

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 645 415**

51 Int. Cl.:

G10L 21/038 (2013.01)

G10L 19/26 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.06.2010 PCT/SE2010/050746**

87 Fecha y número de publicación internacional: **26.05.2011 WO11062535**

96 Fecha de presentación y número de la solicitud europea: **29.06.2010 E 10831864 (3)**

97 Fecha y número de publicación de la concesión europea: **09.08.2017 EP 2502229**

54 Título: **Métodos y disposiciones para la compensación de volumen y nitidez en códecs de audio**

30 Prioridad:

19.11.2009 US 262714 P

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

05.12.2017

73 Titular/es:

**TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)
(100.0%)
164 83 Stockholm, SE**

72 Inventor/es:

**GRANCHAROV, VOLODYA y
SVERRISSON, SIGURDUR**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 645 415 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

Métodos y disposiciones para la compensación de volumen y nitidez en códecs de audio

Campo técnico

5 La presente invención se refiere a la codificación / decodificación de audio, en general, y, particularmente, a un esquema de extensión de ancho de banda en el que se realiza o soporta la compensación de volumen y nitidez en la codificación de audio.

Antecedentes

10 El campo de la psicoacústica se refiere al estudio de la percepción del sonido. Esto incluye cómo escuchan los humanos, sus respuestas fisiológicas y el impacto fisiológico de la música y el sonido sobre el sistema nervioso humano. En particular, para el desarrollo de los sistemas de comunicación modernos, el conocimiento de cómo son procesados los estímulos acústicos por el sistema auditivo es importante en el desarrollo de nuevas tecnologías de audio digital y en el mejoramiento de las tecnologías existentes. Los códecs de audio, que son componentes esenciales en los servicios multimedia y de difusión, dependen del conocimiento de las características del sistema auditivo humano para comprimir la información de audio para una transmisión y almacenamiento eficientes a bajas velocidades de bits. Además, se han desarrollado esquemas objetivos para la medición de la calidad, que también dependen en gran medida del conocimiento psicoacústico, para simular valoraciones subjetivas de la calidad del audio.

20 Casi todos los códecs de audio modernos [1-5] aprovechan el concepto de codificación y transmiten solo parte de los componentes de frecuencia de señal de una señal de audio y reconstruyen las frecuencias restantes de la señal de audio en el decodificador. Normalmente, solo se transmiten las bandas de baja frecuencia (LB) de una señal, y las bandas de alta frecuencia (HB) de la señal se reconstruyen posteriormente mediante la denominada extensión del ancho de banda (BWE). En un esquema BWE típico, el contenido de frecuencia de una señal se extiende traduciendo o volteando los componentes de frecuencia disponibles de una banda vecina (normalmente el LB disponible). Sin embargo, una señal reconstruida de tal manera no tiene un HB que coincida exactamente con el HB de la señal de audio original, debido a ciertas aberraciones que se pueden percibir en la señal reconstruida. Para minimizar el impacto de estas aberraciones, en un esquema BWE, la ganancia del HB reconstruido se mantiene típicamente por debajo de la ganancia original del HB, lo que conduce a una señal reconstruida con propiedades psicoacústicas modificadas. Entre las propiedades más afectadas están la sensación de volumen y la sensación de nitidez. El volumen está relacionado con la intensidad de la señal o la presión sonora de la señal de voz. La nitidez está relacionada con la distribución de energía sobre la frecuencia de la señal de voz, y aumenta con el aumento relativo de los componentes de alta frecuencia. Cuando la señal está limitada en banda o se aplica un esquema BWE convencional, tanto el volumen como la nitidez percibidos de la señal reconstruida disminuyen en comparación con la señal original, lo que conduce a una disminución de la calidad subjetiva. Según la solicitud de patente US2007/0067163A1, se conoce un módulo de extensión de ancho de banda que tiene un módulo de preacentuación para invertir un efecto en una banda de frecuencia intermedia (3400-4000Hz) de un filtro anti-efecto dentado (anti-aliasing, en inglés).

Según la solicitud internacional WO86/03873, se aplica un filtro de preacentuación a muestras de voz antes de codificar en el dominio de la frecuencia. El propósito es igualar el espectro reduciendo los efectos de paso bajo de un filtro de inicialización y la atenuación de la alta frecuencia de los labios.

40 Según la solicitud internacional WO2009/055493, se conoce también la aplicación de un módulo de preacentuación antes de un módulo codificador / decodificador. Por lo tanto, existe una necesidad de métodos y disposiciones que permitan mejorar el volumen y la nitidez percibidos de una señal recibida / decodificada.

Compendio

45 La invención está definida por las reivindicaciones independientes. Las reivindicaciones dependientes proporcionan realizaciones de la invención.

La presente invención se refiere a un esquema de extensión de ancho de banda mejorado. Un objeto de la presente invención es proporcionar un método y un sistema para mejorar la calidad percibida de una señal de voz.

Otro objeto es permitir mejoras del volumen y la nitidez percibidos de una señal de voz reconstruida.

Un objeto específico es proporcionar disposiciones de codificador y decodificador para procesar una señal de voz.

50 Las ventajas de la presente invención incluyen mejorar el volumen y la nitidez percibidos, en general, de una señal de voz reconstruida pre-filtrando parte de la señal de voz.

Breve descripción de los dibujos

La invención, junto con otros objetivos y ventajas de la misma, se puede entender mejor haciendo referencia a la siguiente descripción tomada junto con los dibujos adjuntos, en los que:

- 5 la figura 1 es un diagrama de flujo esquemático de una realización de un método según la presente invención;
la figura 2 es un diagrama de flujo esquemático de otra realización de un método según la presente invención;
la figura 3 es un diagrama de bloques esquemático de los trabajos de la realización de la figura 2;
la figura 4 como un diagrama de flujo esquemático de otra realización más de un método según la presente invención;
- 10 la figura 5 es un diagrama de bloques esquemático de los trabajos de la realización de la figura 4;
la figura 6 es un diagrama de bloques esquemático de realizaciones de disposiciones según la presente invención;
la figura 7 es un gráfico que ilustra la respuesta del oído medio exterior;
la figura 8 es un gráfico que ilustra una comparación entre la técnica anterior y el efecto de la presente invención;
- 15 la figura 9 es un diagrama que ilustra una prueba de escucha comparativa entre la técnica anterior y el efecto de la presente invención;
la figura 10 es un diagrama de bloques esquemático de realizaciones adicionales de disposiciones según la presente invención.
la figura 11 es un diagrama de bloques esquemático de una realización de la presente invención.

Descripción detallada

- 20 La presente descripción se refiere a la codificación / decodificación de voz en sistemas de comunicación, tales como los sistemas que utilizan esquemas de extensión del ancho de banda y métodos y disposiciones para mejorar la calidad percibida en tales sistemas, específicamente para mejorar el volumen y la nitidez percibidos. Un ejemplo de un códec particular que se beneficiaría de las realizaciones de la presente invención es el códec AMR-WB (Adaptive Multi-Rate WideBand). Sin embargo, también otros códecs que utilizan extensión del ancho de banda se beneficiarían de la invención o de realizaciones de la misma.
- 25

- Un objetivo de la presente descripción es proporcionar métodos y disposiciones para adaptar una señal de voz para mejorar el volumen y la nitidez percibidos de la señal, p. ej. la señal reconstruida. Se ha reconocido que es posible adaptar o pre-filtrar solo una parte seleccionada de la señal, de tal manera que se mejora la calidad percibida de toda la señal. Tomando en consideración la respuesta natural del oído humano, es posible mejorar una señal de voz para las frecuencias a las que el oído es, típicamente, más sensible. En consecuencia, el oyente es engañado para percibir toda la señal de voz recombinada o reconstruida con un volumen y una nitidez mejorados.
- 30

- Haciendo referencia a la figura 1, una realización de un método para mejorar el volumen y la nitidez percibidos de una señal de voz, se describirá la señal de voz correspondiente a una señal de voz natural delimitada por un ancho de banda predeterminado, de la presente invención. En esta realización, el método según la invención no está limitado a un nodo o dispositivo de red particular.
- 35

- Inicialmente, se proporciona una señal de voz S10. La señal de voz puede ser proporcionada por cualquier medio convencional. Posteriormente, la señal de voz se separa S20 por lo menos en una primera y una segunda porción de señal en base a una primera y una segunda porciones de ancho de banda del ancho de banda predeterminado, respectivamente. Típicamente, esto se realiza dividiendo el ancho de banda de frecuencia predeterminado en una porción de banda de baja frecuencia (LB) y una porción de banda de alta frecuencia (HB). Sin embargo, también es posible realizar otra separación del ancho de banda. Para un ejemplo particular de la presente invención, el ancho de banda predeterminado corresponde a un intervalo de frecuencia de 0 a 8,0 kHz, donde las bandas de baja frecuencia están representadas por frecuencias de 0 a 6,4 kHz, mientras que las bandas de alta frecuencia están representadas por frecuencias de 6,4 a 8,0 kHz. Sin embargo, otros intervalos de frecuencia son igualmente posibles. Posteriormente, la primera porción de señal es adaptada S30 para enfatizar al menos una frecuencia predeterminada o intervalo de frecuencias dentro de la primera porción del ancho de banda. Para un ejemplo particular, esta frecuencia predeterminada está representada por la frecuencia central de la respuesta del oído interno, p. ej. 3,2 kHz, o toda la gama de frecuencias de 3,2 a 6,4 kHz. Finalmente, se reconstruye S40 la segunda porción de señal o una representación de la misma en base a la primera porción de señal y, posteriormente, la primera porción de señal adaptada y la segunda porción de señal reconstruida se combinan S50 para proporcionar una señal de voz reconstruida con un volumen y una nitidez percibidos globales mejorados.
- 40
- 45
- 50

A modo de ejemplo, la adaptación de la primera porción de la señal de voz separada se realiza de tal manera que al menos parte de la energía de la primera porción de señal se distribuye hacia una frecuencia seleccionada dentro de la primera porción del ancho de banda y, simultáneamente otra parte de la energía de la primera porción de señal se distribuye hacia un intervalo o región de alta frecuencia de la primera porción del ancho de banda. De esta manera, el volumen y la nitidez percibidos de la señal posteriormente reconstruida serán mejorados en comparación con una señal de voz reconstruida en base a la banda de baja frecuencia no filtrada o no adaptada de la señal de voz.

