de voz (VAD) suele ser, en general, una técnica que se proporciona para detectar una actividad de voz en una señal. La detección de actividad de voz se conoce, además, como detección de actividad vocal o simplemente detección vocal. La función de detección VAD es la detección, en canales de comunicación, de la presencia de ausencia de señales activas tales como voz o música. Por lo tanto, las redes pueden decidir comprimir un ancho de banda de transmisión en períodos en donde señales activas están ausentes, o realizar otro procesamiento en función de si existe, o no, una señal activa. En la detección VAD, un parámetro de característica o un conjunto de parámetros de características, extraídos desde la señal de audio de entrada, se pueden comparar con los valores de umbral correspondientes con el fin de determinar si la señal de audio de entrada es una señal activa, o no lo es, sobre la base del resultado de comparación. Se han propuesto numerosos parámetros para la detección VAD. En general, parámetros basados en energía son conocidos por proporcionar un buen rendimiento. Por lo tanto, en los últimos años, los parámetros basados en SNR de sub-banda, como un tipo de parámetros basados en energía, se han utilizado ampliamente para la detección VAD. Sin importar qué parámetro de característica o parámetros de característica son utilizados por un detector de actividad de voz, estos parámetros presentan una característica de habla débil en los desplazamientos de ráfagas de voz, aumentando así la posibilidad de detectar, de forma errónea, los desplazamientos de voz. Habitualmente, con el fin de asegurar una detección correcta de los desplazamientos de voz, un detector de actividad de voz convencional realiza algún procesamiento especial en los desplazamientos de voz. Una forma convencional de poner en práctica este procesamiento especial es aplicar una resonancia "fuerte" a la decisión de VAD en desplazamientos de voz en donde el primer grupo de tramas detectadas como inactivas, por el detector de actividad de voz, en desplazamientos de voz, se fuerza para su activación. Otra posibilidad es aplicar una resonancia "suave" a la decisión de detección de actividad de voz en los desplazamientos de voz. Al aplicar una resonancia suave, el umbral de decisión de detección VAD, en los desplazamientos de voz, se ajusta para favorecer la detección de voz para las primeras varias tramas desplazadas de la señal de audio. En consecuencia, en este detector de actividad de voz convencional cuando la señal de entrada es una señal de desplazamiento no vocal, la decisión de VAD se realiza de forma normal mientras que en un estado de procesamiento la decisión de VAD se realiza de forma que favorezca la detección de voz.

Aunque la aplicación de un proceso de resonancia fuerte, con el fin de asegurar una detección correcta de los desplazamientos de voz, puede ayudar a disminuir la posibilidad de que, al producirse una detección errónea en los desplazamientos de voz, el sistema de resonancia fuerte carezca de eficiencia. Numerosas tramas inactivas reales se verán obligadas a activación de forma innecesaria, disminuyendo así el rendimiento general de la VAD. Por otro lado, aunque un sistema de procesamiento de resonancia suave, tal como el utilizado, a modo de ejemplo, por el detector de actividad de voz normalizado G.718 ITU-T, mejora la eficacia de la resonancia a un nivel más alto, el rendimiento de la VAD puede mejorarse aún más.

45 De conformidad con lo que antecede, un objetivo de la presente invención es dar a conocer un método y un aparato para la detección VAD que proporcionen un más alto rendimiento de VAD que los aparatos y métodos de VAD convencionales.

50 El documento EP2159788 A1 da a conocer un dispositivo de detección de actividad de voz (VAD). El dispositivo 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 evaluación de VAD de entrada, obtener parámetros relacionados con la variación del ruido de fondo y generar estos parámetros; una unidad de ajuste del umbral de VAD, adaptada para: obtener un sesgo del umbral de VAD de acuerdo con los parámetros emitidos por la unidad de análisis de fondo, y generar el sesgo del umbral de VAD; y una unidad de evaluación de VAD, adaptada para: modificar un umbral de VAD para ser modificado de acuerdo con el sesgo de la salida de umbral de VAD por la unidad de ajuste de umbral de VAD, evaluar el ruido de fondo utilizando el umbral de VAD modificado y emitir un resultado de evaluación de VAD. El documento US2008/077400 A1 da a conocer que un detector de duración de la voz incluye una unidad de detección de final-iniciación, que detecta un final de iniciación de una primera duración en donde la característica supera un valor umbral tal como un final de iniciación de una duración de voz, cuando la primera duración continúa para una primera magnitud temporal; una unidad de detección de puesta en cola-final-candidato, que detecta un final de iniciación de una segunda duración, en donde la característica es inferior que el valor umbral como un punto candidato para un final de puesta en cola de la voz, cuando la segunda duración continúa para una segunda magnitud temporal; y una unidad de determinación de puesta en cola-final-candidato, que determina el punto candidato como un final de puesta en cola de la duración de la voz, cuando la segunda duración, en donde la característica supera el valor umbral no continúa para la primera magnitud temporal, mientras transcurre una tercera magnitud temporal desde la medición en el punto candidato.

SUMARIO

Es un objeto de la invención superar las deficiencias en la técnica anterior. Este objeto de la invención se resuelve mediante las reivindicaciones independientes. Formas de realización específicas se definen en las reivindicaciones dependientes.

5 De conformidad con un primer aspecto de la presente invención, se da a conocer un aparato de detección de actividad de voz (VAD) para la determinación de una decisión de VAD (VADD) para una señal de audio de entrada, en donde el aparato de VAD comprende:

10 un detector de estado, adaptado para determinar un estado de funcionamiento actual (WS) de entre al menos dos estados de funcionamiento diferentes del aparato de VAD, que depende de la señal de audio de entrada,

15 en donde cada uno de los al menos dos estados de funcionamiento (WS) diferentes, está asociado con un conjunto de decisión de parámetro de estado de funcionamiento correspondiente (WSPDS) que incluye, al menos, un parámetro de VAD (VADP); y

20 un dispositivo de cálculo de actividad de voz, adaptado para calcular un valor de parámetro VAD, para el parámetro VAD (VADP) del conjunto de decisión de parámetro de estado de funcionamiento (WSPDS) asociado con el estado de funcionamiento actual (WS) y para determinar la decisión VAD (VADD) mediante la comparación del valor del parámetro VAD calculado con un valor umbral.

25 En consecuencia, el aparato de VAD, de conformidad con el primer aspecto de la presente invención, comprende más de un estado de funcionamiento (WS). El aparato de VAD, de conformidad con el primer aspecto de la presente invención utiliza al menos dos parámetros diferentes o dos conjuntos de parámetros diferentes, para tomar decisiones de VAD para diferentes estados de funcionamiento.

30 En una manera de puesta en práctica posible, los parámetros VAD pueden tener la misma forma general, pero pueden comprender diferentes factores. En una forma de realización posible, los diferentes parámetros de VAD pueden comprender parámetros basados en la relación de señal a ruido segmental (SNR) de sub-banda modificada que se procesan de forma no lineal, en una manera diferente.

35 El número de estados de funcionamiento utilizados por el aparato VAD de conformidad con el primer aspecto de la presente invención puede variar. En una posible puesta en práctica del aparato VAD, el aparato comprende dos estados de funcionamiento diferentes, es decir, un estado de funcionamiento normal (NWS) y un estado de funcionamiento de compensación (OWS).

