

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 790 733**

51 Int. Cl.:

G10L 19/025 (2013.01)

G10L 21/038 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **28.01.2014** E 17191504 (4)

97 Fecha y número de publicación de la concesión europea: **01.04.2020** EP 3279894

54 Título: **Codificadores de audio, decodificadores de audio, sistemas, métodos y programas informáticos que utilizan una resolución temporal aumentada en la proximidad temporal de inicios o finales de fricativos o africados**

30 Prioridad:

29.01.2013 US 201361758078 P

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

29.10.2020

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**DISCH, SASCHA;
HELMRICH, CHRISTIAN;
MULTRUS, MARKUS;
SCHNELL, MARKUS y
TRITTHART, ARTHUR**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 790 733 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

Codificadores de audio, decodificadores de audio, sistemas, métodos y programas informáticos que utilizan una resolución temporal aumentada en la proximidad temporal de inicios o finales de fricativos o africados

5 **Campo técnico**

Las realizaciones de acuerdo con la invención están relacionadas con un codificador de audio para proporcionar una información de audio codificada basándose en una información de audio de entrada.

10 Otras realizaciones de acuerdo con la invención están relacionadas con un decodificador de audio para proporcionar una información de audio decodificada basándose en una información de audio codificada.

15 Otras realizaciones de acuerdo con la invención están relacionadas con un sistema que comprende un codificador de audio y un decodificador de audio.

Otras realizaciones de acuerdo con la invención están relacionadas con un método para proporcionar información de audio codificada basándose en una información de audio de entrada.

20 Otras realizaciones de acuerdo con la invención están relacionadas con un método para proporcionar una información de audio decodificada basándose en una información de audio codificada.

Otras realizaciones de acuerdo con la invención están relacionadas con un programa informático para realizar uno de dichos métodos.

25 Otras realizaciones de acuerdo con la invención están relacionadas con un modelado de inicio y final de fricativos o africados en extensión de ancho de banda de audio para la voz.

Antecedentes de la invención

30 En los últimos años ha existido una creciente demanda de almacenamiento digital y transmisión de señales de audio y, en particular, señales de voz. En algunos casos como, por ejemplo, en aplicaciones de comunicaciones móviles, es conveniente obtener una tasa de bits comparativamente baja.

35 Sin embargo, para obtener un buen compromiso entre tasa de bits y calidad de audio (o calidad de voz), hay enfoques para codificar una porción de baja frecuencia de una señal de audio (por ejemplo, una porción de frecuencia de hasta aproximadamente 6 kHz) utilizando una precisión comparativamente alta y para basarse en una extensión de ancho de banda para reconstruir una porción de alta frecuencia del contenido de audio (por ejemplo, superior a aproximadamente 6 o 7 kHz). Por ejemplo, la extensión de ancho de banda se puede basar en una reconstrucción de la porción de alta frecuencia del contenido de audio usando un número comparativamente bajo de parámetros, en el que los parámetros pueden describir, por ejemplo, una envolvente espectral de manera aproximada.

40 Una implementación bien conocida de la extensión de ancho de banda es la replicación de ancho de banda espectral (SBR), que se ha normalizado dentro del MPEG (grupo de expertos en imágenes en movimiento).

Por ejemplo, algunos detalles con respecto a la replicación espectral del ancho de banda se han descrito en las secciones 4.6.18 y 4.6.19 de la Norma Internacional ISO/IEC 14496-3: 200X(E), subparte 4.

50 Además, también se hace referencia al documento US 2011/0099018 A1, que describe un aparato y un método para calcular datos de extensión de ancho de banda utilizando una alineación de tramas espectral con control de distorsión. Dicha solicitud de patente describe un aparato para calcular datos de extensión de ancho de banda de una señal de audio en un sistema de extensión de ancho de banda, en el cual se codifica una primera banda espectral con un primer número de bits y se codifica una segunda banda espectral diferente de la primera banda espectral con un segundo número de bits, siendo el segundo número de bits menor que el primer número de bits. El aparato tiene una calculadora de parámetros controlables de extensión de ancho de banda para calcular parámetros de extensión de ancho de banda para la segunda banda de frecuencia trama a trama para una primera secuencia de tramas de la señal de audio. Cada trama tiene un instante de tiempo de inicio controlable. El aparato incluye además un detector de distorsión espectral para detectar una distorsión espectral en una porción de tiempo de la señal de audio y para señalar un instante de tiempo de inicio para las tramas individuales de la señal de audio dependiendo de la distorsión espectral.