Se puede conseguir una BWE mejorada prefiltrando las bandas de baja frecuencia disponibles (LB) de una señal de voz, de tal manera que el volumen y la nitidez globales de la señal reconstruida se compensen por cualquier pérdida debida al esquema BWE. El prefiltrado no se realiza típicamente en las bandas de alta frecuencia reconstruidas (HB), ya que esto aumentará la cantidad de aberraciones de señal introducidos. El término pre-filtrado se utiliza para referirse al hecho de que el filtrado o adaptación descrito se lleva a cabo antes de reconstruir o recombinar la señal. En consecuencia, el filtrado o adaptación se aplica, preferiblemente, solamente a parte de la señal, pero el impacto o mejora se percibe para toda la señal recombinada o reconstruida.

La etapa de adaptación S30 se basa, típicamente, en el pre-filtrado de las bandas de baja frecuencia, y la etapa de reconstrucción S40 puede basarse en BWE o en el filtrado de paso bajo.

En la siguiente descripción, las etapas funcionales se describirán como distribuidas o compartidas entre dos nodos en una red, p. ej. codificador y decodificador en un respectivo nodo transmisor y receptor en el sistema o red de comunicación. En consecuencia, la etapa de adaptación S30 o de filtrado de la primera porción de señal separada o seleccionada puede realizarse después o antes de transmitir la primera porción de señal o representación de la primera porción de señal, cuyos detalles se describirán a continuación.

Haciendo referencia a la figura 2, una realización de un método en el que el filtrado o adaptación de la primera porción de señal, p. ej. de las bandas de baja frecuencia, de la señal de voz se realiza en una disposición de decodificador o receptor en un primer nodo de red. En consecuencia, algunas de las diversas etapas del procedimiento general se ejecutarán en una disposición de codificador o transmisor, y algunas se ejecutarán en una disposición de decodificador o receptor. En esta realización particular, una señal de voz se codifica de una manera conocida. En consecuencia, las etapas de proporcionar S10 una señal de voz y separar S20 la señal de voz en al menos una primera y una segunda porción de señal en base a una primera y una segunda porciones del ancho de banda de un ancho de banda predeterminado de la señal de voz, se realizan preferiblemente en un codificador. La primera porción de señal separada o seleccionada o una representación de la misma se transmite S24 a y se recibe S25, a continuación, en una disposición de receptor o decodificador en un segundo nodo de la red. Posteriormente, el decodificador adapta S30 la primera porción de señal recibida o su representación para enfatizar una frecuencia o un intervalo de frecuencias predeterminados dentro de la primera porción del ancho de banda. Según las medidas conocidas, la segunda porción de señal o bandas de alta frecuencia de la señal de voz se reconstruye S40 en base a la primera porción de señal recibida. Finalmente, la primera porción de señal adaptada y la segunda porción de señal reconstruida se combinan S50 para proporcionar una señal de voz reconstruida con un volumen y una nitidez percibidos globales mejorados.

Haciendo referencia a la figura 3, se muestran las diversas porciones de la señal de voz proporcionada y su procesamiento durante la ejecución del método descrito. En consecuencia, en la figura 3a, una señal de voz para el procesamiento de voz de audio es proporcionada en una forma adecuada por un proveedor de señal 10. La señal es posteriormente separada por el separador de señales 20 en una primera y una segunda porciones de señal en base a sus bandas de baja frecuencia LB y sus bandas de alta frecuencia HB. La primera porción de señal LB es transmitida, a continuación, por un transmisor 24. Posteriormente, la primera porción de señal transmitida LB se recibe en un receptor 25. Basándose en la primera porción de señal LB recibida, la segunda porción de señal HB o su representación se reconstruye mediante el reconstructor 40 (por ejemplo, preferiblemente usando BWE) y la primera porción de señal se adapta o filtra mediante el adaptador 30 para proporcionar una primera porción de señal filtrada o adaptada LB_f. Finalmente, las dos porciones LB_f y HB son recombinadas por el combinador 50 para formar la señal de voz reconstruida o recombinada mejorada.

Haciendo referencia a la figura 4, se describirá una realización de un método en el que el filtrado o la adaptación de la primera porción de señal, p. ej. las bandas de baja frecuencia, de la señal de voz se realiza en un codificador o una disposición de transmisor. En esta realización, también la disposición de decodificador necesita adaptarse para permitir el aprovechamiento de todos los beneficios de la invención, que se describirán a continuación.

En consecuencia, en el nodo o disposición de codificador o transmisor, se realizan las etapas de proporcionar S10 una señal de voz y separar S20 la señal de voz en al menos una primera y una segunda porción de señal en base a una primera y una segunda porciones de ancho de banda de un ancho de banda predeterminado de la señal de voz. Posteriormente, la disposición de codificador adapta S30 la primera porción de señal proporcionada para enfatizar una frecuencia o un intervalo de frecuencias predeterminados dentro de la primera porción del ancho de banda. La primera porción de señal adaptada o una representación de la misma, a continuación, se transmite S34 a y se recibe S35 en un nodo en la red, p. ej. una disposición de receptor o decodificador. Además, el codificador proporciona información opcional sobre qué tipo de códec se utiliza, o cualquier otra información necesaria para que el decodificador pueda reconstruir S40 la segunda porción de señal o bandas de alta frecuencia en base a al menos la

primera porción de señal adaptada recibida (por ejemplo, bandas de baja frecuencia). Típicamente, esta información de asistencia ya está disponible durante la negociación de sesión entre los dos nodos o se conoce de antemano, en donde se acuerdan el códec y otros parámetros de sesión. Sin embargo, en algunos casos es necesario proporcionar información de asistencia adicional para ayudar a la reconstrucción de la segunda porción de señal. Finalmente, el decodificador es capaz de combinar S50 la primera porción de señal adaptada recibida LB_f y la segunda porción de señal HB reconstruida para proporcionar una señal de voz reconstruida con un volumen y una nitidez percibido globales mejorados. Esto se ilustra adicionalmente en la figura 5.

Haciendo referencia a la figura . 5, se muestran las diversas porciones de la señal de voz proporcionada y su procesamiento durante la ejecución del método descrito. En consecuencia, en la figura 5, un proveedor de señales 10 proporciona una señal de voz, que es separada subsiguientemente por el separador de señales 20 en una primera y una segunda porciones de señal en base a sus bandas de baja frecuencia LB y sus bandas de alta frecuencia HB. La primera porción de señal LB se adapta entonces o se filtra mediante el adaptador 30 para proporcionar una primera porción de señal filtrada o adaptada LB_f . Esto es transmitido entonces por un transmisor 34. Posteriormente, la primera porción de señal adaptada transmitida LB_f se recibe en un receptor 35. Junto con esta señal, o ya durante la inicialización de la sesión o la negociación del códec, se proporciona información que permite la reconstrucción de la segunda porción de señal HB. Basándose en la primera porción de señal adaptada recibida LB_f , la segunda porción de señal HB o su representación se reconstruye mediante el reconstructor 40 (por ejemplo, preferiblemente usando BWE o un filtro de paso bajo). Finalmente, las dos porciones LB_f y HB se combinan mediante el combinador 50 para formar la señal de voz mejorada reconstruida o combinada.

Haciendo referencia a la figura 6, realizaciones de un sistema 100 y disposiciones, p. ej. la disposición de codificador 1, la disposición de decodificador 2, el transmisor / receptor, se describirán los primeros / segundos nodos que soportan el método global. Además, la funcionalidad de la adaptación o filtrado de la primera porción de señal puede proporcionarse como una funcionalidad separada, p. ej. filtrado 30, que puede ser implementado en cualquiera de la disposición de codificador 1 o disposición de decodificador 2, o de algún otro nodo en el sistema 100, como se indica mediante el recuadro de puntos 30.

Una realización de un sistema 100, con referencia a la figura . 6, según la presente invención incluye un proveedor de señal 10 para proporcionar una señal de voz delimitada por un ancho de banda predeterminado. Esta señal puede ser proporcionada desde otro nodo en el sistema, o registrada / generada realmente en una disposición de codificador 1 por medio de un micrófono u otro dispositivo de audio o en alguna otra disposición del sistema. Además, el sistema 100 incluye un separador 20 para separar la señal de voz en al menos dos porciones de señal en base a dos porciones de ancho de banda dentro del ancho de banda predeterminado. Típicamente, las dos porciones de señal corresponden a las bandas de baja frecuencia LB y a las bandas de alta frecuencia HB de la señal, pero se podría realizar otra separación. Además, el sistema 100 incluye un adaptador 30 para filtrar o adaptar la primera porción de señal o LB para enfatizar al menos una frecuencia o intervalo de frecuencias predeterminados dentro de la primera porción de ancho de banda. Finalmente, el sistema 100 incluye un reconstructor 40 para reconstruir la segunda porción de señal o HB de la señal y un combinador 50 para combinar la primera porción de señal adaptada y la segunda porción de señal reconstruida para proporcionar una señal de voz reconstruida con calidad percibida mejorada, p. ej. volumen y nitidez. También, con referencia a la figura 6, el sistema 100 comprende dos nodos en el sistema de comunicación, p. ej. un primer nodo con una disposición de codificador 1 y un segundo nodo con una disposición de decodificador 2, cuyas realizaciones se describirán a continuación.

Según una realización de un codificador 1, la disposición de codificador 1 incluye el proveedor de señales de voz 10 para proporcionar una señal de voz y un separador de señales 20 para separar la señal de voz en las porciones de señal primera y segunda. Además, la disposición de codificador 1 incluye un primer adaptador 30 de porción de señal para adaptar la primera porción de señal según métodos descritos anteriormente en esta descripción. Además, el codificador 1 incluye un transmisor de señales 34, adaptado para transmitir al menos una representación de la primera porción de señal adaptada y, opcionalmente, para ayudar a reconstruir la segunda porción de señal en una disposición de decodificador 2 en el sistema 100.

Según una realización de un decodificador 2, la disposición de decodificador 2 está adaptada para cooperar con la disposición de codificador 1 anteriormente descrita. Por consiguiente, el decodificador 2 incluye un receptor de señal 35, para recibir una representación de una primera porción de señal adaptada junto con cualquier información adicional, proporcionándose la primera porción de señal adaptada por el codificador 1 descrito anteriormente. Además, el decodificador 2 incluye un reconstructor 40, para reconstruir una segunda porción de señal de la señal de voz en base a la primera porción de señal adaptada recibida. Finalmente, el decodificador 2 incluye un combinador 50, para combinar la primera porción de señal adaptada recibida y la segunda porción de señal reconstruida para proporcionar una señal reconstruida con un volumen y una nitidez percibidos mejorados.