40 En una manera de puesta en práctica posible del aparato VAD de conformidad con el primer aspecto de la presente invención, para cada estado de funcionamiento (WS) del aparato de VAD, se proporciona un conjunto de decisión de parámetro de estado de funcionamiento correspondiente (WSPDS), que comprende cada uno al menos un parámetro VAD (VADP). El número y tipo de parámetros de VAD (VADPs) puede variar para los diferentes conjuntos de decisión de parámetro de estado de funcionamiento (WSPDS), de los diferentes estados de funcionamiento (WS), del aparato de VAD de conformidad con el primer aspecto de la presente invención.

45 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, la decisión de VAD (VADD) determinada por dicho dispositivo de cálculo de actividad de voz, se determina o calcula utilizando una relación señal-ruido segmental (SNR) de sub-banda sobre la base de parámetros VAD (VADPs).

50 En una manera de puesta en práctica posible del aparato VAD, de conformidad con el primer aspecto de la presente invención, la decisión VAD (VADD) para dicha señal de audio de entrada se determina mediante dicho dispositivo de cálculo de actividad de voz sobre la base de al menos un parámetro VAD (VADP) del conjunto de decisión de parámetro de funcionamiento (WSPDS), que se proporciona para el estado de funcionamiento actual (WS) de dicho aparato de VAD, utilizando un algoritmo de procesamiento de VAD predeterminado, que se proporciona para el estado de funcionamiento actual (WS) de dicho aparato de VAD. El algoritmo de procesamiento de VAD utilizado se puede reconfigurar o configurar a través de una interfaz, con lo que se proporciona más flexibilidad para el aparato de VAD de conformidad con el primer aspecto de la presente invención.

60 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con la presente invención, se puede adaptar el algoritmo de procesamiento de VAD utilizado para determinar la decisión de VAD (VADD).

65 En una manera adicional de puesta en práctica posible del aparato VAD, de conformidad con el primer aspecto de la presente invención, el aparato de VAD es conmutable entre diferentes estados de funcionamiento (WS), en función de las condiciones de transición de estado de funcionamiento configurables. Esta conmutación se puede realizar en una manera de puesta en práctica posible bajo el control del detector de estado.

En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente

invención, el aparato de VAD comprende un estado de funcionamiento normal (NWS) y un estado de funcionamiento de desplazamiento (OWS) y se puede conmutar entre estos dos estados de funcionamiento diferentes, en función de condiciones de transición de estado de funcionamiento configurables.

5 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, el aparato de VAD detecta un cambio desde la actividad de la voz presente a una actividad de la voz que está ausente y/o se conmuta desde un estado de funcionamiento normal (NWS) a un estado de funcionamiento de desplazamiento (OWS) en dicha señal de audio de entrada, si en el estado de funcionamiento normal (NWS) de dicho aparato de VAD, la decisión de VAD (VADD) se determina sobre la base del al menos un parámetro VAD (VADP) del conjunto de decisión de parámetro de estado de funcionamiento normal (NWSPDS), de dicho estado de funcionamiento normal (NWS), indica que está presente una actividad de voz para una trama anterior y una actividad de voz está ausente en una trama actual de dicha señal de audio de entrada. En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, la decisión VADD que el aparato VAD detecta en su estado de funcionamiento normal (NWS) forma una decisión VADD intermedia (VADD_{int}), que puede formar la VADD o la salida de VADD final por el aparato VAD, en caso de que esta VAD intermedia indique que la actividad de voz está presente en la trama actual. Tal como se describió anteriormente, en caso de que esta VADD intermedia indique que no hay actividad de voz en la trama actual, esta VADD intermedia se puede utilizar para detectar una transición, o cambio, desde un estado de funcionamiento normal a un estado de funcionamiento de desplazamiento, y para conmutar al estado de funcionamiento de desplazamiento en donde el detector de actividad de voz calcula, para la trama actual, un parámetro de detección de actividad de voz del conjunto de decisión de parámetro de estado de funcionamiento de desplazamiento, para determinar la VADD o salida de VADD final, por el aparato de VAD.

25 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, si dicho aparato de VAD detecta, en su estado de funcionamiento normal (NWS), que está presente una actividad de voz en una trama actual de dicha señal de audio de entrada, esta decisión VAD intermedia (VADD_{int}) se proporciona, a la salida, como una decisión de VAD final (VADD_{fin}).

30 En una manera posible de puesta en práctica adicional del aparato de VAD, de conformidad con el primer aspecto de la presente invención, en donde si dicho aparato de VAD detecta, en su estado de funcionamiento normal (NWS), que existe una actividad de voz presente en la trama anterior y que una actividad de voz está ausente en una trama actual de dicha señal de entrada, se cambia desde su estado de funcionamiento normal (NWS) a un estado de funcionamiento de desplazamiento (OWS), en el que la decisión de VAD (VADD) se determina sobre la base del al menos un parámetro VAD del conjunto de decisión de parámetro de estado de funcionamiento de desplazamiento (OWSPDS).

40 En una manera adicional de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, la decisión VAD (VADD), determinada en el estado de funcionamiento de desplazamiento (OWS), de dicho aparato de VAD forma la decisión VADD final o la decisión de VAD (VADD) que se proporciona, a la salida, por el aparato de VAD si la decisión VAD (VADD), determinada sobre la base del al menos un parámetro VAD (VADP), del conjunto de decisión de parámetro de estado de funcionamiento de desplazamiento (OWSPDS) indica que existe una actividad de voz presente en la trama actual la señal de audio de entrada.

45 En otra posible manera de puesta en práctica adicional del aparato de VAD, de conformidad con el primer aspecto de la presente invención, la decisión de VAD (VADD), que se determina en el estado de funcionamiento de desplazamiento (OWS) de dicho aparato de VAD, forma una decisión VAD intermedia (VAD_{int}) si la decisión VAD (VADD), determinada sobre la base del al menos un parámetro VAD (VADP), del conjunto de decisión de parámetro de estado de funcionamiento de desplazamiento (OWSPDS), indica que una actividad de voz está ausente en la trama actual de la señal de audio de entrada.

50 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, la decisión VAD intermedia (VADD_{int}) se somete a un proceso de resonancia fuerte para proporcionar una decisión VAD final (VADD_{fin}).

55 En una manera adicional de puesta en práctica posible de aparato de VAD, de conformidad con el primer aspecto de la presente invención, el aparato de VAD se conmuta desde el estado de funcionamiento normal (NWS) al estado de funcionamiento de desplazamiento (OWS) si la decisión de VAD (VADD) determinada por el dispositivo de cálculo de actividad de voz del aparato de VAD, en el estado de funcionamiento normal (NWS) utilizando un algoritmo de procesamiento VAD y el conjunto de decisión de parámetro de estado de funcionamiento (NWSPDS), que se proporciona para dicho estado de funcionamiento normal (NWS), indica una ausencia de voz en la señal de audio de entrada y, dicho un contador de resonancia suave (SHC) supera un valor umbral de contador predeterminado.

65 En una manera adicional de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, el aparato de VAD se conmuta desde el estado de funcionamiento de desplazamiento (OWS) al estado de funcionamiento normal (NWS) si el contador de resonancia suave (SHC) no supera un valor umbral de contador predeterminado.