Sin embargo, se ha encontrado que muchos de los enfoques convencionales para la extensión de ancho de banda

degradan sustancialmente una impresión auditiva que se obtiene en presencia de fricativos o africados. Por ejemplo, los pre-ecos y post-ecos pueden provocarse por técnicas convencionales de extensión de ancho de banda. Además, los fricativos o africados pueden sonar demasiado agudos cuando se utilizan las técnicas convencionales de extensión de ancho de banda.

5 En vista de esta situación, existe un deseo de crear un concepto de extensión de ancho de banda que permita una calidad de audio mejorada.

10 **Sumario de la invención**

Las realizaciones, de acuerdo con la invención, crean un codificador de audio de acuerdo con la reivindicación 1, un decodificador de audio de acuerdo con la reivindicación 2, un sistema de acuerdo con la reivindicación 3, métodos de acuerdo con las reivindicaciones 4 y 5 y un programa informático de acuerdo con la reivindicación 6.

15 Asimismo, cabe destacar que cualquiera de las reivindicaciones descritas en el presente documento que no comprenden las enseñanzas tal y como se definen en las reivindicaciones independientes o equivalentes de las mismas deberían considerarse ejemplos adicionales.

20 Una realización de acuerdo con la invención crea un codificador de audio para proporcionar una información de audio codificada basándose en una información de audio de entrada. El codificador de audio comprende un proveedor de información de extensión de ancho de banda configurado para proporcionar información de extensión de ancho de banda empleando una resolución temporal variable. El codificador de audio comprende además un detector configurado para detectar un inicio de un fricativo o africado. El codificador de audio está configurado para
25 ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda esté proporcionada con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta un inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado.

30 Esta realización de acuerdo con la invención se basa en el hallazgo de que se puede obtener una buena calidad auditiva si se proporciona información de extensión de ancho de banda con alta resolución temporal durante el entorno total del momento en el cual se detecta el inicio del fricativo o africado. En consecuencia, se codifica un inicio completo de un fricativo o africado, que típicamente comprende una determinada extensión temporal antes del momento en el cual se detecta el inicio del fricativo o africado y un determinado periodo (extensión temporal)
35 después del momento en el cual se detecta en realidad el inicio del fricativo o africado, con elevada resolución temporal (al menos con respecto a la información de extensión de ancho de banda), lo que contribuye a evitar pre-ecos y también contribuye a evitar una impresión auditiva no natural. Típicamente, no se puede detectar con mucha precisión el inicio del fricativo o africado, puesto que la detección del inicio del fricativo o africado se basa con frecuencia en una detección de cruce de un umbral, que naturalmente no aparece al comienzo mismo del inicio del
40 fricativo o africado. En consecuencia, el momento en el cual se detecta (en realidad) el inicio del fricativo o africado es temporalmente después del comienzo mismo (o inicio) del fricativo o africado. En consecuencia, garantizando que la información de extensión de ancho de banda se proporcione con una resolución temporal aumentada (en comparación con una resolución temporal "normal") al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta (en realidad) el inicio del fricativo o africado, se puede obtener también la
45 reproducción de los detalles al comienzo mismo del inicio del fricativo o africado con buena resolución, en el que se ha encontrado que incluso esos detalles del comienzo mismo del inicio del fricativo o africado son importantes para una buena impresión auditiva. Por consiguiente, proporcionando información de extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio del fricativo o africado no solo se contribuye a evitar los pre-ecos sino también permite
50 reproducir los detalles del inicio del fricativo o africado. De modo similar, garantizando que la información de extensión de ancho de banda esté proporcionada con una resolución temporal aumentada durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado permite reproducir los detalles del inicio del fricativo o africado que son importantes para la impresión auditiva.

55 En consecuencia, el concepto descrito en el presente documento permite reproducir un inicio completo de un fricativo o africado con una elevada resolución temporal, lo que contribuye a evitar la degradación de una impresión auditiva, que se produciría, por ejemplo, por una resolución temporal (de la información de extensión de ancho de banda) demasiado basta al comienzo mismo del inicio del fricativo o africado o en una transición del inicio del fricativo o africado a una parte estacionaria de la señal.

