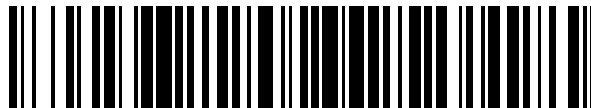


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 787 894**

51 Int. Cl.:

**G10L 25/78** (2013.01)

**G10L 25/18** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.12.2014 PCT/CN2014/092694**

87 Fecha y número de publicación internacional: **17.09.2015 WO15135344**

96 Fecha de presentación y número de la solicitud europea: **01.12.2014 E 14885786 (5)**

97 Fecha y número de publicación de la concesión europea: **12.02.2020 EP 3118852**

54 Título: **Método y dispositivo para detectar la señal de audio**

30 Prioridad:

**12.03.2014 CN 201410090386**

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

**19.10.2020**

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:

**ELZABURU, S.L.P**

ES 2 787 894 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 dispositivo para detectar la señal de audio

**Campo técnico**

5 La presente invención se refiere al campo de las tecnologías de procesamiento de señales, y más específicamente, a un método para detectar una señal de audio y un aparato.

**Antecedentes**

10 La detección de actividad de voz (VAD) es una tecnología clave ampliamente utilizada en campos tales como las comunicaciones de voz y la interacción hombre-máquina. La VAD también puede denominarse detección de actividad de sonido (SAD). La VAD se utiliza para detectar si existe una señal activa en una señal de audio de entrada, donde la señal activa es relativa a una señal inactiva (tal como ruido ambiental de fondo y una voz muda). Las señales activas típicas incluyen una voz, música y similares. Un principio de la VAD es que uno o más parámetros de características se extraen de una señal de audio de entrada, uno o más valores de características se determinan según uno o más parámetros de características, y a continuación uno o más valores de características se comparan con uno o más umbrales.

15 En la técnica anterior, un método de detección de señal activa basado en una relación señal/ruido segmental (SSNR) incluye: dividir una señal de audio de entrada en múltiples señales de subbanda en una banda de frecuencia, calcular energía de la señal de audio en cada subbanda, y comparar la energía de la señal de audio en cada subbanda con la energía estimada de una señal de ruido de fondo en cada subbanda, con el fin de obtener una relación señal/ruido (SNR) de la señal de audio en cada subbanda; y a continuación determinar una SSNR según una SNR de subbanda de cada subbanda, y comparar la SSNR con un umbral de decisión de VAD predeterminado, donde si la SSNR excede el umbral de decisión de VAD, la señal de audio es una señal activa, o si la SSNR no excede el umbral de decisión de VAD, la señal de audio es una señal inactiva.

Un método típico para calcular la SSNR es sumar todos los SNR de subbanda de la señal de audio, y un resultado obtenido es la SSNR. Por ejemplo, la SSNR puede determinarse utilizando la fórmula 1.1:

$$25 \quad SSNR = \sum_{k=0}^{N-1} snr(k) \quad \text{Fórmula 1.1}$$

donde k indica la sub- banda k-ésima, snr(k) indica una SNR de subbanda de la subbanda k-ésima, y N indica una cantidad total de subbanda de subbandas en las que se divide la señal de audio.

Cuando el método anterior para calcular la SSNR se utiliza para detectar una voz activa, puede producirse una detección errónea de una voz activa.

30 El documento US2013/191117A1 describe que en sistemas de procesamiento de voz, la compensación se hace para cambios repentinos en el ruido de fondo en el cálculo de la relación señal/ruido (SNR) promedio. El filtrado atípico de SNR puede utilizarse, solo o junto con la ponderación de la SNR promedio. Los pesos adaptativos pueden aplicarse sobre las SNR por banda antes de calcular la SNR promedio. La función de ponderación puede ser una función de nivel de ruido, tipo de ruido y/o valor de SNR instantáneo.

35 El documento US 2013/304464A1 describe un método y un aparato para detectar de forma adaptativa una actividad de voz en una señal de audio de entrada compuesta de tramas. El método comprende las operaciones de: determinar una característica de ruido de la señal de entrada basándose en una trama recibida de la señal de audio de entrada; derivar un parámetro de detección de actividad de voz (VAD) basándose en la característica de ruido de la señal de audio de entrada; y comparar el parámetro VAD derivado con el valor umbral para proporcionar una decisión de detección de actividad de voz.

40 WEIWU JIANG Y COL.: "a new voice activity detection method using maximized sub-band SNR" (un nuevo método de detección de actividad de voz que utilizan SNR de subbanda maximizada) describe un método VAD que utiliza un Valor Máximo de la SNR de subbanda (MVSS) como la característica de detección. La nueva característica MVSS propuesta tiene diferentes distribuciones entre la señal de voz y la señal que no son de voz, que es útil para separar la señal de voz del ruido fuerte. Un umbral adaptativo se aplica para mejorar las precisiones VAD y seguir la señal ruidosa rápidamente sin cálculos complejos.

**Compendio**

La presente invención proporciona un método para detectar una señal de audio y un aparato, que puede distinguir con precisión entre una voz activa y una voz inactiva.

50 La invención se define en las reivindicaciones adjuntas. A continuación, cados de la palabra "realización (realizaciones)", si se refiere a combinaciones de característica diferentes de las definidas por las reivindicaciones, relacionadas con ejemplos que originalmente se han archivado pero que no representan realizaciones de la invención

actualmente reivindicada; estos ejemplos se muestran aún, solamente para propósitos ilustrativos.

5 Según un primer aspecto, la presente invención proporciona un método para detectar una señal de audio, donde el método incluye: determinar una señal de audio de entrada como una señal de audio que se ha de determinar si se determina que la señal de audio es una señal sin voz; determinar una relación señal/ruido segmental mejorada (SSNR) de la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia; y comparar la SSNR mejorada con un umbral de decisión de detección de actividad de voz (VAD) para determinar si la señal de audio es una señal activa, en donde la determinación de la SSNR mejorada de la señal de audio comprende: determinar la SSNR de referencia de la señal de audio; y determinar la SSNR mejorada según la SSNR de referencia de la señal de audio.

10 Con referencia al primer aspecto, en una primera manera posible de implementación del primer aspecto, la determinación de la SSNR mejorada según la SSNR de referencia de la señal de audio incluye: determinar la SSNR mejorada utilizando la siguiente fórmula:  $SSNR' = x * SSNR + y$ , donde SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y x e y indican los parámetros de mejora.

15 Según un segundo aspecto, la presente invención proporciona un método para detectar una señal de audio, donde el método incluye: determinar una señal de audio de entrada como una señal de audio que se ha de determinar si se determina que la señal de audio es una señal sin voz; determinar un peso de una SNR de subbanda de cada subbanda en la señal de audio, donde un peso de una SNR de subbanda de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que un primer umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda; determinar una SSNR mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia; y comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa.

20 Según un tercer aspecto, la presente invención proporciona un aparato, donde el aparato incluye: una primera unidad de determinación, configurada para determinar una señal de audio de entrada como una señal de audio que se ha de determinar si se determina que la señal de audio es una señal sin voz; una segunda unidad de determinación, configurada para determinar una SSNR mejorada de la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia; y una tercera unidad de determinación, configurada para comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa, en donde la segunda unidad de determinación está configurada específicamente para determinar la SSNR de referencia de la señal de audio y determinar la SSNR mejorada según la SSNR de referencia de la señal de audio.

25 Con referencia al cuarto aspecto, de una primera manera posible de implementación del cuarto aspecto, la segunda unidad de determinación está configurada específicamente para determinar la SSNR mejorada utilizando la siguiente fórmula:  $SSNR' = x * SSNR + y$ , donde SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y x e y indican los parámetros de mejora.

30 Según un cuarto aspecto, la presente invención proporciona un aparato, donde el aparato incluye: una primera unidad de determinación, configurada para determinar una señal de audio de entrada como una señal de audio que se ha de determinar si se determina que la señal de audio es una señal sin voz; una segunda unidad de determinación, configurada para determinar un peso de una SNR de subbanda de cada subbanda en la señal de audio, donde un peso de una SNR de subbanda de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que un primer umbral predeterminado que es mayor que un peso de una SNR de subbanda de otra subbanda, y determina una SSNR mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia; y una tercera unidad de determinación, configurada para comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa.

35 Según el método de la presente invención, se puede determinar una característica de una señal de audio, se determina una SSNR mejorada de la manera correspondiente según la característica de la señal de audio, y la SSNR mejorada se compara con un umbral de decisión de VAD, de manera que pueda reducirse una proporción de detección errónea de una señal activa.

**Breve descripción de los dibujos**

40 Para describir las soluciones técnicas de la presente invención más claramente, a continuación se introducen brevemente los dibujos adjuntos que describen las realizaciones de la presente invención. Aparentemente, los dibujos adjuntos en la siguiente descripción muestran simplemente algunas realizaciones de la presente invención.

La fig. 1 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización de la presente invención;

45 La fig. 2 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización de la presente invención;

La fig. 3 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización

de la presente invención;

La fig. 4 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización de la presente invención;

La fig. 5 es un diagrama de bloques estructural de un aparato según una realización de la presente invención;

5 La fig. 6 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención;

La fig. 7 es un diagrama de bloques estructural de un aparato según una realización de la presente invención;

La fig. 8 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención;

La fig. 9 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención; y

La fig. 10 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención.

## 10 Descripción de realizaciones

Lo siguiente describe clara y completamente las soluciones técnicas de la presente invención con referencia a los dibujos adjuntos que muestran las realizaciones preferidas de la presente invención. Aparentemente, las realizaciones descritas son simplemente algunas pero no todas las realizaciones de la presente invención.

15 La fig. 1 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización de la presente invención.

101. Determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

102. Determinar una SSNR mejorada de la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia.

20 103. Comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa.

En esta realización de la presente invención, cuando la SSNR mejorada se compara con el umbral de decisión de VAD, se puede utilizar un umbral de decisión de VAD de referencia, o se puede utilizar un umbral de decisión de VAD reducido obtenido después de que se reduzca un umbral de decisión de VAD de referencia utilizando un algoritmo predeterminado. El umbral de decisión de VAD de referencia puede ser un umbral de decisión de VAD por defecto, y el umbral de decisión de VAD de referencia puede almacenarse previamente, o puede obtenerse temporalmente a través del cálculo, donde el umbral de decisión de VAD de referencia puede calcularse utilizando una tecnología existente bien conocida. Cuando el umbral de decisión de VAD de referencia se reduce utilizando el algoritmo predeterminado, el algoritmo predeterminado puede estar multiplicando el umbral de decisión de VAD de referencia por un coeficiente que sea menor que 1, o se puede utilizar otro algoritmo. Esta realización de la presente invención no impone limitación en un algoritmo específico utilizado.