5 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, la señal de audio de entrada consiste en una secuencia de tramas de señal de audio y el contador de resonancia suave (SHC) se disminuye en el estado de funcionamiento de desplazamiento (OWS), de dicho aparato de VAD, para cada trama de señal de audio recibida hasta que se alcanza el valor de contador de umbral predeterminado.

10 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, si se detecta un número predeterminado de tramas de señal de audio activa consecutivas de la señal de audio de entrada, el contador de resonancia suave (SHC) se restablece a un valor de contador que depende de una relación de señal/ruido a largo plazo (ISNR) de la señal de audio de entrada.

15 En una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, se detecta una trama de señal de audio activa si una métrica de voz calculada, de la señal de audio, supera un valor umbral de métrica de voz predeterminado y una estabilidad de tono, de dicha trama de señal de audio, está por debajo de un valor umbral de estabilidad predeterminado.

20 En una manera de puesta en práctica del aparato de VAD, de conformidad con el primer aspecto de la presente invención, los parámetros VAD de un conjunto de decisión de parámetro de estado de funcionamiento (WSPDS), de un estado de funcionamiento de dicho aparato de detección de actividad, comprenden parámetros de decisión basados en energía y/o parámetros basados en envolvente espectral y/o parámetros de decisión basados en entropía y/o parámetros de decisión basados en estadística.

25 En una manera adicional de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención, una decisión VAD intermedia ($VADD_{int}$), que se determina por dicho dispositivo de cálculo de actividad de voz, de dicho aparato de VAD, se aplica a una unidad de procesamiento de resonancia fuerte que realiza una resonancia fuerte de dicha decisión de VAD intermedia aplicada ($VADD_{int}$).

30 Según un segundo aspecto de la presente invención, se da a conocer un dispositivo de procesamiento de señal de audio, que comprende un aparato de VAD, de conformidad con el primer aspecto de la presente invención, y que comprende una unidad de procesamiento de señal de audio controlada por una decisión VAD ($VADD$) generada por dicho aparato de VAD.

35 De conformidad con un tercer aspecto de la presente invención, se proporciona un método para realizar una detección VAD, en donde una decisión de VAD ($VADD$) se calcula mediante un aparato de VAD para una señal de audio de entrada utilizando al menos un parámetro de VAD ($VADP$) de un conjunto de decisión de parámetro de estado de funcionamiento (WSPDS), de un estado de funcionamiento actual, que se detecta por un detector de estado de dicho aparato de VAD.

40 **BREVE DESCRIPCIÓN DE LOS DIBUJOS**

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

45 La Figura 1 ilustra un diagrama de bloques de un aparato de VAD de conformidad con una manera de puesta en práctica posible del aparato de VAD, de conformidad con el primer aspecto de la presente invención.

50 La Figura 2 ilustra un diagrama de bloques de una manera de puesta en práctica posible de un aparato de procesamiento de señal de audio, de conformidad con un segundo aspecto de la presente invención.

DESCRIPCIÓN DE LAS FORMAS DE REALIZACIÓN

55 La Figura 1 ilustra un diagrama de bloques de una manera de puesta en práctica posible de un aparato de VAD 1, de conformidad con un primer aspecto de la presente invención. Tal como se puede ver en la Figura 1, el aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, comprende en la puesta en práctica a modo de ejemplo, un detector de estado 2 y un dispositivo de cálculo de actividad de voz 3. El aparato de VAD 1 se proporciona para determinar una decisión de VAD, $VADD$, para una señal de audio de entrada recibida aplicada a una entrada 4 del aparato de VAD 1. La decisión de VAD determinada $VADD$ se envía a una salida 5 del aparato de VAD 1. El detector de estado 2 está adaptado para determinar un estado de funcionamiento actual WS, del aparato de VAD 1, que depende de la señal de audio de entrada aplicada a la entrada 4. El aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, comprende al menos dos estados de funcionamiento diferentes WS. En una manera de puesta en práctica posible, el aparato de VAD 1 comprende, a modo de ejemplo, dos estados de funcionamiento WS. Cada uno de los al menos dos estados de funcionamiento WS diferentes, está asociado con un correspondiente conjunto de decisión de parámetro de estado de funcionamiento WSPDS, que incluye al menos un parámetro de VAD, $VADP$.

El aparato de VAD 1 comprende, en la puesta en práctica ilustrada de la Figura 1, además, un dispositivo de cálculo de actividad de voz 3 que está adaptado para calcular un valor de parámetro VAD para el al menos un parámetro VAD, VADP, del conjunto de decisión de parámetro de estado de funcionamiento WSPDS, asociado con el estado de funcionamiento actual WS del aparato de VAD 1. Este cálculo se realiza con el fin de determinar una decisión VAD, VADD, mediante la comparación del valor del parámetro VAD calculado del al menos un parámetro VAD, con un valor umbral correspondiente.

El detector de estado 2, así como el dispositivo de cálculo de actividad de voz 3, del aparato de VAD 1, pueden ponerse en práctica mediante hardware o software. El aparato de VAD 1, según el primer aspecto de la presente invención, tiene más de un estado de funcionamiento. El aparato de VAD 1 utiliza al menos dos parámetros VAD diferentes, o dos conjuntos diferentes de parámetros VAD, para generar la decisión VAD, VADD, para diferentes estados de funcionamiento WS.

La decisión de VAD, VADD, determinada para dicha señal de audio de entrada, por dicho dispositivo de cálculo de actividad de voz 3, se determina en una manera de puesta en práctica posible, sobre la base del al menos un parámetro de VAD, VADP, del conjunto de decisión de parámetro de estado de funcionamiento WSPDS, que se proporciona para el estado de funcionamiento actual WS del aparato de VAD 1, utilizando un algoritmo de procesamiento de VAD predeterminado, que se proporciona para el estado de funcionamiento actual WS del aparato VAD 1. El detector de estado 2 detecta el estado de funcionamiento actual WS del aparato de VAD 1. Se realiza la determinación del estado de funcionamiento actual WS por el detector de estado 2, que depende de la señal de audio de entrada recibida. En una manera de puesta en práctica posible, el aparato de VAD 1 es conmutable entre diferentes estados de funcionamiento WS de conformidad con condiciones de transición de estado de funcionamiento configurables. En una manera de puesta en práctica posible, el aparato de VAD 1 comprende dos estados de funcionamiento, es decir, un estado de funcionamiento normal NWS y un estado de funcionamiento de desplazamiento OWS.

En una manera de puesta en práctica posible del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, el aparato de VAD 1 detecta un cambio desde una actividad de voz presente a una actividad de voz ausente en la señal de audio de entrada si se cumple una condición correspondiente. Si, en el estado de funcionamiento normal NWS de dicho aparato de VAD 1, la decisión de VAD, VADD, determinada por el dispositivo de cálculo de actividad de voz 3 de dicho aparato de VAD 1, sobre la base de al menos un parámetro VAD, VADP, del conjunto de decisión de parámetro de estado de funcionamiento normal NWSPDS, de dicho estado de funcionamiento normal NWS, indica que está presente una actividad de voz para una trama previa y falta actividad de voz en una trama actual de dicha señal de audio de entrada, el aparato de VAD 1 detecta un cambio desde la actividad de voz presente en la señal de audio de entrada, a una actividad de voz ausente en la señal de audio de entrada.