Según una realización adicional de un codificador 1, la disposición de codificador 1 incluye simplemente un proveedor de señales de voz 10 para proporcionar la señal de voz, un separador de señales 20 para separar la señal de voz en una primera y segunda porciones de señal y, finalmente, una unidad 24 para transmitir la primera porción de señal o al menos una representación de la misma a un segundo nodo en la red de comunicación.

Según una realización adicional de un decodificador 2, la disposición de decodificador 2 incluye un receptor de señal 25 para recibir una primera porción de señal de la disposición de codificador 1 descrita anteriormente. Además, el decodificador 2 incluye un primer adaptador 30 de porción de señal para adaptar o filtrar la primera porción de señal recibida, un reconstructor 40 para reconstruir una segunda porción de señal en base a la primera porción de señal recibida, y un combinador 50 para combinar la primera porción de señal adaptada y la segunda porción de señal reconstruida para proporcionar una señal reconstruida con un volumen y una nitidez percibidos globales mejorados.

A continuación, se indican algunos ejemplos de cómo puede realizarse la adaptación o filtrado de la primera porción de señal con el fin de proporcionar el énfasis deseado de una frecuencia o intervalo de frecuencias predeterminados dentro de la primera porción de ancho de banda. Estos son simples ejemplos, es evidente para el experto que las expresiones matemáticas reales pueden ser modificadas o expresadas de manera diferente manteniendo el mismo impacto global sobre el volumen y la nitidez percibidos.

El énfasis de las frecuencias medias LB (típicamente alrededor de 3,2 kHz para una realización particular) se puede conseguir con el siguiente tipo de filtro:

$$H(z) = \alpha \cdot z^{-2} + \beta \cdot z^{-1} - \gamma + \beta \cdot z^{+1} + \alpha \cdot z^{+2} \quad (1)$$

con coeficientes $\alpha = 0,1$, $\beta = 0$ y $\gamma = 0,85$

Implementación alternativa de filtro, que afecta a la inclinación de la señal LB:

$$H(z) = \alpha \cdot z^{-1} - \beta + \alpha \cdot z^{+1} \quad (2)$$

con coeficientes $\alpha = 0,06$ y $\beta = 0,66$

o

$$H(z) = 1 - \mu \cdot z^{-1} \quad (3)$$

con coeficiente $\mu = 0,2$.

Según las realizaciones de la invención, se activa un módulo de pre-filtrado para pre-filtrar la parte LB de la señal, si el HB de la señal ha sido reconstruido mediante el esquema BWE o se le ha aplicado un filtrado de paso bajo. En este contexto, el término pre-filtrado se refiere al hecho de que el filtrado se realiza antes de reconstruir la señal de voz. De este modo, solo se filtra una parte de la señal, pero el filtrado tiene un efecto sobre la calidad percibida de toda la señal reconstruida. El pre-filtrado de las realizaciones de la presente invención apunta a enfatizar las frecuencias medias o altas de la LB.

Como se mencionó anteriormente, considérese una LB típica, que consiste en componentes de frecuencia de 0 a 6,4 kHz, y un HB reconstruido que consiste en componentes de frecuencia de 6,4 a 8 kHz. En ese escenario, el pre-filtrado hará hincapié en las frecuencias centradas alrededor de 3,2 kHz, o en toda la gama de 3,2 a 6,4 kHz. La frecuencia de énfasis se determina típicamente en relación con la respuesta del oído medio exterior de un sujeto de prueba de audición normal, véase la figura 7. Sin embargo, también se pueden aplicar otros criterios para seleccionar la frecuencia o rango de frecuencias de énfasis. Por ejemplo, la adaptación podría adaptarse en función del perfil auditivo real de un cliente (desactivado o no).

La figura 8 muestra una ilustración del efecto de la invención. En este ejemplo, la línea continua muestra la señal de voz original. La línea de puntos corresponde a una señal reconstruida que ha sido sometida a un esquema BWE convencional y filtrada en paso bajo. Finalmente, la línea discontinua corresponde a una señal reconstruida según la presente invención. Tanto las señales de trazos como las de puntos tienen baja energía en la región por encima de 6 kHz, en comparación con la señal original. A pesar de ello, la señal de trazos será percibida como más fuerte y más nítida que la señal de puntos, debido al énfasis de frecuencia en la región de 3 a 4 kHz. En otras palabras, la nitidez y el volumen que tienen mucha energía en altas frecuencias se pueden reconstruir amplificando el LB de la señal en lugar del HB: Esto evita de manera efectiva, dar lugar a aberraciones de señal.

Para entender cómo afecta el pre-filtrado anterior a las sensaciones o a la percepción de *volumen* y *nitidez* (mejorando así la calidad percibida), es beneficioso examinar sus respectivos modelos psicoacústicos. Vamos a definir el volumen específico en la banda crítica k por $\tilde{N}(k)$, entonces, el volumen y la nitidez se pueden definir como [6]:

$$N = \sum_k \tilde{N}(k) \quad , \quad (4)$$

$$S \propto \frac{\sum_k k \times f(k) \times \tilde{N}(k)}{\sum_k \tilde{N}(k)} \quad (5)$$

La suma es sobre todas las bandas críticas del ancho de banda de la señal, y la función $F(k)$ es igual a uno para las bandas de baja frecuencia y aumenta para las últimas bandas de frecuencia críticas. El volumen específico se define como:

$$\tilde{N}(k) \propto (0.5 + 0.5 \times E(k) \times E^*(k))^{0.23}, \quad (6)$$

donde el factor de normalización E^* Puede estar relacionado con el inverso del umbral en la respuesta de frecuencia del oído medio o externo, véase la figura 7. La excitación E puede calcularse transformando la forma de onda de la señal en el dominio de la frecuencia, seguido por la agrupación de intervalos de frecuencia en bandas de frecuencia críticas.

De la ecuación (4), (6) y de la figura 7 se puede concluir que la sensación de volumen puede aumentarse distribuyendo la energía de señal disponible hacia la región de 3,2 kHz, incluso si se conserva la intensidad total de la señal.

De la ecuación (5) se puede concluir que la sensación de nitidez puede ser aumentada distribuyendo energía de baja a alta frecuencias en el LB - bandas más altas tienen un peso mayor en la suma, debido al aumento de k y de $f(k)$.

Los inventores han realizado extensas pruebas de audición según el bien establecido sistema MUSHRA [7], cuyos resultados se presentan en la figura 9. La columna blanca es la señal de referencia, la columna gris es el resultado de la presente invención, y la columna negra es un resultado de la técnica anterior. Como puede verse en el diagrama, la adaptación de la señal según la presente invención produce una señal que está más próxima a la señal de referencia que los métodos de la técnica anterior, proporcionando así una experiencia de escucha mejorada en comparación con la técnica anterior.

Además, la figura 10 ilustra ejemplos de la funcionalidad de un codificador y un decodificador según la presente invención.

Las etapas, funciones, procedimientos y/o bloques descritos anteriormente pueden implementarse en hardware usando cualquier tecnología convencional, tal como tecnología de circuitos discretos o tecnología de circuitos integrados, incluyendo circuitos electrónicos de uso general y circuitos específicos para una aplicación.

Alternativamente, al menos algunas de las etapas, funciones, procedimientos y/o bloques descritos anteriormente pueden implementarse en un software para su ejecución por un dispositivo de procesamiento adecuado, tal como un microprocesador, un procesador de señal digital (DSP) y/o cualquier dispositivo lógico programable, tal como un dispositivo de matriz de puertas programable por campo (FPGA).

También debe entenderse que sería posible reutilizar las capacidades generales de procesamiento de los nodos de red. Por ejemplo, esto puede realizarse reprogramando el software existente o añadiendo nuevos componentes de software.

El software puede realizarse como un programa informático, que normalmente se transporta en un medio legible por ordenador. El software puede así ser cargado en la memoria de funcionamiento de un ordenador para su ejecución por el procesador del ordenador. El ordenador / procesador no tiene que estar dedicado para ejecutar solamente las etapas, funciones, procedimientos y/o bloques descritos anteriormente, sino que también puede ejecutar otras tareas de software.

A continuación, se describirá un ejemplo de implementación en ordenador con referencia a la figura 11. Un ordenador 200 comprende un procesador 210, una memoria de funcionamiento 220 y una unidad de entrada / salida 230. En este ejemplo particular, al menos algunas de las etapas, funciones, procedimientos y/o bloques descritos anteriormente se implementan en el software 225, que se carga en la memoria de funcionamiento 220 para su ejecución por el procesador 210. El procesador 210 y la memoria 220 están interconectados entre sí a través de un bus de sistema para permitir la ejecución de software normal. La unidad de E/S 230 puede estar interconectada con el procesador 210 y/o con la memoria 220 a través de un bus de E/S, para permitir la entrada y/o salida de datos relevantes, tales como un parámetro o parámetros de entrada y/o un parámetro o parámetros de salida resultantes).

El esquema propuesto para la compensación parcial de volumen y nitidez mejora la calidad perceptual, al tiempo que se preservan los requisitos de la velocidad de bits y las limitaciones de complejidad. El concepto es aplicable a casi cualquier códec de audio moderno o esquema BWE. El filtrado enfatiza las frecuencias medias o altas de la porción LB de la señal para mejorar la sensación de volumen y nitidez para toda la señal reconstruida. En otras palabras, un filtrado parcial de la señal proporciona una mejor calidad percibida para toda la señal.

Referencias

- [1] 3GPP TS 26.190, "Adaptive Multi-Rate-Wideband (AMR-WB) speech codec; Transcoding functions", 2008
- [2] 3GPP TS 26.290 "Extended Adaptive Multi-Rate- Wideband (AMR-WB+) speech codec; Transcoding functions", 2005
- 5 [3] 3GPP TS 26.404 "Enhanced aacPlus encoder SBR part", 2007
- [4] ITU-T Rec. G.729.1, "G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729", 2006
- [5] ITU-T Rec. G.718, "Frame error robust narrowband and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s", 2008
- 10 [6] H. Fastl and E. Zwicker, "Psychoacoustics: Facts and Models," Chapter 8.7.1 and 9.2, Springer, 2007
- [7] G. Stoll and F. Kozamernik, "EBU listening tests on Internet audio codecs", EBU Technical Review, June 2000.

