

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 609 958**

51 Int. Cl.:

**G10L 25/78** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.10.2010 PCT/CN2010/077726**

87 Fecha y número de publicación internacional: **21.04.2011 WO11044842**

96 Fecha de presentación y número de la solicitud europea: **14.10.2010 E 10821452 (9)**

97 Fecha y número de publicación de la concesión europea: **28.09.2016 EP 2346027**

54 Título: **Método y aparato para realizar una detección de actividad vocal**

30 Prioridad:

**15.10.2009 CN 200910207311**

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

**25.04.2017**

73 Titular/es:

**HUAWEI TECHNOLOGIES CO., LTD. (100.0%)  
B1-3A Intellectual Property Department, Huawei  
Administration Building, Bantian, Longgang  
District  
Shenzhen, Guangdong 518129, CN**

72 Inventor/es:

**WANG, ZHE y  
ZHANG, QING**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

**ES 2 609 958 T3**

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

**DESCRIPCIÓN**

Método y aparato para realizar una detección de actividad vocal

5 **CAMPO DE LA INVENCION**

La presente invención se refiere a tecnologías de comunicaciones y en particular, a un método y un aparato para Detección de Actividad Vocal (VAD).

10 **ANTECEDENTES DE LA INVENCION**

En un sistema de comunicaciones, en particular en un sistema de comunicaciones inalámbricas o un sistema de comunicaciones móviles, el ancho de banda del canal es un recurso no frecuente. En conformidad con la estadística, en una llamada bidireccional, el tiempo de conversación para ambas partes de la llamada solamente representa aproximadamente la mitad del tiempo de conversación total, y la llamada en la otra mitad del tiempo de conversación total está en un estado de silencio. Puesto que el sistema de comunicaciones solamente transmite señales cuando una persona está hablando e interrumpe la transmisión de señales en el estado de silencio, pero no puede asignar el ancho de banda ocupado en el estado de silencio a otros servicios de comunicaciones, lo que hace que se desperdicie, en gran medida, los recursos de ancho de banda de canal limitados.

Para hacer pleno uso de los recursos del canal, en la técnica anterior, el tiempo cuando las dos partes de la llamada comienzan a hablar y cuando dejan de hablar se detecta utilizando una tecnología de detección VAD, es decir, el momento cuando se activa la voz es objeto de adquisición, con el fin de asignar el ancho de banda del canal a otros servicios de comunicaciones cuando la voz no está activada. Con el desarrollo de la red de comunicaciones, la tecnología de detección VAD puede detectar también señales de entrada, tales como tonos de llamada. En un sistema de detección VAD basado en la tecnología de VAD, se suele determinar que las señales de entrada son señales de primer plano o ruido de fondo en conformidad con el criterio de decisión preestablecido que incluye parámetros de decisión y lógica de decisión. Las señales de primer plano incluyen señales vocales, señales musicales y señales de Multifrecuencia de Doble Tono (DTMF) y los ruidos de fondo no incluyen las señales. Dicho proceso de determinación se denomina también una decisión VAD.

En la etapa inicial del desarrollo de la tecnología VAD, se adopta un criterio de decisión de carácter estático, es decir, sin importar cuáles sean las características de una señal de entrada, los parámetros de decisión y la lógica de decisión de la VAD permanecen sin cambiar. A modo de ejemplo, en la tecnología VAD basada en la norma G.729, sea cual fuere el tipo de la señal de entrada, la relación de señal a ruido (SNR) es, y las características del ruido de fondo, el mismo grupo de parámetros de decisión que se utilizan para realizar la decisión VAD con el mismo grupo de lógica de decisión y umbrales de decisión. Puesto que la tecnología VAD basada en la norma G.729 está diseñada y presentada sobre la base de una condición de alta relación SNR, el rendimiento de la tecnología VAD es peor en una condición de relación SNR baja. Con el desarrollo de la tecnología VAD, se ofrece un criterio de decisión de carácter dinámico, en donde la tecnología VAD puede seleccionar diferentes parámetros de decisión y/o diferentes umbrales de decisión en conformidad con características diferentes de la señal de entrada y determinar que la señal de entrada es una señal de primer plano o un ruido de fondo. Puesto que el criterio de decisión de carácter dinámico se adopta para determinar los parámetros de decisión o la lógica de decisión en conformidad con las características específicas de la señal de entrada, el proceso de decisión se optimiza y se mejora la eficiencia de la decisión y la exactitud de la decisión, con lo que se mejora también el rendimiento de la decisión de VAD. Además, si se adopta el criterio de decisión de carácter dinámico, diferentes salidas de VAD pueden establecerse para la señal de entrada con diferentes características, en función de las demandas de aplicación específicas. A modo de ejemplo, cuando un operador espera transmitir información de fondo sobre algunas personas que hablan en el sistema VAD en alguna medida, una tendencia de decisión de VAD puede establecerse en el caso de que el ruido de fondo contenga una mayor cantidad de información, con el fin de facilitar la determinación de que el ruido de fondo que contiene una gran cantidad de información es también una estructura vocal. Actualmente, la decisión dinámica se ha conseguido en un codificador vocal de tipo multi-tasa adaptativo (AMR, en forma abreviada). El codificador AMR puede ajustar dinámicamente el umbral de decisión, la duración de la resaca de voz y la condición de iniciación de resaca de la detección VAD en conformidad con el nivel de ruido de fondo en la señal de entrada.

Sin embargo, cuando el AMR existente realiza la decisión de VAD, el AMR solamente puede ser adaptativo al nivel del ruido de fondo, pero no puede ser adaptativo para la fluctuación del ruido de fondo. De este modo, el rendimiento de la decisión de VAD para la señal de entrada que tiene diferentes tipos de ruido de fondo puede ser bastante diferente. A modo de ejemplo, bajo el nivel del mismo ruido de fondo, el AMR tiene mucho más alto rendimiento de decisión de VAD en el caso de que el ruido de fondo sea un ruido de vehículo, pero el rendimiento de decisión de VAD se reduce en gran medida en el caso de que el ruido de fondo sea un ruido de balbuceo, lo que da lugar a un gran desperdicio de los recursos de ancho de banda del canal.

El documento US6453291 B1 da a conocer un aparato y un método para detección de actividad vocal en un sistema de comunicaciones. Con el fin de que la decisión del Detector de Actividad Vocal (VAD) supere el problema de ser hipersensible a la fluctuación, las condiciones de ruido de fondo no estacionarias, se utiliza un factor de polarización

para aumentar el umbral en el que está basada la decisión de VAD. Este factor de sesgo se deriva de una estimación de la variabilidad de la estimación del ruido de fondo. La estimación de variabilidad se basa, además, en valores negativos de la relación SNR instantánea.

5 El documento EP2159788 da a conocer un método y dispositivo de detección de actividad vocal (VAD), de modo que el umbral de VAD pueda ser adaptativo para la variación del ruido de fondo. El dispositivo VAD incluye: una unidad analizadora del ruido de fondo, adaptada para: analizar las características del ruido de fondo de una señal actual en conformidad con un resultado de determinación de VAD de entrada, para obtener parámetros relacionados con la variación del ruido de fondo y proporcionar, a la salida, estos parámetros; una unidad de ajuste de VAD, adaptada para: obtener un sesgo del umbral de VAD en conformidad con la salida de parámetros proporcionada por la unidad analizadora del ruido de fondo y proporcionar el sesgo del umbral de VAD; y una unidad de determinación de VAD, adaptada para: modificar un umbral de VAD para modificarse en conformidad con el sesgo del umbral de VAD proporcionado por la unidad de ajuste del umbral de VAD, determinar el ruido de fondo utilizando el umbral de VAD modificado, y proporcionar un resultado de determinación de VAD.

15 El documento "Detección de actividad acústica independiente de códec basada en la entropía con actualización del ruido adaptativa" (9TH IEEE INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING (ICSP), 26 octubre 2008, páginas 549-552) da a conocer un método de detección de actividad acústica independiente de los *codecs* (codificador/decodificador) de audio. Una característica de entropía establecida con la actualización de estimación del ruido adaptativa se ofrece para mejorar el rendimiento de la entropía en la detección de la voz y de la música. Un sistema de detección de actividad acústica independiente de códec se construye mediante la integración de características basadas en la relación SNR con la entropía propuesta.

#### SUMARIO DE LA INVENCION

25 Las formas de realización de la presente invención dan a conocer un método y un aparato para que la detección de actividad vocal VAD sea adaptativa para la fluctuación de un ruido de fondo para realizar una decisión de VAD, con lo que se mejora el rendimiento de la decisión de VAD.

30 Una forma de realización de la presente invención da a conocer un método para VAD, según se establece en la reivindicación independiente 1.

Una forma de realización de la presente invención da a conocer un aparato para la detección VAD, según se establece en la reivindicación independiente 2. Otra forma de realización de la presente invención da a conocer un soporte de memorización legible por ordenador, según se establece en la reivindicación independiente 4.

35 Sobre la base del método para la detección VAD y el aparato para VAD en conformidad con las formas de realización de la presente invención, cuando un asen es un ruido de fondo, un valor de característica fluctuante que se utiliza para representar la fluctuación del ruido de fondo es objeto de adquisición; se realiza un ajuste adaptativo sobre un parámetro relacionado con el criterio de decisión de VAD en conformidad con el valor de característica fluctuación y la decisión de VAD se realiza sobre la señal de entrada utilizando el parámetro relacionado con el criterio de decisión sobre el que se realiza el ajuste adaptativo. En comparación con la técnica anterior, la solución técnica de la presente invención puede conseguir un más alto rendimiento de decisión de VAD en el caso de diferentes tipos de ruido de fondo. Esta circunstancia mejora la eficiencia de decisión de VAD y la exactitud de la decisión.

La solución técnica de la presente invención se describe, en mayor detalle, haciendo referencia a los dibujos adjuntos y a las formas de realización.

#### 50 BREVE DESCRIPCION DE LOS DIBUJOS

Para ilustrar las soluciones técnicas en conformidad con las formas de realización de la presente invención o en la técnica anterior, con mayor claridad, los dibujos adjuntos se introducen, de forma concisa, a continuación. Evidentemente, los dibujos adjuntos en la descripción siguiente son solamente algunas formas de realización de la presente invención o sus ejemplos.

La Figura 1 es un diagrama de flujo de una forma de realización de un método para la detección VAD en conformidad con la presente invención;

60 La Figura 2 es un diagrama de flujo de un ejemplo de adquisición de un valor de característica fluctuante de un ruido de fondo;

La Figura 3 es un diagrama de flujo de una forma de realización de la adquisición del valor de característica fluctuante del ruido de fondo en conformidad con la presente invención;

65 La Figura 4 es un diagrama de flujo de otro ejemplo de adquisición del valor de característica fluctuante del ruido de

fondo;

La Figura 5 es un diagrama de flujo de un ejemplo de ajuste dinámico de un parámetro relacionado con el criterio de decisión de VAD, en conformidad con un nivel del ruido de fondo;

La Figura 6 es una vista estructural esquemática de una forma de realización de un aparato para la detección VAD en conformidad con la presente invención;

La Figura 7 es una vista estructural esquemática de un ejemplo de un aparato para la detección VAD;

La Figura 8 es una vista estructural esquemática de otro ejemplo de un aparato para la detección VAD;

La Figura 9 es una vista estructural esquemática detallada de la forma de realización del aparato para la detección VAD en conformidad con la presente invención;

La Figura 10 es una vista estructural esquemática de otro ejemplo de un aparato para la detección VAD;

La Figura 11 es una vista estructural esquemática de otro ejemplo de un aparato para la detección VAD;

La Figura 12 es una vista estructural esquemática de otro ejemplo de un aparato para la detección VAD;

La Figura 13 es otra vista estructural esquemática detallada de la forma de realización del aparato para la detección VAD con una unidad de control opcional adicional, en conformidad con la presente invención;

La Figura 14 es una vista estructural esquemática de otro ejemplo de un aparato para la detección VAD;

La Figura 15 es una vista estructural esquemática de otro ejemplo de un aparato para la detección VAD; y

La Figura 16 es una vista estructural esquemática de otro ejemplo de un aparato para la detección VAD.

#### DESCRIPCIÓN DETALLADA DE LAS FORMAS DE REALIZACIÓN

La solución técnica de la presente invención está descrita, de forma clara y completa, a continuación haciendo referencia a los dibujos adjuntos. Es evidente que las formas de realización que se describen son solamente una parte y no la totalidad de todas las formas de realización de la presente invención.

La Figura 1 es un diagrama de flujo de una forma de realización de un método para la detección VAD en conformidad con la presente invención. Según se ilustra en la Figura 1, el método para la detección VAD en conformidad con esta forma de realización incluye las etapas siguientes:

Etapa 101: Adquirir un valor de característica fluctuante de un ruido de fondo cuando una señal de entrada es el ruido de fondo, en donde el valor de característica fluctuante se utiliza para representar la fluctuación del ruido de fondo.

Etapa 102: Realizar un ajuste adaptativo sobre un parámetro vinculado con el criterio de decisión de detección VAD, en conformidad con el valor de característica fluctuante del ruido de fondo.

Etapa 103: Realizar decisión de detección VAD sobre la señal de entrada utilizando el parámetro vinculado con el criterio de decisión sobre el que se realiza el ajuste adaptativo.

Con el método para la detección VAD en conformidad con la forma de realización de la presente invención, cuando una señal de entrada es un ruido de fondo, se adquiere un valor de característica fluctuante utilizado para representar la fluctuación del ruido de fondo, se realiza un ajuste adaptativo sobre un parámetro vinculado con el criterio de decisión de detección VAD en conformidad con el valor de característica fluctuante, con el fin de hacer que el parámetro vinculado con el criterio de decisión de detección VAD sea adaptativo para la fluctuación del ruido de fondo. De este modo, cuando se realiza una decisión de detección VAD sobre la señal de entrada utilizando el parámetro vinculado con el criterio de decisión sobre el que se realiza el ajuste adaptativo, puede conseguirse un más alto rendimiento de la decisión de detección VAD en el caso de diferentes tipos de ruido de fondos, lo que mejora la eficiencia de decisión de detección VAD y su exactitud de decisión, aumentando así la utilización de los recursos de ancho de banda de canal limitados.

El parámetro vinculado con el criterio de decisión de detección VAD puede incluir cualquiera o más de un umbral de decisión primario, una condición de iniciación del resaca, una longitud de resaca y una tasa de actualización de un parámetro a largo plazo relacionado con el ruido de fondo.

Cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye el umbral de decisión primario,

en conformidad con una forma de realización de la presente invención, la etapa 102 puede ponerse en práctica concretamente en las formas siguientes:

Se busca un mapeado de correspondencia entre un valor de característica fluctuante y un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  y se adquiere un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente al valor de característica fluctuante del ruido de fondo, en donde el sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  se utiliza para representar un valor de sesgo del umbral bajo un ruido de fondo con diferente fluctuación y el mapeado de correspondencia puede establecerse con anterioridad o actualmente, o puede adquirirse desde otras entidades de la red.

Un umbral de decisión primario de VAD  $vad\_thr$  se adquiere utilizando la fórmula  $vad\_thr = f_1(snr) + f_2(snr) \cdot thr\_bias\_noise$ , en donde  $f_1(snr)$  es un umbral de referencia correspondiente a una relación señal a ruido SNR  $snr$  de una trama de ruido de fondo actual, y  $f_2(snr)$  es un coeficiente de ponderación de un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. Más concretamente, forma de función de  $f_1(snr)$  y  $f_2(snr)$  para  $snr$  puede establecerse en conformidad con valores empíricos.

El umbral de decisión primario en el parámetro vinculado con el criterio de decisión de detección VAD se actualiza al umbral de decisión primario adquirido  $vad\_thr$ , con el fin de realizar un ajuste adaptativo sobre el umbral de decisión primario de VAD  $vad\_thr$  en conformidad con el valor de característica fluctuante del ruido de fondo.

Cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye la condición de iniciación de la denominada resaca vocal, en conformidad con una forma de realización de la presente invención, la etapa 102 puede ponerse en práctica concretamente en los modos siguientes:

Se busca un valor de la longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl$ [valor de característica fluctuante] correspondiente al valor de característica fluctuante del ruido de fondo a partir de una tabla de mapeado de correspondencia de fluctuación de ruido de longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl$ [], y se busca un umbral vocal determinado  $burst\_thr\_noise\_tbl$ [valor de característica fluctuante] correspondiente al valor de característica fluctuante del ruido de fondo a partir de una tabla de sesgo del umbral vocal determinado en conformidad con la fluctuación del ruido  $burst\_thr\_noise\_tbl$ [], en donde la tabla de mapeado de correspondencia de fluctuación de ruido de longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl$ [] y la tabla de sesgo del umbral vocal determinado en conformidad con la fluctuación del ruido  $burst\_thr\_noise\_tbl$ [] puede establecerse también con anterioridad o actualmente o adquirirse a partir de otras entidades de la red.