35 Cuando se utiliza un método de cálculo SSNR convencional para calcular las SSNR de algunas señales de audio, las SSNR de estas señales de audio pueden ser inferiores a un umbral de decisión de VAD predeterminado. Sin embargo, en realidad, estas señales de audio son señales de audio activas. Esto es provocado por las características de estas señales de audio. Por ejemplo, en un caso en el que una SNR ambiental es relativamente baja, una SNR de subbanda de una parte de alta frecuencia se reduce significativamente. Además, ya que una teoría psicoacústica se utiliza generalmente para realizar la división de subbanda, la SNR de subbanda de la parte de alta frecuencia tiene una contribución relativamente baja a una SSNR. En este caso, para algunas señales, tal como una señal sin voz, cuya energía se centraliza principalmente en una parte de frecuencia relativamente alta, una SSNR obtenida a través del cálculo utilizando el método de cálculo SSNR convencional puede ser inferior al umbral de decisión de VAD, lo que provoca una detección errónea de una señal activa. Para otro ejemplo, para algunas señales de audio, la distribución de energía de estas señales de audio es relativamente plana en un espectro, pero la energía general de estas señales de audio es relativamente baja. Por lo tanto, en el caso en que una SNR ambiental es relativamente baja, una SSNR obtenida a través del cálculo utilizando el método de cálculo SSNR convencional puede ser inferior al umbral de decisión de VAD. En el método mostrado en la FIG. 1, se utiliza una manera de aumentar adecuadamente una SSNR, de manera que la SSNR puede ser mayor que un umbral de decisión de VAD. Por lo tanto, puede reducirse de manera eficaz una proporción de detección errónea de una señal activa.

40 La fig. 2 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización de la presente invención.

201. Determinar una SNR de subbanda de una señal de audio de entrada.

50 Un espectro de la señal de audio de entrada se divide en N subbandas, donde N es un entero positivo mayor que 1. Específicamente, una teoría psicoacústica puede utilizarse para dividir el espectro de la señal de audio. En un caso

en el que la teoría psicoacústica se utiliza para dividir el espectro de la señal de audio, cuanto menor es la frecuencia de la subbanda, más estrecho es el ancho de banda de la subbanda. Cuando mayor es la frecuencia de la subbanda, mayor es el ancho de banda de la subbanda. Ciertamente, el espectro de la señal de audio también puede dividirse de otra manera, por ejemplo, una manera de dividir uniformemente el espectro de la señal de audio en N subbandas.

5 Se calcula una SNR de subbanda de cada subbanda de la señal de audio de entrada, donde la SNR de subbanda es una relación de energía de la subbanda a energía de ruido de fondo en la subbanda. La energía del ruido de fondo en la subbanda generalmente es un valor estimado obtenido por estimación por un estimador de ruido de fondo. Cómo utilizar el estimador de ruido de fondo para estimar la energía de ruido de fondo correspondiente a cada subbanda es una tecnología bien conocida de este campo. Por lo tanto, no se necesitan describir detalles en la presente memoria.  
 10 Un experto en la técnica puede comprender que la SNR de subbanda puede ser una relación de energía directa, o puede ser otra forma de expresión de una relación de energía directa, tal como una SNR de subbanda logarítmica. Además, un experto en la técnica puede comprender además que la SNR de subbanda también puede ser una SNR de subbanda obtenida después de que se realice un procesamiento lineal o no lineal en una SNR de subbanda directa, o puede ser otra transformación de la SNR de subbanda. La relación de energía directa de la SNR de subbanda se muestra en la siguiente fórmula:  
 15

$$snr(k) = E(k) / En(k)$$

Fórmula 1.2

donde  $snr(k)$  indica una SNR de subbanda de la subbanda k-ésima, y  $E(k)$  y  $En(k)$  indican respectivamente la energía de la subbanda k-ésima y la energía del ruido de fondo en la subbanda k-ésima. Una SNR de subbanda logarítmica se puede indicar como:  $snr_{log}(k) = 10 \times \log_{10} snr(k)$ , donde  $snr_{log}(k)$  indica una SNR de subbanda logarítmica de la subbanda k-ésima, y  $snr(k)$  indica una SNR de subbanda que es de la subbanda k-ésima y se obtiene a través del cálculo utilizando la fórmula 1.2. Un experto en la técnica puede comprender además que la energía de subbanda utilizada para calcular una SNR de subbanda puede ser energía de la señal de audio de entrada en una subbanda, o puede ser energía obtenida después de la energía del ruido de fondo en una subbanda se resta de la energía de la señal de audio de entrada en la subbanda. El cálculo de la SNR es correcto sin desviarse del significado de la SNR.

25 202. Determinar la señal de audio de entrada como una señal de audio que se ha de determinar.

Opcionalmente, en una realización, la determinación de la señal de audio de entrada como una señal de audio que se ha de determinar puede incluir: determinar la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda que es de la señal de audio y determinado en la etapa 201.

30 Opcionalmente, en una realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de la señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una primera cantidad.

35 Opcionalmente, en otra realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de la señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una  
 40 segunda cantidad, y una cantidad de subbandas de extremidad de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad. En esta realización de la presente invención, una porción de alta frecuencia y una extremidad de baja frecuencia de una trama de señal de audio son relativas, es decir, una parte que tiene una frecuencia relativamente alta es la porción de alta frecuencia, y una parte que tiene una frecuencia relativamente baja es la extremidad de baja frecuencia.  
 45

Opcionalmente, en otra realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de la señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas que se encuentran en la señal de audio y cuyos valores de las SNR de subbanda son mayores que un tercer umbral predeterminado es mayor que una cuarta cantidad.  
 50

El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, las estadísticas sobre las SNR de subbanda de las subbandas de porción de alta frecuencia se recopilan en una gran cantidad de muestras sin voz incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, las estadísticas sobre las SNR de subbanda de las subbandas de extremidad de baja frecuencia se recopilan en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de  
 55

las subbandas de extremidad de frecuencia baja en estas muestras sin voz son menores que el segundo umbral predeterminado.

5 El tercer umbral predeterminado también se obtiene mediante la recopilación de estadísticas. Específicamente, el tercer umbral predeterminado se determina según las SNR de subbanda de una gran cantidad de señales de ruido, de manera que las SNR de subbanda de la mayoría de las subbandas en estas señales de ruido son menores que el tercer umbral predeterminado.

10 La primera cantidad, la segunda cantidad, la tercera cantidad y la cuarta cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que el primera cantidad. Un método para adquirir la segunda cantidad es similar a un método para adquirir la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o la segunda cantidad puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de extremidad de baja frecuencia cuyas SNR de subbanda son menores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidad de baja frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son menores que el segundo umbral predeterminado es mayor que la tercera cantidad. Para la cuarta cantidad, en una gran cantidad de tramas de señal de ruido, se recopilan estadísticas sobre una cantidad de subbandas cuyas SNR de subbanda son menores que el tercer umbral predeterminado, y la cuarta cantidad se determina según la cantidad, de manera que una cantidad de subbandas que se encuentran en la mayoría de estas tramas de muestra de ruido y cuyas SNR de subbanda son menores que el tercer umbral predeterminado es mayor que la cuarta cantidad.

15 Opcionalmente, en otra realización, si la señal de audio de entrada es una señal de audio que se ha de determinar puede determinarse determinando si la señal de audio de entrada es una señal sin voz. En este caso, no es necesario determinar la SNR de subbanda de la señal de audio cuando se determina si la señal de audio es una señal de audio que se ha de determinar. En otras palabras, la operación 201 no necesita realizarse cuando se determina si la señal de audio es una señal de audio que se ha de determinar. Específicamente, la determinación de la señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que se determina que la señal de audio es una señal sin voz. Específicamente, un experto en la técnica puede comprender que puede haber múltiples métodos para detectar si la señal de audio es una señal sin voz. Por ejemplo, si la señal de audio es una señal sin voz puede determinarse detectando una tasa de cruce por cero en el dominio del tiempo (ZCR) de la señal de audio. Específicamente, en un caso en el que la ZCR de la señal de audio es mayor que un umbral de ZCR, se determina que la señal de audio es una señal sin voz, donde el umbral de ZCR se determina según una gran cantidad de experimentos.

203. Determinar una SSNR mejorada de la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia.

40 La SSNR de referencia puede ser una SSNR obtenida a través del cálculo utilizando la fórmula 1.1. A partir de la fórmula 1.1 se puede ver que el procesamiento de ponderación no se realiza en una SNR de subbanda de ninguna subbanda cuando se está calculando la SSNR de referencia, es decir, los pesos de las SNR de subbanda de todas las subbandas son iguales cuando SSNR de referencia se está calculando.

45 Opcionalmente, en una realización, en un caso en el que la cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que la primera cantidad, o en un caso en el que la cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que la segunda cantidad, y la cantidad de subbandas de extremidad de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que el segundo umbral predeterminado es mayor que la tercera cantidad, la determinación de una SSNR mejorada de la señal de audio incluye: determinar un peso de una SNR de subbanda de cada subbanda banda en la señal de audio, donde un peso de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que el primer umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda; y determinar la SSNR mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio.

55 Por ejemplo, si la señal de audio se divide en 20 subbandas, es decir, la subbanda 0 a la subbanda 19, según la teoría psicoacústica, y las relaciones de señal/ruido de la subbanda 18 y la subbanda 19 son ambas mayores que una primera valor T1 predeterminado, se pueden agregar cuatro subbandas, es decir, la subbanda 20 a la subbanda 23. Específicamente, la subbanda 18 y la subbanda 19 cuyas relaciones señal/ruido son mayores que T1 pueden dividirse respectivamente en la subbanda 18a, la subbanda 18b y la subbanda 18c; y la subbanda 19a, la subbanda 19b y la subbanda 19c. En este caso, la subbanda 18 puede considerarse como una subbanda madre de la subbanda 18a, la

subbanda 18b y la subbanda 18c, y la subbanda 19 puede considerarse como una subbanda madre de subbanda -  
 banda 19a, subbanda 19b y subbanda 19c. Los valores de las relaciones señal/ruido de la subbanda 18a, la subbanda  
 18b y la subbanda 18c son los mismos que un valor de la relación señal/ruido de su subbanda madre y los valores de  
 las relaciones señal/ruido de la subbanda 19a, la subbanda 19b y la subbanda 19c son las mismas que un valor de la  
 5 relación señal/ruido de su subbanda madre. De esta manera, las 20 subbandas que se han obtenido originalmente a  
 través de la división se vuelven a dividir en 24 subbandas. Ya que la VAD todavía está diseñada según las 20  
 subbandas durante la detección de señal activa, las 24 subbandas necesitan asignarse de nuevo a las 20 subbandas  
 para determinar la SSNR mejorada. En conclusión, cuando la SSNR mejorada se determina aumentando la cantidad  
 de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral  
 10 predeterminado, el cálculo puede realizarse utilizando la siguiente fórmula:

$$SSNR' = \frac{20}{24} \times \left[ 2 \times (snr(18) + snr(19)) + \sum_{k=0}^{19} snr(k) \right]$$

Fórmula 1.3

donde SSNR' indica la SSNR mejorada, y snr(k) indica una SNR de subbanda de la subbanda k-ésima.

Si una SSNR obtenida a través del cálculo utilizando la fórmula 1.1 es la SSNR de referencia, la SSNR de referencia  
 obtenida a través del cálculo es  $\sum_{k=0}^{19} snr(k)$ . Obviamente, para una señal de audio de un primer tipo, un valor de  
 15 la SSNR mejorada obtenida a través del cálculo utilizando la fórmula 1.3 es mayor que un valor de la SSNR de  
 referencia obtenida a través del cálculo utilizando la fórmula 1.1.