En una manera de puesta en práctica posible del aparato de VAD 1, de conformidad con el primer aspecto, si el aparato de VAD 1 detecta, en su estado de funcionamiento normal NWS, que una actividad de voz está presente en una trama actual de la señal de audio de entrada, esta decisión VAD intermedia VADD_{int} se puede proporcionar, a la salida, como una decisión de VAD final VADD_{fin} en la salida 5 del aparato de VAD 1 para su posterior procesamiento.

En una manera adicional de puesta en práctica posible del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, si dicho aparato de VAD 1 detecta, en su estado de funcionamiento normal NWS, que una actividad de voz está presente en la trama previa de la señal de audio de entrada, y que una actividad de voz está ausente en una trama actual de la señal de audio de entrada, se conmuta automáticamente desde su estado de funcionamiento normal NWS, a un estado de funcionamiento de desplazamiento OWS. En el estado de funcionamiento de desplazamiento OWS, la decisión de VAD, VADD, se determina por el dispositivo de cálculo de actividad de voz 3 sobre la base del al menos un parámetro de VAD, VADP, del conjunto de decisión de parámetro de estado de funcionamiento de desplazamiento OWSPDS. Los parámetros de VAD, VADPs, de los diferentes conjuntos de decisión de parámetro de estado de funcionamiento WSPDS se pueden memorizar, en una manera de puesta en práctica posible, en una memoria de configuración del aparato de VAD 1.

En una manera de puesta en práctica posible del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, la decisión VAD, VADD, que se determina por el dispositivo de cálculo de actividad de voz 3, en el estado de funcionamiento de desplazamiento OWS, forma una decisión intermedia VAD, VADD_{int} si la decisión VAD, VADD, determinada sobre la base del al menos un parámetro de VAD, VADP, del conjunto de decisión de parámetro de estado de desplazamiento OWSPDS, indica que una actividad de voz está ausente en la trama actual de la señal de audio de entrada. En una manera de puesta en práctica posible, esta decisión de VAD intermedia generada se somete a un proceso de resonancia fuerte antes de que se emita como una decisión de VAD final VADD_{fin} en la salida 5 del aparato de VAD 1.

En una manera de puesta en práctica posible del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, el aparato de VAD 1 se conmuta, automáticamente, desde el estado de funcionamiento normal NWS al estado de funcionamiento de desplazamiento OWS si la decisión de VAD, VADD, determinada por el dispositivo de cálculo de actividad de voz 3 del aparato de VAD 1, en el estado de funcionamiento normal NWS utilizando un algoritmo de procesamiento VAD y el conjunto de decisión de parámetro de estado de funcionamiento WSPDS, que se proporciona para este estado de funcionamiento normal NWS, indica una ausencia de voz en la señal

de audio de entrada y, si un contador de resonancia suave SHC supera, al mismo tiempo, un valor umbral de contador predeterminado.

5 En una manera adicional de puesta en práctica posible del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, el aparato de VAD 1 se conmuta desde el estado de funcionamiento de desplazamiento OWS al estado de funcionamiento normal NWS si un contador de resonancia suave SHC no supera, al mismo tiempo, un valor umbral de contador predeterminado.

10 La señal de audio de entrada aplicada a la entrada 4 del aparato de VAD 1 consiste en una manera de puesta en práctica posible de una secuencia de tramas de señal de audio, en donde el contador de resonancia suave SHC, utilizado por el aparato de VAD 1, es objeto de disminución en el estado de funcionamiento de desplazamiento OWS de dicho aparato de VAD 1, para cada trama de señal de audio recibida, hasta que se alcanza el valor umbral de contador predeterminado. En una manera de puesta en práctica posible, si se detecta un número predeterminado de señales de audio activas consecutivas de la señal de audio de entrada, el contador de resbalamiento suave SHC se restablece a un valor de contador dependiendo de una relación señal a ruido a largo plazo (ISNR) de la señal de audio de entrada recibida. Esta relación señal a ruido a largo plazo (ISNR) se puede calcular mediante una unidad de estimación de relación señal a ruido a largo plazo del aparato de VAD 1. En una manera de puesta en práctica posible del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, se detecta una trama de señal de audio activa si una métrica de voz calculada de la trama de señal de audio supera un valor umbral de métrica de voz predeterminado y una estabilidad de tono de la trama de señal de audio está por debajo de un valor umbral de estabilidad predeterminado.

25 En una manera de puesta en práctica posible del aparato de VAD 1, según el primer aspecto de la presente invención, los parámetros VAD, VADP, de un conjunto de decisión de parámetro de estado de funcionamiento WSPDS, de un estado de funcionamiento WS del aparato de VAD 1, pueden comprender parámetros de decisión basados en energía y/o parámetros de decisión basados en envolvente espectral y/o parámetros de decisión basados en entropía y/o parámetros de decisión basados en estadística. En una manera de puesta en práctica específica del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, la decisión VAD, VADD, determinada por el dispositivo de cálculo de actividad de voz 3, utiliza parámetros de VAD, VADPs, basados en la relación de señal a ruido segmental (SNR) de sub-banda.

35 En una manera adicional de puesta en práctica posible del aparato de VAD 1, se puede aplicar una decisión de VAD intermedia, VADD, determinada por el dispositivo de cálculo de actividad de voz 3 del aparato de VAD 1, a otra unidad de procesamiento de resonancia fuerte, que realiza una resonancia fuerte de la decisión VAD intermedia aplicada, VADD.

40 El aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, puede comprender en una manera de puesta en práctica posible, dos estados de funcionamiento, en los que el aparato de VAD 1 opera en un estado de funcionamiento normal NWS o en un estado de funcionamiento de desplazamiento OWS. Un desplazamiento de voz es un período corto al final de la ráfaga de voz dentro de la señal de audio recibida. De este modo, un desplazamiento de voz contiene energía vocal relativamente baja. Una ráfaga de voz es un período vocal de la señal de audio de entrada entre dos pausas de voz adyacentes. La longitud de un desplazamiento de voz se suele extender sobre varias tramas de señal continuas y puede ser dependiente de la muestra. El aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, identifica de forma continua, los inicios de los desplazamientos de voz en la señal de audio de entrada y cambia del estado de funcionamiento normal NWS al estado de funcionamiento de desplazamiento OWS cuando se detecta un desplazamiento de voz, y cambia, de nuevo, al estado de funcionamiento normal NWS cuando finaliza el estado de desplazamiento de voz. El aparato de VAD 1 selecciona un parámetro de VAD, o un conjunto de parámetros para el estado de funcionamiento normal NWS y otro parámetro de VAD, o conjunto de parámetros, para el estado de funcionamiento de desplazamiento OWS. Por consiguiente, con un aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, se realizan diferentes operaciones VAD para diferentes partes de la señal de audio recibida y se realizan operaciones específicas de VAD para cada estado de funcionamiento WS. El aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, realiza una ráfaga de voz y detección de desplazamiento, en la señal de entrada de audio recibida, en donde la detección de desplazamiento se puede realizar de diferentes maneras según diferentes puestas en práctica del aparato de VAD 1.