REIVINDICACIONES

1. Un método para mejorar el volumen y la nitidez percibidos de una señal de voz reconstruida delimitada por un ancho de banda predeterminado, **que comprende las etapas de:**

proporcionar (S10) una señal de voz;

5 separar (S20) dicha señal de voz en al menos una primera porción de señal en base a una primera porción de ancho de banda de dicho ancho de banda predeterminado, y una segunda porción de señal en base a una segunda porción de ancho de banda de dicho ancho de banda predeterminado, dicha primera porción de ancho de banda corresponde a bandas de baja frecuencia (LB) de dicha señal de voz proporcionada, y dicha segunda porción de ancho de banda corresponde a bandas de alta frecuencia (HB) de dicha señal de voz proporcionada;

10 adaptar (S30) dicha primera porción de señal para enfatizar al menos una frecuencia o intervalo de frecuencias predeterminados dentro de dicha primera porción de ancho de banda, estando el método **caracterizado por que** dicha etapa de adaptación (S30) comprende la etapa de filtrar dicha primera porción de señal según cualquiera de las siguientes funciones de filtro $H(z)$:

$$H(z) = \alpha \cdot z^{-2} + \beta \cdot z^{-1} - \gamma + \beta \cdot z^{+1} + \alpha \cdot z^{+2}$$

15 con coeficientes $\alpha = 0,1$, $B = 0$, $\gamma = 0,85$, o

$$H(z) = \alpha \cdot z^{-1} - \beta + \alpha \cdot z^{+1}$$

con coeficientes $\alpha = 0,06$ y $\beta = 0,66$, o

$$H(z) = 1 - \mu \cdot z^{-1}$$

20 con coeficiente $\mu = 0,2$, con lo cual al menos parte de la energía de la primera porción de señal se distribuye hacia una frecuencia seleccionada en dicha primera parte de ancho de banda y, simultáneamente, al menos otra parte de la energía de dicha primera porción de señal se distribuye hacia un intervalo de alta frecuencia seleccionado de dicha primera porción de ancho de banda;

reconstruir (S40) dicha segunda porción de señal en base a al menos dicha primera porción de señal o dicha primera porción de señal adaptada;

25 combinar (S50) dicha primera porción de señal adaptada y dicha segunda porción de señal reconstruida para proporcionar una señal de voz reconstruida con un volumen y una nitidez percibidos globales mejorados.

2. El método según la reivindicación 1, **en donde** dicha etapa de adaptación (S30) se basa en la etapa de pre-filtrado de las bandas de baja frecuencia (LB), y dicha etapa de reconstrucción (S40) de dicha segunda porción de señal se basa en la extensión de ancho de banda (BWE) o en el filtrado de paso bajo.

30 3. Un sistema para mejorar el volumen y la nitidez percibidos de una señal de voz reconstruida delimitada por un ancho de banda predeterminado, **que comprende:**

medios (10) configura dos para proporcionar una señal de voz;

medios (20) configura dos para separar dicha señal de voz en al menos una primera porción de señal en base a una primera porción de ancho de banda de dicho ancho de banda predeterminado, y una segunda porción de señal en base a una segunda porción de ancho de banda de dicho ancho de banda predeterminado, dicha primera porción de ancho de banda corresponde a bandas de baja frecuencia (LB) de dicha señal de voz proporcionada, y dicha segunda porción de ancho de banda corresponde a bandas de alta frecuencia (HB) de dicha señal de voz proporcionada; estando el sistema **caracterizado por** que comprende, además:

medios (30) configura dos para adaptar dicha primera porción de señal para enfatizar al menos una frecuencia o intervalo de frecuencias predeterminados dentro de dicha primera porción de ancho de banda, dichos medios (30) están configura dos para filtrar dicha primera porción de señal según cualquiera de las siguientes funciones de filtro $H(z)$:

$$H(z) = \alpha \cdot z^{-2} + \beta \cdot z^{-1} - \gamma + \beta \cdot z^{+1} + \alpha \cdot z^{+2}$$

con coeficientes $\alpha = 0,1$, $\beta = 0$, $\gamma = 0,85$, o

$$H(z) = \alpha \cdot z^{-1} - \beta + \alpha \cdot z^{+1}$$

45

con coeficientes $\alpha = 0,06$ y $\beta = 0,66$, o

$$H(z) = 1 - \mu \cdot z^{-1}$$

con coeficiente $\mu = 0,2$, con lo cual, al menos parte de la energía de la primera porción de señal se distribuye hacia una frecuencia seleccionada en dicha primera porción de ancho de banda y, simultáneamente, al menos otra parte de la energía de dicha primera porción de señal se distribuye hacia un intervalo de alta frecuencia seleccionado de dicha primera porción de ancho de banda;

medios (40) configura dos para reconstruir dicha segunda porción de señal en base a al menos dicha primera porción de señal o dicha primera porción de señal adaptada;

medios (50) configura dos para combinar dicha primera porción de señal adaptada y dicha segunda porción de señal reconstruida para proporcionar una señal de voz reconstruida con un volumen y una nitidez globales percibidos mejorados.

4. El sistema según la reivindicación 3, **en donde** dicho medio (30) está configurado para adaptar dicha primera porción de señal mediante pre-filtrado, donde dicha primera porción de señal corresponde a bandas de baja frecuencia (LB) de dicha señal de voz, y dichos medios (40) están configurados para reconstruir bandas de alta frecuencia (HB) de dicha extensión de ancho de banda en base a la señal de voz (BWE) o el filtrado de paso bajo

5. Una disposición de codificador (1) para procesar una señal de voz delimitada por un ancho de banda predeterminado, **que comprende:**

medios (10) configura dos para proporcionar dicha señal de voz;

medios (20) configura dos para separar dicha señal de voz en al menos una primera porción de señal en base a una primera porción de ancho de banda de dicho ancho de banda predeterminado, y una segunda porción de señal en base a una segunda porción de ancho de banda de dicho ancho de banda predeterminado, dicha primera porción de ancho de banda corresponde a bandas de baja frecuencia (LB) de dicha señal de voz proporcionada, y dicha segunda porción de ancho de banda corresponde a bandas de alta frecuencia (HB) de dicha señal de voz proporcionada; estando la disposición de codificador **caracterizada por que** comprende, además:

medios (30) configura dos para adaptar dicha primera porción de señal para enfatizar al menos una frecuencia o intervalo de frecuencias predeterminados dentro de dicha primera porción de ancho de banda, para mejorar un volumen y una nitidez percibidos de dicha señal de voz, estando configurados dichos medios (30) para filtrar dichas primeras porciones de señal según cualquiera de las siguientes funciones de filtro $H(z)$:

$$H(z) = \alpha \cdot z^{-2} + \beta \cdot z^{-1} - \gamma + \beta \cdot z^{+1} + \alpha \cdot z^{+2}$$

con coeficientes $\alpha = 0,1$, $\beta = 0$, $\gamma = 0,85$, o

$$H(z) = \alpha \cdot z^{-1} - \beta + \alpha \cdot z^{+1}$$

con coeficientes $\alpha = 0,06$ y $\beta = 0,66$, o

$$H(z) = 1 - \mu \cdot z^{-1}$$

con coeficiente $\mu = 0,2$, con lo cual, al menos parte de la energía de la primera porción de señal se distribuye hacia una frecuencia seleccionada en dicha primera porción de ancho de banda y, simultáneamente, al menos otra parte de la energía de dicha primera porción de señal se distribuye hacia un intervalo de alta frecuencia seleccionado de dicha primera porción de ancho de banda;

medios (34) configura dos para transmitir al menos dicha primera porción de señal adaptada a otro nodo de un sistema de comunicación.

6. Disposición de codificador (1) según la reivindicación 5, **en donde** dichos medios (30) están adaptados para pre-filtrar bandas de baja frecuencia (LB) de la señal de voz

7. Disposición de decodificador (1) para procesar una señal de voz delimitada por un ancho de banda predeterminado, **que comprende:**

medios (25) configura dos para recibir una primera porción de señal desde una disposición de codificador, dicha primera porción de señal que se origina desde la separación de una señal de voz proporcionada en al menos una primera porción de señal en base a una primera porción de ancho de banda de dicho ancho de banda predeterminado y una segunda porción de señal en una segunda porción de ancho de banda de dicho ancho de banda predeterminado, dicha primera porción de ancho de banda corresponde a bandas de baja frecuencia (LB) de dicha señal de voz proporcionada, y dicha segunda porción de ancho de banda corresponde a bandas de alta

ES 2 645 415 T3

frecuencia (HB) de dicha señal de voz proporcionada; estando la disposición de decodificador **caracterizada por** que comprende, además:

- 5 medios (30) configura dos para adaptar dicha primera porción de señal recibida para enfatizar al menos una frecuencia o intervalo de frecuencias predeterminados dentro de dicha primera porción de ancho de banda, dichos medios (30) están configura dos para filtrar dicha primera porción de señal según cualquiera de las siguientes funciones de filtro $H(z)$:

$$H(z) = \alpha \cdot z^{-2} + \beta \cdot z^{-1} - \gamma + \beta \cdot z^{+1} + \alpha \cdot z^{+2}$$

con coeficientes $\alpha = 0,1$, $\beta = 0$, $\gamma = 0,85$, o

$$H(z) = \alpha \cdot z^{-1} - \beta + \alpha \cdot z^{+1}$$

- 10 con coeficientes $\alpha = 0,06$ y $\beta = 0,66$, o

$$H(z) = 1 - \mu \cdot z^{-1}$$

- 15 con coeficiente $\mu = 0,2$, con lo cual, al menos parte de la energía de la primera porción de señal se distribuye hacia una frecuencia seleccionada en dicha primera parte de ancho de banda y, simultáneamente, al menos otra parte de la energía de dicha primera porción de señal se distribuye hacia un intervalo de alta frecuencia seleccionado de dicha primera porción de ancho de banda;

medios (40) configura dos para reconstruir dicha segunda porción de señal en base a al menos dicha primera porción de señal;

- 20 medios (50) configura dos para combinar dicha primera porción de señal adaptada y dicha segunda porción de señal reconstruida para proporcionar una señal de voz reconstruida con un volumen y una nitidez globales percibidos mejorados.