Para otro ejemplo, si la señal de audio se divide en 20 subbandas, es decir, la subbanda 0 a la subbanda 19, según  
 la teoría psicoacústica, snr(18) y snr(19) son ambas mayores que un primer el valor T1 predeterminado y snr(0) a  
 snr(17) son todas menores que un segundo umbral T2 predeterminado, la SSNR mejorada se puede determinar  
 20 utilizando lo siguiente:

$$SSNR' = a_1 \times snr(18) + a_2 \times snr(19) + \sum_{k=0}^{17} snr(k)$$

Fórmula 1.4

donde SSNR' indica la SSNR mejorada, snr(k) indica una SNR de subbanda de la subbanda k-ésima, a1 y a2 son  
 parámetros de aumento de peso y valores de a1 y a2 hacen a1 x snr(18) + a2 x snr(19) mayor que snr(18) + snr(19).  
 25 Obviamente, un valor de la SSNR mejorada obtenido a través del cálculo utilizando la fórmula 1.4 es mayor que el  
 valor de la SSNR de referencia obtenido a través del cálculo utilizando la fórmula 1.1.

Opcionalmente, en otra realización, la determinación de una SSNR mejorada de la señal de audio incluye: determinar  
 una SSNR de referencia de la señal de audio y determinar la SSNR mejorada según la SSNR de referencia de la señal  
 de audio.

Opcionalmente, la SSNR mejorada se puede determinar utilizando la siguiente fórmula:

$$SSNR' = x * SSNR + y$$

Fórmula 1.5

donde SSNR indica la SSNR de referencia de la señal de audio, SSNR' indica la SSNR mejorada, y x e y indican  
 parámetros de mejora. Por ejemplo, un valor de x puede ser 1,05, y un valor de y puede ser 1. Un experto en la técnica  
 puede comprender que, los valores de x e y pueden ser otros valores adecuados que hacen que la SSNR mejorada  
 sea mayor que la SSNR de referencia correctamente.

Opcionalmente, la SSNR mejorada se puede determinar utilizando la siguiente fórmula:

$$SSNR' = f(x) * SSNR + h(y)$$

Fórmula 1.6

donde SSNR indica una SSNR original de la señal de audio, SSNR' indica la SSNR mejorada, y f(x) y h(y) indican  
 funciones de mejora. Por ejemplo, f(x) y h(y) pueden ser funciones relacionadas con una relación señal/ruido a largo  
 plazo (LSNR) de la señal de audio, donde la LSNR de la señal de audio es una SNR promedio o una SNR ponderada  
 40 dentro de un período de tiempo relativamente largo. Por ejemplo, cuando la lsnr es mayor que 20, f(lsnr) puede ser  
 igual a 1.1, e y(lsnr) puede ser igual a 2; cuando la lsnr es menor que 20 y mayor que 15, f(lsnr) puede ser igual a  
 1,05, e y(lsnr) puede ser igual a 1; y cuando la lsnr es menor que 15, f(lsnr) puede ser igual a 1 e y(lsnr) puede ser  
 igual a 0. Un experto en la técnica puede comprender que, f(x) y h(y) pueden estar en otras formas adecuadas que  
 hacen que la SSNR mejorada sea mayor que la SSNR de referencia correctamente.

204. Comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una  
 señal activa.

Específicamente, cuando la SSNR mejorada se compara con el umbral de decisión de VAD, si la SSNR mejorada es mayor que el umbral de decisión de VAD, se determina que la señal de audio es una señal activa; o si la SSNR mejorada no es mayor que el umbral de decisión de VAD, se determina que la señal de audio es una señal inactiva.

5 Opcionalmente, en otra realización, antes de comparar la SSNR mejorada con un umbral de decisión de VAD, el método puede incluir además: utilizar un algoritmo predeterminado para reducir el umbral de decisión de VAD, con el fin de obtener un umbral de decisión de VAD reducido. En este caso, la comparación de la SSNR mejorada con un umbral de decisión de VAD incluye específicamente: comparar la SSNR mejorada con el umbral de decisión de VAD reducido para determinar si la señal de audio es una señal activa. Un umbral de decisión de VAD de referencia puede ser un umbral de decisión de VAD por defecto, y el umbral de decisión de VAD de referencia puede almacenarse  
10 previamente o puede obtenerse temporalmente a través del cálculo, donde el umbral de decisión de VAD de referencia puede calcularse utilizando una tecnología existente bien conocida. Cuando el umbral de decisión de VAD de referencia se reduce utilizando el algoritmo predeterminado, el algoritmo predeterminado puede estar multiplicando el umbral de decisión de VAD de referencia por un coeficiente que sea menor que 1, o se puede utilizar otro algoritmo. Esta realización de la presente invención no impone limitación en un algoritmo específico utilizado. El umbral de  
15 decisión de VAD puede reducirse adecuadamente utilizando el algoritmo predeterminado, de manera que la SSNR mejorada sea mayor que el umbral de decisión de VAD reducido. Por lo tanto, puede reducirse una proporción de detección errónea de una señal activa.

20 Según el método mostrado en la fig. 2, se determina una característica de una señal de audio, se determina una SSNR mejorada de manera correspondiente según la característica de la señal de audio, y la SSNR mejorada se compara con un umbral de decisión de VAD. De esta manera, puede reducirse una proporción de detección errónea de una señal activa.

La fig. 3 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización de la presente invención.

301. Determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

25 302. Determinar un peso de una SNR de subbanda de cada subbanda en la señal de audio, donde un peso de una SNR de subbanda de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que un primer umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda.

303. Determinar una SSNR mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia.

30 La SSNR de referencia puede ser una SSNR obtenida a través del cálculo utilizando la fórmula 1.1. A partir de la fórmula 1.1 se puede ver que el procesamiento de ponderación no se realiza sobre una SNR de subbanda de ninguna subbanda cuando se calcula la SSNR de referencia, es decir, los pesos de las SNR de subbanda de todas las subbandas son iguales cuando SSNR de referencia se está calculando.

35 Por ejemplo, si la señal de audio se divide en 20 subbandas, es decir, la subbanda 0 a la subbanda 19, según una teoría psicoacústica, y las relaciones señal/ruido de la subbanda 18 y la subbanda 19 son ambas mayores que un primer valor T1 predeterminado, se pueden agregar cuatro subbandas, es decir, la subbanda 20 a la subbanda 23. Específicamente, la subbanda 18 y la subbanda 19 cuyas relaciones señal/ruido son mayores que T1 pueden dividirse respectivamente en la subbanda 18a, la subbanda 18b y la subbanda 18c; y la subbanda 19a, la subbanda 19b y la subbanda 19c. En este caso, la subbanda 18 puede considerarse como una subbanda madre de la subbanda 18a, la subbanda 18b y la subbanda 18c, y la subbanda 19 puede considerarse como una subbanda madre de subbanda -  
40 banda 19a, subbanda 19b y subbanda 19c. Los valores de las relaciones señal/ruido de la subbanda 18a, la subbanda 18b y la subbanda 18c son los mismos que un valor de la relación señal/ruido de su subbanda madre y los valores de las relaciones señal/ruido de la subbanda 19a, la subbanda 19b y la subbanda 19c son los mismos que un valor de la relación de señal a ruido de su subbanda madre. De esta manera, las 20 subbandas que se han obtenido originalmente a través de la división se vuelven a dividir en 24 subbandas. Ya que la VAD está diseñada aún según las 20 subbandas  
45 durante la detección de señal activa, las 24 subbandas necesitan asignarse de nuevo a las 20 subbandas para determinar la SSNR mejorada. En conclusión, cuando la SSNR mejorada se determina aumentando una cantidad de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, el cálculo puede realizarse utilizando la siguiente fórmula:

$$SSNR' = \frac{20}{24} \times \left[ 2 \times (snr(18) + snr(19)) + \sum_{k=0}^{19} snr(k) \right]$$

Fórmula 1.3

donde SSNR' indica la SSNR mejorada, y snr(k) indica una SNR de subbanda de la subbanda k-ésima.

Si una SSNR obtenida a través del cálculo utilizando la fórmula 1.1 es la SSNR de referencia, la SSNR de referencia obtenida a través del cálculo es  $\sum_{k=0}^{19} snr(k)$ . Obviamente, para una señal de audio de un primer tipo, un valor de



la SSNR mejorada obtenido a través del cálculo utilizando la fórmula 1.3 es mayor que un valor de la SSNR de referencia obtenido a través del cálculo utilizando la fórmula 1.1.

5 Para otro ejemplo, si la señal de audio se divide en 20 subbandas, es decir, la subbanda 0 a la subbanda 19, según la teoría psicoacústica,  $snr(18)$  y  $snr(19)$  son ambas mayores que un primer valor  $T1$  predeterminado y  $snr(0)$  a  $snr(17)$  son todos menores que un segundo umbral  $T2$  predeterminado, la SSNR mejorada se puede determinar utilizando la siguiente fórmula:

$$SSNR' = a_1 \times snr(18) + a_2 \times snr(19) + \sum_{k=0}^{17} snr(k)$$

Fórmula 1.4

donde  $SSNR'$  indica la SSNR mejorada,  $snr(k)$  indica una SNR de subbanda de la subbanda  $k$ -ésima,  $a_1$  y  $a_2$  son parámetros que aumentan de peso y valores de  $a_1$  y  $a_2$  hacen  $a_1 \times snr(18) + a_2 \times snr(19)$  mayor que  $snr(18) + snr(19)$ .

10 Obviamente, un valor de la SSNR mejorada obtenido a través del cálculo utilizando la fórmula 1.4 es mayor que el valor de la SSNR de referencia obtenido a través del cálculo utilizando la fórmula 1.1.

304. Comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa.

15 Específicamente, cuando la SSNR mejorada se compara con el umbral de decisión de VAD, si la SSNR mejorada es mayor que el umbral de decisión de VAD, se determina que la señal de audio es una señal activa; o si la SSNR mejorada no es mayor que el umbral de decisión de VAD, se determina que la señal de audio es una señal inactiva.

20 Según el método mostrado en la fig. 3, se puede determinar una característica de una señal de audio, se determina una SSNR mejorada de una manera correspondiente según la característica de la señal de audio, y la SSNR mejorada se compara con un umbral de decisión de VAD. Por lo tanto, puede reducirse una proporción de detección errónea de una señal activa.

Además, la determinación de una señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar según una SNR de subbanda de la señal de audio.

25 Opcionalmente, en una realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de la señal de audio como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que una primera cantidad

30 Opcionalmente, en otra realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de una señal de audio como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que una segunda cantidad, y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad.

40 El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, las estadísticas sobre las SNR de subbanda de las subbandas de porción de alta frecuencia se recopilan en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas finales de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, las estadísticas sobre las SNR de subbanda de las subbandas de extremidades de baja frecuencia se recopilan en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

50 La primera cantidad, la segunda cantidad y la tercera cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que el primera cantidad. Un método para adquirir la segunda cantidad es similar a

un método para adquirir la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o la segunda cantidad puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son menores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son menores que el segundo umbral predeterminado es mayor que la tercera cantidad.

En realizaciones de la fig. 1 a la fig. 3, si una señal de audio de entrada es una señal activa se determina como quien dice utilizando una SSNR mejorada. En un método mostrado en la fig. 4, si una señal de audio de entrada es una señal activa se determina como quien dice reduciendo un umbral de decisión de VAD.

La fig. 4 es un diagrama de flujo esquemático de un método para detectar una señal de audio según una realización de la presente invención.

401. Determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

Opcionalmente, en una realización, la determinación de una señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda que es de la señal de audio y determinado en la operación 201.