60 En una realización preferida, el codificador de audio está configurado para conmutar de una primera resolución temporal para la provisión de la información de extensión de ancho de banda a una segunda resolución temporal para la provisión de la información de extensión de ancho de banda en respuesta a la detección del inicio del

fricativo o africado, en el que la segunda resolución temporal es más elevada que la primera resolución temporal. En consecuencia, se realiza una conmutación entre dos resoluciones temporales diferentes para la provisión de la información de extensión de ancho de banda, en el que dicha conmutación está controlada por la detección del inicio del fricativo o africado. En consecuencia, se crea un esquema de control sencillo, que puede implementarse fácilmente en un codificador de audio o en un decodificador de audio.

En una realización preferida, el proveedor de información de extensión de ancho de banda está configurado para proporcionar la información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda esté asociada a intervalos de tiempo temporalmente regulares de igual longitud temporal (que pueden formar una cuadrícula de tiempo fundamental - aunque sub-divisible para la provisión de la información de extensión de ancho de banda). El proveedor de información de extensión de ancho de banda está configurado para proporcionar un único conjunto de información de extensión de ancho de banda para un intervalo de tiempo de una duración temporal dada cuando se utiliza una primera resolución temporal (por ejemplo, una resolución temporal comparativamente baja). Asimismo, el proveedor de información de extensión de ancho de banda puede estar configurado para proporcionar una pluralidad de conjuntos de información de extensión de ancho de banda asociados a sub-intervalos de tiempo durante un intervalo de tiempo de la duración temporal dada cuando se utiliza una segunda resolución temporal (por ejemplo, una resolución temporal comparativamente más elevada).

Mediante el uso de intervalos de tiempo temporalmente regulares de igual longitud (por ejemplo, tramas) como una cuadrícula de tiempo (fundamental) para la provisión de la información de extensión de ancho de banda, se puede implementar fácilmente un codificador de audio. Por ejemplo, solo es necesario conmutar el proveedor de información de extensión de ancho de banda entre dos resoluciones temporales discretas, lo que puede implementarse sin excesivo esfuerzo. Por ejemplo, puede ser solamente necesario implementar el proveedor de información de extensión de ancho de banda para proporcionar un único conjunto de información de extensión de ancho de banda basándose en un intervalo de tiempo de la duración temporal dada, y para proporcionar múltiples conjuntos de información de extensión de ancho de banda basándose en un número predeterminado (y fijo) de sub-intervalos del intervalo de tiempo (de igual longitud) de la longitud temporal dada. En consecuencia, puede bastar, por ejemplo, que el proveedor de información de extensión de ancho de banda esté configurado para proporcionar como alternativa un único conjunto de información de extensión de ancho de banda basándose en un intervalo de tiempo de la longitud temporal dado o para proporcionar cuatro conjuntos de información de extensión de ancho de banda basándose en cuatro sub-intervalos de tiempo, teniendo cada uno de los sub-intervalos de tiempo una longitud que es igual a un cuarto de la duración temporal dada. Además, utilizando un concepto de este tipo, se puede mantener pequeño un esfuerzo de señalización, que puede requerirse para la señalización con respecto a qué intervalos de tiempo se produce la información de extensión de ancho de banda, puesto que solo se debe elegir entre una "resolución basta" (por ejemplo, un único conjunto de información de extensión de ancho de banda durante un intervalo de tiempo de la duración temporal dada) y la "resolución precisa" (por ejemplo, n conjuntos de información de extensión de ancho de banda asociados a n sub-intervalos de tiempo de igual longitud). Por consiguiente, se presenta un concepto particularmente eficiente para la provisión de la información de extensión de ancho de banda.

En una realización preferida, el codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que al menos un sub-intervalo de tiempo al cual está asociada un conjunto de información de extensión de ancho de banda, precede inmediatamente a otro sub-intervalo de tiempo al cual está asociado otro conjunto de información de extensión de ancho de banda y durante el cual se detecta otro sub-intervalo de tiempo del inicio de un fricativo o africado, de manera que la resolución temporal se utiliza en al menos un sub-intervalo de tiempo anterior al sub-intervalo de tiempo en el cual se detecta el inicio de un fricativo o africado. En consecuencia, es posible proporcionar la información de extensión de ancho de banda con una elevada resolución temporal incluso al comienzo mismo del inicio del fricativo o africado, es decir, incluso antes de que el inicio del fricativo o africado sea realmente detectable.