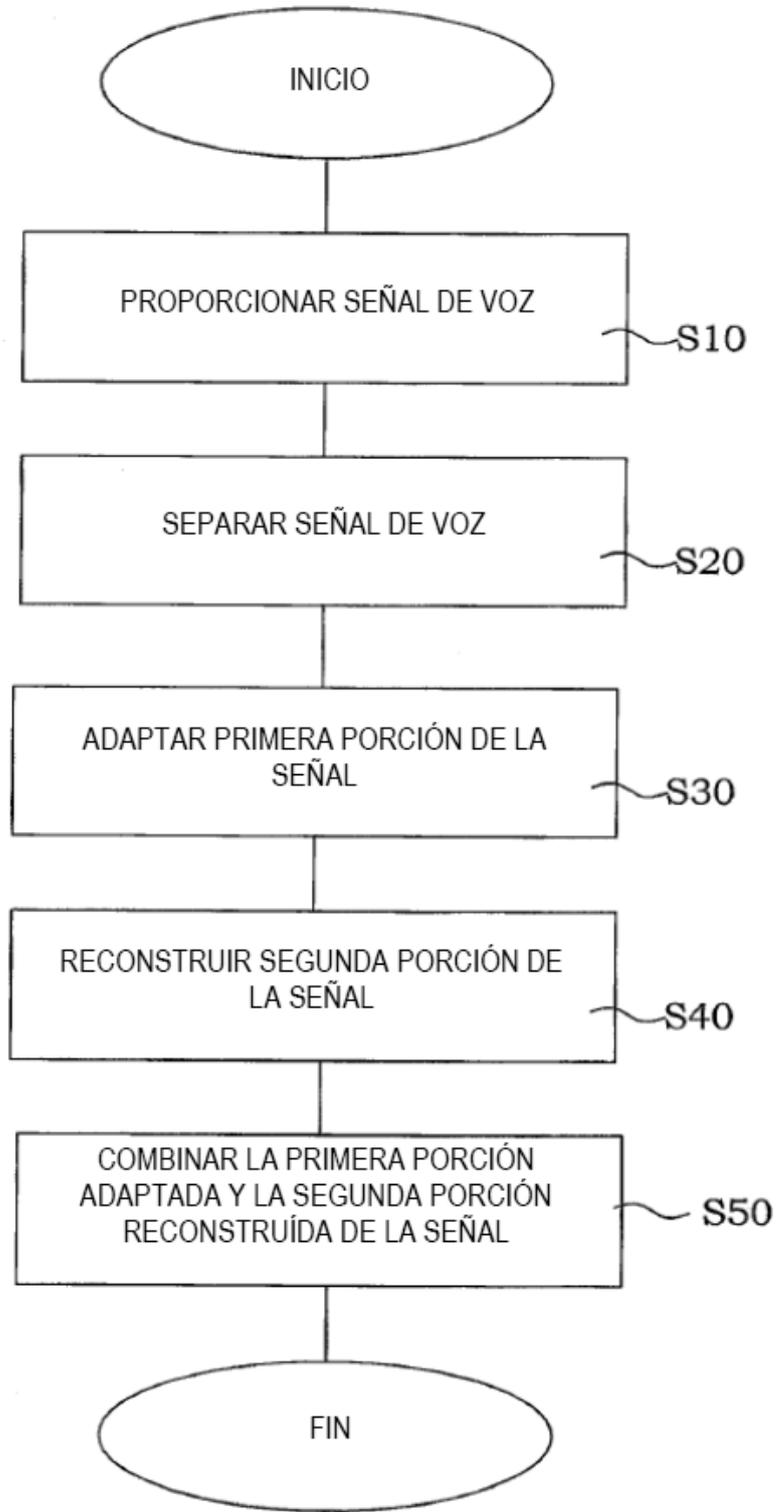


Fig. 1

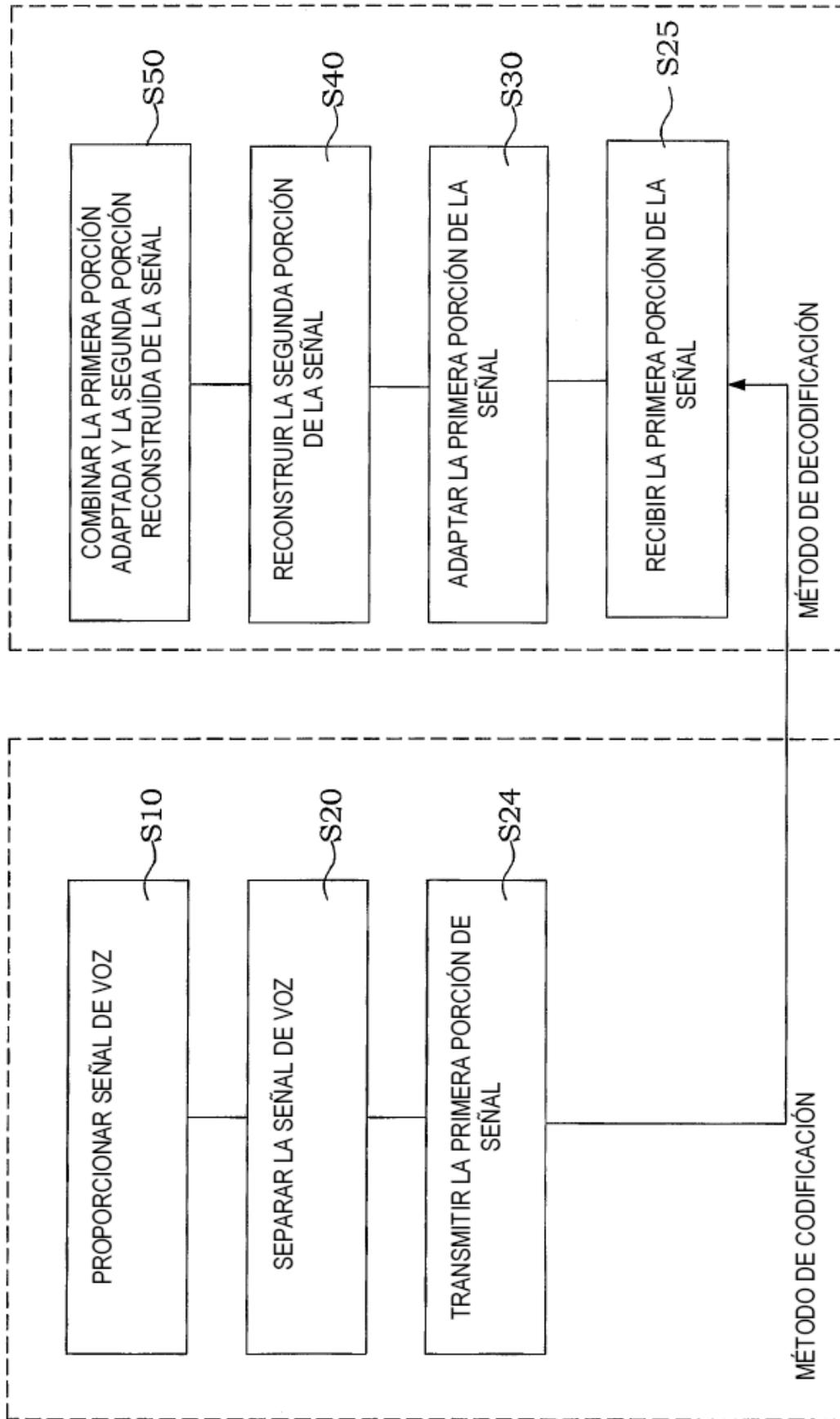


Fig. 2

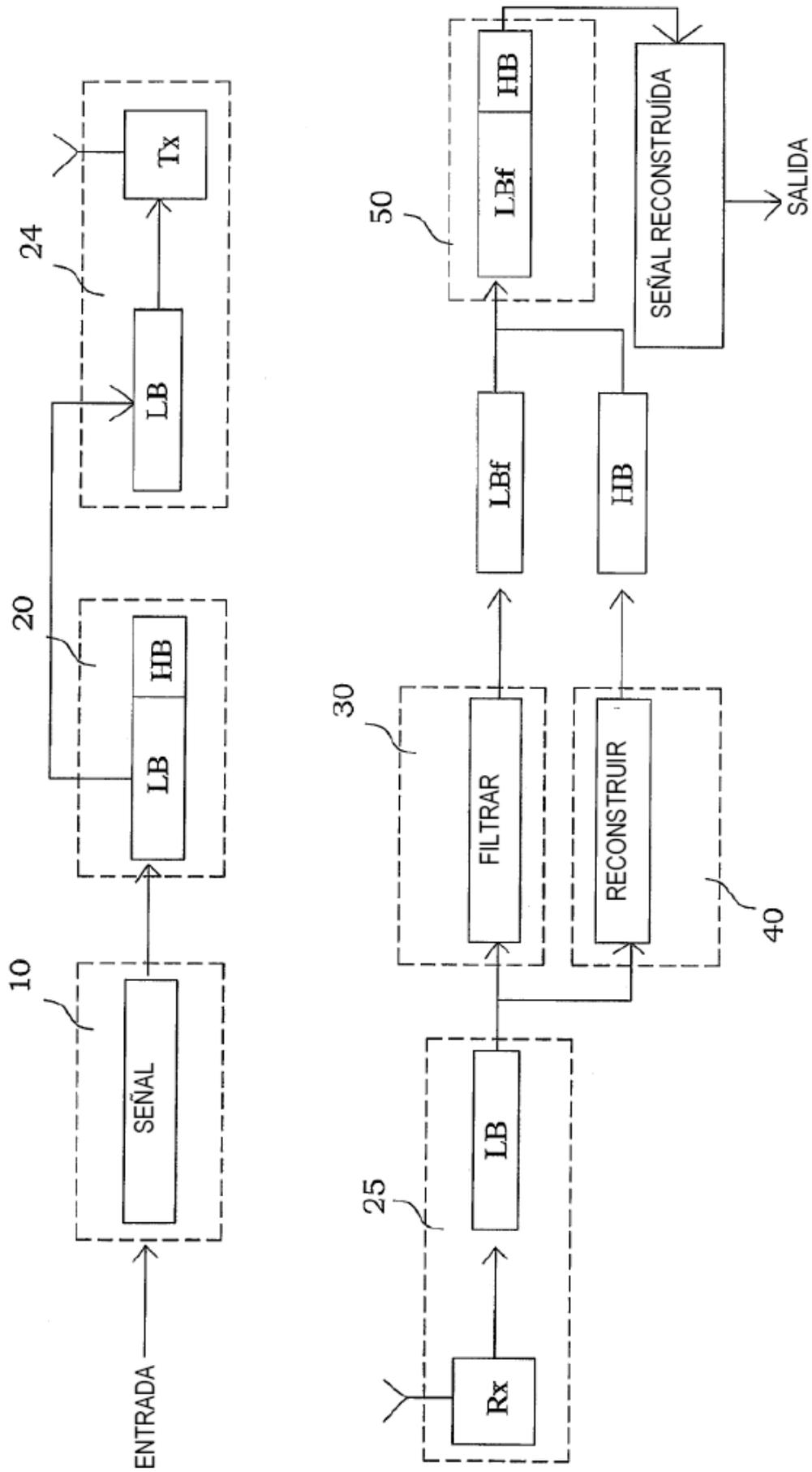