Opcionalmente, en una realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de una señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una primera cantidad.

Opcionalmente, en otra realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de una señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una segunda cantidad, y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad.

Opcionalmente, en otra realización, en un caso en el que la señal de audio se determina como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la determinación de una señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas que se encuentran en la señal de audio y cuyos valores de las SNR de subbanda son mayores que un tercer umbral predeterminado es mayor que una cuarta cantidad.

El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, las estadísticas sobre las SNR de subbanda de las subbandas de porción de alta frecuencia se recopilan en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, las estadísticas sobre las SNR de subbanda de las subbandas de extremidades de baja frecuencia se recopilan en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

El tercer umbral predeterminado también se obtiene mediante la recopilación de estadísticas. Específicamente, el tercer umbral predeterminado se determina según las SNR de subbanda de una gran cantidad de señales de ruido, de manera que las SNR de subbanda de la mayoría de las subbandas en estas señales de ruido son menores que el tercer umbral predeterminado.

La primera cantidad, la segunda cantidad, la tercera cantidad y la cuarta cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que la primera cantidad. Un método para adquirir la segunda cantidad es similar a un método para adquirir la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad,

o la segunda cantidad puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son menores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son menores que el segundo umbral predeterminado es mayor que la tercera cantidad. Para la cuarta cantidad, en una gran cantidad de tramas de señal de ruido, se recopilan estadísticas sobre una cantidad de subbandas cuyas SNR de subbanda son menores que el tercer umbral predeterminado, y la cuarta cantidad se determina según la cantidad, de manera que una cantidad de subbandas que se encuentran en la mayoría de estas tramas de muestra de ruido y cuyas SNR de subbanda son menores que el tercer umbral predeterminado es mayor que la cuarta cantidad

Opcionalmente, en otra realización, si la señal de audio de entrada es una señal de audio que se ha de determinar puede determinarse determinando si la señal de audio de entrada es una señal sin voz. En este caso, no es necesario determinar la SNR de subbanda de la señal de audio cuando se determina si la señal de audio es una señal de audio que se ha de determinar. En otras palabras, la operación 201 no necesita realizarse cuando se determina si la señal de audio es una señal de audio que se ha de determinar. Específicamente, la determinación de una señal de audio de entrada como una señal de audio que se ha de determinar incluye: determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que se determina que la señal de audio es una señal sin voz. Específicamente, un experto en la técnica puede comprender que puede haber múltiples métodos para detectar si la señal de audio es una señal sin voz. Por ejemplo, si la señal de audio es una señal sin voz puede determinarse detectando una ZCR de dominio de tiempo de la señal de audio. Específicamente, en un caso en el que la ZCR de la señal de audio es mayor que un umbral de ZCR, se determina que la señal de audio es una señal sin voz, donde el umbral de ZCR se determina según una gran cantidad de experimentos.

402. Adquirir una SSNR de referencia de la señal de audio.

Específicamente, la SSNR de referencia puede ser una SSNR obtenida a través del cálculo utilizando la fórmula 1.1.

403. Utilizar un algoritmo predeterminado para reducir un umbral de decisión de VAD de referencia, con el fin de obtener un umbral de decisión de VAD reducido.

Específicamente, el umbral de decisión de VAD de referencia puede ser un umbral de decisión de VAD por defecto, y el umbral de decisión de VAD de referencia puede almacenarse previamente o puede obtenerse temporalmente a través del cálculo, donde el umbral de decisión de VAD de referencia puede calcularse utilizando una tecnología existente bien conocida. Cuando el umbral de decisión de VAD de referencia se reduce utilizando el algoritmo predeterminado, el algoritmo predeterminado puede estar multiplicando el umbral de decisión de VAD de referencia por un coeficiente que sea menor que 1, o se puede utilizar otro algoritmo. Esta realización de la presente invención no impone limitación en un algoritmo específico utilizado. El umbral de decisión de VAD puede reducirse adecuadamente utilizando el algoritmo predeterminado, de manera que una SSNR mejorada sea mayor que el umbral de decisión de VAD reducido. Por lo tanto, puede reducirse una proporción de detección errónea de una señal activa.

404. Comparar la SSNR de referencia con el umbral de decisión de VAD reducido para determinar si la señal de audio es una señal activa.

Cuando se utiliza un método de cálculo de SSNR convencional para calcular las SSNR de algunas señales de audio, las SSNR de estas señales de audio pueden ser inferiores a un umbral de decisión de VAD predeterminado. Sin embargo, en realidad, estas señales de audio son señales de audio activas. Esto es provocado por las características de estas señales de audio. Por ejemplo, en un caso en el que una SNR ambiental es relativamente baja, una SNR de subbanda de una parte de alta frecuencia se reduce significativamente. Además, ya que una teoría psicoacústica se utiliza generalmente para realizar una división de subbanda, la SNR de subbanda de la parte de alta frecuencia tiene una contribución relativamente baja a una SSNR. En este caso, para algunas señales, como una señal sin voz, cuya energía se centraliza principalmente en una parte de frecuencia relativamente alta, una SSNR obtenida a través del cálculo utilizando el método de cálculo de SSNR convencional puede ser inferior al umbral de decisión de VAD, lo que provoca una detección errónea de una señal activa. Para otro ejemplo, para algunas señales de audio, la distribución de energía de estas señales de audio es relativamente plana en un espectro, pero la energía general de estas señales de audio es relativamente baja. Por lo tanto, en el caso en que una SNR ambiental es relativamente baja, una SSNR obtenida a través del cálculo utilizando el método de cálculo de SSNR convencional puede ser inferior al umbral de decisión de VAD. En el método mostrado en la fig. 4, se utiliza una manera de reducir un umbral de decisión de VAD, de manera que una SSNR obtenida a través del cálculo utilizando el método de cálculo de SSNR convencional es mayor que el umbral de decisión de VAD. Por lo tanto, puede reducirse de manera eficaz una proporción de detección errónea de una señal activa.

La fig. 5 es un diagrama de bloques estructural de un aparato según una realización de la presente invención. El aparato mostrado en la fig. 5 puede realizar todas las operaciones mostradas en la fig. 1 o en la fig. 2. Como se muestra en la fig. 5, un aparato 500 incluye una primera unidad 501 de determinación, una segunda unidad 502 de determinación y una tercera unidad 503 de determinación.

La primera unidad 501 de determinación está configurada para determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

La segunda unidad 502 de determinación está configurada para determinar una relación señal/ruido segmental (SSNR) mejorada de la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia.

- 5 La tercera unidad 503 de determinación está configurada para comparar la SSNR mejorada con un umbral de decisión de detección de actividad de voz (VAD) para determinar si la señal de audio es una señal activa.

10 El aparato 500 mostrado en la fig. 5 puede determinar una característica de una señal de audio de entrada, determinar una SSNR mejorada de una manera correspondiente según la característica de la señal de audio, y comparar la SSNR mejorada con un umbral de decisión de VAD, de manera que pueda reducirse una proporción de detección errónea de una señal activa.

Opcionalmente, en una realización, la primera unidad 501 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar según una SNR de subbanda de la señal de audio.

15 Opcionalmente, en una realización, en un caso en el que la primera unidad 501 de determinación determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la primera unidad 501 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una primera cantidad.

20 Opcionalmente, en otra realización, en un caso en el que la primera unidad 501 de determinación determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la primera unidad 501 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una segunda cantidad y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad.

30 Opcionalmente, en otra realización, en un caso en el que la primera unidad 501 de determinación determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la primera unidad 501 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas que se encuentran en la señal de audio y cuyos valores de las SNR de subbanda son mayores que un tercer umbral predeterminado es mayor que un cuarto cantidad.

35 Opcionalmente, en otra realización, la primera unidad 501 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que se determina que la señal de audio es una señal sin voz. Específicamente, un experto en la técnica puede comprender que puede haber múltiples métodos para detectar si la señal de audio es una señal sin voz. Por ejemplo, si la señal de audio es una señal sin voz puede determinarse detectando una ZCR de dominio de tiempo de la señal de audio. Específicamente, en un caso en el que la ZCR de la señal de audio es mayor que un umbral de ZCR, se determina que la señal de audio es una señal sin voz, donde el umbral de ZCR se determina según una gran cantidad de experimentos.

45 El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, las estadísticas sobre las SNR de subbanda de las subbandas de porción de alta frecuencia se recopilan en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, las estadísticas sobre las SNR de subbanda de las subbandas de extremidades de baja frecuencia se recopilan en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

55 El tercer umbral predeterminado también se obtiene mediante la recopilación de estadísticas. Específicamente, el tercer umbral predeterminado se determina según las SNR de subbanda de una gran cantidad de señales de ruido, de manera que las SNR de subbanda de la mayoría de las subbandas en estas señales de ruido son menores que el tercer umbral predeterminado.

La primera cantidad, la segunda cantidad, la tercera cantidad y la cuarta cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como ejemplo, donde en una gran cantidad de muestras

de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que la primera cantidad. Un método para determinar la segunda cantidad es similar a un método para determinar la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son mayores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el segundo umbral predeterminado es mayor que la tercera cantidad. Para la cuarta cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas cuyas SNR de subbanda son mayores que el tercer umbral predeterminado, y la cuarta cantidad se determina según la cantidad, por manera que una cantidad de subbandas que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el tercer umbral predeterminado es mayor que la cuarta cantidad.

Además, la segunda unidad 502 de determinación está configurada específicamente para determinar un peso de una SNR de subbanda de cada subbanda en la señal de audio, donde el peso de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que el primer umbral predeterminado es mayor que el peso de una SNR de subbanda de otra subbanda, y determinar la SSNR mejorada según la SNR de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio.

Opcionalmente, en una realización, la segunda unidad 502 de determinación está configurada específicamente para determinar una SSNR de referencia de la señal de audio, y determinar la SSNR mejorada según la SSNR de referencia de la señal de audio.

La SSNR de referencia puede ser una SSNR obtenido a través del cálculo utilizando la fórmula 1.1. Cuando se calcula la SSNR de referencia, los pesos de los SNR de subbanda que son de todas las subbandas y que están incluidos en la SSNR son los mismos en la SSNR.

Opcionalmente, en otra realización, la segunda unidad 502 de determinación está configurada específicamente para determinar la SSNR mejorada utilizando la siguiente fórmula:

$$SSNR' = x * SSNR + y \quad \text{Fórmula 1.7}$$

donde SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y x e y indican los parámetros de mejora. Por ejemplo, un valor de x puede ser 1,05, y un valor de y puede ser 1. Un experto en la técnica puede comprender que los valores de x e y pueden ser otros valores adecuados que hacen que la SSNR mejorada sea mayor que la SSNR de referencia correctamente.

Opcionalmente, en otra realización, la segunda unidad 502 de determinación está configurada específicamente para determinar la SSNR mejorada utilizando la siguiente fórmula:

$$SSNR' = f(x) * SSNR + h(y) \quad \text{Fórmula 1.8}$$

donde SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y f(x) y h(y) indican funciones de mejora. Por ejemplo, f(x) y h(y) pueden ser funciones relacionadas con LSNR de la señal de audio, donde la LSNR de la señal de audio es una SNR promedio o una SNR ponderada dentro de un período de tiempo relativamente largo. Por ejemplo, cuando la lsnr es mayor que 20, f(lsnr) puede ser igual a 1.1, e y(lsnr) puede ser igual a 2; cuando lsnr es menor que 20 y mayor que 15, f(lsnr) puede ser igual a 1,05, e y(lsnr) puede ser igual a 1; y cuando lsnr es menor que 15, f(lsnr) puede ser igual a 1, e y(lsnr) puede ser igual a 0. Un experto en la técnica puede comprender que f(x) y h(y) pueden estar en otras formas adecuadas que hacen que la SSNR mejorada sea mayor que la SSNR de referencia correctamente.