En una realización preferida, el codificador de audio está configurado para subdividir un determinado intervalo de tiempo de la duración temporal dada en cuatro sub-intervalos de tiempo de igual longitud, si se utiliza una resolución temporal aumentada para proporcionar información de extensión de ancho de banda correspondiente al intervalo de tiempo dado de la duración temporal dada, de tal manera que se generen cuatro conjuntos de información de extensión de ancho de banda (por ejemplo, cuatro conjuntos de parámetros de extensión de ancho de banda, cada uno de los cuales está asociado uno de los sub-intervalos de tiempo) se proporcionan para el intervalo de tiempo dado de la duración temporal dada. En consecuencia, se puede obtener una elevada resolución temporal de la información de extensión de ancho de banda, ya que los cuatro conjuntos de información de extensión de ancho de banda pueden describir por separado, por ejemplo, envolventes de una porción de alta frecuencia de la señal del contenido de audio para los cuatro sub-intervalos. Por consiguiente, se pueden considerar las diferencias de las envolventes espectrales de la porción de alta frecuencia de la señal de los cuatro sub-intervalos de tiempo, puesto que cada uno de los conjuntos de información de extensión de ancho de banda puede representar la envolvente de frecuencia (o la envolvente espectral) de la porción de alta frecuencia de uno de los sub-intervalos de tiempo.

En una realización preferida, el codificador de audio está configurado para utilizar selectivamente una resolución temporal aumentada para proporcionar información de extensión de ancho de banda para un primer intervalo de tiempo de una longitud temporal dada anterior a un segundo intervalo de tiempo de la duración temporal dada, si se detecta un inicio de un fricativo o africado dentro del segundo intervalo de tiempo y si se detecta una distancia temporal entre un momento en el cual se detecta el inicio del fricativo o africado y un límite entre el primer intervalo de tiempo y el segundo intervalo de tiempo es menor que una distancia temporal predeterminada. En consecuencia, se proporciona la información de extensión de ancho de banda de un primer intervalo de tiempo (por ejemplo, una primera trama) de resolución temporal aumentada (en comparación con una resolución temporal "normal") incluso si el momento en el cual se detecta el inicio del fricativo o africado se encuentra dentro de un segundo intervalo de tiempo subsiguiente (por ejemplo, una segunda trama subsiguiente), si se presume que el comienzo mismo del inicio del fricativo o africado (que típicamente se encuentra antes del momento en el cual se detecta realmente el inicio del fricativo o africado) se encuentra dentro del primer intervalo de tiempo. En consecuencia, el inicio completo del fricativo o africado, incluyendo el comienzo mismo del inicio del fricativo o africado y posiblemente incluso una cierta cantidad de tiempo antes del inicio del fricativo o africado, se evalúa con elevada resolución temporal cuando se proporciona la información de extensión de ancho de banda, lo que trae aparejada una reproducción satisfactoria de la voz. En lugar de simplemente evitar los pre-ecos, se puede reproducir con precisión el inicio del fricativo o africado, sin agudos excesivos ni otras alteraciones sustanciales.

En una realización preferida, el codificador de audio está configurado para realizar una anticipación temporal, por lo que se utiliza una resolución temporal aumentada para proporcionar información de extensión de ancho de banda para un primer intervalo de tiempo de una longitud temporal dada anterior a un segundo intervalo de tiempo de la duración temporal dada en respuesta a una detección del inicio de un fricativo o africado en el segundo intervalo de tiempo. En consecuencia, es posible proporcionar la información de extensión de ancho de banda de resolución temporal aumentada para un inicio completo del fricativo o africado (y posiblemente incluso durante un corto periodo de tiempo antes del inicio del fricativo o africado), lo que contribuye a una calidad de audio mejorada.

En una realización preferida, el codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de manera tal que la información de extensión de ancho de banda se proporciona de una misma resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado. Utilizando una resolución temporal igual, se simplifica la provisión de la información de extensión de ancho de banda en comparación con casos en los cuales se utilizan diferentes resoluciones temporales antes y después del momento en el cual se detecta el inicio del fricativo o africado. Además, se reduce el esfuerzo de señalización utilizando una misma resolución temporal aumentada para el periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado.