Fig. 3

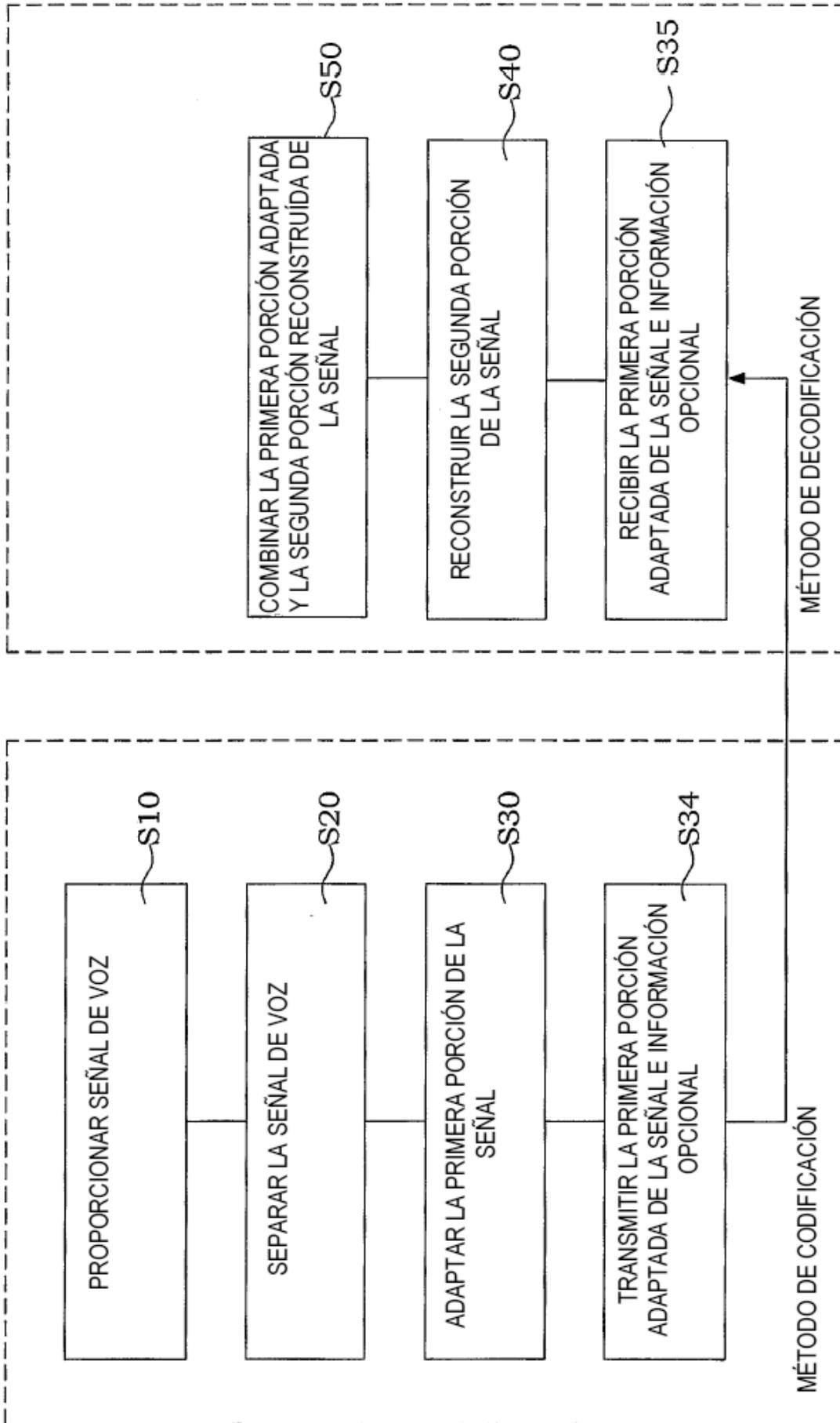


Fig. 4

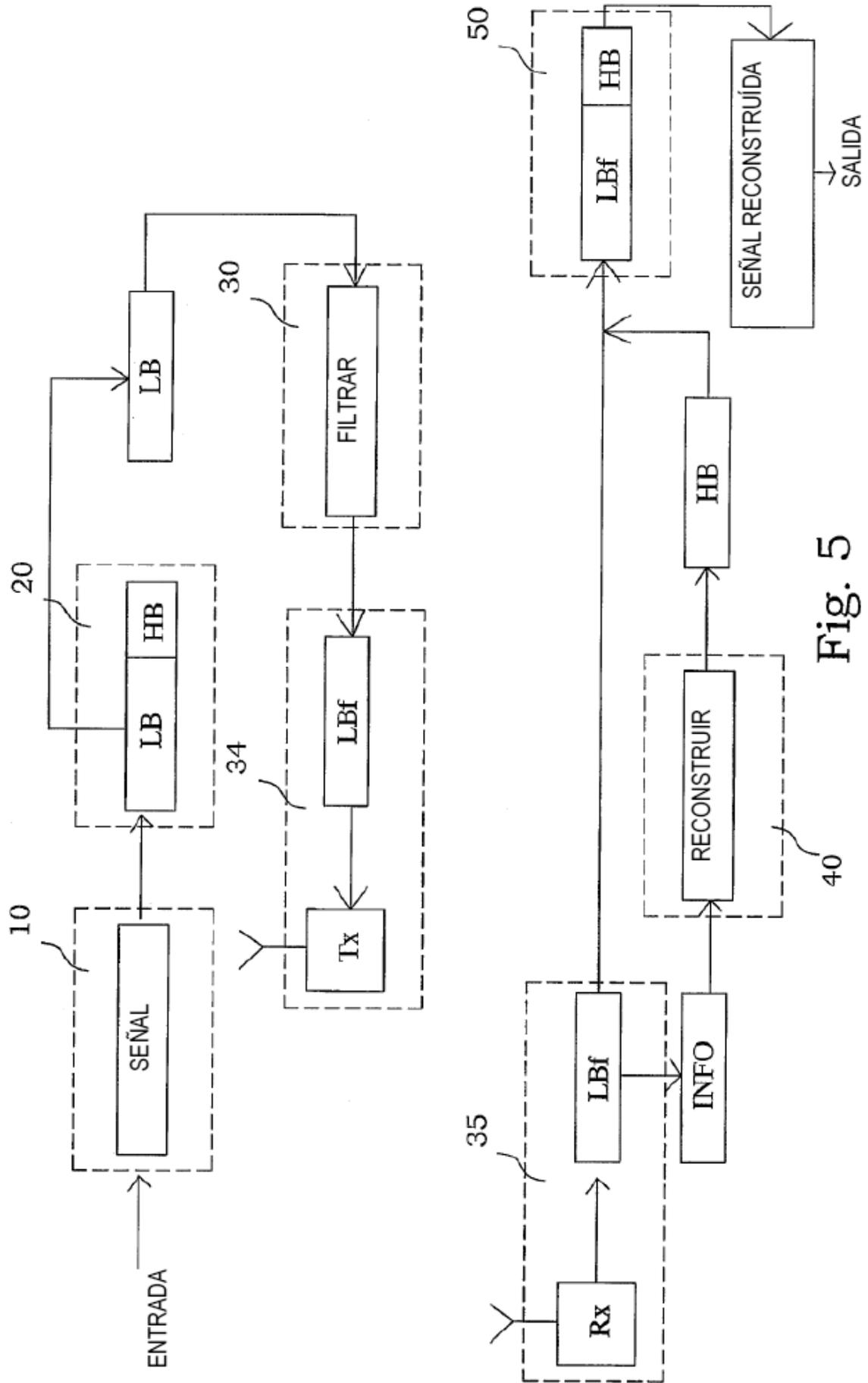


Fig. 5

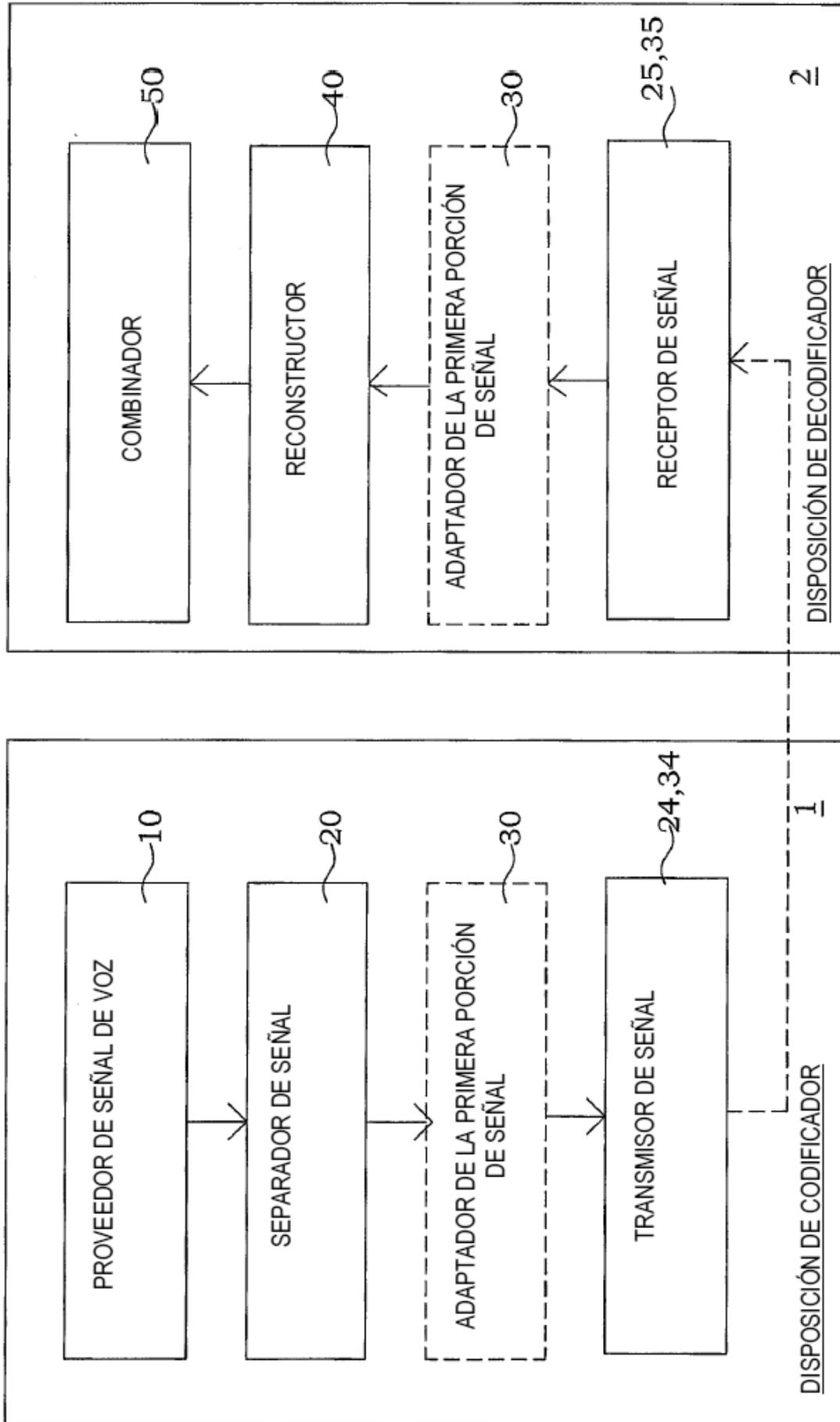