La tercera unidad 503 de determinación está configurada específicamente para comparar la SSNR mejorada con el umbral de decisión de VAD para determinar, según un resultado de la comparación, si la señal de audio es una señal activa. Específicamente, si la SSNR mejorada es mayor que el umbral de decisión de VAD, se determina que la señal de audio es una señal activa, o si la SSNR mejorada es menor que el umbral de decisión de VAD, se determina que la señal de audio es una señal inactiva.

Opcionalmente, en otra realización, también se puede utilizar un algoritmo predeterminado para reducir un umbral de decisión de VAD de referencia para obtener un umbral de decisión de VAD reducido, y el umbral de decisión de VAD reducido se utiliza para determinar si la señal de audio es una señal activa. En este caso, el aparato 500 puede incluir además una cuarta unidad 504 de determinación, donde la cuarta unidad 504 de determinación está configurada para utilizar un algoritmo predeterminado para reducir el umbral de decisión de VAD, con el fin de obtener un umbral de

decisión de VAD reducido. En este caso, la tercera unidad 503 de determinación está configurada específicamente para comparar la SSNR mejorada con el umbral de decisión de VAD reducido para determinar si la señal de audio es una señal activa.

5 La fig. 6 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención. El aparato mostrado en la fig. 6 puede realizar todas las operaciones mostradas en la fig. 3. Como se muestra en la fig. 6, un aparato 600 incluye una primera unidad 601 de determinación, una segunda unidad 602 de determinación y una tercera unidad 603 de determinación.

La primera unidad 601 de determinación está configurada para determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

10 La segunda unidad 602 de determinación está configurada para determinar un peso de una SNR de subbanda de cada subbanda en la señal de audio, donde un peso de una SNR de subbanda de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que un primer umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda y determinar una SSNR mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia.

15 La tercera unidad 603 de determinación está configurada para comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa.

20 El aparato 600 mostrado en la fig. 6 puede determinar una característica de una señal de audio de entrada, determinar una SSNR mejorada de una manera correspondiente según la característica de la señal de audio y comparar la SSNR mejorada con un umbral de decisión de VAD, de manera que pueda reducirse una proporción de detección errónea de una señal activa.

Además, la primera unidad 601 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar según una SNR de subbanda de la señal de audio.

25 Opcionalmente, en una realización, la primera unidad 601 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que una primera cantidad.

30 Opcionalmente, en otra realización, la primera unidad 601 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que una segunda cantidad, y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad.

35 El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, las estadísticas sobre las SNR de subbanda de las subbandas de porción de alta frecuencia se recopilan en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, las estadísticas sobre las SNR de subbanda de las subbandas de extremidades de baja frecuencia se recopilan en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

45 La primera cantidad, la segunda cantidad y la tercera cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que el primer cantidad. Un método para adquirir la segunda cantidad es similar a un método para adquirir la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o la segunda cantidad puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son menores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son menores que el segundo umbral predeterminado es mayor que la tercera cantidad.

La fig. 7 es un diagrama de bloques estructural de un aparato según una realización de la presente invención. El aparato mostrado en la fig. 7 puede realizar todas las operaciones mostradas en la fig. 1 o en la fig. 2. Como se muestra en la fig. 7, un aparato 700 incluye un procesador 701 y una memoria 702. El procesador 701 puede ser un procesador de propósito general, un procesador de señal digital (DSP), un circuito integrado específico de aplicación (ASIC), una matriz de puertas lógicas programables en campo (FPGA) u otro componente lógico programable, una puerta discreta o un componente lógico de transistor, o un componente de hardware discreto, que puede implementar o realizar los métodos, las operaciones y los diagramas de bloques lógicos descritos en las realizaciones de la presente invención. El procesador de propósito general puede ser un microprocesador o el procesador puede ser cualquier procesador convencional o similar. Las operaciones de los métodos descritos en las realizaciones de la presente invención pueden ejecutarse directamente mediante un procesador de decodificación de hardware, o ejecutarse mediante una combinación de módulos de hardware y software en un procesador de decodificación. El módulo de software puede ubicarse en un medio de almacenamiento maduro en la técnica, tal como una memoria de acceso aleatorio (RAM), una memoria flash, una memoria de solo lectura (ROM), una memoria de solo lectura programable, una memoria programable que se puede borrar eléctricamente, o un registro. El medio de almacenamiento está ubicado en la memoria 702. El procesador 701 lee una instrucción de la memoria 702 y completa las operaciones de los métodos anteriores en combinación con el hardware.

El procesador 701 está configurado para determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

El procesador 701 está configurado para determinar una SSNR mejorada de la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia.

El procesador 701 está configurado para comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa.

El aparato 700 mostrado en la fig. 7 puede determinar una característica de una señal de audio de entrada, determinar una SSNR mejorada de una manera correspondiente según la característica de la señal de audio, y comparar la SSNR mejorada con un umbral de decisión de VAD, de manera que pueda reducirse una proporción de detección errónea de una señal activa.

Opcionalmente, en una realización, el procesador 701 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar según una SNR de subbanda de la señal de audio.

Opcionalmente, en una realización, en un caso en el que el procesador 701 determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, el procesador 701 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una primera cantidad.

Opcionalmente, en otra realización, en un caso en el que el procesador 701 determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, el procesador 701 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una segunda cantidad, y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad.

Opcionalmente, en otra realización, en un caso en el que el procesador 701 determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, el procesador 701 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas que se encuentran en la señal de audio y cuyos valores de las SNR de subbanda son mayores que un tercer umbral predeterminado es mayor que una cuarta cantidad.

Opcionalmente, en otra realización, el procesador 701 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que se determina que la señal de audio es una señal sin voz. Específicamente, un experto en la técnica puede comprender que puede haber múltiples métodos para detectar si la señal de audio es una señal sin voz. Por ejemplo, si la señal de audio es una señal sin voz puede determinarse detectando una ZCR de dominio de tiempo de la señal de audio. Específicamente, en un caso en el que la ZCR de la señal de audio es mayor que un umbral de ZCR, se determina que la señal de audio es una señal sin voz, donde el umbral de ZCR se determina según una gran cantidad de experimentos.

El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, se recopilan estadísticas sobre las SNR de subbanda de las subbandas de porción de alta frecuencia en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, se recopilan estadísticas sobre las SNR de subbanda de las

subbandas de extremidades de baja frecuencia en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

5 El tercer umbral predeterminado también se obtiene mediante la recopilación de estadísticas. Específicamente, el tercer umbral predeterminado se determina según las SNR de subbanda de una gran cantidad de señales de ruido, de manera que las SNR de subbanda de la mayoría de las subbandas en estas señales de ruido son menores que el tercer umbral predeterminado.

10 La primera cantidad, la segunda cantidad, la tercera cantidad y la cuarta cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que la primera cantidad. Un método para determinar la segunda cantidad es similar a un método para determinar la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son mayores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el segundo umbral predeterminado es mayor que la tercera cantidad. Para la cuarta cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas cuyas SNR de subbanda son mayores que el tercer umbral predeterminado, y la cuarta cantidad se determina según la cantidad, de manera que una cantidad de subbandas que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el tercer umbral predeterminado es mayor que la cuarta cantidad.

25 Además, el procesador 701 está configurado específicamente para determinar un peso de una SNR de subbanda de cada subbanda en la señal de audio, donde un peso de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que la primera el umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda y determinar la SSNR mejorada según la SNR de cada subbanda y el peso de la SNR de subbanda de cada subbanda en el audio señal.

Opcionalmente, en una realización, el procesador 701 está configurado específicamente para determinar una SSNR de referencia de la señal de audio, y determinar la SSNR mejorada según la SSNR de referencia de la señal de audio.

35 La SSNR de referencia puede ser una SSNR obtenida a través del cálculo utilizando la fórmula 1.1. Cuando se calcula la SSNR de referencia, los pesos de los SNR de subbanda que son de todas las subbandas y que están incluidos en la SSNR son los mismos en la SSNR.

Opcionalmente, en otra realización, el procesador 701 está configurado específicamente para determinar la SSNR mejorada utilizando la siguiente fórmula:

$$SSNR' = x * SSNR + y \tag{Fórmula 1.7}$$

40 donde SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y x e y indican los parámetros de mejora. Por ejemplo, un valor de x puede ser 1,07, y un valor de y puede ser 1. Un experto en la técnica puede comprender que los valores de x e y pueden ser otros valores adecuados que hacen que la SSNR mejorada sea mayor que la SSNR de referencia correctamente.

Opcionalmente, en otra realización, el procesador 701 está configurado específicamente para determinar la SSNR mejorada utilizando la siguiente fórmula:

$$45 \quad SSNR' = f(x) * SSNR + h(y) \tag{Fórmula 1.8}$$

50 donde SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y f(x) y h(y) indican funciones de mejora. Por ejemplo, f(x) y h(y) pueden ser funciones relacionadas con una LSNR de la señal de audio, donde la LSNR de la señal de audio es una SNR promedio o una SNR ponderada dentro de un período de tiempo relativamente largo. Por ejemplo, cuando la lsnr es mayor que 20, f(lsnr) puede ser igual a 1.1, e y(lsnr) puede ser igual a 2; cuando la lsnr es menor que 20 y mayor que 17, f(lsnr) puede ser igual a 1,07, e y(lsnr) puede ser igual a 1; y cuando la lsnr es menor que 17, f(lsnr) puede ser igual a 1 e y(lsnr) puede ser igual a 0. Un experto en la técnica puede comprender que f(x) y h(y) pueden estar en otras formas adecuadas que hacen que la SSNR mejorada sea mayor que la SSNR de referencia correctamente.

El procesador 701 está configurado específicamente para comparar la SSNR mejorada con el umbral de decisión de



VAD para determinar, según un resultado de la comparación, si la señal de audio es una señal activa. Específicamente, si la SSNR mejorada es mayor que el umbral de decisión de VAD, se determina que la señal de audio es una señal activa, o si la SSNR mejorada es menor que el umbral de decisión de VAD, se determina que la señal de audio es una señal inactiva.

- 5 Opcionalmente, en otra realización, también se puede utilizar un algoritmo predeterminado para reducir un umbral de decisión de VAD de referencia para obtener un umbral de decisión de VAD reducido, y el umbral de decisión de VAD reducido se utiliza para determinar si la señal de audio es una señal activa. En este caso, el procesador 701 puede configurarse adicionalmente para utilizar un algoritmo predeterminado para reducir el umbral de decisión de VAD, con el fin de obtener un umbral de decisión de VAD reducido. En este caso, el procesador 701 está configurado  
10 específicamente para comparar la SSNR mejorada con el umbral de decisión de VAD reducido para determinar si la señal de audio es una señal activa.