En una realización preferida, el codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que se proporcionan conjuntos de información de extensión de ancho de banda con las mismas resoluciones temporales aumentadas al menos para un primer sub-intervalo, un segundo sub-intervalo y un tercer sub-intervalo, en el que el primer sub-intervalo de tiempo precede inmediatamente al segundo sub-intervalo de tiempo, en el que se detecta un inicio de un fricativo o africado en el segundo sub-intervalo de tiempo y en el que el tercer sub-intervalo de tiempo sigue inmediatamente al segundo sub-intervalo de tiempo. En consecuencia, el primer sub-intervalo de tiempo y el tercer sub-intervalo de tiempo, que "embeben" el segundo sub-intervalo de tiempo durante el cual se detecta el inicio del fricativo o africado, se procesan con una misma resolución temporal cuando se proporcionan los conjuntos de información de extensión de ancho de banda. En consecuencia, se trata una parte sustancial de un inicio de un fricativo o africado, o incluso un inicio completo de un fricativo o africado, con una resolución temporal elevada cuando se proporciona la información de extensión de ancho de banda. Además, utilizando la misma resolución temporal (aumentada o "elevada" para el primer sub-intervalo de tiempo, el segundo sub-intervalo de tiempo y el tercer sub-intervalo de tiempo, la codificación y decodificación es sencilla y el gasto de señalización (para señalar una resolución temporal) es pequeño.

En una realización preferida, el detector está configurado para detectar un final de un fricativo o africado. En este caso, el codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda se proporcione con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado. Esta realización de acuerdo con la invención se basa en el hallazgo de que la extensión de ancho de banda también se puede realizar con elevada resolución temporal para un final de un fricativo o africado. Se ha descubierto que la audición humana

también es sensible a los finales de fricativos o africados, de tal manera que justifica el consumo de tasa de bits para codificar el final del fricativo o africado con resolución temporal elevada (con respecto a la información de extensión de ancho de banda). Además, se ha descubierto que el suministro de información de extensión de ancho de banda con baja resolución temporal durante un final de un fricativo o africado da como resultado, típicamente, una
5 impresión auditiva inadecuadamente aguda del final del fricativo o africado, que se percibe como alteración.

Además, se debe tener en cuenta que cualquiera de los conceptos antes mencionados con respecto al ajuste de la resolución temporal utilizado por el proveedor de información de extensión de ancho de banda en respuesta a un inicio de un fricativo o africado se pueden aplicar también ventajosamente en respuesta a una detección de un final
10 de un fricativo o africado. En otras palabras, el concepto antes descrito se puede aplicar de manera análoga, en el que se reemplaza el "inicio de un fricativo o africado" por el "final de un fricativo o africado".

En una realización preferida, el detector está configurado para evaluar una tasa de cruce en cero y/o una relación de energía y/o una distorsión espectral para detectar un inicio de un fricativo o africado. Se ha descubierto que la evaluación de una o más de las cantidades antes mencionadas (tasa de cruce en cero, relación de energía, distorsión espectral) permite una detección razonablemente precisa del inicio de un fricativo o africado. Por ejemplo, se puede comparar uno o más de los valores antes mencionados, o un valor derivado de una combinación de las cantidades antes mencionadas, con un valor umbral para detectar la presencia de un fricativo o africado.
15

En una realización preferida el codificador está configurado para ajustar selectivamente una resolución temporal usada por el proveedor de información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda se proporcione con una resolución temporal aumentada en respuesta a una detección del inicio de un fricativo o africado solo para una porción de la señal de voz, pero no para una porción de señal de música. Este concepto se basa en el hallazgo de que los fricativos o africados son más importantes para la percepción de la voz que para la percepción de porciones de señales de música. En consecuencia, se puede evitar el consumo de la tasa de bits, que puede provocarse por el uso de una resolución temporal aumentada para la provisión de información de extensión de ancho de banda para porciones de señales de música, lo que contribuye a reducir una tasa de bits total, o que ayuda a centrarse en la codificación de características de mayor importancia perceptual para las porciones de señales de música.
20
25
30