55 En una manera de puesta en práctica posible del aparato de VAD 1, la señal de audio de entrada se segmenta en tramas de señal y se introduce en el aparato de VAD 1 en la entrada 4. La señal de audio de entrada puede comprender, a modo de ejemplo, tramas de señal de 20 ms de longitud. En una manera de puesta en práctica posible específica, para cada trama de señal de entrada se puede realizar un análisis de tono de bucle abierto dos veces, cada una para una sub-trama que tenga 10 ms. Los retardos de tono buscados para las dos sub-tramas de cada trama de entrada se indican como T(0), T(1), respectivamente, y las correlaciones correspondientes se indican, respectivamente, como sonorización(0) y sonorización(1). La métrica de sonorización(V) de la trama de señal de audio V(0) se calcula mediante:

$$V(0) = (\text{sonorización}(-1) + \text{sonorización}(0) + \text{sonorización}(1))/3 + \text{corr_shift}$$

en donde la sonorización(-1) representa la correlación correspondiente como un retardo de tono de la segunda sub-trama, de la trama de señal de entrada anterior, y en donde corr_shift es un valor de compensación que depende del nivel de ruido de fondo.

5

La estabilidad de tono(S) de dicha trama de señal de audio se puede calcular mediante:

$$Sr(0)=[abs(T(-1)-T(-2)) + abs(T(0)-T(-1)) + abs(T(1)-T(0))]/3$$

10 en donde T(-1), T(-2) son el primer y segundo retardos de tono de la trama de señal de entrada anterior y abs() significa el valor absoluto. En una manera de puesta en práctica posible específica, la trama de entrada se considera como una trama de voz o trama activa cuando se cumple la siguiente condición:

$$V(0) > 0.65 \ \&\& \ S_T(0) < 14$$

15

En una manera de puesta en práctica posible, si se detectan tres tramas activas consecutivas, se detecta una ráfaga sonora de la señal de audio de entrada, y un contador de resonancia suave SHC se restablece a un valor distinto de cero, determinado en función de la señal SNR a largo plazo, lSNR. Cuando el aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, opera en un estado de funcionamiento normal NWS y la decisión de VAD intermedia determinada VADD cae después de que las tramas anteriores hayan sido clasificadas o determinadas como activa a inactiva para una trama de señal actual, y si el contador de resonancia suave SHC es mayor que 0, se supone que la señal de audio de entrada entra en un desplazamiento de voz y el aparato de VAD 1 pasa desde el estado de funcionamiento normal NWS al estado de funcionamiento de desplazamiento OWS. La longitud del contador de resonancia suave SHC define la longitud de la VAD del estado de funcionamiento de desplazamiento OWS. En una manera de puesta en práctica posible, el contador de resonancia suave SHC se reduce o disminuye en uno en cada trama de señal dentro de la detección VAD del estado de funcionamiento de desplazamiento de voz OWS. El estado de funcionamiento de desplazamiento de voz OWS, del aparato de VAD 1, finaliza cuando el contador de resonancia suave SHC disminuye a un valor umbral predeterminado tal como 0, y el aparato de VAD 1 cambia de nuevo a su estado de funcionamiento normal NWS al mismo tiempo.

20

En una manera de puesta en práctica posible específica, se utilizan tres parámetros por el aparato de VAD 1 para realizar una decisión de VAD intermedia VADD_{int}. Un parámetro es la métrica de sonorización(V-1) de la trama anterior y los otros dos parámetros vienen dados por:

25

$$mSSNR_{nur} = \begin{cases} \sum_i^N (snr(i) + \alpha)^4 & snr(i) + \alpha \geq 1, lsnr > 18 \\ \sum_i^N (snr(i) + \alpha)^{10} & snr(i) + \alpha \geq 1, 8 < lsnr \leq 18 \\ \sum_i^N (snr(i) + \alpha)^{15} & snr(i) + \alpha \geq 1, lsnr \leq 8 \\ \sum_i^N (snr(i) + \alpha)^9 & \text{de otro modo} \end{cases}$$

30

$$mSSNR_{off} = \begin{cases} \sum_i^N (snr(i) + \alpha + \beta)^4 & snr(i) + \alpha \geq 1, lsnr > 18 \\ \sum_i^N (snr(i) + \alpha + \beta)^{10} & snr(i) + \alpha \geq 1, 8 < lsnr \leq 18 \\ \sum_i^N (snr(i) + \alpha + \beta)^{15} & snr(i) + \alpha \geq 1, lsnr \leq 8 \\ \sum_i^N (snr(i) + \alpha + \beta)^9 & \text{de otro modo} \end{cases}$$

35 en donde snr(i) es la relación SNR de registro modificada de la i-ésima sub-banda espectral de la trama de señal de entrada,

40

N es el número de sub-bandas por trama,

lsnr es la estimación de SNR a largo plazo y

α , β son dos coeficientes configurables.

- 5 El primer coeficiente α puede determinarse en una manera de puesta en práctica posible mediante:

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

- 10 en donde $a(i)$ y $b(i)$ son dos números reales o flotantes determinados por el índice de sub-banda i . El segundo coeficiente β puede determinarse mediante la métrica de sonorización $V(-1)$, en la que si $V(-1) > 0.65$ $\beta = 0.2$ y si $V(-1) \leq 0.65$ $\beta = 0.1$. En una manera de puesta en práctica posible, el cálculo de la relación SNR de cada sub-banda $\text{snr}(i)$ viene dada por:

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

- 15 en donde $E(i)$ es la energía de la i -ésima sub-banda de la trama de entrada,

$E_n(i)$ es la energía de la i -ésima sub-banda de la estimación de ruido de fondo.

- 20 En una manera de puesta en práctica posible, la energía de cada sub-banda de la estimación de ruido de fondo se puede estimar moviendo el promedio de las energías de cada sub-banda entre las tramas de ruido de fondo detectadas de la siguiente manera:

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

- 25 en donde $E(i)$ es la energía de la i -ésima sub-banda de la trama detectada como ruido de fondo,

- 30 λ es un factor de olvido usualmente en un rango entre 0.9 - 0.99. El espectro de potencia relacionado en el cálculo anterior se puede obtener en una manera de puesta en práctica posible mediante una Transformación de Fourier Rápida FFT.

- 35 En el estado de funcionamiento normal NWS, el aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, el aparato utiliza la relación SNR segmental modificada $\text{mssnr}_{\text{nor}}$ para realizar una decisión VAD intermedia VADD_{int} . Esta decisión VAD intermedia VADD_{int} , se puede realizar mediante la comparación de la relación SNR segmental modificada $\text{mssnr}_{\text{nor}}$ calculada con un valor umbral thr que puede determinarse por:

$$\text{thr} = \begin{cases} 135 & \text{lsnr} > 18 \\ 35 & 8 < \text{lsnr} \leq 18 \\ 10 & \text{lsnr} \leq 8 \end{cases}$$

- 40 La decisión de VAD intermedia VADD_{int} está activa si la relación SNR modificada es $\text{mssnr}_{\text{nor}} > \text{thr}$, de lo contrario, la decisión intermedia de VAD, VADD_{int} está inactiva.

- 45 En el estado de desplazamiento de voz, el aparato de VAD 1 utiliza, en una manera de puesta en práctica posible, tanto la relación SNR modificada $\text{mssnr}_{\text{off}}$ como la métrica de voz $V(-1)$ para tomar una decisión de VAD intermedia VADD_{int} . La decisión de VAD intermedia VADD_{int} se establece como activa si la relación SNR segmental modificada $\text{mssnr}_{\text{off}} > \text{thr}$ o la métrica de voz $V(-1) >$ que un valor umbral configurable de, p.ej., 0.7, de no ser así, la decisión de VAD intermedia VADD_{int} se establece como inactiva.