Un umbral de magnitud de trama vocal sucesiva  $M$  se adquiere utilizando la fórmula  $M = f_3(snr) + f_4(snr) \cdot burst\_cnt\_noise\_tbl$ [valor de característica fluctuante] y un umbral de trama vocal determinado  $burst\_thr$  se adquiere utilizando la fórmula  $burst\_thr = f_5(snr) + f_6(snr) \cdot burst\_thr\_noise\_tbl$ [valor de característica fluctuante], en donde  $f_3(snr)$  es un umbral de magnitud de referencia correspondiente a una relación SNR  $snr$  de una trama de ruido de fondo actual,  $f_4(snr)$  es un coeficiente de ponderación de la longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl$ [valor de característica fluctuante] correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual,  $f_5(snr)$  es un umbral de trama vocal de referencia correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual, y  $f_6(snr)$  es un coeficiente de ponderación del umbral vocal determinado  $burst\_thr\_noise\_tbl$ [valor de característica fluctuante] correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. Más concretamente, las formas funcionales de  $f_3(snr)$ ,  $f_4(snr)$ ,  $f_5(snr)$ , y  $f_6(snr)$  a  $snr$  pueden establecerse en conformidad con valores empíricos. Como una forma de realización específica, las formas funcionales específicas de  $f_3(snr)$ ,  $f_4(snr)$ ,  $f_5(snr)$ , y  $f_6(snr)$  a  $snr$  pueden habilitar el umbral de magnitud de trama vocal sucesiva  $M$  y el umbral de trama vocal determinado  $burst\_thr$  para aumentar con la reducción del valor de característica fluctuante adquirido.

La condición de iniciación de la resaca vocal en el parámetro vinculado con el criterio de decisión de detección VAD se actualiza en conformidad con el umbral de magnitud de trama vocal sucesiva  $M$  adquirido y el umbral de trama vocal determinado  $burst\_thr$ , con el fin de realizar un ajuste adaptativo sobre la condición de iniciación de la resaca vocal de la detección VAD en conformidad con el valor de característica fluctuante del ruido de fondo.

En conformidad con las formas de realización de la presente invención establecidas en las reivindicaciones 1 y 2, el parámetro vinculado con el criterio de decisión de detección VAD incluye la longitud de la resaca vocal y la etapa 102 se realiza concretamente en la forma siguiente:

Se busca una longitud de la resaca vocal  $hangover\_nosie\_tbl$ [valor de característica fluctuante] correspondiente al valor de característica fluctuante del ruido de fondo a partir de una tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca  $hangover\_noise\_tbl$ [], en donde la tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca vocal  $hangover\_noise\_tbl$ [] puede establecerse con anterioridad o actualmente, o adquirirse desde otras entidades de la red.

Un valor máximo de reposición del contador de resaca vocal  $hangover\_max$  se busca utilizando la fórmula

$hangover\_max = f_7(snr) + f_8(snr) \cdot hangover\_nosie\_tbl$  [valor de característica fluctuante], en donde  $f_7(snr)$  es un valor de reposición de referencia correspondiente a una relación SNR  $snr$  de una trama de ruido de fondo actual, y  $f_8(snr)$  es un coeficiente de ponderación de una longitud de resaca  $hangover\_nosie\_tbl$  [valor de característica fluctuante] correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. Más concretamente, una forma de función de  $f_7(snr)$  y  $f_8(snr)$  a  $snr$  puede establecerse en conformidad con valores empíricos. La forma de función específica de  $f_7(snr)$  y  $f_8(snr)$  a  $snr$  puede permitir aumentar el valor máximo de reposición del contador de resaca vocal  $hangover\_max$  con el incremento del valor de característica fluctuante adquirido.

La longitud de la resaca vocal en el parámetro vinculado con el criterio de decisión de detección VAD se actualiza al valor máximo de reposición del contador de la resaca que se adquiere  $hangover\_max$ , con el fin de realizar un ajuste adaptativo sobre la longitud de resaca sobre la detección VAD en conformidad con el valor de característica fluctuante del ruido de fondo.

Una media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo puede adoptarse para representar la fluctuación del ruido de fondo. La Figura 2 es un diagrama de flujo de un ejemplo de adquisición de un valor de característica fluctuante de un ruido de fondo. En este caso, el valor de característica fluctuante es concretamente un valor cuantizado  $idx$  de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo. Según se ilustra en la Figura 2, el proceso en conformidad con este ejemplo incluye las etapas siguientes:

Etapla 201: Recibir una trama actual de la señal de entrada.

Etapla 202: Dividir la trama actual de la señal de entrada en N sub-bandas en un dominio de la frecuencia, en donde N es un número entero mayor que 1, a modo de ejemplo, N puede ser 32, y calcular las energías  $enrg(i)$  (en donde  $i=0, 1, \dots, N-1$ ) de las N sub-bandas respectivamente.

Más concretamente, las N sub-bandas pueden ser de anchura igual o de anchura desigual o cualquier número de sub-bandas en las N sub-bandas puede ser de anchura igual.

Etapla 203: Decidir si la trama actual es una trama de ruido de fondo en conformidad con el criterio de decisión de detección VAD. Si la trama actual es una trama de ruido de fondo, realizar la etapa 204; si la trama actual no es una trama de ruido de fondo, no realizar los procedimientos subsiguientes de este ejemplo.

Etapla 204: Calcular una energía de media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente en las N sub-bandas utilizando la fórmula  $enrg\_n(i) = \alpha \cdot enrg\_n + (1 - \alpha) \cdot enrg(i)$ , en donde  $\alpha$  es coeficiente de olvido para controlar una tasa de actualización de la energía de la media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente de las N sub-bandas, y  $enrg\_n$  es una energía de la trama de ruido de fondo.

Etapla 205: Blanquear un espectro de la trama de ruido de fondo actual utilizando la fórmula  $enrg\_w(i) = enrg(i)/enrg\_n(i)$ , y una energía  $enrg\_w(i)$  del ruido blanqueado de fondo en una i-ésima sub-band se adquiere a este respecto.

Etapla 206: Adquirir una entropía espectral del ruido blanqueado de fondo  $hb$  utilizando la fórmula

$$hb = - \sum_{i=0}^{N-1} p_i \cdot \log p_i, \text{ en donde } p_i = enrg\_w(i) / \sum_{i=0}^{N-1} enrg\_w(i).$$

Etapla 207: Adquirir una media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo utilizando la fórmula  $hb\_noise\_mov = \beta \cdot hb\_noise\_mov + (1 - \beta) \cdot hb$ , en donde  $\beta$  es un factor de olvido para controlar la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo-

En esta realización ejemplo, la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo representa la fluctuación del ruido de fondo. Cuanto mayor es el valor  $hb\_noise\_mov$  es, tanto menor será la fluctuación del ruido de fondo; por el contrario, cuanto más pequeño es el valor  $hb\_noise\_mov$  es, tanto mayor será la fluctuación del ruido de fondo.

Etapla 208: Cuantizar la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo utilizando la fórmula  $idx = \lfloor (hb\_noise\_mov - A) / B \rfloor$ , con el fin de adquirir un valor cuantizado  $idx$ , en donde A y B son valores preestablecidos, a modo de ejemplo, A puede ser un valor empírico 3.11, y B puede ser un valor empírico 0.05.

En correspondencia con el ejemplo ilustrado en la Figura 2, cuando el valor de característica fluctuante es concretamente el valor cuantizado  $idx$  de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del

ruido blanqueado de fondo, la tasa de actualización del parámetro a largo plazo relacionado con el ruido de fondo puede incluir la tasa de actualización de una de media móvil a largo plazo  $enrg\_n(i)$  del ruido de fondo. En correspondencia, la etapa 102 puede realizarse concretamente en las formas siguientes:

- 5 Se consulta una tabla de tasa de actualización ruido de fondo  $alpha\_tbl[]$  y un coeficiente de olvido  $\alpha$  de la tasa de actualización de energía de la media móvil a largo plazo  $enrg\_n(i)$  correspondiente al valor cuantizado  $idx$  del ruido de fondo se adquiere a este respecto. Más concretamente, la tabla de tasa de actualización del ruido de fondo  $alpha\_tbl[]$  puede establecerse con anterioridad o actualmente, o puede adquirirse desde otras entidades de la red. Como una forma de realización específica, el establecimiento de la tabla de tasa de actualización del ruido de fondo  $alpha\_tbl[]$  puede permitir que el coeficiente de olvido  $\alpha$  de la tasa de actualización se la energía de la media móvil a largo plazo  $enrg\_n(i)$  para disminuir con la reducción del valor cuantizado  $idx$  del ruido de fondo.

15 El coeficiente de olvido adquirido  $\alpha$  se utiliza como un coeficiente de olvido para controlar la tasa de actualización de la energía de la media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente en las  $N$  sub-bandas, con el fin de realizar un ajuste adaptativo sobre la tasa de actualización de la energía de media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente en las  $N$  sub-bandas en conformidad con el valor de característica fluctuante del ruido de fondo.

20 Además, en correspondencia con el ejemplo ilustrado en la Figura 2, cuando el valor de característica fluctuante es concretamente el valor cuantizado  $idx$  de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo, la tasa de actualización del parámetro a largo plazo relacionado con el ruido de fondo puede incluir también la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo. En correspondencia, la etapa 102 puede realizarse concretamente en las formas siguientes:

25 Se consulta una tabla de tasa de actualización de fluctuación de ruido de fondo  $beta\_tbl[]$ , y un factor de olvido  $\beta$  de la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  correspondiente al valor cuantizado  $idx$  del ruido de fondo, se adquiere a este respecto. Más concretamente, el tabla de tasa de actualización de fluctuación de ruido de fondo  $beta\_tbl[]$  puede establecerse con anterioridad o actualmente, o puede adquirirse a partir de otras entidades de la red. Como una forma de realización específica, el establecimiento específico de la tabla de tasa de actualización de fluctuación de ruido de fondo  $beta\_tbl[]$  puede permitir que el factor de olvido  $\beta$  de la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  aumente con la disminución del valor cuantizado  $idx$  del ruido de fondo.

35 El factor de olvido adquirido  $\beta$  se utiliza como un factor de olvido para controlar la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo, con el fin de realizar el ajuste adaptativo sobre la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo en conformidad con el valor de característica fluctuante del ruido de fondo.

40 Con respecto el ruido de fondo con diferentes valores de característica fluctuantes, la energía de la media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente en las  $N$  sub-bandas y de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo se actualizan con diferentes tasas, lo que puede mejorar la tasa de detección para el ruido de fondo de forma efectiva.

45 En conformidad con el método para la detección VAD de la presente invención, se utiliza una media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr\_n\_mov$  como un valor de característica fluctuante del ruido de fondo, con el fin de representar la fluctuación del ruido de fondo. La Figura 3 es un diagrama de flujo de una forma de realización de la adquisición del valor de característica fluctuante del ruido de fondo en conformidad con la presente invención. En esta forma de realización, el valor de característica fluctuante del ruido de fondo es concretamente la media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr\_n\_mov$ . Según se ilustra en la Figura 3, el proceso en conformidad con la forma de realización establecida en la reivindicación 1 incluye las etapas siguientes:

55 Etapa 301: Recibir una trama actual de la señal de entrada.

Etapa 302: Decidir si la trama actual es una trama de ruido de fondo en conformidad con el criterio de decisión de detección VAD. Si la trama actual es una trama de ruido de fondo, realizar la etapa 303; si la trama actual no es una trama de ruido de fondo, no realizar los procedimientos subsiguientes de esta forma de realización.

60 Etapa 303: Adquirir una media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr\_n\_mov$  utilizando la fórmula  $snr\_n\_mov = k \cdot snr\_n\_mov + (1 - k) \cdot snr$ .

$snr$  es una relación SNR de la trama de ruido de fondo actual y  $k$  es un factor de olvido para controlar una tasa de actualización de la media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr\_n\_mov$ .

65 En correspondencia con la forma de realización ilustrada en la Figura 3, cuando el valor de característica fluctuante

del ruido de fondo es concretamente la media móvil a largo plazo de la relación SNR de trama de ruido de fondo  $snr_n\_mov$ , la tasa de actualización del parámetro a largo plazo relacionado con el ruido de fondo puede incluir la tasa de actualización de la media móvil a largo plazo  $snr_n\_mov$ . En correspondencia, la etapa 102 puede realizarse concretamente en las formas siguientes: establecimiento de valores diferentes para el factor de olvido  $k$  para controlar la tasa de actualización de la media móvil a largo plazo de la relación SNR de trama de ruido de fondo  $snr_n\_mov$  cuando la relación SNR  $snr$  de la trama de ruido de fondo actual es mayor que un valor medio  $snr_n$  de SNRs de las últimas  $n$  tramas de ruido de fondos, y cuando la relación SNR  $snr$  de la trama de ruido de fondo actual es más pequeña que el valor medio  $snr_n$  de la relación SNR de las últimas  $n$  tramas de ruido de fondos. A modo de ejemplo, cuando  $snr_n\_mov < snr$ ,  $k$  se establece para ser  $x$ , y cuando  $snr_n\_mov \geq snr$ ,  $k$  se establece para ser  $y$ .

La una media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr_n\_mov$  se actualiza en sentido ascendente y descendente con diferentes tasas de actualización, lo que puede evitar que la media móvil a largo plazo de la relación SNR de trama de ruido de fondo  $snr_n\_mov$  sea afectada por un cambio brusco, con el fin de hacer que la media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr_n\_mov$  sea más estable. Antes de la tasa de actualización del parámetro a largo plazo relacionado con el ruido de fondo actualizado por la relación SNR  $snr$  de la trama de ruido de fondo actual puede incluir la media móvil a largo plazo  $snr_n\_mov$ , la relación SNR  $snr$  de la trama de ruido de fondo actual puede limitarse a un margen preestablecido, a modo de ejemplo, cuando el valor de SNR  $snr$  de la trama de ruido de fondo actual sea menor que 10, la relación SNR  $snr$  de la trama de ruido de fondo actual está limitada a 10.

En conformidad con otro ejemplo del método para la detección VAD, una media móvil a largo plazo  $flux_{bgd}$  de la relación SNR segmental modificada a largo plazo de la trama de ruido de fondo (MSSNR) puede utilizarse como el valor de característica fluctuante del ruido de fondo para representar la fluctuación del ruido de fondo. La Figura 4 es un diagrama de flujo de otro ejemplo de adquisición del valor de característica fluctuante del ruido de fondo. En este ejemplo, el valor de característica fluctuante del ruido de fondo es concretamente la media móvil a largo plazo de la trama de ruido de fondo MSSNR  $flux_{bgd}$ . Según se ilustra en la 4, el proceso en conformidad con esta realización ejemplo incluye las etapas siguientes:

Etapa 401: Recibir una trama actual de la señal de entrada.

Etapa 402: Decidir si la trama actual es una trama de ruido de fondo en conformidad con el criterio de decisión de detección VAD. Si la trama actual es una trama de ruido de fondo, realizar etapa 403; si la trama actual no es una trama de ruido de fondo, no realizar los procedimientos subsiguientes de esta forma de realización.

Etapa 403: Dividir un espectro de la Transformada de Fourier Rápida (FFT) de la trama de ruido de fondo actual en  $H$  sub-bandas, en donde  $H$  es un número entero mayor que 1, y calcular las energías de  $i$  sub-bandas  $E_{band}(i)$ ,

$$E_{band}(i) = \frac{p}{h(i) - l(i) + 1} \sum_{j=l(i)}^{h(i)} S_j + (1 - p) E_{band\_old}(i),$$

siendo  $i=0, 1, \dots, H-1$  respectivamente, utilizando la fórmula

en donde  $l(i)$  y  $h(i)$  representan un punto de frecuencia FFT con la más baja frecuencia y un punto de frecuencia FFT con la más alta frecuencia en una  $i$ -ésima sub-banda respectivamente,  $S_j$  representa una energía de un  $j$ -ésimo punto de frecuencia en el impuesto de FFT,  $E_{band\_old}(i)$  representa una energía de la  $i$ -ésima sub-banda en una trama anterior de la trama de ruido de fondo actual y  $P$  es una constante preestablecida. En una forma de realización, el valor de  $P$  es 0.55. Como una instancia de aplicación específica de la presente invención, el valor de  $H$  puede ser 16.

Etapa 404: Calcular una relación SNR  $snr(i)$  de la  $i$ -ésima sub-banda en la trama de ruido de fondo actual

respectivamente, utilizando la fórmula 
$$snr(i) = 10 \log(E_{band}(i) / \overline{E_{band\_n}(i)}).$$

$\overline{E_{band\_n}(i)}$  es una media móvil a largo plazo del ruido de fondo que puede adquirirse concretamente actualizando la media móvil a largo plazo del ruido de fondo  $\overline{E_{band\_n}(i)}$  utilizando la energía de la  $i$ -ésima sub-banda en una trama de ruido de fondo anterior, utilizando la fórmula  $\overline{E_{band\_n}(i)} = q \cdot \overline{E_{band\_n}(i)} + (1 - q) \cdot E_{band}(i)$ , en donde  $q$  es una constante preestablecida. En una forma de realización, el valor  $q$  es 0.95.