La fig. 8 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención. El aparato mostrado en la fig. 8 puede realizar todas las operaciones mostradas en la fig. 3. Como se muestra en la fig. 8, un aparato 800 incluye un procesador 801 y una memoria 802. El procesador 801 puede ser un procesador de propósito general, un DSP, un ASIC, una FPGA u otro componente lógico programable, una puerta discreta o un componente lógico de transistor, o un componente de hardware discreto, que puede implementar o realizar los métodos, las operaciones y los diagramas de bloques lógicos descritos en las realizaciones de la presente invención. El procesador de propósito general puede ser un microprocesador o el procesador puede ser cualquier procesador convencional o similar. Las operaciones de los métodos descritos en las realizaciones de la presente invención pueden  
15 ejecutarse directamente mediante un procesador de decodificación de hardware, o ejecutarse mediante una combinación de módulos de hardware y software en un procesador de decodificación. El módulo de software puede ubicarse en un medio de almacenamiento maduro en la técnica, tal como una RAM, una memoria flash, una ROM, una memoria de solo lectura programable, una memoria programable que se puede borrar eléctricamente, o un registro. El medio de almacenamiento se encuentra en la memoria 802. El procesador 801 lee una instrucción de la memoria 802 y completa las operaciones de los métodos anteriores en combinación con el hardware.  
20  
25

El procesador 801 está configurado para determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

El procesador 801 está configurado para determinar un peso de una SNR de subbanda de cada subbanda en la señal de audio, donde un peso de una SNR de subbanda de una subbanda de porción de alta frecuencia cuya SNR de subbanda es mayor que un primer umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda y determinar una SSNR mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia.  
30

El procesador 801 está configurado para comparar la SSNR mejorada con un umbral de decisión de VAD para determinar si la señal de audio es una señal activa.

35 El aparato 800 mostrado en la fig. 8 puede determinar una característica de una señal de audio de entrada, determinar una SSNR mejorada de manera correspondiente según la característica de la señal de audio y comparar la SSNR mejorada con un umbral de decisión de VAD, de manera que pueda reducirse una proporción de detección errónea de una señal activa.

40 Además, el procesador 801 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar según una SNR de subbanda de la señal de audio.

Opcionalmente, en una realización, el procesador 801 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que una primera cantidad.

45 Opcionalmente, en otra realización, el procesador 801 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que una segunda cantidad, y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad.  
50

El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, se recopilan estadísticas sobre las SNR de subbanda de subbandas de porción de alta frecuencia en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, se recopilan estadísticas sobre las SNR de subbanda de las subbandas de extremidades de baja frecuencia en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de  
55

extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

La primera cantidad, la segunda cantidad y la tercera cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que el primera cantidad. Un método para adquirir la segunda cantidad es similar a un método para adquirir la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o la segunda cantidad puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de tramas de muestra sin voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son menores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas tramas de muestra sin voz y cuyas SNR de subbanda son menores que el segundo umbral predeterminado es mayor que la tercera cantidad.

La fig. 9 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención. Un aparato 900 mostrado en la fig. 9 puede realizar todas las operaciones mostradas en la fig. 4. Como se muestra en la fig. 9, el aparato 900 incluye una primera unidad 901 de determinación, una segunda unidad 902 de determinación, una tercera unidad 903 de determinación y una cuarta unidad 904 de determinación.

La primera unidad 901 de determinación está configurada para determinar una señal de audio de entrada como una señal de audio que se ha de determinar.

La segunda unidad 902 de determinación está configurada para adquirir una SSNR de referencia de la señal de audio.

Específicamente, la SSNR de referencia puede ser una SSNR obtenida a través del cálculo utilizando la fórmula 1.1.

La tercera unidad 903 de determinación está configurada para utilizar un algoritmo predeterminado para reducir un umbral de decisión de VAD de referencia, con el fin de obtener un umbral de decisión de VAD reducido.

Específicamente, el umbral de decisión de VAD de referencia puede ser un umbral de decisión de VAD por defecto, y el umbral de decisión de VAD de referencia puede almacenarse previamente o puede obtenerse temporalmente a través del cálculo, donde el umbral de decisión de VAD de referencia puede calcularse utilizando una tecnología existente bien conocida. Cuando el umbral de decisión de VAD de referencia se reduce utilizando el algoritmo predeterminado, el algoritmo predeterminado puede estar multiplicando el umbral de decisión de VAD de referencia por un coeficiente que sea menor que 1, o se puede utilizar otro algoritmo. Esta realización de la presente invención no impone limitación en un algoritmo específico utilizado. El umbral de decisión de VAD puede reducirse adecuadamente utilizando el algoritmo predeterminado, de manera que la SSNR mejorada sea mayor que el umbral de decisión de VAD reducido. Por lo tanto, puede reducirse una proporción de detección errónea de una señal activa.

La cuarta unidad 904 de determinación está configurada para comparar la SSNR de referencia con el umbral de decisión de VAD reducido para determinar si la señal de audio es una señal activa.

Opcionalmente, en una realización, la primera unidad 901 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar según una SNR de subbanda de la señal de audio.

Opcionalmente, en una realización, en un caso en el que la primera unidad 901 de determinación determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la primera unidad 901 de determinación está configurada específicamente para determinar el audio señal como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una primera cantidad .

Opcionalmente, en una realización, en un caso en el que la primera unidad 901 de determinación determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la primera unidad 901 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una segunda cantidad, y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son inferiores a un segundo umbral predeterminado es mayor que una tercera cantidad.

Opcionalmente, en una realización, en un caso en el que la primera unidad 901 de determinación determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, la primera unidad 901 de determinación está configurada específicamente para determinar la señal de audio como una señal de

audio que se ha de determinar en un caso en el que una cantidad de subbandas que se encuentran en la señal de audio y cuyos valores de las SNR de subbanda son mayores que un tercer umbral predeterminado es mayor que una cuarta cantidad.

5 Opcionalmente, en una realización, la primera unidad 901 de determinación está configurada específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que se determina que la señal de audio es una señal sin voz. Específicamente, Un experto en la técnica puede comprender que puede haber múltiples métodos para detectar si la señal de audio es una señal sin voz. Por ejemplo, si la señal de audio es una señal sin voz puede determinarse detectando una ZCR de la señal de audio. Específicamente, en un caso en el que la ZCR de la señal de audio es mayor que un umbral de ZCR, se determina que la señal de audio es una señal sin voz, donde el umbral de ZCR se determina según una gran cantidad de experimentos.

10 El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, se recopilan estadísticas sobre las SNR de subbanda de las subbandas de porción de alta frecuencia en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, se recopilan estadísticas sobre las SNR de subbanda de las subbandas de extremidades de baja frecuencia en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

20 El tercer umbral predeterminado también se obtiene mediante la recopilación de estadísticas. Específicamente, el tercer umbral predeterminado se determina según las SNR de subbanda de una gran cantidad de señales de ruido, de manera que las SNR de subbanda de la mayoría de las subbandas en estas señales de ruido son menores que el tercer umbral predeterminado.

25 La primera cantidad, la segunda cantidad, la tercera cantidad y la cuarta cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que la primera cantidad. Un método para determinar la segunda cantidad es similar a un método para determinar la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son mayores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el segundo umbral predeterminado es mayor que la tercera cantidad. Para la cuarta cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas cuyas SNR de subbanda son mayores que el tercer umbral predeterminado, y la cuarta cantidad se determina según la cantidad, de manera que una cantidad de subbandas que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el tercer umbral predeterminado es mayor que la cuarta cantidad.

45 El aparato 900 mostrado en la fig. 9 puede determinar una característica de una señal de audio de entrada, reducir un umbral de decisión de VAD de referencia según la característica de la señal de audio y comparar una SSSNR mejorada con un umbral de decisión de VAD reducido, de manera que pueda reducirse una proporción de detección errónea de una señal activa.

La fig. 10 es un diagrama de bloques estructural de otro aparato según una realización de la presente invención. Un aparato 1000 mostrado en la fig. 10 puede realizar todas las operaciones mostradas en la fig. 4. Como se muestra en la FIG. 10, el aparato 1000 incluye un procesador 1001 y una memoria 1002. El procesador 1001 puede ser un procesador de propósito general, un DSP, un ASIC, una FPGA u otro componente lógico programable, una puerta discreta o un componente lógico de transistor, o un componente de hardware discreto, que puede implementar o realizar los métodos, las operaciones y los diagramas de bloques lógicos descritos en las realizaciones de la presente invención. El procesador de propósito general puede ser un microprocesador o el procesador puede ser cualquier procesador convencional o similar. Las operaciones de los métodos descritos en las realizaciones de la presente invención pueden ejecutarse directamente por un procesador de decodificación de hardware, o ejecutarse por una combinación de módulos de hardware y software en un procesador de decodificación. El módulo de software puede ubicarse en un medio de almacenamiento maduro en la técnica, tal como una RAM, una memoria flash, una ROM, una memoria de solo lectura programable, una memoria programable que se puede borrar eléctricamente, o un registro. El medio de almacenamiento está ubicado en la memoria 1002. El procesador 1001 lee una instrucción de la memoria 1002 y completa las operaciones de los métodos anteriores en combinación con el hardware.

60 El procesador 1001 está configurado para determinar una señal de audio de entrada como una señal de audio que se

ha de determinar.

El procesador 1001 está configurado para adquirir una SSNR de referencia de la señal de audio.

Específicamente, la SSNR de referencia puede ser una SSNR obtenida a través del cálculo utilizando la fórmula 1.1.

5 El procesador 1001 está configurado para utilizar un algoritmo predeterminado para reducir un umbral de decisión de VAD de referencia, con el fin de obtener un umbral de decisión de VAD reducido.

10 Específicamente, el umbral de decisión de VAD de referencia puede ser un umbral de decisión de VAD por defecto, y el umbral de decisión de VAD de referencia puede almacenarse previamente o puede obtenerse temporalmente a través del cálculo, donde el umbral de decisión de VAD de referencia puede calcularse utilizando una tecnología existente bien conocida. Cuando el umbral de decisión de VAD de referencia se reduce utilizando el algoritmo predeterminado, el algoritmo predeterminado puede estar multiplicando el umbral de decisión de VAD de referencia por un coeficiente que sea menor que 1, o se puede utilizar otro algoritmo. Esta realización de la presente invención no impone limitación en un algoritmo específico utilizado. El umbral de decisión de VAD puede reducirse adecuadamente utilizando el algoritmo predeterminado, de manera que una SSNR mejorada sea mayor que el umbral de decisión de VAD reducido. Por lo tanto, puede reducirse una proporción de detección errónea de una señal activa.

15 El procesador 1001 está configurado para comparar la SSNR de referencia con el umbral de decisión de VAD reducido para determinar si la señal de audio es una señal activa.

Opcionalmente, en una realización, el procesador 1001 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar según una SNR de subbanda de la señal de audio.

20 Opcionalmente, en una realización, en un caso en el que el procesador 1001 determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, el procesador 1001 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una primera cantidad.

25 Opcionalmente, en una realización, en un caso en el que el procesador 1001 determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, el procesador 1001 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son mayores que un primer umbral predeterminado es mayor que una segunda cantidad, y una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la señal de audio y cuyas SNR de subbanda son menores que un segundo umbral predeterminado es mayor que una tercera cantidad.

30 Opcionalmente, en una realización, en un caso en el que el procesador 1001 determina la señal de audio como una señal de audio que se ha de determinar según la SNR de subbanda de la señal de audio, el procesador 1001 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que una cantidad de subbandas que se encuentran en la señal de audio y cuyos valores de las SNR de subbanda son mayores que un tercer umbral predeterminado es mayor que una cuarta cantidad.