- 50 En una manera de puesta en práctica posible, una resonancia fuerte se puede aplicar, de forma opcional, a la decisión de VAD intermedia VADD_{int} . En esta puesta en práctica específica, si un contador de resonancia fuerte HHC es mayor que un valor umbral predeterminado, tal como 0, y si la decisión de VAD intermedia VADD_{int} está inactiva, la decisión final de VAD, VADD_{fin} se fuerza a activarse y el contador de resonancia fuerte HHC se reduce en 1. En una manera de puesta en práctica posible, el contador de resonancia fuerte HHC se restablece a su valor máximo de conformidad con la misma regla aplicada al restablecimiento del contador de resonancia suave SHC.

- 55 En una manera de puesta en práctica posible adicional del aparato de VAD 1, de conformidad con el primer aspecto de la presente invención, el aparato de VAD 1 selecciona, en esta puesta en práctica específica, solamente dos parámetros VAD para su decisión VAD intermedia, es decir, $\text{mssnr}_{\text{nor}}$ y $\text{mssnr}_{\text{off}}$.

$$mssnr_{nor} = \begin{cases} \sum_i^N (snr(i) + \alpha)^4 & snr(i) + \alpha \geq 1, lsnr > 18 \\ \sum_i^N (snr(i) + \alpha)^9 & snr(i) + \alpha \geq 1, 8 < lsnr \leq 18 \\ \sum_i^N (snr(i) + \alpha)^{13} & snr(i) + \alpha \geq 1, lsnr \leq 8 \end{cases}$$

$$mssnr_{off} = \begin{cases} \sum_i^N (snr(i) + \alpha + \beta)^5 & lsnr > 18 \\ \sum_i^N (snr(i) + \alpha + \beta)^{11} & 8 < lsnr \leq 18 \\ \sum_i^N (snr(i) + \alpha + \beta)^{15} & lsnr \leq 8 \end{cases}$$

5 en donde la relación SNR segmental modificada $mssnr_{nor}$ se utiliza en el estado de funcionamiento normal NWS y la SNR segmental modificada $mssnr_{off}$ se utiliza en el estado de funcionamiento de desplazamiento OWS. El coeficiente β se determina, en esta manera de puesta en práctica, no solamente por la métrica $V(-1)$ sino también por el índice de sub-banda i , en donde para el índice de sub-banda i , mayor que un valor entero de m , si $V(-1) > 0.65$, el coeficiente β se establece en 0.2; de lo contrario, el coeficiente β se establece en 0.1. Además, para el índice de sub-banda i , que no sea mayor que m , si $V(-1) > 0.65$, el segundo coeficiente β se establece en $\beta = 0.2 / + 1.5$; de no ser así, el segundo coeficiente β se establece en $0.1 \cdot 1.5$. En esta forma de realización específica, se define otro conjunto de umbrales para que el estado de funcionamiento de desplazamiento OWS sea diferente del conjunto de umbrales para el estado de funcionamiento normal NWS.

15 La invención da a conocer, además, como un segundo aspecto, un aparato de procesamiento de señal de audio tal como se ilustra en la Figura 2 que comprende un aparato de VAD 1, que proporciona una decisión de VAD final, VADD, a una unidad de procesamiento de señal de audio 7 del aparato de procesamiento de señal de audio 6. En consecuencia, la unidad de procesamiento de señal de audio 7 se controla por una decisión de VAD, VADD, que se genera por el aparato de VAD 1. La unidad de procesamiento de señal de audio 7 puede realizar diferentes clases de procesamientos de señal de audio sobre la señal de audio aplicada, tal como una codificación de voz, dependiendo de la decisión de VAD.

De conformidad con un tercer aspecto, la presente invención da a conocer un método para realizar una detección VAD, en donde la decisión de VAD, VADD, se calcula mediante un aparato de VAD para una señal de audio de entrada utilizando al menos un parámetro de VAD, VADP, de un conjunto de decisión de parámetro de estado de funcionamiento WSPDS, de un estado de funcionamiento actual WS, que se detecta por un detector de estado de dicho aparato de VAD. De conformidad con una manera de puesta en práctica posible del método, se recibe una trama de entrada de la señal de audio de entrada aplicada. A continuación, se puede identificar un tipo de señal de la señal de entrada a partir de un conjunto de tipos de señal predefinidos. En una etapa posterior, se selecciona o elige un estado de funcionamiento WS del aparato de VAD entre varios estados de funcionamiento posibles WS de conformidad con el tipo de señal de entrada identificado. En una etapa adicional, los parámetros de VAD se seleccionan de conformidad con el estado de funcionamiento WS seleccionado del aparato de VAD entre un conjunto más grande de parámetros de decisión de VAD predefinidos. Por último, se realiza una decisión VAD, VADD, en función de los parámetros VAD elegidos o seleccionados.

35 En una manera de puesta en práctica posible del método de conformidad con un tercer aspecto de la presente invención, el conjunto de tipos de señal predefinidos puede estar constituido de un tipo de desplazamiento de voz y un tipo sin desplazamiento de voz. Varios estados de funcionamiento posibles WS pueden incluir un estado para desplazamiento de voz definido como un período corto de la señal de audio aplicada al final de las ráfagas de voz. El desplazamiento de voz puede identificarse, en condiciones normales, mediante unas pocas tramas inmediatamente después de que la decisión intermedia del aparato de VAD que opera en el estado de funcionamiento sin desplazamiento de voz pase a inactivo, desde activo, en una ráfaga de voz. Se puede detectar una ráfaga de voz p.ej., cuando se detecta una señal de voz activa de más de 60 ms de longitud. En una manera de puesta en práctica posible del método de conformidad con el tercer aspecto de la presente invención, el conjunto de parámetros de VAD predefinidos puede incluir parámetros basados en SNR segmental de sub-banda con diferentes formas. En una manera de puesta en práctica posible, los parámetros basados en la SNR segmental de sub-banda, con diferentes formas, son parámetros de SNR segmental de sub-banda, que se procesan por diferentes funciones no lineales.

REIVINDICACIONES

1. Un aparato de detección de actividad de voz (1) para determinar una decisión de detección de actividad de voz, VADD, para una señal de audio de entrada, en donde el aparato de detección de actividad de voz (1) comprende:

un detector de estado (2), adaptado para determinar un estado de funcionamiento actual WS de entre dos estados de funcionamiento diferentes del aparato de detección de actividad de voz (1), dependiente de la señal de audio de entrada, en donde cada uno de los dos estados de funcionamiento diferentes, está asociado con un conjunto de decisión de parámetro de estado de funcionamiento correspondiente, WSPDS, que incluye al menos un parámetro de decisión de actividad de voz, VADP, los dos estados de funcionamiento diferentes comprenden un estado de funcionamiento normal y un estado de funcionamiento de desplazamiento, y el al menos un VADP está basado en la relación señal a ruido segmental de sub-banda, SNR; y

un dispositivo de cálculo de actividad de voz (3), adaptado para calcular un valor de parámetro de detección de actividad de voz para al menos un VADP del conjunto WSPDS asociado con el estado de funcionamiento actual, y para determinar el VADD comparando el valor del parámetro de detección de actividad de voz calculado del VADP con respecto a un valor umbral,

en donde dicho aparato de detección de actividad de voz (1) es adaptado para estar conmutado desde el estado de funcionamiento normal al estado de funcionamiento de desplazamiento, si el VADD indica una ausencia de voz en la señal de audio de entrada y un contador de resonancia suave, SHC, supera un valor umbral de contador predeterminado.