Etapa 405: Modificar la relación SNR  $snr(i)$  de la  $i$ -ésima sub-banda en la trama de ruido de fondo actual respectivamente, utilizando la fórmula



$$msnr(i) = \begin{cases} \text{MAX}[\text{MIN}[\frac{snr(i)^3}{C1}, 1], 0] & , i \in \text{primer conjunto} \\ \text{MAX}[\text{MIN}[\frac{snr(i)^3}{C2}, 1], 0] & , i \in \text{segundo conjunto} \end{cases}$$

en donde  $msnr(i)$  es la relación SNR de la  $i$ -ésima sub-banda modificada,  $C1$  y  $C2$  son constantes reales preestablecidas mayores que 0 y los valores en el primer conjunto y en el segundo conjunto forman un conjunto  $[0, H-1]$ .

$$MSSNR = \sum_{i=0}^{H-1} msnr(i).$$

Etapa 406: Adquirir una trama de ruido de fondo actual  $MSSNR$  utilizando la fórmula

Etapa 407: Calcular una media móvil a largo plazo  $MSSNR$  de la trama de ruido de fondo actual  $flux_{bgd}$  utilizando la fórmula  $flux_{bgd} = r \cdot flux_{bgd} + (1-r) \cdot MSSNR$ , en donde  $r$  es un coeficiente de olvido para controlar una tasa de actualización de la media móvil a largo plazo  $MSSNR$  de la trama de ruido de fondo actual  $flux_{bgd}$ .

En una forma de realización, el valor de  $r$  puede establecerse concretamente en las formas siguientes: en un periodo inicial preestablecido desde una primera trama de la señal de entrada y cuando se tiene  $MSSNR > flux_{bgd}$ ,  $r=0.955$ ; en el periodo inicial preestablecido de la primera trama de la señal de entrada y cuando  $MSSNR \leq flux_{bgd}$ ,  $r=0.995$ ; después del periodo inicial preestablecido desde la primera trama a la señal de entrada y cuando se tiene  $MSSNR > flux_{bgd}$ ,  $r=0.997$ ; y después del periodo inicial preestablecido de la primera trama de la señal de entrada y cuando se tiene  $MSSNR \leq flux_{bgd}$ ,  $r=0.9997$ .

En correspondencia con el ejemplo ilustrado en la Figura 4, cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye el umbral de decisión primario, la etapa 102 puede realizarse concretamente en las formas siguientes:

Un mapeado de correspondencia entre un valor de característica fluctuante y un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  es objeto de consulta, y un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente al valor de característica fluctuante del ruido de fondo es objeto de adquisición, en donde el sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  se utiliza para representar un valor de sesgo del umbral bajo un ruido de fondo con diferente fluctuación, y el mapeado de correspondencia puede establecerse con anterioridad o actualmente, o puede adquirirse a partir de otras entidades de la red.

Un umbral de decisión primario de VAD  $vad\_thr$  se adquiere utilizando la fórmula  $vad\_thr = f_1(snr) + f_2(snr) \cdot thr\_bias\_noise$ , en donde  $f_1(snr)$  es un umbral de referencia correspondiente a una relación SNR  $snr$  de una trama de ruido de fondo actual y  $f_2(snr)$  es un coeficiente de ponderación de la sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. Más concretamente, una forma de función de  $f_1(snr)$  y  $f_2(snr)$  a SNR puede establecerse en conformidad con un valor empírico.

El umbral de decisión primario en el parámetro vinculado con el criterio de decisión de detección VAD se actualiza al umbral de decisión primario  $vad\_thr$ .

Además, en correspondencia con la realización ejemplo ilustrada en la Figura 4, cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye el umbral de decisión primario, la etapa 102 puede realizarse concretamente en las formas siguientes.

Un nivel de fluctuación  $flux\_idx$  correspondiente a la media móvil a largo plazo de  $MSSNR$  de la trama de ruido de fondo actual  $flux_{bgd}$  es objeto de adquisición y un nivel  $snr\_idx$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual es también objeto de adquisición.

Un umbral de decisión primario  $thr\_tbl[snr\_idx][flux\_idx]$  correspondiente al nivel de fluctuación adquirido  $flux\_idx$  y el nivel de relación SNR  $snr\_idx$  simultáneamente es objeto de búsqueda.

El umbral de decisión primario en el parámetro vinculado con el criterio de decisión se actualiza al umbral de decisión primario buscado  $thr\_tbl[snr\_idx][flux\_idx]$ .

Después de la media móvil a largo plazo de  $MSSNR$  de la trama de ruido de fondo actual  $flux_{bgd}$  y la relación SNR  $snr$  correspondan a niveles correspondientes, el aparato para la detección VAD solamente necesita memorizar el mapeado de correspondencia entre el nivel de fluctuación, el nivel de relación SNR y el umbral de decisión primario. La cantidad de datos del nivel de fluctuación y del nivel de SNR es mucho más pequeña que los datos de  $flux_{bgd}$  y  $snr$  que pueden cubrirse, con el fin de reducir el espacio de almacenamiento del aparato VAD ocupado por el mapeado de correspondencia en mayor medida y utilizar eficientemente el espacio de almacenamiento.

A modo de ejemplo, la media móvil a largo plazo de MSSNR de la trama de ruido de fondo actual  $flux_{bgd}$  puede dividirse en tres niveles de fluctuación en conformidad con valores, en los que  $flux\_idx$  representa el nivel de fluctuación de  $flux_{bgd}$ , y  $flux\_idx$  puede establecerse a 0, 1, y 2, que representan una baja fluctuación, fluctuación media y alta fluctuación, respectivamente. En conformidad con un ejemplo, el valor de  $flux\_idx$  se determina en la forma siguiente:

Si  $flux_{bgd} < 3.5$ ,  $flux\_idx = 0$ .

Si  $3.5 \leq flux_{bgd} < 6$ ,  $flux\_idx = 1$ .

Si  $flux_{bgd} \geq 6$ ,  $flux\_idx = 2$ .

Análogamente, una relación SNR  $snr$  de trama de ruido de fondo actual a largo plazo de la señal SNR  $snr$  se divide en cuatro niveles de relación de SNR en conformidad con los valores, en donde  $snr\_idx$  representa un nivel de relación SNR  $snr$ , y  $snr\_idx$  puede establecerse a 0, 1, 2, y 3 para representar SNR baja, SNR media, SNR alta y SNR más alta, respectivamente.

Además, el nivel de fluctuación  $flux\_idx$  correspondiente a la media móvil a largo plazo de MSSNR de la trama de ruido de fondo actual  $flux_{bgd}$  es objeto de adquisición y una tendencia de decisión  $op\_idx$  correspondiente al rendimiento de trabajo actual del aparato para VAD que realiza la de decisión de detección VAD sobre la señal de entrada puede adquirirse también cuando el nivel de SNR  $snr\_idx$  correspondiente a la relación SNR de la trama de ruido de fondo actual, es decir, con tendencia a decidir que la trama actual es una trama focal o una trama de ruido de fondo. Más concretamente, el rendimiento de trabajo actual del aparato para VAD puede incluir la economía de ancho de banda por la calidad de codificación vocal después del inicio de VAD y la VAD propiamente dicha. En correspondencia, un umbral de decisión primario  $vad\_thr = thr\_tbl[snr\_idx][flux\_idx][op\_idx]$  correspondiente al nivel de fluctuación  $flux\_idx$ , el nivel de SNR  $snr\_idx$ , y el nivel de rendimiento  $op\_idx$  pueden ser objeto de consulta y el umbral de decisión primario en el parámetro vinculado con el criterio de decisión de detección VAD se actualiza para el umbral de decisión primario  $vad\_thr = thr\_tbl[snr\_idx][flux\_idx][op\_idx]$ .

La actualización adaptativa se realiza, además, sobre el umbral de decisión primario en el parámetro vinculado con el criterio de decisión de detección VAD en combinación con la tendencia de decisión correspondiente al rendimiento de trabajo actual del aparato para VAD, con el fin de hacer el criterio de decisión de detección VAD más aplicable a un aparato específico para VAD, con lo que se adquiere un rendimiento de decisión VAD más alto que es más aplicable a un entorno específico, mejorando todavía más la eficiencia de la decisión de VAD y la exactitud de dicha decisión y aumentando la utilización de recursos de ancho de banda de canal limitados.

Cualquiera o más de los parámetros vinculados con el criterio de decisión de detección VADs: el umbral de decisión primario, la longitud de resaca y la condición de iniciación de la resaca de voz pueden ser ajustados dinámicamente, además, en conformidad con el nivel de ruido de fondo en la señal de entrada. FIG. 5 es un diagrama de flujo de un ejemplo de ajuste dinámico de un parámetro vinculado con el criterio de decisión de detección VAD en conformidad con un nivel de ruido de fondo y este ejemplo puede realizarse concretamente por un AMR. Según se ilustra en la Figura 5, el proceso incluye las etapas siguientes:

Eta 501: Dividir la señal de entrada en N sub-bandas en el dominio de la frecuencia, y calcular los niveles  $level(i)$  (en donde  $i=0, 1, 2...N-1$ ) en cada sub-banda respectivamente para cada señal de entrada de trama. Al mismo tiempo, los niveles  $bckr\_level(i)$  (en donde  $i=0, 1, 2...N-1$ ) del ruido de fondo en la señal de entrada en cada sub-banda se estiman de forma continua.

$$noise\_level = \frac{1}{N} \sum_{i=0}^{N-1} bckr\_level(i)$$

Representa el nivel de la trama de ruido de fondo actual.

Eta 502: Calcular una relación SNR  $snr(i)$  de la trama actual en cada sub-banda utilizando la fórmula.

$$snr(i) = level(i)^2 / bckr\_level(i)^2$$

Eta 503: Adquirir una suma de SNR de la trama actual  $snr\_sum$  utilizando la fórmula  $snr\_sum = \sum snr(i)$ , y el valor SNR de la trama actual  $snr\_sum$  es el parámetro de decisión primario de la detección VAD. Al mismo tiempo, la condición de iniciación de la resaca y la longitud de resaca sobre la detección VAD se ajustan en conformidad con el nivel de ruido de fondo  $noise\_level$ .

Un resultado de la decisión medio (o denominado un primer resultado de la decisión) sobre la detección VAD puede adquirirse comparando la suma de SNR de la trama actual  $snr\_sum$  con un umbral de decisión preestablecido

vad\_thr. Más concretamente, si la suma de SNR de la trama actual snr\_sum es mayor que el umbral de decisión vad\_thr, el resultado de decisión medio de VAD es 1, es decir, la trama actual se decide como que es una trama vocal; si la suma de SNR de la trama actual snr\_sum es más pequeña o igual al umbral de decisión tvad\_thr, el resultado de decisión medio de VAD es 0, es decir, la trama actual se decide que es una trama de ruido de fondo.

El umbral de decisión vad\_thr se controla por el nivel de ruido de fondo noise\_level, que se decide concretamente utilizando la fórmula  $vad\_thr = [(VAD\_THR\_HIGH - VAD\_THR\_LOW)/(p2 - p1)] \cdot (noise\_level - p1) + VAD\_THR\_HIGH$ , en donde VAD\_THR\_HIGH y VAD\_THR\_LOW son límites superiores e inferiores de una gama de valores del umbral de decisión vad\_thr respectivamente, y p2 y p1 representan niveles de ruido de fondo correspondientes a los límites superior e inferior del umbral de decisión vad\_thr respectivamente. De este modo, resulta evidente que el umbral de decisión vad\_thr es interpolado entre los límites superior e inferior en conformidad con el valor de nivel de ruido de fondo noise\_level, y está en una relación lineal con el nivel de ruido noise\_level. Cuanto más alto sea el ruido de fondo noise\_level, tanto menor será el umbral de decisión thr\_vad, de modo que una exactitud de VAD suficiente puede garantizarse también en el caso de un ruido de fondo mayor.

La condición de iniciación de la resaca vocal de la VAD se controla también por el nivel de ruido de fondo noise\_level. La así denominada condición de iniciación de la resaca vocal significa que el contador de la resaca vocal puede establecerse para tener una longitud máxima de resaca cuando se satisfaga la condición de iniciación de la resaca vocal correspondiente. Cuando el resultado de decisión medio es 0, si se realiza una resaca vocal se determina en función de si el contador de la resaca es mayor que 0. Si el contador es mayor que cero, una salida final del VAD se cambia desde 0 a 1 y el contador de resaca resta 1. Si el contador de resaca es menor o igual a 0, la salida final del VAD se mantiene como 0. En el VAD del AMR, la condición de iniciación de la resaca es si el número N de tramas vocales sucesivas actuales es mayor que un umbral preestablecido. Si el número N de tramas vocales sucesivas actuales es mayor que el umbral preestablecido, se satisface la condición de reiniciación de la resaca y es objeto de reposición el contador de resaca. Cuando el nivel de ruido noise\_level es mayor que otro umbral preestablecido, se considera que el ruido de fondo actual es mayor y N es la condición de inicialización que se establecer para ser un valor más pequeño, de modo que permita una ocurrencia más fácil de la resaca de voz. De no ser así, cuando el nivel de ruido noise\_level no es mayor que el otro umbral preestablecido, se considera que el ruido de fondo actual es más pequeño y N se establece para ser un valor mayor, lo que hace difícil la ocurrencia de la resaca.

Además, la longitud de resaca, es decir, el valor máximo del contador de la resaca vocal, se controla también por el nivel de ruido de fondo noise\_level. Cuando el nivel de ruido de fondo noise\_level es mayor que otro umbral preestablecido, se considera que el ruido de fondo es mayor, y cuando se inicia una resaca vocal, el contador de resaca puede establecerse a un valor mayor. De no ser así, cuando el nivel de ruido de fondo noise\_level no es mayor que el umbral preestablecido adicional, se considera que el ruido de fondo es más pequeño, y cuando se inicia una resaca vocal, el contador de resaca puede establecerse para ser un valor más pequeño.

La Figura 6 es una vista estructural esquemática de una forma de realización de un aparato de VAD en conformidad con la presente invención. El aparato para VAD en conformidad con esta forma de realización está configurado para realizar el método para VAD en conformidad con la forma de realización de la presente invención. Según se ilustra en la Figura 6, el aparato para VAD en conformidad con esta forma de realización incluye un módulo de adquisición 601, de módulo de ajuste 602 y un módulo de decisión 603.

El módulo de adquisición 601 está configurado para adquirir un valor de característica fluctuante de un ruido de fondo cuando una señal de entrada es el ruido de fondo, en donde el valor de característica fluctuante se utiliza para representar la fluctuación del ruido de fondo. El módulo de ajuste 602 está configurado para realizar un ajuste adaptativo de un parámetro vinculado con el criterio de decisión de detección VAD en conformidad con el valor de característica fluctuante adquirido por el módulo de adquisición 601. El módulo de decisión 603 está configurado para realizar una decisión de detección VAD sobre la señal de entrada utilizando el parámetro vinculado con el criterio de decisión en donde el ajuste adaptativo se realiza por el módulo de ajuste 602.

Además, haciendo referencia a la Figura 6, el aparato para VAD en conformidad con esta forma de realización de la presente invención incluye también un módulo de memorización 604, configurado para memorizar el parámetro vinculado con el criterio de decisión de detección VAD, en donde el parámetro vinculado con el criterio de decisión puede incluir cualquiera o más de un umbral de decisión primario, una condición de iniciación del resaca, una longitud de resaca y una tasa de actualización de un parámetro a largo plazo relacionado con el ruido de fondo. En correspondencia, el módulo de ajuste 602 está configurado para realizar un ajuste adaptativo sobre el parámetro vinculado con el criterio de decisión de detección VAD memorizado en el módulo de memorización 604; y el módulo de decisión 603 realiza una decisión de detección VAD sobre la señal de entrada utilizando el parámetro vinculado con el criterio de decisión memorizado en el módulo de memorización 604 en el que se realiza el ajuste adaptativo.

La Figura 7 es una vista estructural esquemática de un ejemplo del aparato para VAD. En comparación con la forma de realización ilustrada en la Figura 6, en el aparato ejemplo para VAD en conformidad con este ejemplo, cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye el umbral de decisión primario, el módulo de ajuste 602 incluye una primera unidad de memorización 701, una primera unidad de búsqueda 702, una primera

unidad de adquisición 703, y una primera unidad de actualización 704. La primera unidad de memorización 701 está configurada para memorizar un mapeado de correspondencia entre un valor de característica fluctuante y un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$ . La primera unidad de búsqueda 702 está configurada para consultar el mapeado de correspondencia entre el valor de característica fluctuante y el sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  a partir de la primera unidad de memorización 701, y adquirir un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente a un valor de característica fluctuante de un ruido de fondo, en donde el sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  se utiliza para representar un valor de sesgo del umbral bajo un ruido de fondo con diferente fluctuación. La primera unidad de adquisición 703 está configurada para adquirir un umbral de decisión primario  $vad\_thr$  utilizando la fórmula  $vad\_thr = f_1(snr) + f_2(snr) \cdot thr\_bias\_noise$ , en donde  $f_1(snr)$  es un umbral de referencia correspondiente a una relación SNR  $snr$  de una trama de ruido de fondo actual, y  $f_2(snr)$  es un coeficiente de ponderación del sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. La primera unidad de actualización 704 está configurada para actualizar el umbral de decisión primario en el parámetro vinculado con el criterio de decisión de detección VAD para el umbral de decisión primario  $vad\_thr$  adquirido por la primera unidad de adquisición 703.