35 Opcionalmente, en una realización, el procesador 1001 está configurado específicamente para determinar la señal de audio como una señal de audio que se ha de determinar en un caso en el que se determina que la señal de audio es una señal sin voz. Específicamente, Un experto en la técnica puede comprender que puede haber múltiples métodos para detectar si la señal de audio es una señal sin voz. Por ejemplo, si la señal de audio es una señal sin voz puede determinarse detectando una ZCR de la señal de audio. Específicamente, en un caso en el que la ZCR de la señal de audio es mayor que un umbral de ZCR, se determina que la señal de audio es una señal sin voz, donde el umbral de ZCR se determina según una gran cantidad de experimentos.

40 El primer umbral predeterminado y el segundo umbral predeterminado pueden obtenerse mediante la recopilación de estadísticas según una gran cantidad de muestras de voz. Específicamente, se recopilan estadísticas sobre las SNR de subbanda de subbandas de porción de alta frecuencia en una gran cantidad de muestras sin voz, incluyendo el ruido de fondo, y el primer umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de porción de alta frecuencia en estas muestras sin voz son mayores que el primer umbral predeterminado. De manera similar, se recopilan estadísticas sobre las SNR de subbanda de las subbandas de extremidades de baja frecuencia en estas muestras sin voz, y el segundo umbral predeterminado se determina según las SNR de subbanda, de manera que las SNR de subbanda de la mayoría de las subbandas de extremidades de baja frecuencia en estas muestras sin voz son menores que el segundo umbral predeterminado.

45 El tercer umbral predeterminado también se obtiene mediante la recopilación de estadísticas. Específicamente, el tercer umbral predeterminado se determina según las SNR de subbanda de una gran cantidad de señales de ruido, de manera que las SNR de subbanda de la mayoría de las subbandas en estas señales de ruido son menores que el tercer umbral predeterminado.

La primera cantidad, la segunda cantidad, la tercera cantidad y la cuarta cantidad también se obtienen mediante la recopilación de estadísticas. La primera cantidad se utiliza como un ejemplo, donde en una gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de porción de alta frecuencia cuyas SNR de subbanda son mayores que el primer umbral predeterminado, y la primera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de porción de alta frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el primer umbral predeterminado es mayor que la primera cantidad. Un método para determinar la segunda cantidad es similar a un método para determinar la primera cantidad. La segunda cantidad puede ser la misma que la primera cantidad, o puede ser diferente de la primera cantidad. De manera similar, para la tercera cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas de subbandas de extremidades de baja frecuencia cuyas SNR de subbanda son mayores que el segundo umbral predeterminado, y la tercera cantidad se determina según la cantidad, de manera que una cantidad de subbandas de extremidades de baja frecuencia que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el segundo umbral predeterminado es mayor que la tercera cantidad. Para la cuarta cantidad, en la gran cantidad de muestras de voz, incluyendo el ruido, se recopilan estadísticas sobre una cantidad de subbandas cuyas SNR de subbanda son mayores que el tercer umbral predeterminado, y la cuarta cantidad se determina según la cantidad, de manera que una cantidad de subbandas que se encuentran en la mayoría de estas muestras de voz y cuyas SNR de subbanda son mayores que el tercer umbral predeterminado es mayor que la cuarta cantidad.

El aparato 1000 mostrado en la fig. 10 puede determinar una característica de una señal de audio de entrada, reducir un umbral de decisión de VAD de referencia según la característica de la señal de audio y comparar una SSNR mejorada con un umbral de decisión de VAD reducido, de manera que pueda reducirse una proporción de detección errónea de una señal activa.

Un experto en la técnica puede comprender claramente que, con el propósito de una descripción conveniente y breve, para un proceso de trabajo detallado del sistema, aparato y unidad anteriores, se puede hacer referencia a un proceso correspondiente en las realizaciones del método anterior, y los detalles no se describen en la presente memoria nuevamente.

En las diversas realizaciones proporcionadas en la presente solicitud, debería comprenderse que el sistema, el aparato y el método descritos pueden implementarse de otras maneras. Por ejemplo, la realización del aparato descrita es simplemente ejemplar. Por ejemplo, la división de unidades es simplemente división de función lógica y puede ser otra división en la implementación real. Por ejemplo, una pluralidad de unidades o componentes pueden combinarse o integrarse en otro sistema, o algunas características pueden ignorarse o no realizarse. Además, los acoplamientos mutuos mostrados o descritos o los acoplamientos directos o las conexiones de comunicación pueden implementarse utilizando algunas interfaces. Los acoplamientos indirectos o las conexiones de comunicación entre los aparatos o unidades pueden implementarse de forma electrónica, mecánica o de otras formas.

Las unidades descritas como partes separadas pueden estar o no físicamente separadas, y las partes mostradas como unidades pueden ser o no unidades físicas, pueden estar ubicadas en una posición o pueden distribuirse en una pluralidad de unidades de red. Algunas o todas las unidades pueden seleccionarse según las necesidades reales para lograr los objetivos de las soluciones de las realizaciones.

Además, las unidades funcionales en las realizaciones de la presente invención pueden integrarse en una unidad de procesamiento, o cada una de las unidades puede existir solo físicamente, o dos o más unidades están integradas en una unidad.

Cuando las funciones se implementan en forma de una unidad funcional de software y se venden o utilizan como un producto independiente, las funciones pueden almacenarse en un medio de almacenamiento legible por ordenador. Basándose en tal comprensión, las soluciones técnicas de la presente invención esencialmente, o la parte que contribuye a la técnica anterior, o una parte de las soluciones técnicas pueden implementarse en forma de un producto de software. El producto de software se almacena en un medio de almacenamiento e incluye varias instrucciones para instruir a un dispositivo informático (que puede ser un ordenador personal, un servidor o un dispositivo de red) o un procesador para realizar todas o parte de las operaciones de los métodos descritos en las realizaciones de la presente invención. El medio de almacenamiento anterior incluye: cualquier medio que pueda almacenar código de programa, tal como una unidad flash USB, un disco duro extraíble, una ROM, una RAM, un disco magnético o un disco óptico.

Las descripciones anteriores son simplemente realizaciones específicas de la presente invención, pero no pretenden limitar la presente invención.

**REIVINDICACIONES**

1. Un método para detectar una señal activa, en donde el método comprende:

cuando se determina que una señal de audio es una señal sin voz,

5 determinar (102) una relación señal/ruido segmental, SSNR, mejorada de la señal de audio, en donde la SSNR mejorada es mayor que una SSNR de referencia y la SSNR de referencia se calcula sumando todas las SNR de subbanda de la señal de audio; y

comparar (103) la SSNR mejorada con un umbral de decisión de detección de actividad de voz, VAD, para determinar si la señal de audio es una señal activa,

en donde determinar (102) la SSNR mejorada de la señal de audio comprende:

10 determinar la SSNR de referencia de la señal de audio; y

determinar la SSNR mejorada según la SSNR de referencia de la señal de audio.

2. El método según la reivindicación 1, en donde la determinación de la SSNR mejorada según la SSNR de referencia de la señal de audio comprende:

Determinar la SSNR mejorada utilizando la siguiente fórmula

$$15 \quad SSNR' = x * SSNR + y ,$$

en donde

SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y x e y indican los parámetros de mejora.

3. Un método para detectar una señal activa, en donde el método comprende:

Cuando se determinar que una señal de audio es una señal sin voz,

20 determinar (302) un peso de una relación señal/ruido, SNR, de subbanda, de cada subbanda en la señal de audio, en donde un peso de una SNR de subbanda de una subbanda de porción de alta frecuencia cuya SNR es mayor que un primer umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda;

25 determinar (303) una relación señal/ruido segmental, SSNR, mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia y la SSNR de referencia se calcula sumando todas las SNR de subbanda de la señal de audio;

comparar (304) la SSNR mejorada con un umbral de decisión de detección de actividad de voz, VAD, para determinar si la señal de audio es una señal activa.

4. Un aparato para detectar una señal activa, comprendiendo el aparato:

30 una primera unidad (501) de determinación, configurada para determinar si una señal de audio es una señal sin voz;

una segunda unidad (502) de determinación configurada para determinar una relación señal/ruido segmental, SSNR, mejorada de la señal de audio, donde la SSNR mejorada es mayor que una SSNR de referencia, y la SSNR de referencia se calcula sumando todas las SNR de subbanda de la señal de audio; y

35 una tercera unidad (503) de determinación, configurada para comparar la SSNR mejorada con un umbral de decisión de detección de actividad de voz, VAD, para determinar si la señal de audio es una señal activa, si la señal de audio es una señal sin voz;

en donde la segunda unidad (502) de determinación está configurada específicamente para determinar la SSNR de referencia de la señal de audio, y determinar la SSNR mejorada según la SSNR de referencia de la señal de audio.

40 5. El aparato según la reivindicación 4, en donde la segunda unidad (502) de determinación está configurada específicamente para determinar la SSNR mejorada utilizando la siguiente fórmula:

$$SSNR' = x * SSNR + y ,$$

en donde

SSNR indica la SSNR de referencia, SSNR' indica la SSNR mejorada, y x e y indican los parámetros de mejora.

6. Un aparato para detectar una señal activa, comprendiendo el aparato:

una primera unidad (601) de determinación, configurada para determinar una señal de audio es una señal sin voz;

- 5 una segunda unidad (602) de determinación, configurada para determinar un peso de una relación señal/ruido, SNR, de subbanda de cada subbanda en la señal de audio, en donde un peso de una SNR de subbanda de una subbanda de porción de alta frecuencia cuya SNR es mayor que un primer umbral predeterminado es mayor que un peso de una SNR de subbanda de otra subbanda y determinar una relación señal/ruido segmental, SSNR, mejorada según la SNR de subbanda de cada subbanda y el peso de la SNR de subbanda de cada subbanda en la señal de audio, en donde la SSNR mejorada es mayor que una SSNR de referencia, y la SSNR de referencia se calcula sumando todas las SNR de subbanda de la señal de audio; y
- 10 una tercera unidad (603) de determinación, configurada para comparar la SSNR mejorada con un umbral de decisión de detección de actividad de voz, VAD, para determinar si la señal de audio es una señal activa, si la señal de audio es una señal sin voz.