2. El aparato de detección de actividad de voz según la reivindicación 1,

en donde dicho aparato de detección de actividad de voz (1) es configurado para estar conmutado entre diferentes estados de funcionamiento de conformidad con condiciones de transición de estado de funcionamiento configurables.

3. El aparato de detección de actividad de voz según la reivindicación 1 o 2,

en donde la decisión VADD determinada en el estado de funcionamiento de compensación forma una decisión de detección de actividad de voz intermedia (VADD_{int}) si la VADD determinada sobre la base de al menos un parámetro VADP del conjunto de decisión de parámetros de estado de funcionamiento de compensación indica que una actividad de voz está ausente en el estado actual de la señal de audio de entrada.

4. El aparato de detección de actividad de voz según la reivindicación 3,

en donde la VADD_{int} se somete a un procesamiento de mantenimiento duro para proporcionar una decisión final de detección de actividad de voz (VADD_{fin}).

5. El aparato de detección de actividad de voz según la reivindicación 1,

en donde dicho aparato de detección de actividad de voz (1) está configurado para conmutarse del estado de funcionamiento de compensación al estado de funcionamiento normal si un contador de mantenimiento suave (SHC) no supera un valor de contador de umbral predeterminado.

6. El aparato de detección de actividad de voz según la reivindicación 4 o 5, en donde dicha señal de audio de entrada consiste en una secuencia de tramas de señal de audio y el contador SHC se disminuye en el estado de funcionamiento de desplazamiento de dicho aparato de detección de actividad de voz (1), para cada trama de señal de audio recibida hasta que se alcanza el valor umbral de contador predeterminado.

7. El aparato de detección de actividad de voz según una de las reivindicaciones precedentes 4 a 6, en donde si se detecta un número predeterminado de tramas de señal de audio activa consecutiva, de la señal de audio de entrada, el SHC se restablece a un valor de contador en función de una relación de señal a ruido a largo plazo, ISNR, de la señal de audio de entrada.

8. El aparato de detección de actividad de voz según una de las reivindicaciones precedentes 4 a 7, en donde se detecta una trama de señal de audio activa si una métrica de sonorización calculada (V), de la trama de señal de audio, supera un valor umbral de métrica de voz predeterminado y una estabilidad de tono (S), de dicha trama de señal de audio es inferior a un valor umbral de estabilidad predeterminado.

9. El aparato de detección de actividad de voz según la reivindicación 8, en donde la métrica de sonorización (V) se calcula de la siguiente manera:

$$V(0) = (\text{sonorización}(-1) + \text{sonorización}(0) + \text{sonorización}(1))/3 + \text{corr_shift}$$

en donde sonorización(-1) representa la correlación correspondiente en tanto como un retardo tonal de la segunda sub-trama de la trama de señal de entrada precedente, sonorización(0) y sonorización(1) representan, respectivamente, las correlaciones correspondientes en tanto como retardos tonales buscados para las dos sub-

tramas de la trama de entrada actual, y corr_shift es un valor de compensación que depende del nivel de ruido de fondo.

5 **10.** El aparato de detección de actividad de voz según la reivindicación 8 o 9, en donde la estabilidad tonal (S) de dicha trama de señal de audio se calcula mediante:

$$S_r(0)=[abs[T(-1)-T(-2))+abs(T(0)-T(-1))+abs[T(1)-T(0)]]/3$$

10 en donde T(-1), T(-2) son los primeros y segundos retardos tonales de la trama de señal de entrada precedente, T(0), T(1) son los retardos tonales buscados para las dos sub-tramas de la entrada de entrada actual y abs () significa el valor absoluto.

15 **11.** El aparato de detección de actividad de voz según una de las reivindicaciones 8 a 10 anteriores, en donde el valor de umbral de métrica vocal predeterminado es 0,65 y el valor de umbral de estabilidad predeterminado es 14.

12. El aparato de detección de actividad de voz según una de las reivindicaciones precedentes 1 a 11, en donde al menos un parámetro VADP comprende:
parámetros de decisión basados en la energía,
parámetros de decisión basados en la envolvente espectral,
20 y/o parámetros de decisión basados en estadísticas.

13. El aparato de detección de actividad de voz según la reivindicación 1, en donde una decisión de detección de actividad de voz intermedia (VADD_{int}), determinada por dicho dispositivo de cálculo de actividad de voz (3), se aplica a una unidad de procesamiento de resonancia fuerte que realiza una resonancia fuerte de dicha decisión de detección de actividad de voz intermedia (VADD_{int}) aplicada.
25

14. El aparato de detección de actividad de voz según una de las reivindicaciones anteriores 1 a 13, en donde los parámetros basados en la relación SNR segmental de sub-banda con diferentes formas son parámetros de SNR segmental de sub-banda procesados por diferentes funciones no lineales.
30

15. Un dispositivo de procesamiento de señal de audio (6), que comprende un aparato de detección de actividad de voz (1), según una de las reivindicaciones precedentes 1 a 14, y una unidad de procesamiento de señal de audio (7), configurado para estar controlada por una decisión de detección de actividad de voz, VADD, generada por dicho aparato de detección de actividad de voz (1).
35

16. Un método para realizar una detección de actividad de voz, en donde el método comprende:
un aparato de detección de actividad de voz (1) determina un estado de funcionamiento actual de dos estados de funcionamiento diferentes del aparato de detección de actividad de voz (1) en función de la señal de audio de entrada, en donde cada uno de los dos estados de funcionamiento diferentes está asociado con un conjunto de decisiones de parámetros de estado de funcionamiento correspondientes, WSPDS, que incluye al menos un parámetro de decisión de actividad de voz, VADP, los dos estados de funcionamiento diferentes comprenden un estado de funcionamiento normal y un estado de funcionamiento de compensación, y el al menos un VADP se basa en la relación de señal a ruido segmental de sub-banda, SNR; y el aparato de detección de actividad de voz (1) calcula un valor de parámetro de detección de actividad de voz para al menos un VADP del WSPDS asociado con el estado de funcionamiento actual y para determinar la decisión de detección de actividad de voz, VADD, comparando el valor del parámetro de detección de actividad de voz calculado del respectivo VADP con un umbral,
40 en donde dicho aparato de detección de actividad de voz (1) se conmuta desde el estado de funcionamiento normal al estado de funcionamiento de compensación si la VADD indica una ausencia de voz en la señal de audio de entrada y un contador de mantenimiento suave, SHC, supera un valor de contador de umbral predeterminado.
45
50

17. El método según la reivindicación 16, en donde los parámetros basados en la relación SNR segmental de sub-banda con diferentes formas son parámetros SNR segmental de sub-banda procesados por diferentes funciones no lineales.