La Figura 8 es una vista estructural esquemática de otro ejemplo de un aparato para VAD. En comparación con la forma de realización ilustrada en la Figura 6, en el aparato para VAD en conformidad con este ejemplo, cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye la condición de iniciación del resaca, el módulo de ajuste 602 incluye un segundo módulo de memorización 711, una segunda unidad de búsqueda 712, una segunda unidad de adquisición 713 y una segunda unidad de actualización 714. El segundo módulo de memorización 711 está configurado para memorizar una tabla de mapeado de correspondencia de fluctuación de longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl[]$  y una tabla de valor de sesgo de fluctuación del umbral vocal determinado  $burst\_thr\_noise\_tbl[]$ , en donde la tabla de mapeado de correspondencia de fluctuación de longitud de trama  $burst\_cnt\_noise\_tbl[]$  incluye un mapeado de correspondencia entre un valor de característica fluctuante y una longitud de trama vocal sucesiva, y la tabla de valor de sesgo de fluctuación del umbral vocal determinado  $burst\_thr\_noise\_tbl[]$  incluye un mapeado de correspondencia entre un valor de característica fluctuante y un umbral vocal determinado. La segunda unidad de búsqueda 712 está configurada para consultar una longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl[\text{valor de característica fluctuante}]$  correspondiente al valor de característica fluctuante del ruido de fondo a partir de la tabla de mapeado de correspondencia de fluctuación de longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl[]$  memorizada por la segunda unidad de memorización 711 y se consulta un umbral vocal determinado  $burst\_thr\_noise\_tbl[\text{valor de característica fluctuante}]$  correspondiente al valor de característica fluctuante del ruido de fondo a partir de la tabla de sesgo del umbral en conformidad con la fluctuación del ruido  $burst\_thr\_noise\_tbl[]$ . La segunda unidad de adquisición 713 está configurada para adquirir un umbral de magnitud de trama vocal sucesiva  $M$  utilizando la fórmula  $M = f_3(snr) + f_4(snr) \cdot burst\_cnt\_noise\_tbl[\text{valor de característica fluctuante}]$ , y para adquirir un umbral de trama vocal determinado  $burst\_thr$  utilizando la fórmula  $burst\_thr = f_5(snr) + f_6(snr) \cdot burst\_thr\_noise\_tbl[\text{valor de característica fluctuante}]$ , en donde  $f_3(snr)$  es un umbral de magnitud de referencia correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual,  $f_4(snr)$  es un coeficiente de ponderación de la longitud de trama vocal sucesiva  $burst\_cnt\_noise\_tbl[\text{valor de característica fluctuante}]$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual,  $f_5(snr)$  es un umbral de trama vocal de referencia correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual, y  $f_6(snr)$  es un coeficiente de ponderación del umbral vocal determinado  $burst\_thr\_noise\_tbl[\text{valor de característica fluctuante}]$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. La segunda unidad de actualización 714 está configurada para actualizar la condición de iniciación de la resaca vocal en el parámetro vinculado con el criterio de decisión de detección VAD en conformidad con el umbral de magnitud de trama vocal sucesiva  $M$  y el umbral de trama vocal determinado  $burst\_thr$  adquirido por la segunda unidad de adquisición 713.

La Figura 9 es una vista estructural esquemática detallada de la forma de realización del aparato para VAD en conformidad con la presente invención. El parámetro vinculado con el criterio de decisión de detección VAD incluye la longitud de resaca, y el módulo de ajuste 602 incluye una tercera unidad de memorización 721, una tercera unidad de búsqueda 722, una tercera unidad de adquisición 723 y una tercera unidad de actualización 724. La tercera unidad de memorización 721 está configurada para memorizar una tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca  $hangover\_noise\_tbl[]$ , en donde la tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca  $hangover\_noise\_tbl[]$  incluye un mapeado de correspondencia entre un valor de característica fluctuante y a longitud de resaca. La tercera unidad de búsqueda 722 está configurada para consultar una longitud de resaca  $hangover\_nosie\_tbl[\text{valor de característica fluctuante}]$  correspondiente al valor de característica fluctuante del ruido de fondo a partir de la tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca  $hangover\_noise\_tbl[]$  que se memoriza por la tercera unidad de memorización 721. La tercera unidad de adquisición 723 está configurada para adquirir un valor máximo de reposición del contador de resaca  $hangover\_max$  utilizando la fórmula  $hangover\_max = f_7(snr) + f_8(snr) \cdot hangover\_nosie\_tbl[\text{valor de característica fluctuante}]$ , en donde  $f_7(snr)$  es un valor de reposición de referencia correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual, y  $f_8(snr)$  es un coeficiente de ponderación de la longitud de resaca  $hangover\_nosie\_tbl[\text{valor de característica fluctuante}]$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. La tercera unidad de actualización 724 está configurada para actualizar la longitud de resaca en el parámetro vinculado con el criterio de decisión de detección VAD para el valor máximo de reposición del contador de resaca calculado  $hangover\_max$  adquirido por la tercera unidad de adquisición 723.

La Figura 10 es una vista estructural esquemática de otro ejemplo de un aparato para VAD. El aparato PA 223A VAD en conformidad con este ejemplo, puede configurarse para poner en práctica el método para VAD del ejemplo ilustrado en la Figura 2. En este ejemplo, el valor de característica fluctuante es concretamente un valor cuantizado  $idx$  de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo. En correspondencia, el módulo de adquisición 601 incluye una unidad de recepción 731, una primera unidad de procesamiento de división 732, una unidad de decisión 733, una primera unidad de cálculo 734, una unidad de blanqueado 735, una cuarta unidad de adquisición 736, una quinta unidad de adquisición 737 y una unidad de procesamiento de cuantización 738. La unidad de recepción 731 está configurada para recibir una trama actual de la señal de entrada. La primera unidad de procesamiento de división 732 está configurada para dividir la trama actual de la señal de entrada recibida por la unidad de recepción 731 en  $N$  sub-bandas en un dominio de la frecuencia, en donde  $N$  es un número entero mayor que 1, y las energías  $enrg(i)$  (en donde  $i=0, 1, \dots, N-1$ ) de las  $N$  sub-bandas se calculan respetivamente. La unidad de decisión 733 está configurada para decidir si la trama actual de la señal de entrada recibida por la unidad de recepción 731 es una trama de ruido de fondo en conformidad con el criterio de decisión de detección VAD. La primera unidad de cálculo 734 está configurada para calcular una energía de media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente en las  $N$  sub-bandas utilizando la fórmula  $enrg\_n(i) = \alpha \cdot enrg\_n + (1 - \alpha) \cdot enrg(i)$  cuando la trama actual es una trama de ruido de fondo, en donde  $\alpha$  es un coeficiente de olvido para controlar una tasa de actualización de la energía de media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente en las  $N$  sub-bandas y  $enrg\_n$  es una energía de la trama de ruido de fondo. La unidad de blanqueado 735 está configurada para blanquear un espectro de la trama de ruido de fondo actual utilizando la fórmula  $enrg\_w(i) = enrg(i) / enrg\_n(i)$ , y para adquirir una energía  $enrg\_w(i)$  del ruido blanqueado de fondo en una  $i$ -ésima sub-banda. La cuarta unidad de adquisición 736 está configurada para adquirir

$$hb = - \sum_{i=0}^{N-1} p_i \cdot \log p_i ,$$

una entropía espectral del ruido blanqueado de fondo  $hb$  utilizando la fórmula en donde

$$p_i = enrg\_w(i) / \sum_{i=0}^{N-1} enrg\_w(i) .$$

La quinta unidad de adquisición 737 está configurada para adquirir una media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo utilizando la fórmula  $hb\_noise\_mov = \beta \cdot hb\_noise\_mov + (1 - \beta) \cdot hb$ , en donde  $\beta$  es un factor de olvido para controlar una tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo. La unidad de procesamiento de cuantización 738 está configurada para cuantizar la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo utilizando la fórmula  $idx = \lfloor (hb\_noise\_mov - A) / B \rfloor$ , con el fin de adquirir un valor cuantizado  $idx$ , en donde  $A$  y  $B$  son valores preestablecidos y pueden ser valor empíricos seleccionados en conformidad con las demandas reales.

La Figura 11 es una vista estructural esquemática de otro ejemplo de un aparato para VAD. Cuando una tasa de actualización del parámetro a largo plazo relacionado con el ruido de fondo incluye la tasa de actualización de una energía media móvil a largo plazo  $enrg\_n(i)$  del ruido de fondo, en comparación con el ejemplo ilustrado en la Figura 10, en el aparato para VAD en conformidad con este ejemplo, el módulo de ajuste 602 incluye una cuarta unidad de memorización 741, una cuarta unidad de búsqueda 742 y una cuarta unidad de actualización 743. La cuarta unidad de memorización 741 está configurada para memorizar una tabla de tasa de actualización de ruido de fondo  $alpha\_tbl[]$ , en donde la tabla de tasa de actualización del ruido de fondo  $alpha\_tbl[]$  incluye un mapeado de correspondencia entre el valor cuantizado y el coeficiente de olvido de la tasa de actualización de la energía de media móvil a largo plazo  $enrg\_n(i)$ . La cuarta unidad de búsqueda 742 está configurada para consultar la tabla de tasa de actualización del ruido de fondo  $alpha\_tbl[]$  desde la cuarta unidad de memorización 741, y para adquirir un coeficiente de olvido  $\alpha$  de la tasa de actualización de la energía de media móvil a largo plazo  $enrg\_n(i)$  correspondiente al valor cuantizado  $idx$  del ruido de fondo. La cuarta unidad de actualización 743 está configurada para utilizar el coeficiente de olvido  $\alpha$  adquirido por la cuarta unidad de búsqueda 742 como un coeficiente de olvido para controlar la tasa de actualización de la energía de media móvil a largo plazo  $enrg\_n(i)$  de la trama de ruido de fondo respectivamente en las  $N$  sub-bandas.

La Figura 12 es una vista estructural esquemática de otro ejemplo de un aparato para la VAD. Cuando la tasa de actualización del parámetro a largo plazo relacionado con el ruido de fondo incluye una tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  de una entropía espectral del ruido blanqueado de fondo, en comparación con el ejemplo ilustrado en la Figura 10, en el aparato para VAD en conformidad con este ejemplo, el módulo de ajuste 602 incluye una quinta unidad de memorización 744, una quinta unidad de búsqueda 745 y una quinta unidad de actualización 746. La quinta unidad de memorización 744 está configurada para memorizar una tabla de tasa de actualización de fluctuación de ruido de fondo  $beta\_tbl[]$ , en donde el tabla de tasa de actualización de fluctuación de ruido de fondo  $beta\_tbl[]$  incluye un mapeado de correspondencia entre el valor cuantizado y el factor de olvido de la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$ . La quinta unidad de búsqueda 745 está configurada para consultar la tabla de tasa de actualización de fluctuación de ruido de fondo  $beta\_tbl[]$  desde la quinta unidad de memorización 744, y para adquirir un factor de olvido  $\beta$  de la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  correspondiente al valor cuantizado  $idx$  del ruido de fondo. La quinta unidad de actualización 746 está configurada para usar el factor de olvido  $\beta$  adquirido por la quinta unidad de búsqueda 745 como un factor de olvido para controlar la tasa de actualización de la media móvil a largo plazo  $hb\_noise\_mov$  de

una entropía espectral del ruido blanqueado de fondo.

La Figura 13 es una vista estructural esquemática de una octava forma de realización del aparato para VAD en conformidad con la presente invención. El aparato para VAD en conformidad con esta forma de realización puede configurarse para poner en práctica el método para VAD en la forma de realización ilustrada en la Figura 3 de la presente invención. En esta forma de realización, el valor de característica fluctuante es concretamente una media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr_n$  mov. En correspondencia, el módulo de adquisición 601 incluye la unidad de recepción 731, la unidad de decisión 733 y una sexta unidad de adquisición 751. La unidad de recepción 731 está configurada para recibir una trama actual de la señal de entrada. La unidad de decisión 733 está configurada para decidir si la trama actual de la señal de entrada recibida por la unidad de recepción 731 es una trama de ruido de fondo en conformidad con el criterio de decisión de detección VAD. La sexta unidad de adquisición 751 está configurada para adquirir una media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr_n$  mov en conformidad con la fórmula  $snr_n mov = k \cdot snr_n mov + (1 - k) \cdot snr$  en función de un resultado de decisión de la unidad de decisión 733 cuando la trama actual es una trama de ruido de fondo, en donde  $snr$  es una relación SNR de la trama de ruido de fondo actual, y  $k$  es un factor de olvido para controlar una tasa de actualización de la media móvil a largo plazo de la relación SNR de trama de ruido de fondo  $snr_n$  mov.

Además, haciendo referencia a la Figura 13, cuando la tasa de actualización del parámetro a largo plazo relacionado con el ruido de fondo incluye la tasa de actualización de la media móvil a largo plazo  $snr_n$  mov, el módulo de ajuste 602 puede incluir una unidad de control 752, configurada para establecer diferentes valores para el factor de olvido  $k$  para controlar la tasa de actualización de la media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr_n$  mov cuando la relación SNR  $snr$  de la trama de ruido de fondo actual es mayor que una media  $snr_n$  de SNRs de las últimas  $n$  tramas de ruido de fondo y cuando la relación SNR  $snr$  de la trama de ruido de fondo actual es menor que el valor medio  $snr_n$  de SNRs de las últimas  $n$  tramas de ruido de fondo.

La Figura 14 es una vista estructural esquemática de otro ejemplo de un aparato para VAD. El aparato para VAD en conformidad con este ejemplo puede configurarse para poner en práctica el método para VAD en el ejemplo ilustrado en la Figura 4. En este ejemplo, el valor de característica fluctuante es concretamente una trama de ruido de fondo MSSNR media móvil a largo plazo  $flux_{bgd}$ . En correspondencia, el módulo de adquisición 601 incluye la unidad de recepción 731, la unidad de decisión 733, una segunda unidad de procesamiento de división 761, una segunda unidad de cálculo 762, una tercera unidad de cálculo 763, una unidad de modificación 764, una séptima unidad de adquisición 765 y una cuarta unidad de cálculo 766. La unidad de recepción 731 está configurada para recibir una trama actual de la señal de entrada. La unidad de decisión 733 está configurada para decidir si la trama actual de la señal de entrada recibida por la unidad de recepción 731 es una trama de ruido de fondo en conformidad con el criterio de decisión de detección VAD. La segunda unidad de procesamiento de división 761 está configurada para dividir el espectro de FFT de la trama de ruido de fondo actual en  $H$  sub-bandas en conformidad con el resultado de decisión de la unidad de decisión 733 cuando la trama actual es una trama de ruido de fondo, en donde  $H$  es un número entero mayor que 1 y para calcular las energías  $E_{band}(i)$  (en donde  $i=0, 1, \dots, H-1$ )

$$E_{band}(i) = \frac{P}{h(i) - l(i) + 1} \sum_{j=l(i)}^{h(i)} S_j + (1 - p) E_{band\_old}(i),$$

de  $i$  sub-bandas respetivamente utilizando la fórmula en donde  $l(i)$  y  $h(i)$  representan un punto de secuencia de FFT con la más baja frecuencia y un punto de frecuencia de FFT con la más alta frecuencia en una  $i$ -ésima sub-banda respectivamente,  $S_j$  representa una energía de un  $j$ -ésimo punto de frecuencia en el espectro de FFT,  $E_{band\_old}(i)$  representa una energía de la  $i$ -ésima sub-banda en una trama anterior de la trama de ruido de fondo actual y  $P$  es una constante preestablecida, que puede establecerse concretamente en conformidad con valores empíricos. La segunda unidad de cálculo 762 está configurada para

actualizar una media móvil a largo plazo de ruido de fondo  $\overline{E_{band\_n}(i)}$  usando la energía de la  $i$ -ésima sub-banda en una trama de ruido de fondo anterior utilizando la fórmula  $\overline{E_{band\_n}(i)} = q \cdot \overline{E_{band\_n}(i)} + (1 - q) \cdot E_{band}(i)$ , en donde  $q$  es una constante preestablecida y puede establecerse concretamente en función de valores empíricos. La tercera unidad de cálculo 763 está configurada calcular una relación SNR  $snr(i)$  de  $i$ -ésima sub-banda en la trama de ruido de fondo actual respetivamente utilizando la fórmula  $snr(i) = 10 \log(E_{band}(i) / \overline{E_{band\_n}(i)})$ . La unidad de modificación 764 está configurada para modificar la relación  $snr(i)$  de la  $i$ -ésima sub-banda en la trama de ruido de fondo actual respetivamente utilizando la fórmula

$$msnr(i) = \begin{cases} \text{MAX}[\text{MIN}[\frac{snr(i)^3}{C1}, 1], 0] & , i \in \text{primer conjunto} \\ \text{MAX}[\text{MIN}[\frac{snr(i)^3}{C2}, 1], 0] & , i \in \text{segundo conjunto} \end{cases}$$

en donde  $msnr(i)$  es la relación SNR  $snr$  de la  $i$ -ésima sub-banda modificada,  $C1$  y  $C2$  son constantes reales preestablecidas mayores que 0 y los valores en el primer conjunto y en el segundo conjunto forman un conjunto  $[0,$

H-1]. La séptima unidad de adquisición 765 está configurada para adquirir una trama de ruido de fondo actual

$$MSSNR = \sum_{i=0}^{H-1} msnr(i).$$

MSSNR utilizando la fórmula La cuarta unidad de cálculo 766 está configurada para calcular una media móvil a largo plazo de MSSNR de la trama de ruido de fondo actual  $flux_{bgd}$  utilizando la fórmula  $flux_{bgd} = r \cdot flux_{bgd} + (1 - r) \cdot MSSNR$ , en donde r es un coeficiente de olvido para controlar una tasa de actualización de la media móvil a largo plazo de MSSNR de la trama de ruido de fondo actual  $flux_{bgd}$ .