Fig. 6

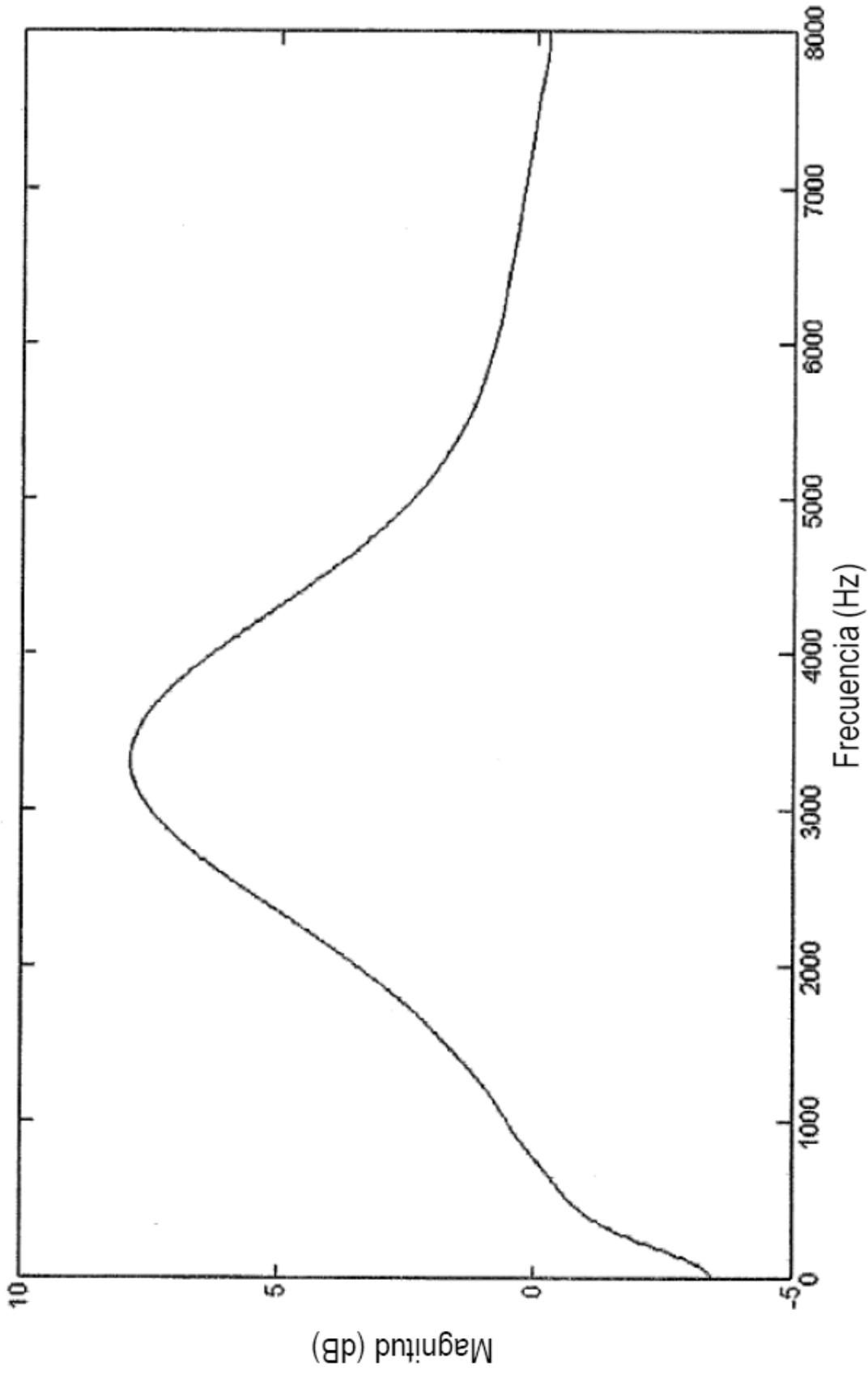


Fig. 7

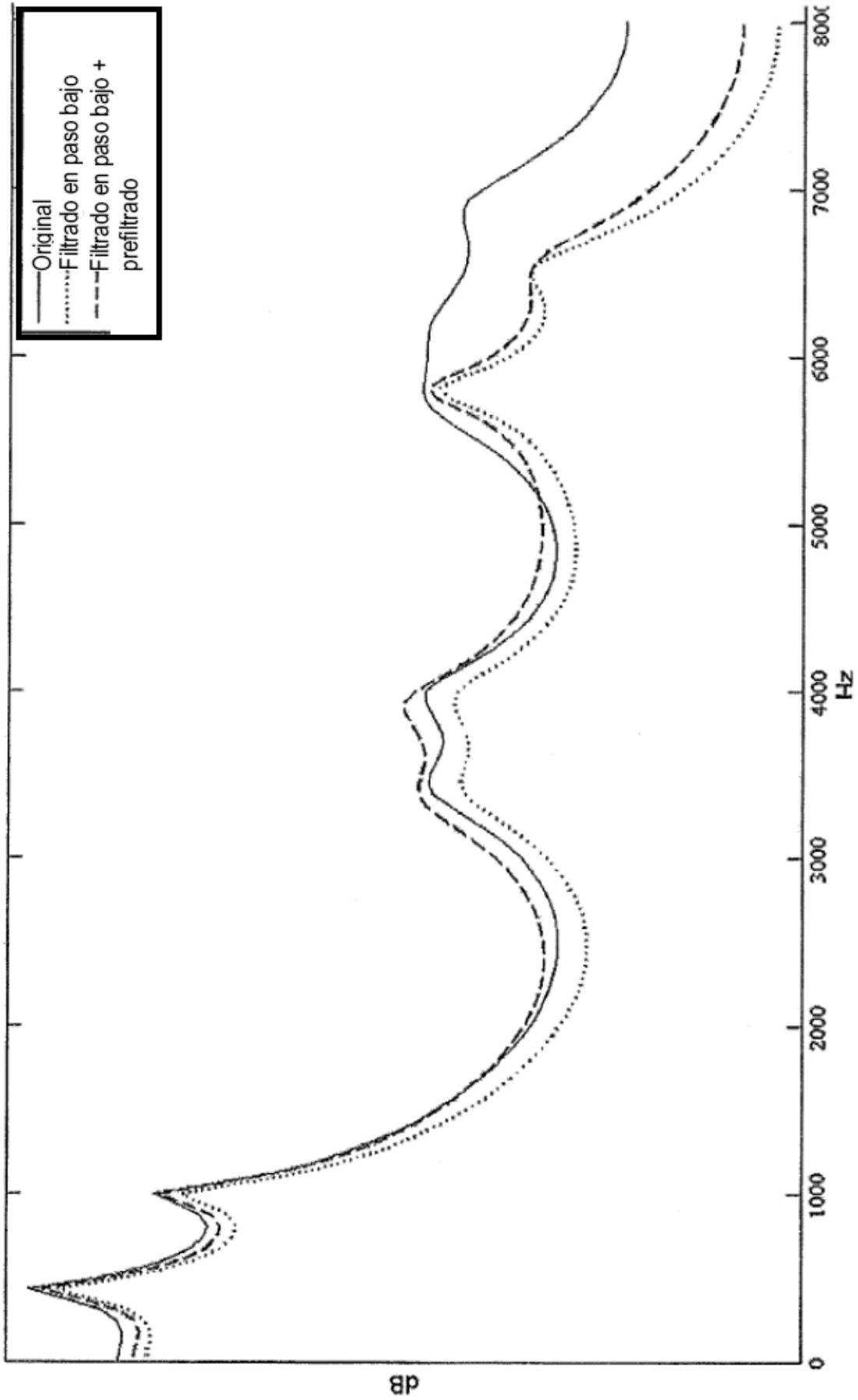


Fig. 8