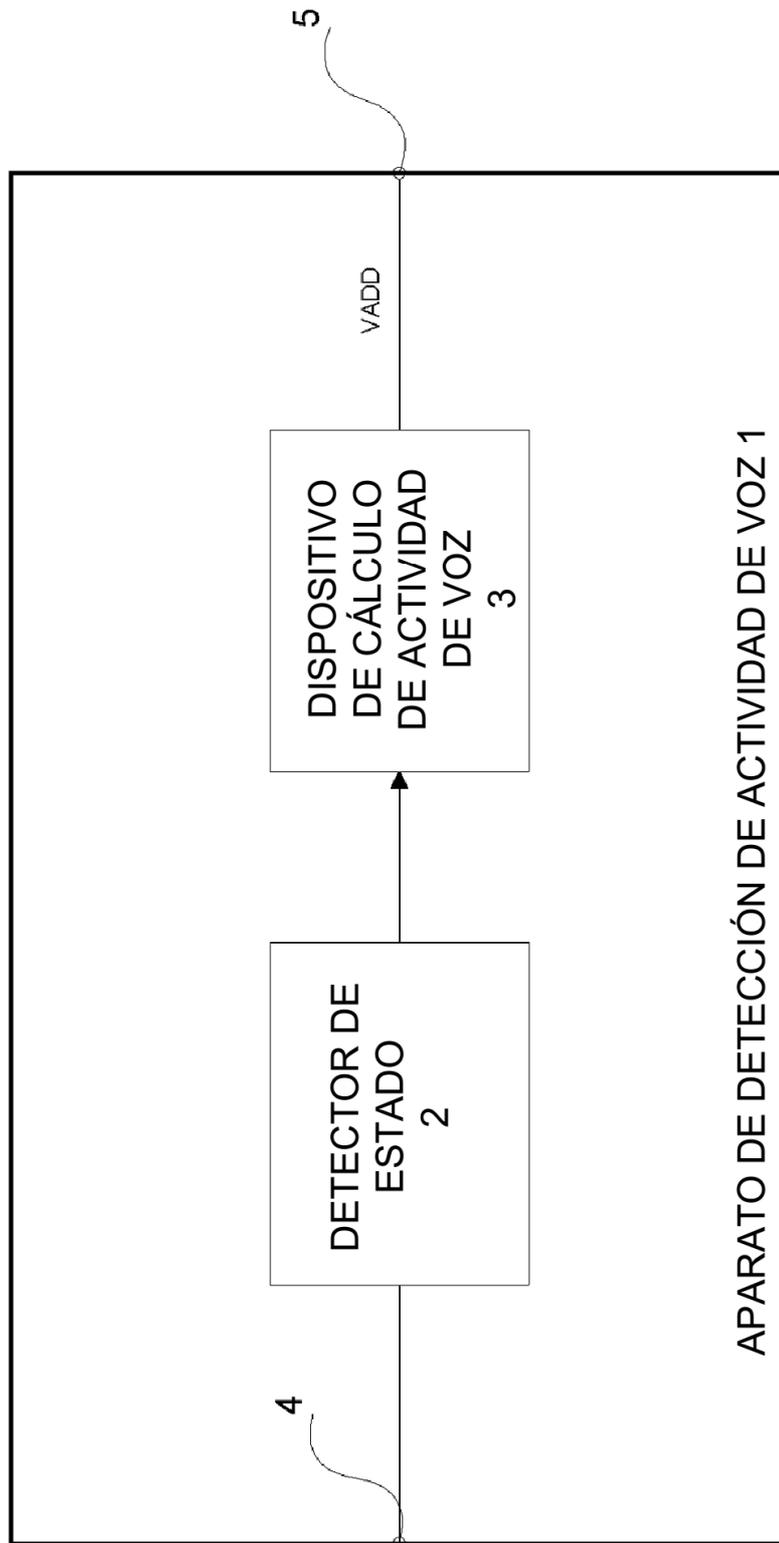


Fig. 1

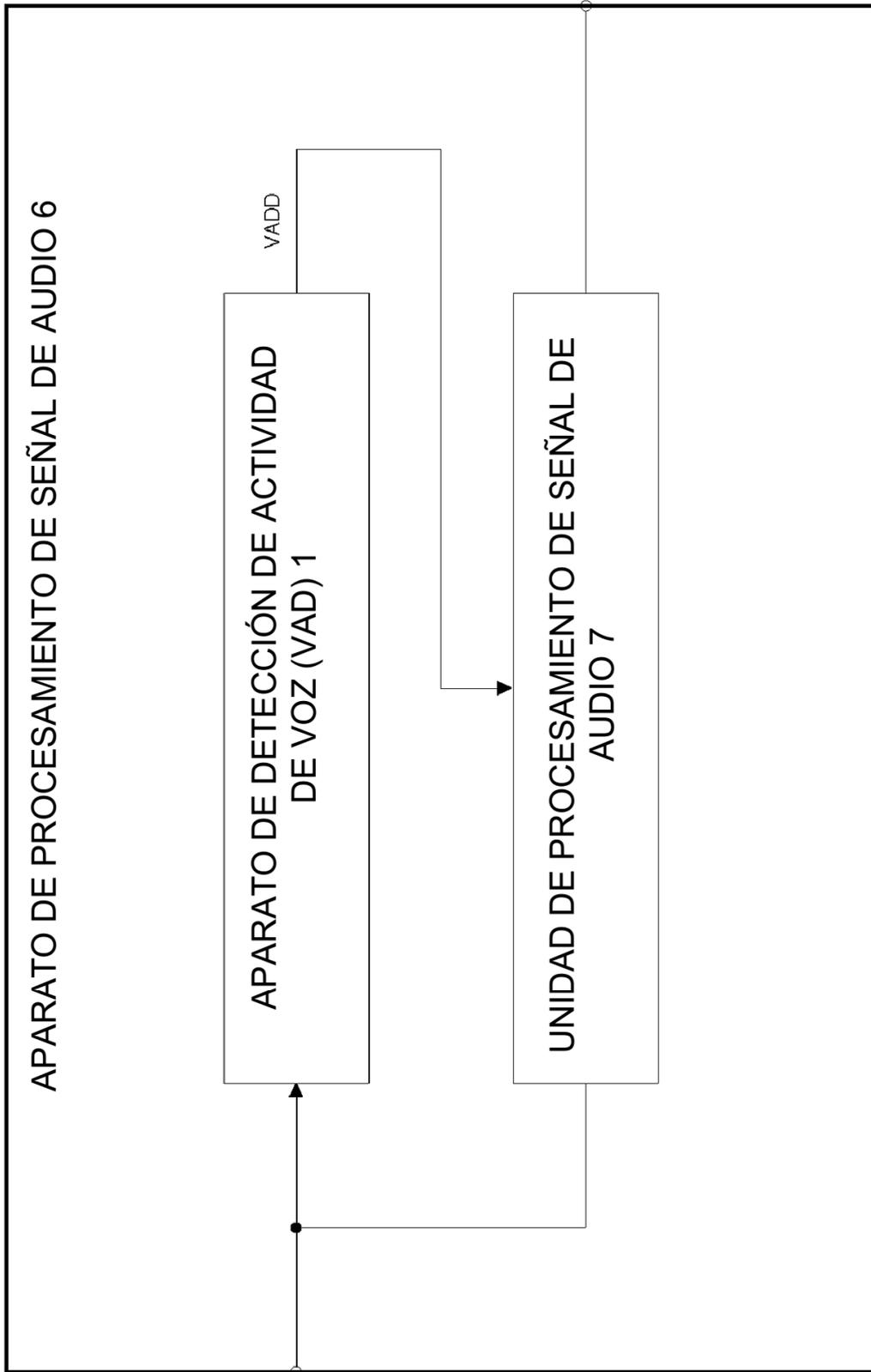


Fig. 2