La Figura 15 es una vista estructural esquemática de otro ejemplo de un aparato para VAD. En comparación con el aparato para VAD en el ejemplo ilustrado en la Figura 14, en el aparato para VAD en conformidad con este ejemplo, cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye el umbral de decisión primario, el módulo de ajuste 602 incluye la primera unidad de memorización 701, la primera unidad de búsqueda 702, la primera unidad de adquisición 703 y la primera unidad de actualización 704. La primera unidad de memorización 701 está configurada para memorizar un mapeado de correspondencia entre un valor de característica fluctuante y un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$ . La primera unidad de búsqueda 702 está configurada para consultar el mapeado de correspondencia entre el valor de característica fluctuante y el sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  desde la primera unidad de memorización 701 y para adquirir un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente a un valor de característica fluctuante de un ruido de fondo, en donde el sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  se utiliza para representar un valor de sesgo del umbral bajo un ruido de fondo con diferentes fluctuación. La primera unidad de adquisición 703 está configurada para adquirir un umbral de decisión primario  $vad\_thr$  utilizando la fórmula  $vad\_thr = f_1(snr) + f_2(snr) \cdot thr\_bias\_noise$ , en donde  $f_1(snr)$  es un umbral de referencia correspondiente a una relación SNR  $snr$  de una trama de ruido de fondo actual, y  $f_2(snr)$  es un coeficiente de ponderación de un sesgo de fluctuación del ruido del umbral de decisión  $thr\_bias\_noise$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. La primera unidad de actualización 704 está configurada para actualizar el umbral de decisión primario en el parámetro vinculado con el criterio de decisión de detección VAD para el umbral de decisión primario  $vad\_thr$  adquirido por la primera unidad de adquisición 703.

La Figura 16 es una vista estructural esquemática de otro ejemplo para un aparato para VAD. En comparación con el aparato para VAD en el ejemplo ilustrado en la Figura 14, el aparato para VAD en conformidad con este ejemplo, cuando el parámetro vinculado con el criterio de decisión de detección VAD incluye el umbral de decisión primario, el módulo de ajuste 602 incluye una sexta unidad de memorización 767, una octava unidad de adquisición 768, una sexta unidad de búsqueda 769 y una sexta unidad de actualización 770. La sexta unidad de memorización 767 está configurada para memorizar una tabla de umbral de decisión primario  $thr\_tbl[]$ , en donde la tabla de umbral de decisión primario  $thr\_tbl[]$  incluye un mapeado de correspondencia entre el nivel de fluctuación, el nivel de SNR, y el umbral de decisión primario  $vad\_thr$ . La octava unidad de adquisición 768 está configurada para adquirir el nivel de fluctuación  $flux\_idx$  correspondiente a la media móvil a largo plazo de MSSNR de la trama de ruido de fondo actual  $flux_{bgd}$  calculada por la cuarta unidad de cálculo 766, y adquirir el nivel de SNR  $snr\_idx$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual. La sexta unidad de búsqueda 769 está configurada para buscar un umbral de decisión primario  $thr\_tbl[snr\_idx][flux\_idx]$  que corresponde simultáneamente al nivel de fluctuación  $flux\_idx$  y el nivel SNR  $snr\_idx$  a partir de la tabla del umbral de decisión primario  $thr\_tbl[]$  memorizada por la sexta unidad de memorización 767. La sexta unidad de actualización 770 está configurada para actualizar el umbral de decisión primario en el parámetro vinculado con el criterio de decisión para el umbral de decisión primario  $thr\_tbl[snr\_idx][flux\_idx]$  buscado por la sexta unidad de búsqueda.

Además, en el aparato para VAD ilustrado en la Figura 16, la tabla del umbral de decisión primario  $thr\_tbl[]$  puede incluir concretamente un mapeado entre el nivel de fluctuación, el nivel de SNR, la tendencia de la decisión, y el umbral de decisión primario  $vad\_thr$ . En correspondencia, la octava unidad de adquisición 768 está configurada, además, para adquirir una tendencia de decisión  $op\_idx$  correspondiente al rendimiento de trabajo actual del aparato para VAD que realiza la decisión de VAD, es decir, está propenso a decidir la trama actual para ser una trama vocal o una trama de ruido de fondo. Más concretamente, el rendimiento del trabajo actual del aparato para VAD puede incluir la economía de ancho de banda por la calidad de codificación de voz después del inicio de VAD y de la VAD propiamente dicha. La sexta unidad de búsqueda 769 está configurada concretamente para buscar un umbral de decisión primario  $vad\_thr = thr\_tbl[snr\_idx][flux\_idx][op\_idx]$  correspondiente al nivel de fluctuación  $flux\_idx$ , el nivel de  $snr\_idx$ , y el nivel de rendimiento  $op\_idx$  simultáneamente a partir de la tabla del umbral de decisión primario  $thr\_tbl[]$  memorizado por la sexta unidad de memorización 767. La sexta unidad de actualización 770 está configurada concretamente para actualizar el umbral de decisión primario en el parámetro vinculado con el criterio de decisión para el umbral de decisión primario  $vad\_thr = thr\_tbl[snr\_idx][flux\_idx][op\_idx]$  buscado por la sexta unidad de búsqueda 769.

Además, en el aparato para VAD en conformidad con una forma de realización preferida de la presente invención, está incluido un módulo de control 605, configurado para ajustar dinámicamente el parámetro vinculado con el criterio de decisión de VAD que es: la longitud de resaca vocal en conformidad con el nivel de ruido de fondo en la señal de entrada. La Figura 16 ilustra una realización ejemplo. Más concretamente, cualquiera o más parámetros vinculados con el criterio de decisión de detección de VAD: el umbral de decisión primario, la longitud de resaca vocal, y la condición de iniciación de resaca vocal pueden ajustarse dinámicamente con el proceso en la forma de

realización ilustrada en la Figura 5.

Un codificador puede incluir concretamente el aparato para VAD en conformidad con cualquier forma de realización ejemplo ilustrada en las Figuras 6 a 16 de la presente invención.

5 Los expertos en esta técnica deben entender que la totalidad o una parte de las etapas del método en conformidad con la forma de realización de la presente invención pueden ponerse en práctica mediante un programa que proporcione instrucciones a un hardware pertinente. El programa puede memorizarse en un soporte de memorización legible por ordenador. Cuando se ejecuta el programa se realizan las etapas del método en  
10 conformidad con la forma de realización de la presente invención. El soporte de memorización puede ser cualquier soporte que sea capaz de memorizar códigos de programas, tales como una memoria ROM, una memoria RAM, un disco magnético y un disco óptico.

15 En conformidad con las formas de realización de la presente invención, cuando una señal de entrada es un ruido de fondo, se adquiere un valor de característica fluctuante utilizado para representar la fluctuación del ruido de fondo, se realiza un ajuste adaptativo sobre un parámetro vinculado con un criterio de decisión de VAD en función del valor de característica fluctuante, y se realiza una decisión de VAD sobre la señal de entrada utilizando el parámetro vinculado con el criterio de decisión sobre el que se realiza el ajuste adaptativo. En comparación con la técnica anterior, se puede conseguir un más alto rendimiento de la decisión de VAD en el caso de diferentes tipos de ruidos  
20 de fondo, lo que mejora la eficiencia de decisión de VAD y la exactitud de la decisión, lo que permite aumentar la utilización de los recursos de ancho de banda de canal limitados.

25 Por último, conviene señalar que las anteriores formas de realización y ejemplos se proporcionan simplemente para describir las soluciones técnicas de la presente invención, pero no están previstas para limitar el alcance de la presente invención.



## REIVINDICACIONES

1. Un método para Detección de Actividad Vocal (VAD), que comprende:

- 5 la adquisición (101) de un valor de característica fluctuante de un ruido de fondo cuando una señal de entrada es el ruido de fondo, en donde el valor de característica fluctuante se utiliza para representar la fluctuación del ruido de fondo;
- 10 la realización (102) de un ajuste adaptativo sobre un parámetro vinculado con el criterio de decisión de detección VAD en conformidad con el valor de característica fluctuante, en donde el parámetro vinculado con el criterio de decisión comprende una longitud de resaca vocal; y
- 15 la realización (103) de una decisión de detección VAD de la señal de entrada utilizando el parámetro vinculado con el criterio de decisión después de que se realice el ajuste adaptativo,
- en donde el valor de la característica fluctuante es concretamente una media móvil a largo plazo de la relación SNR de trama de ruido de fondo  $snr_n_{mov}$ ; y
- 20 la adquisición (101) del valor de característica fluctuante del ruido de fondo cuando la señal de entrada es el ruido de fondo comprende:
- la recepción (301) de una trama actual de la señal de entrada;
- 25 la decisión (302) de si la trama actual es una trama de ruido de fondo en conformidad con un criterio de decisión de detección VAD; y
- la adquisición (303) de una media móvil a largo plazo de la relación SNR de trama de ruido de fondo  $snr_n_{mov}$  utilizando la fórmula  $snr_n_{mov} = k \cdot snr_n_{mov} + (1-k) \cdot snr$ , cuando la trama actual es la trama de ruido de fondo, en donde  $snr$  es una Relación de Señal a Ruido SNR, de una trama de ruido de fondo actual, y  $k$  es un factor de olvido para controlar una tasa de actualización de la media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr_n_{mov}$ ,
- en donde la realización (102) del ajuste adaptativo sobre el parámetro vinculado con el criterio de decisión de detección VAD en conformidad con el valor de característica fluctuante comprende:
- 35 la búsqueda de una longitud de resaca vocal  $hangover\_nosie\_tb$ [valor de característica fluctuante] que corresponde al valor de la característica fluctuante del ruido de fondo a partir de una tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca  $hangover\_noise\_tb1$ ];
- 40 la adquisición de un valor máximo de reposición del contador de resaca  $hangover\_max$  utilizando la fórmula  $hangover\_max = f_7(snr) + f_8(snr) \cdot hangover\_nosie\_tb$ [valor de característica fluctuante], en donde  $f_7(snr)$  es un valor de reposición de referencia correspondiente a una Relación de Señal a Ruido, SNR,  $snr$  de una trama de ruido de fondo actual, y  $f_8(snr)$  es un coeficiente de ponderación de una longitud de resaca  $hangover\_nosie\_tb$ [valor de característica fluctuante] correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual; y
- 45 la actualización de la longitud de resaca en el parámetro vinculado con el criterio de decisión para el valor máximo de reposición del contador de resaca vocal adquirido  $hangover\_max$ .

2. Un aparato para la Detección de Actividad Vocal (VAD), que comprende:

- 50 un módulo de adquisición (601), configurado para adquirir un valor de característica fluctuante de un ruido de fondo cuando una señal de entrada es el ruido de fondo, en donde el valor de característica fluctuante se utiliza para representar la fluctuación del ruido de fondo;
- 55 un módulo de ajuste (602), configurado para realizar un ajuste adaptativo en un parámetro vinculado con el criterio de decisión de detección VAD en conformidad con el valor de característica fluctuante;
- un módulo de decisión (603), configurado para realizar una decisión de detección VAD sobre la señal de entrada utilizando el parámetro vinculado con el criterio de decisión después de que se haya efectuado el ajuste adaptativo;
- 60 y
- un módulo de memorización (604), configurado para memorizar el parámetro vinculado con el criterio de decisión de detección VAD, en donde el parámetro vinculado con el criterio de decisión comprende una longitud de resaca vocal,
- 65 en donde el valor de característica fluctuante es concretamente una media móvil a largo plazo de la relación SNR de la trama de ruido de fondo  $snr_n_{mov}$ ;

el módulo de adquisición (601) comprende:

una unidad de recepción (731), configurada para recibir una trama actual de la señal de entrada;

una unidad de decisión (733), configurada para decidir si la trama actual de la señal de entrada es una trama de ruido de fondo en conformidad con un criterio de decisión de detección VAD; y

una sexta unidad de adquisición (751), configurada para adquirir una media móvil a largo plazo de la Relación de Señal a Ruido, SNR, de la trama de ruido de fondo  $snr_n_{mov}$  utilizando la fórmula  $snr_n_{mov} = k \cdot snr_n_{mov} + (1-k) \cdot snr$  en función de un resultado de decisión de la unidad de decisión cuando la trama actual es una trama de ruido de fondo, en donde  $snr$  es una relación SNR de la trama de ruido de fondo actual y en donde  $k$  representa un factor de olvido que permite controlar una tasa de actualización de la media móvil a largo plazo de la relación SNR de trama de ruido de fondo  $snr_n_{mov}$ ,

en donde el módulo de ajuste comprende:

una tercera unidad de memorización (721), configurada para memorizar una tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca vocal  $hangover\_noise\_tbl[]$ , en donde la tabla de mapeado de la fluctuación de ruido de la longitud de resaca vocal  $hangover\_noise\_tbl[]$  comprende un mapeado de correspondencia entre el valor de característica fluctuante y la longitud de resaca;

una tercera unidad de búsqueda (722), configurada para buscar una longitud de resaca vocal  $hangover\_nosie\_tbl[\text{valor de característica fluctuante}]$  que corresponde al valor de característica fluctuante del ruido de fondo a partir de la tabla de mapeado de correspondencia de la fluctuación de ruido de longitud de resaca vocal  $hangover\_noise\_tbl[]$ ;

una tercera unidad de adquisición (723), configurada para adquirir un valor máximo de reposición del contador de resaca vocal  $hangover\_max$  utilizando la fórmula  $hangover\_max = f_7(snr) + f_8(snr) \cdot hangover\_nosie\_tbl[\text{valor de característica fluctuante}]$ , en donde  $f_7(snr)$  es un valor de reposición de referencia correspondiente a una Relación de Señal a Ruido, SNR,  $snr$  de la trama de ruido de fondo actual, y  $f_8(snr)$  es un coeficiente de ponderación de la longitud de resaca  $hangover\_nosie\_tbl[\text{valor de característica fluctuante}]$  correspondiente a la relación SNR  $snr$  de la trama de ruido de fondo actual; y

una tercera unidad de actualización (724), configurada para actualizar la longitud de resaca en el parámetro vinculado con el criterio de decisión para el valor máximo de reposición del contador de resaca vocal  $hangover\_max$  adquirido por la tercera unidad de adquisición.

**3.** El aparato según la reivindicación 2 que comprende, además:

un módulo de control (605), configurado para ajustar dinámicamente el parámetro vinculado con el criterio de decisión que es la longitud de resaca vocal en conformidad con un nivel del ruido de fondo en la señal de entrada.

**4.** Un soporte de memorización legible por ordenador, que comprende códigos de programa informático que cuando se ejecutan por un procesador de ordenador, hacen que el procesador de ordenador ejecute las etapas en conformidad con lo dispuesto en la reivindicación 1.