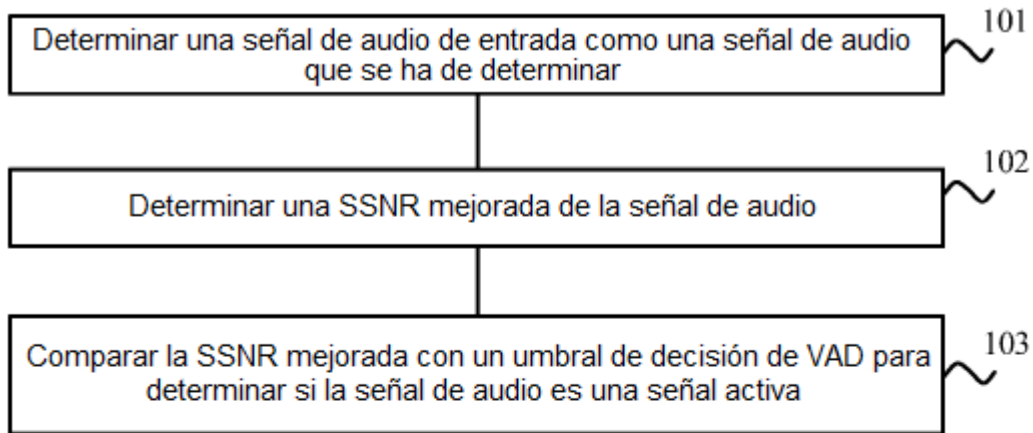


FIG. 1

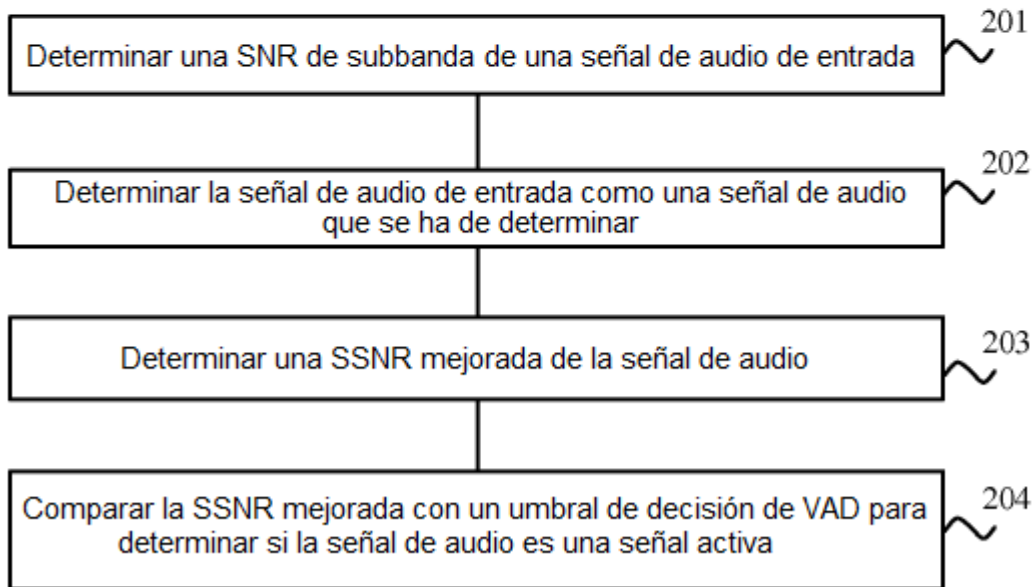


FIG. 2



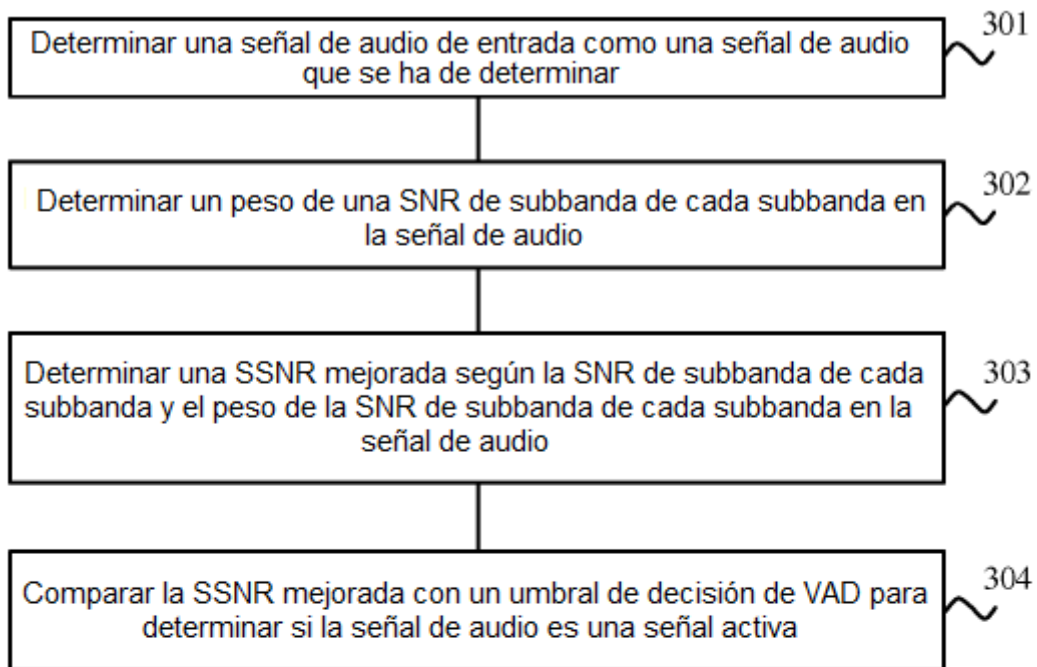


FIG. 3

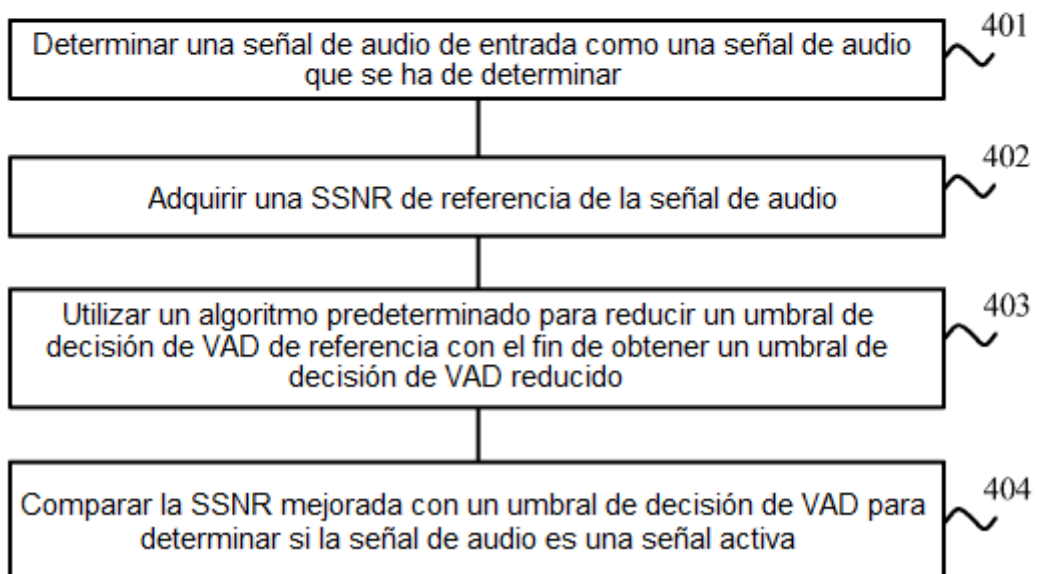


FIG. 4

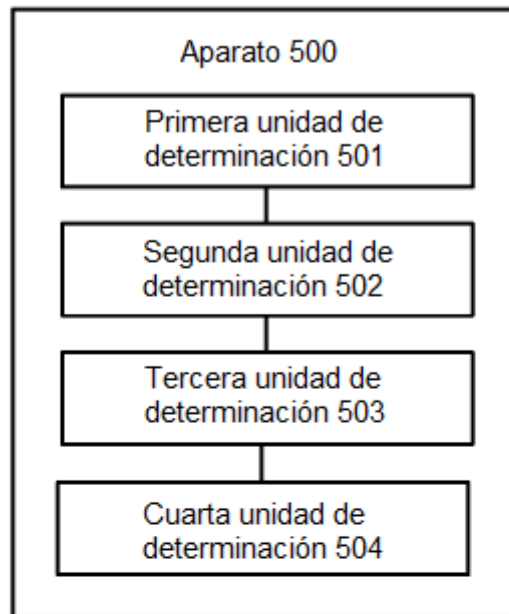


FIG. 5

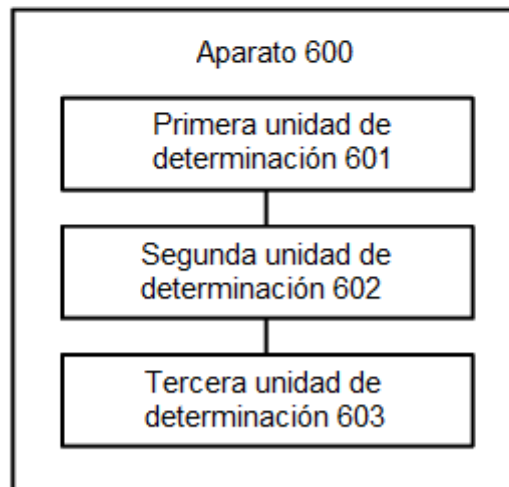


FIG. 6

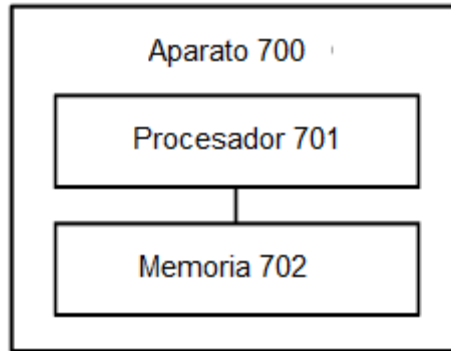


FIG. 7

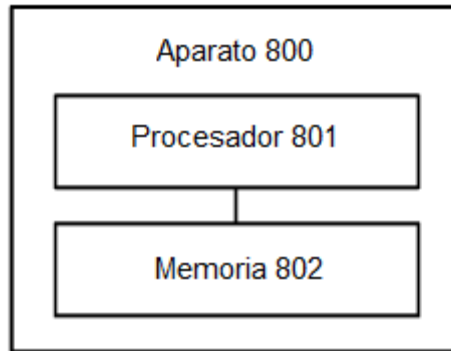


FIG. 8

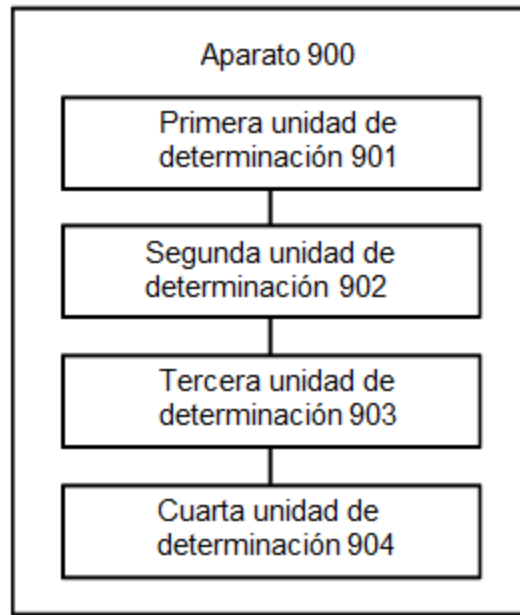


FIG. 9

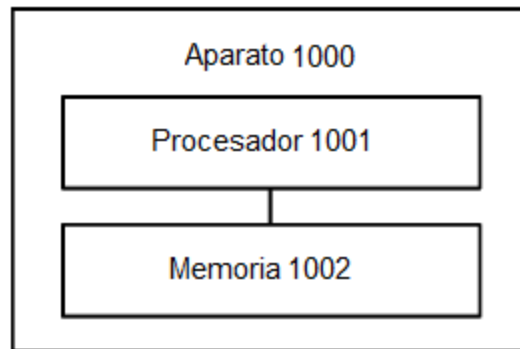


FIG. 10