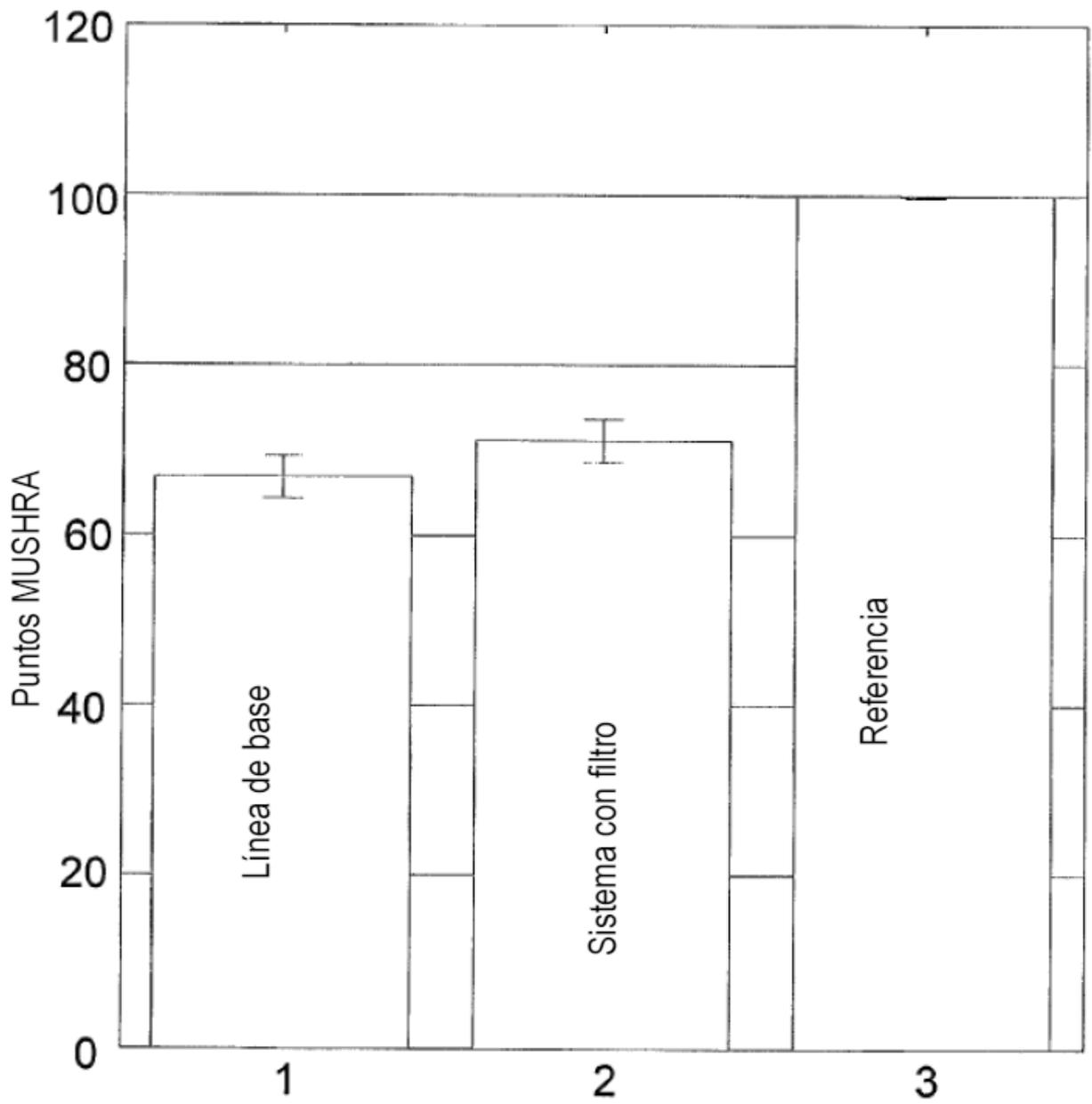


Fig. 9

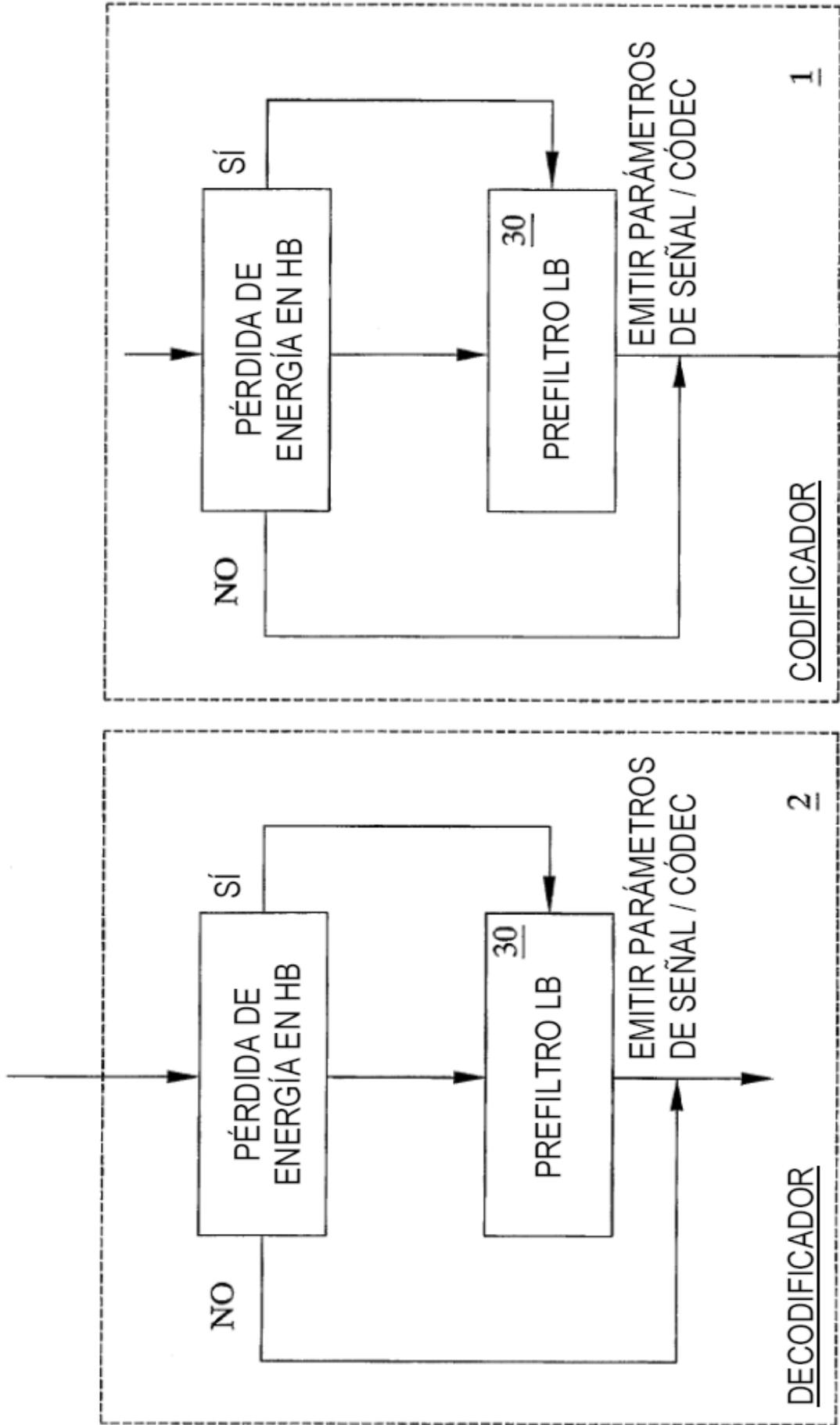


Fig. 10

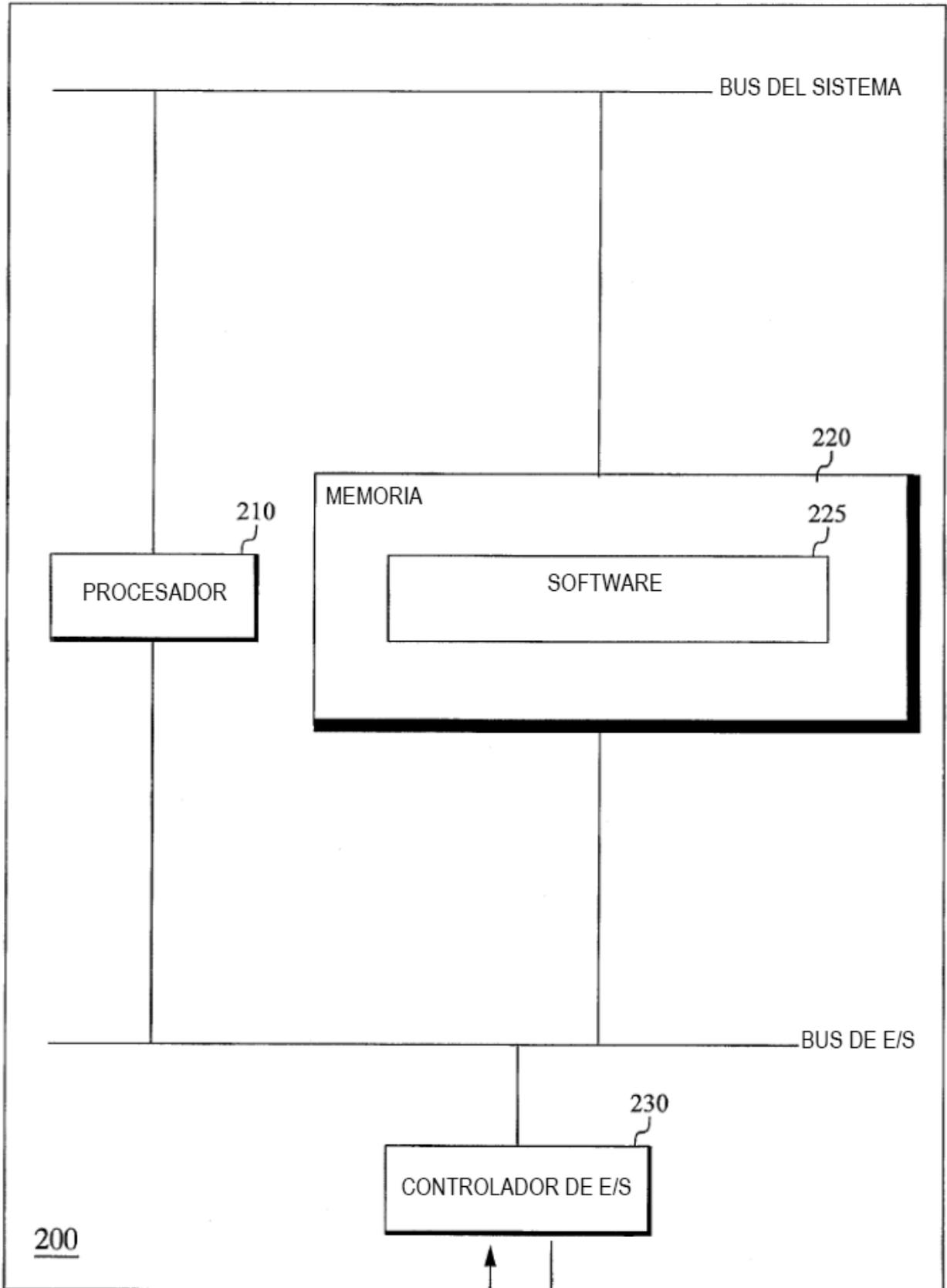


Fig. 11