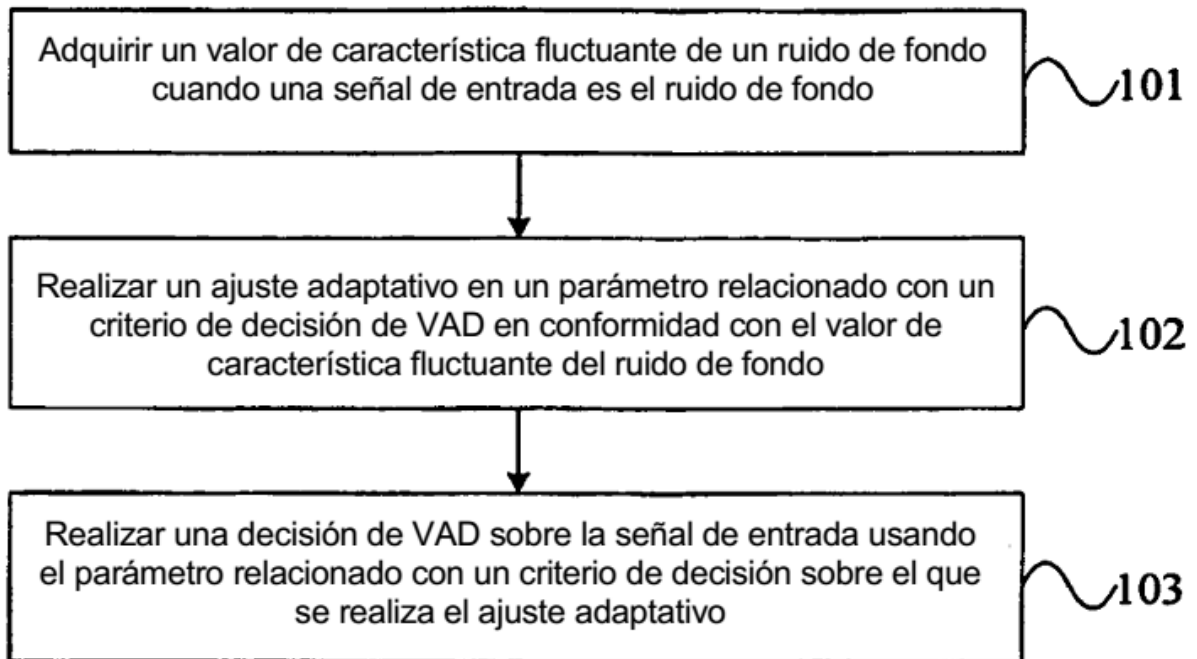


FIG. 1

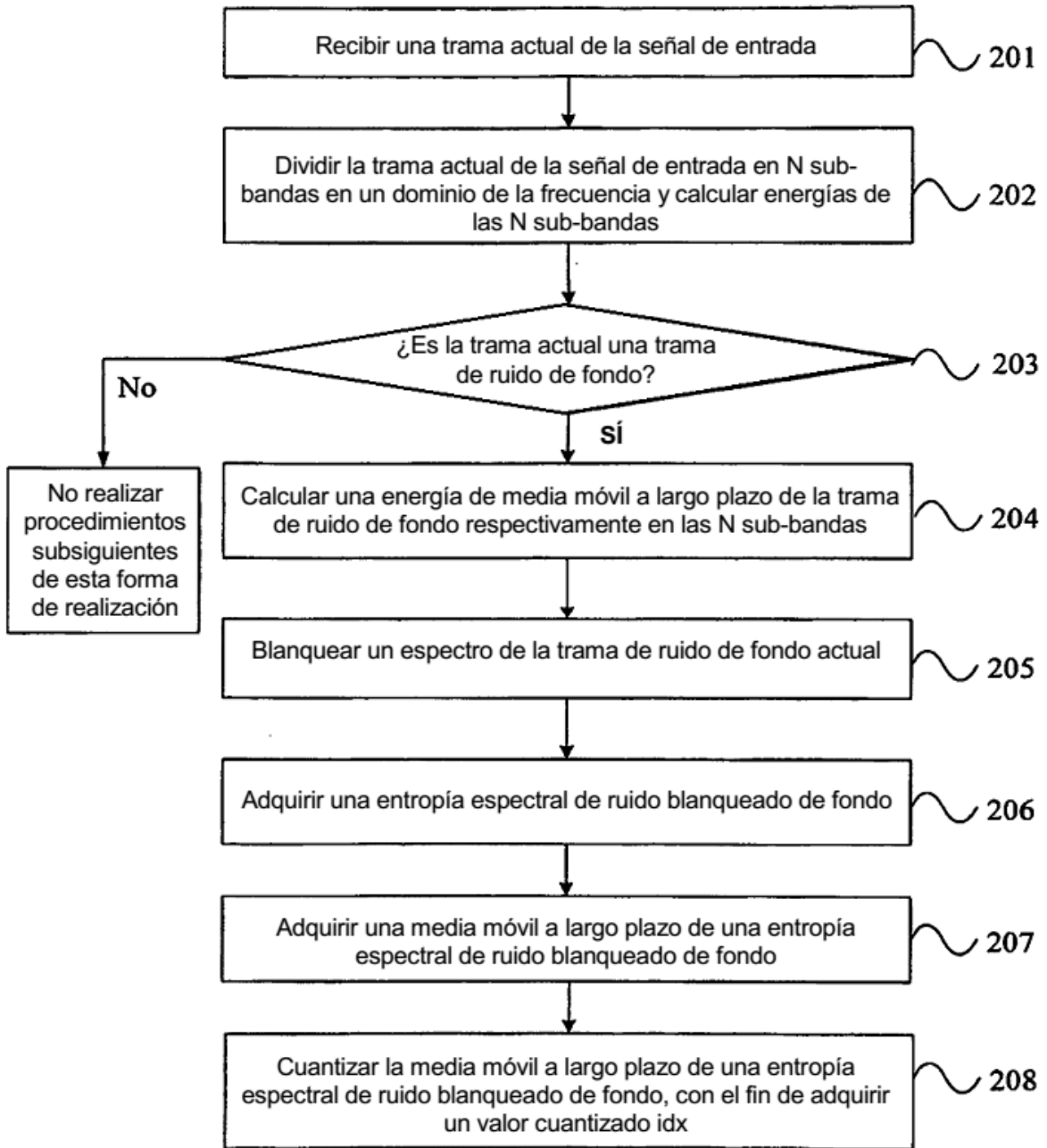


FIG. 2

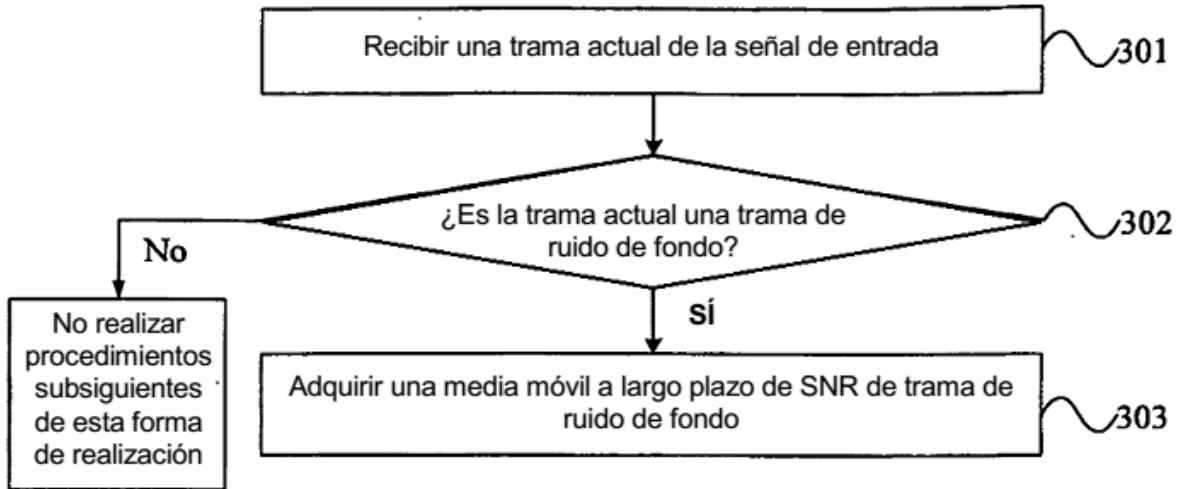


FIG. 3

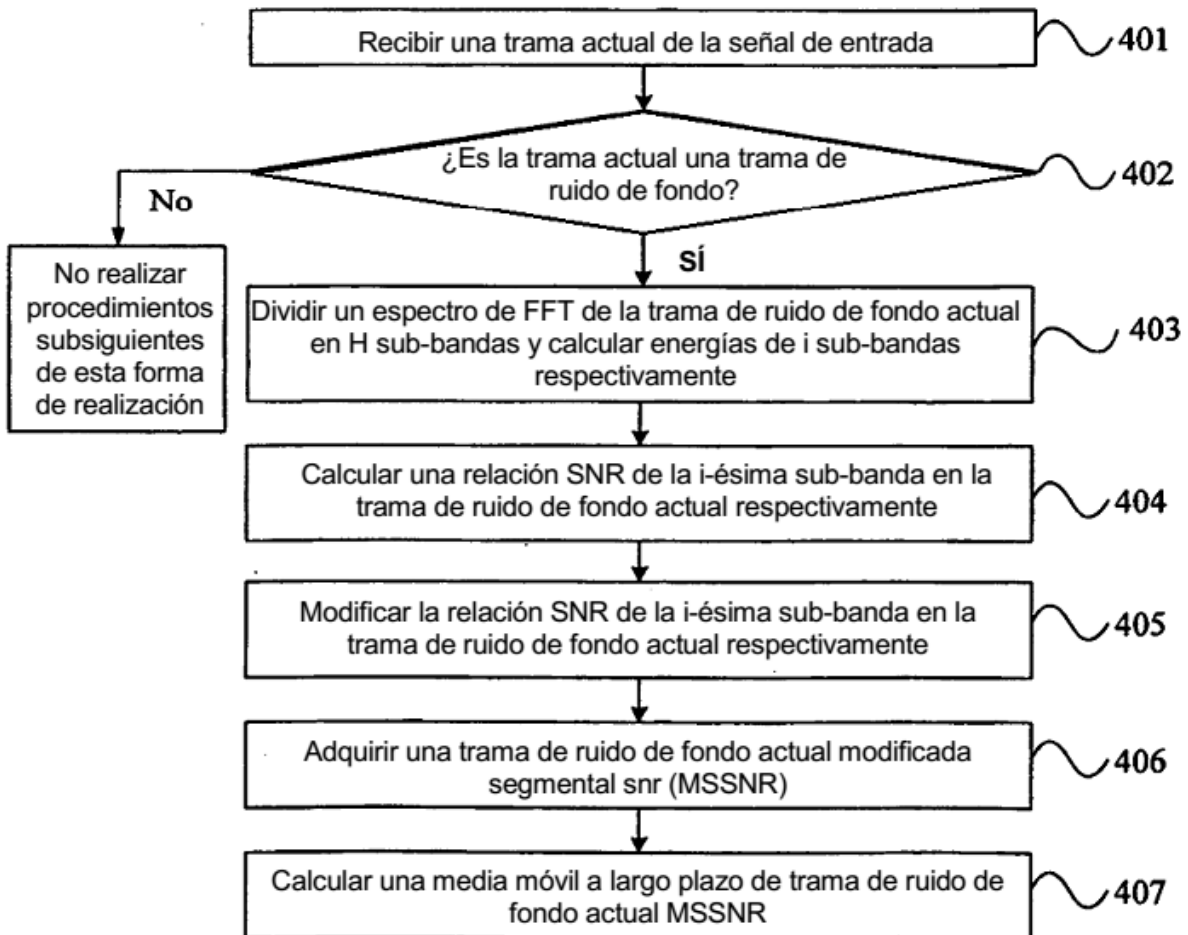


FIG. 4

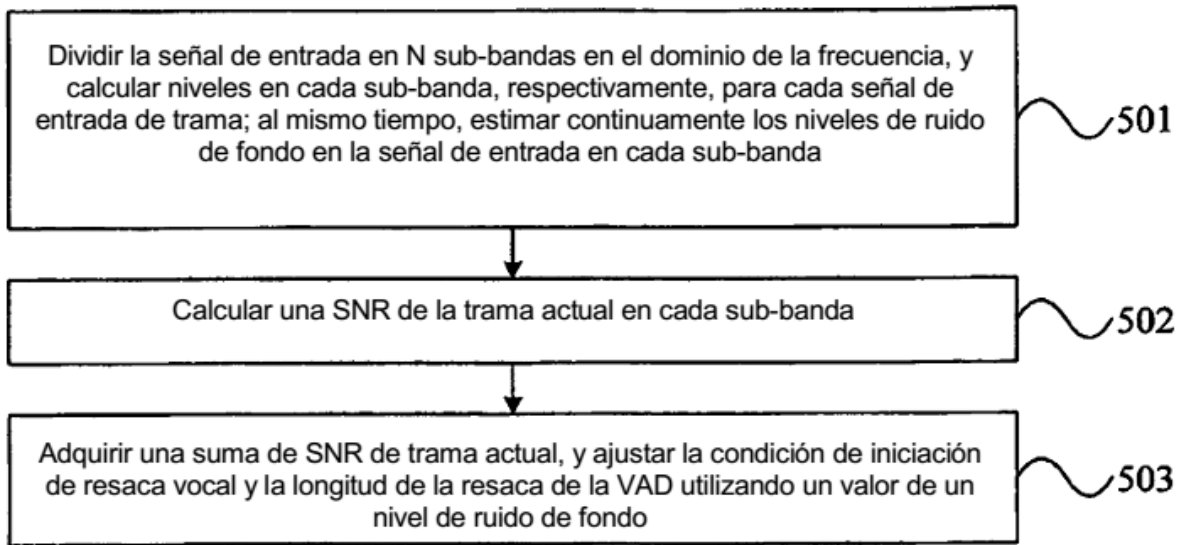


FIG. 5

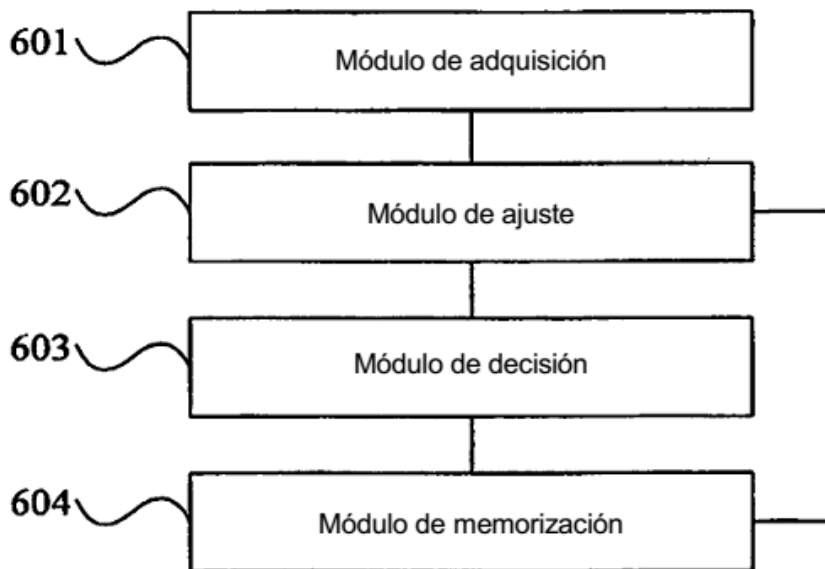


FIG. 6

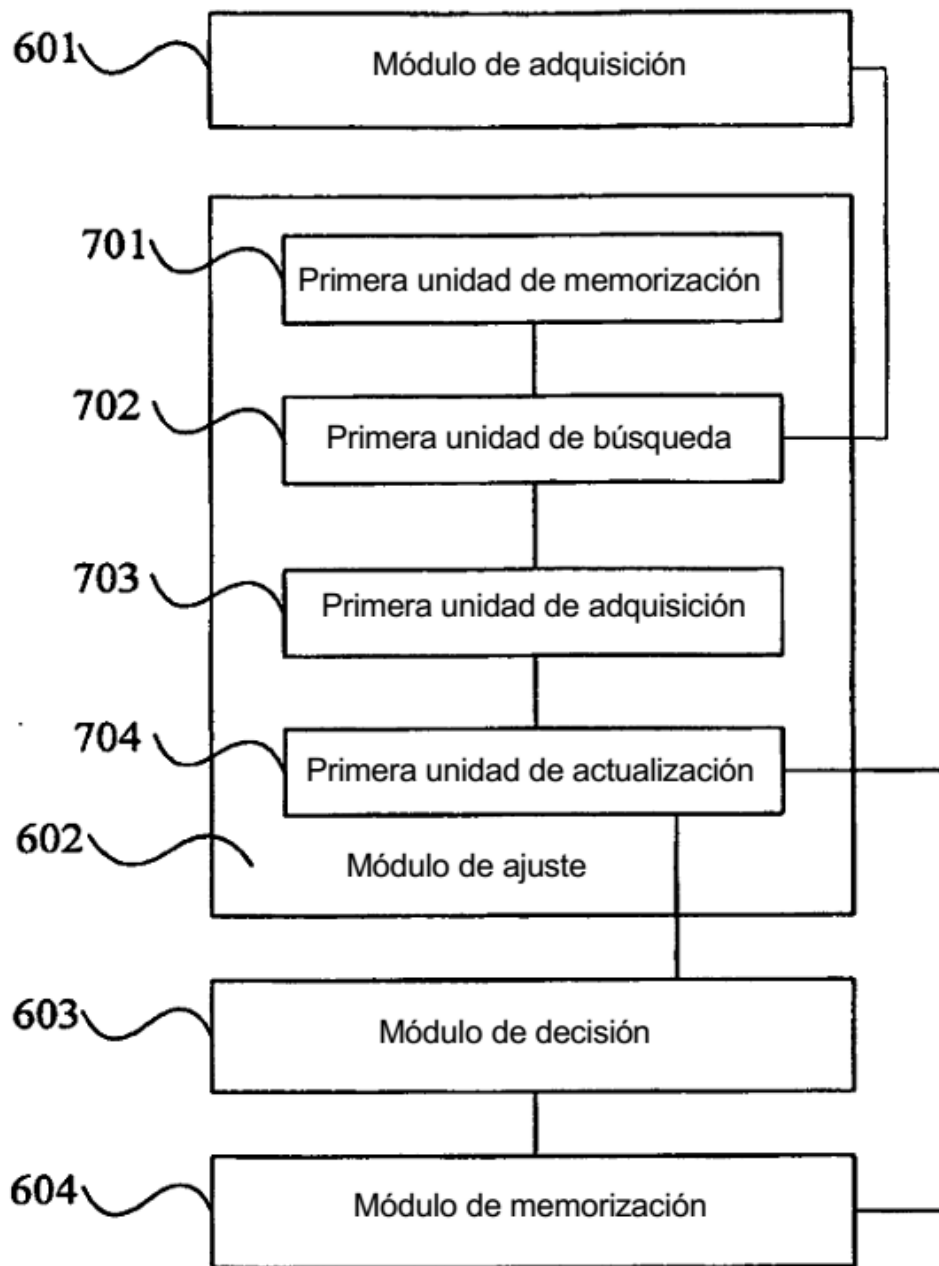


FIG. 7

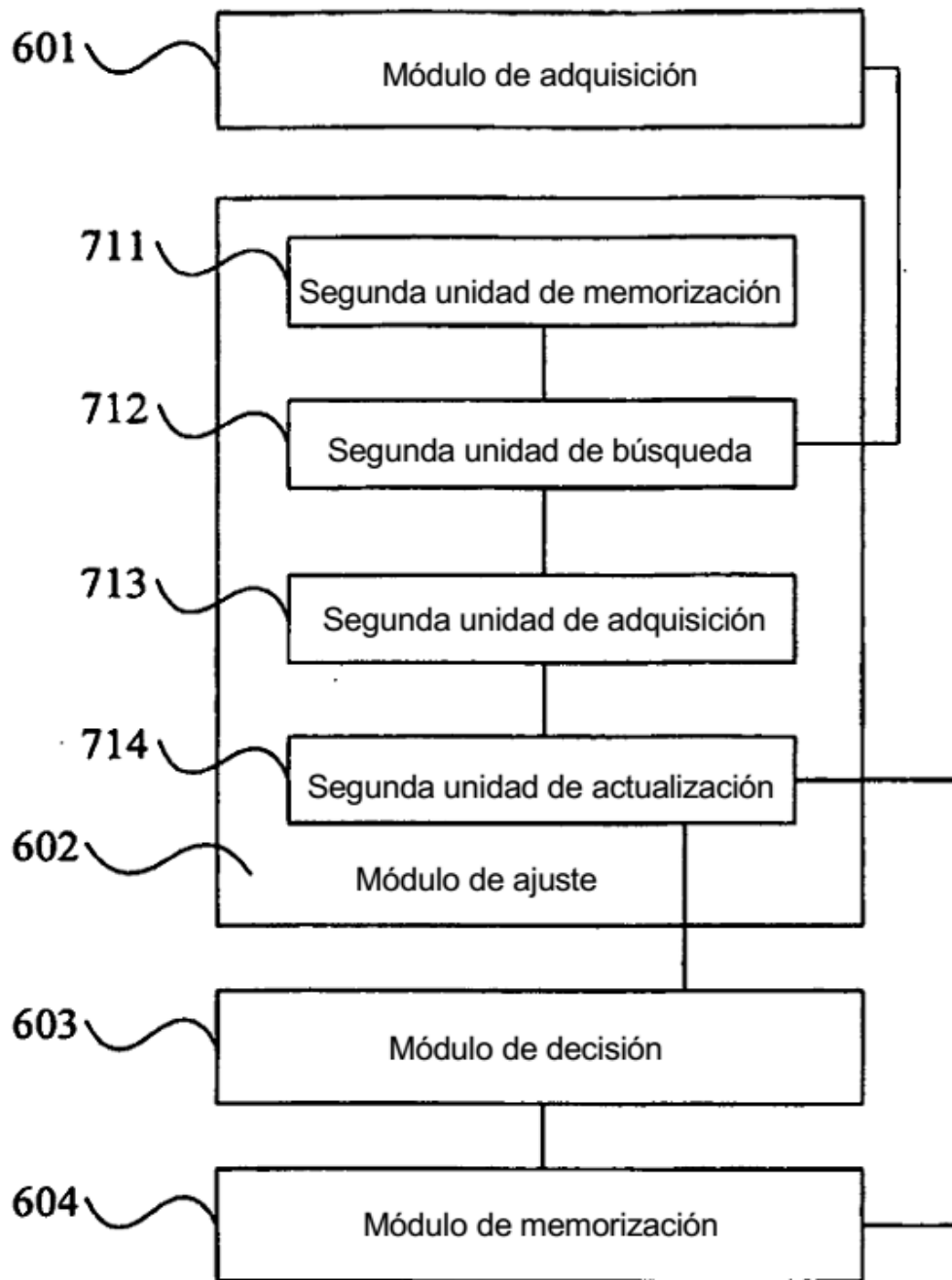


FIG. 8



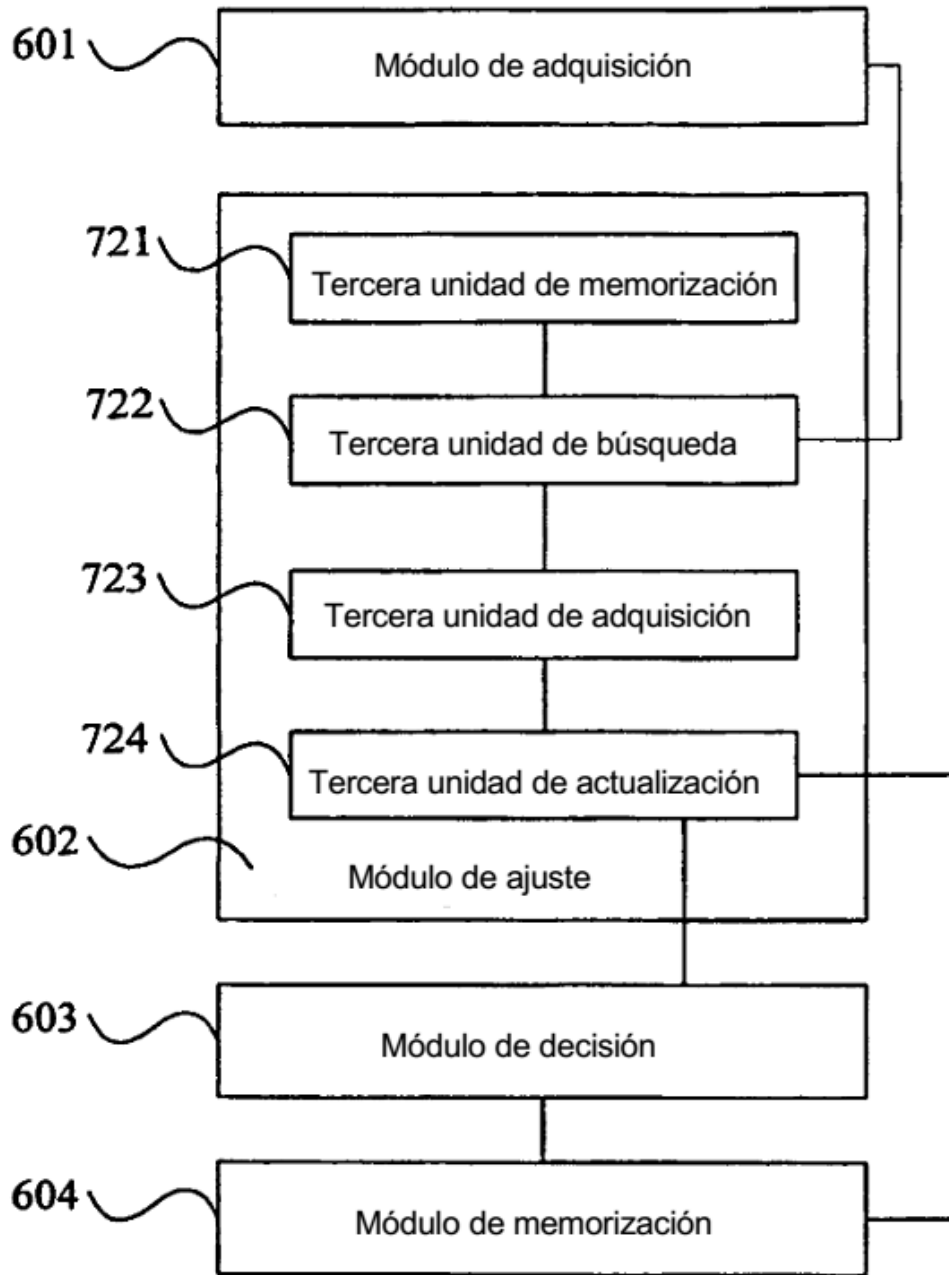


FIG. 9

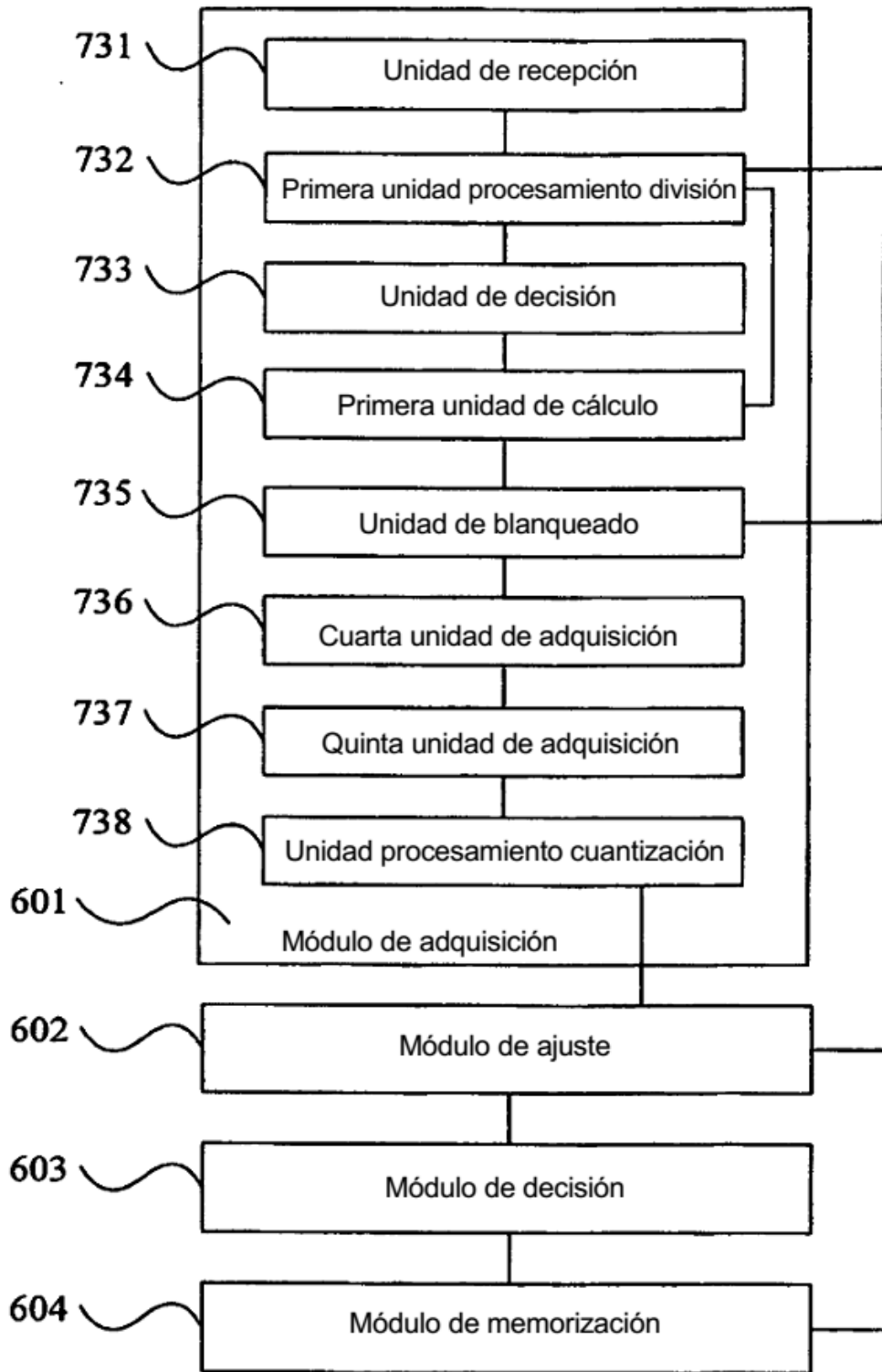


FIG. 10

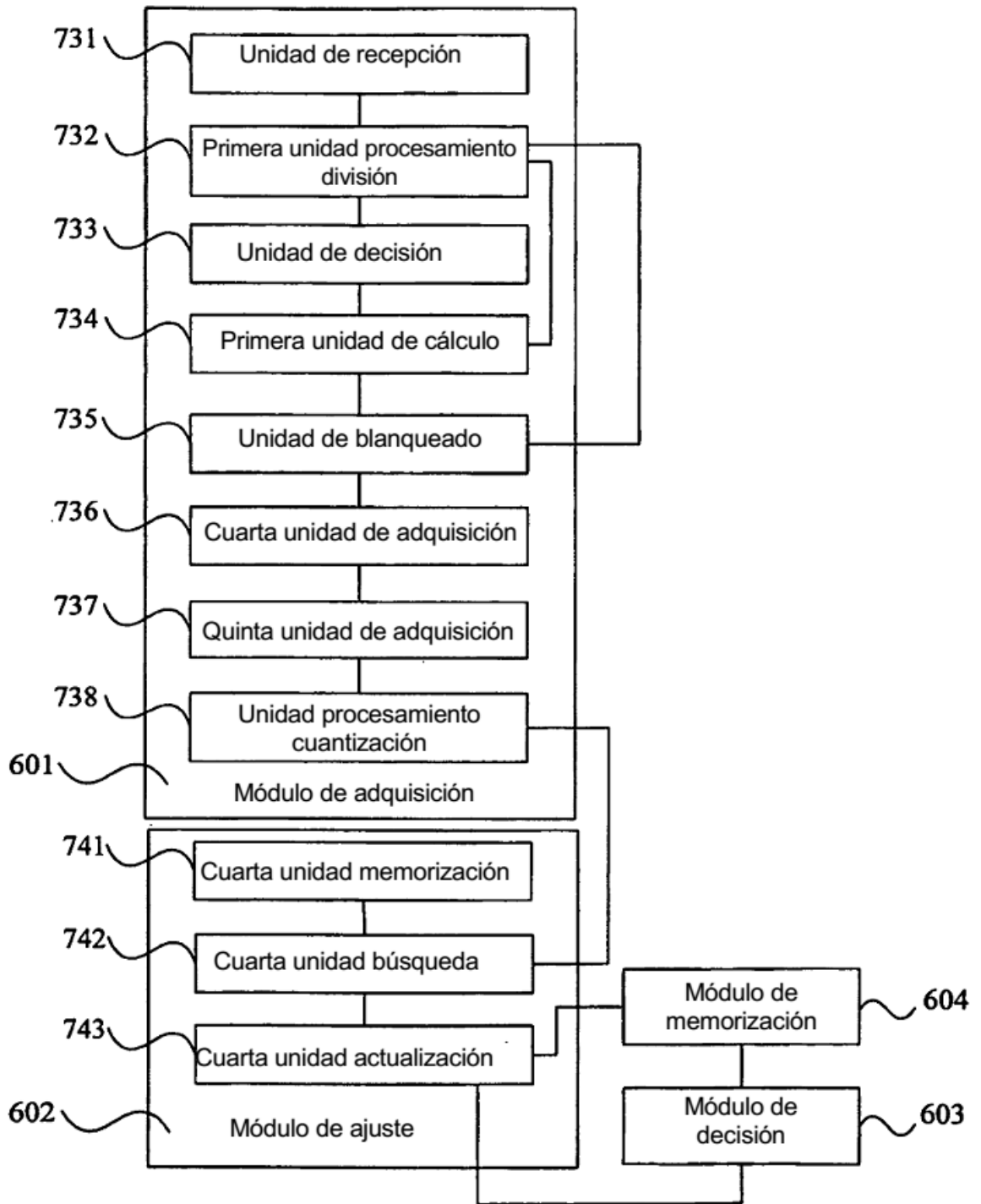


FIG. 11

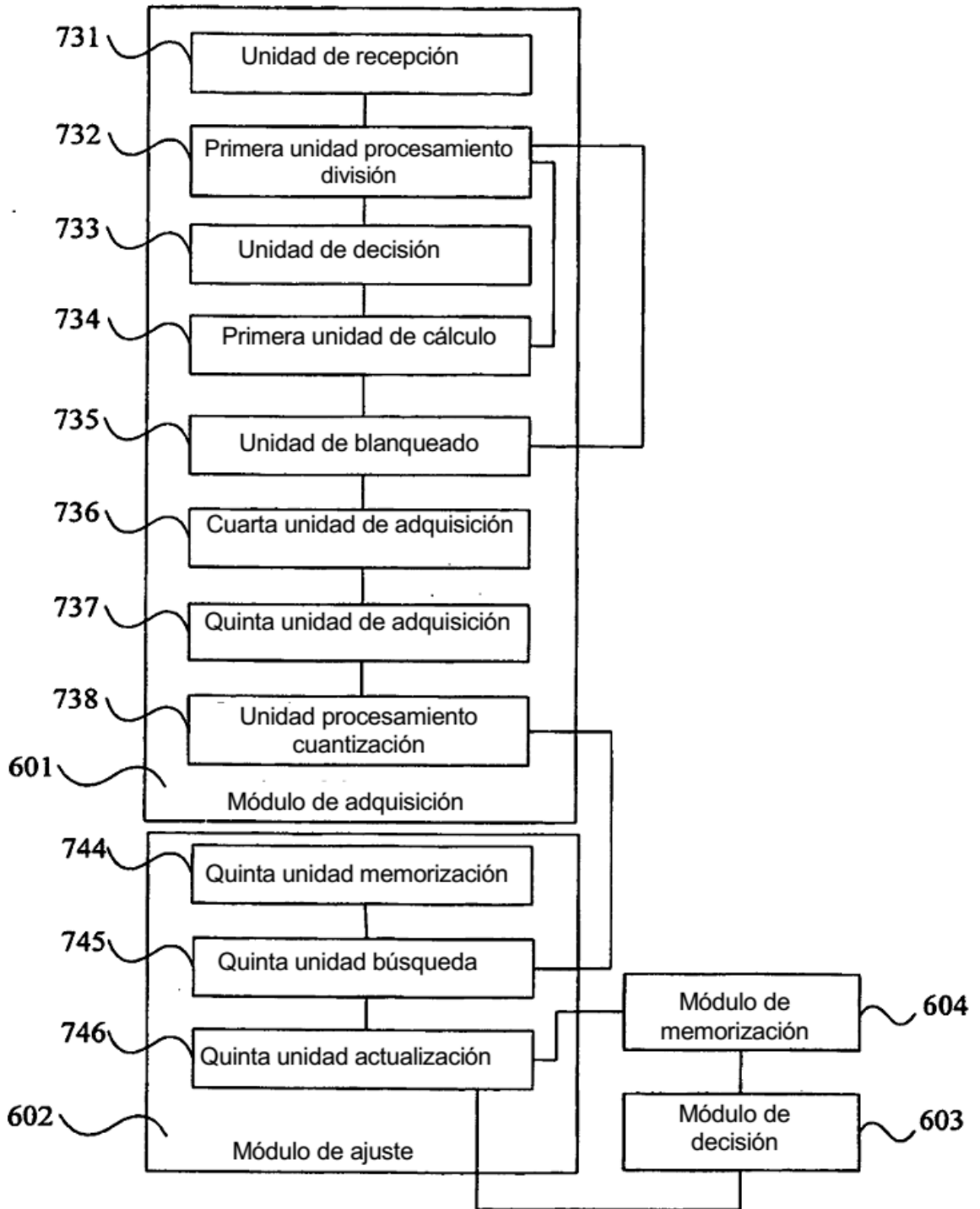


FIG. 12

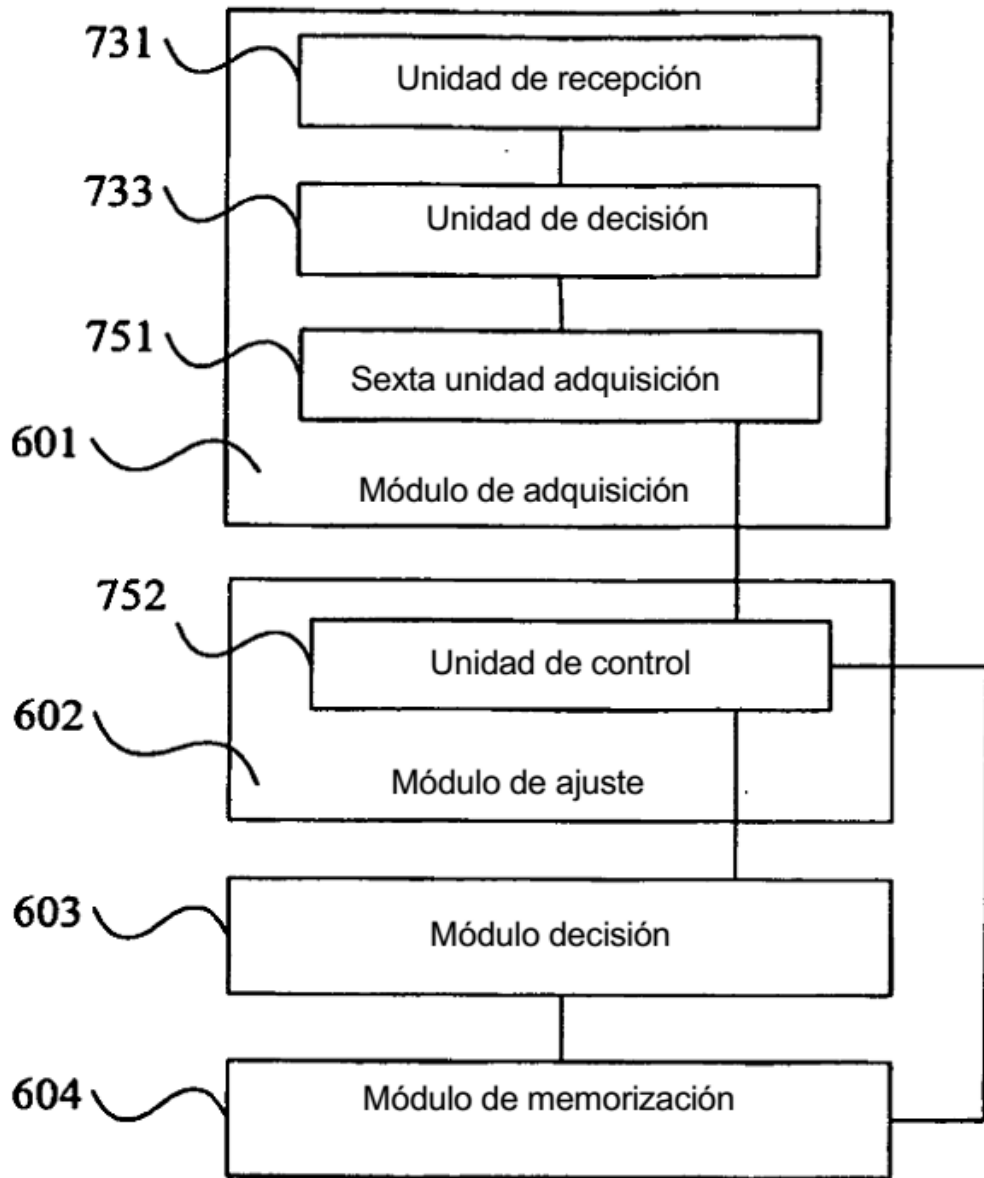


FIG. 13

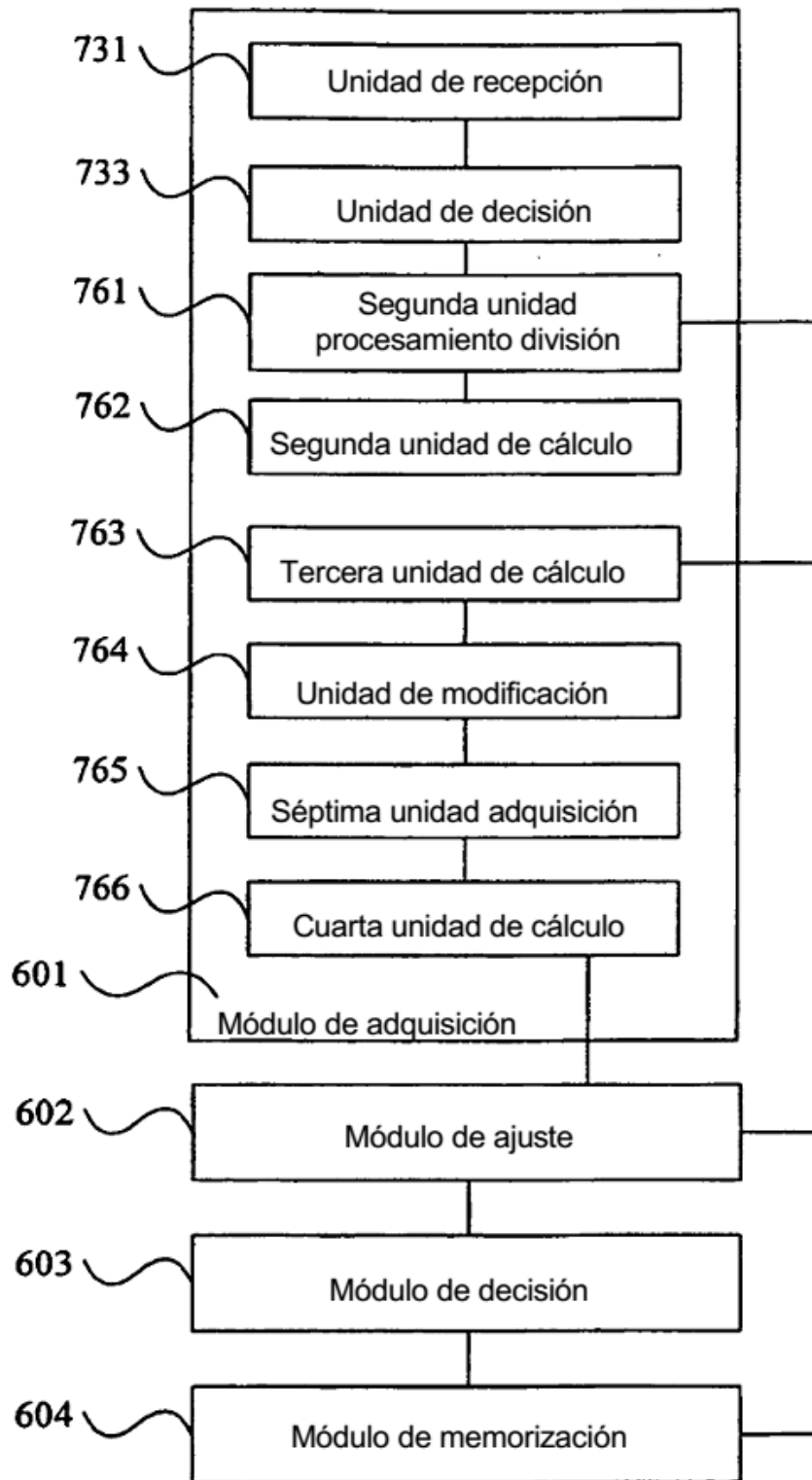


FIG. 14

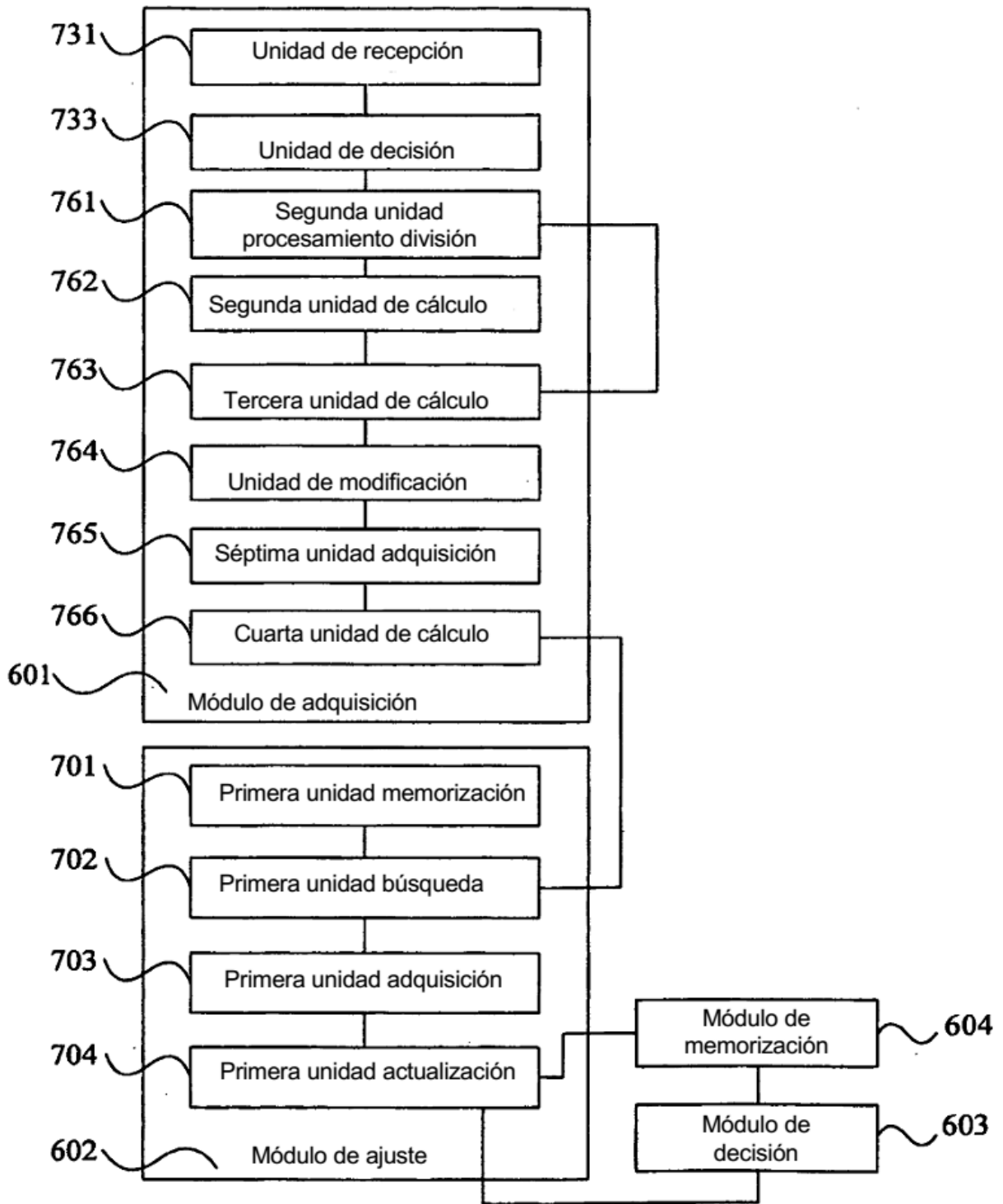


FIG. 15

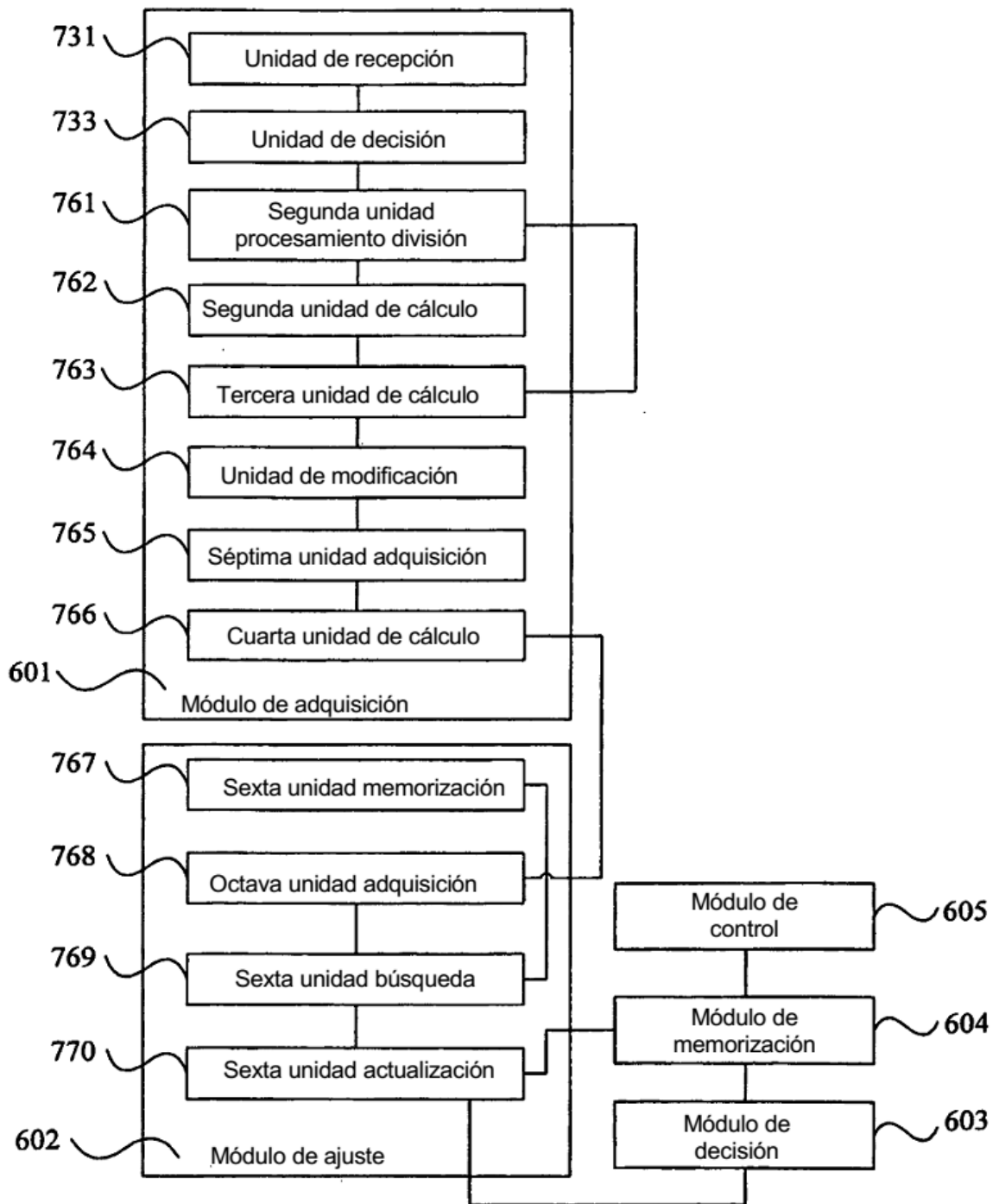


FIG. 16