En una realización preferida, el codificador de audio está configurado para utilizar selectivamente una resolución temporal aumentada para proporcionar información de extensión de ancho de banda para una pluralidad de intervalos de tiempo subsiguientes que abarcan la totalidad de un inicio de un fricativo o africado detectado. En consecuencia, se codifica el inicio de un fricativo o africado con gran precisión incluso cuando se utiliza una extensión de ancho de banda, de manera que el uso de la extensión de ancho de banda no degrada sustancialmente la impresión auditiva.
35

Otra realización de acuerdo con la invención crea un codificador de audio para proporcionar una información de audio codificada basándose en una información de audio de entrada. El codificador de audio comprende un proveedor de información de extensión de ancho de banda configurado para proporcionar información de extensión de ancho de banda empleando una resolución temporal variable. El codificador de audio comprende asimismo un detector configurado para detectar un final de un fricativo o africado. El codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda se proporcione con una resolución temporal aumentada en respuesta a una detección de un final de un fricativo o africado.
40
45

Esta realización de acuerdo con la invención se basa en el hallazgo de que los finales de fricativos o africados también son importantes para una percepción de un contenido de audio y por lo tanto deben codificarse con elevada resolución temporal. En particular, esta realización de acuerdo con la invención se basa en el hallazgo de que un final de un fricativo o africado se percibe típicamente como "demasiado agudo" si el final del fricativo o africado se codifica con una resolución temporal insuficiente de una información de extensión de ancho de banda. Por consiguiente, aumentando la resolución temporal empleada por un proveedor de información de extensión de ancho de banda, se puede mejorar sustancialmente la calidad del audio, por ejemplo, de las señales de voz.
50

En una realización preferida, el codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que una información de extensión de ancho de banda esté proporcionada con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado. En consecuencia, es posible codificar un final completo de un fricativo o africado con resolución temporal aumentada, incluso aunque un detector típicamente solo pueda detectar el centro de un final de un fricativo o africado o similar.
55
60

Otra realización de acuerdo con la invención crea un decodificador de audio para proporcionar una información de

audio decodificada basándose en una información de audio codificada. El decodificador de audio está configurado para realizar una extensión de ancho de banda basándose en una información de extensión de ancho de banda proporcionada por un codificador de audio, de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado. En consecuencia, el decodificador de audio tiene capacidad para reproducir una porción sustancial de un inicio de un fricativo o africado, o incluso un inicio completo de un fricativo o africado, con elevada resolución temporal. En consecuencia, la extensión de ancho de banda, que se realiza por el decodificador de audio, se puede adaptar bien a la presencia del fricativo o africado, de tal manera que los cambios de la envolvente espectral de la porción de alta frecuencia del contenido de audio, que tienen lugar durante el inicio del fricativo o africado, se pueden reproducir con buena calidad perceptual. En consecuencia, se obtiene una impresión auditiva favorable.

En una realización preferida, el decodificador de audio puede comprender un detector que está configurado para detectar un inicio de un fricativo o africado basándose en una información de audio decodificada, que representa una porción de baja frecuencia de un contenido de audio y decidir por sí mismo sobre el ajuste de la resolución temporal utilizada para la extensión de ancho de banda. Cualquiera de los criterios para detectar un inicio de un fricativo o africado aquí descritos con respecto a un codificador de audio se puede aplicar asimismo al decodificador de audio (siempre que la información requerida esté disponible en el lado del decodificador de audio).

Por otro lado, sin embargo, el decodificador de audio puede estar configurado para ajustar la resolución temporal utilizada para la extensión de ancho de banda basándose en una información complementaria de la información de audio codificada.

Otra realización de acuerdo con la invención genera un decodificador de audio para proporcionar una información de audio decodificada basándose en una información de audio codificada. El decodificador de audio está configurado para ejecutar una extensión de ancho de banda basándose en una información de extensión de ancho de banda proporcionada por un codificador de audio, de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

Esta realización de acuerdo con la invención se basa en la idea de que se puede obtener una buena calidad de audio realizando una extensión de ancho de banda con una resolución temporal aumentada durante un final de un fricativo o africado. Además, la realización se basa en la idea de que el final del fricativo o africado se extiende típicamente durante un determinado periodo de tiempo, en el que momento en el cual se detecta el final del fricativo o africado radica típicamente dentro de dicho cierto periodo de tiempo.

Otra realización de acuerdo con la invención crea un sistema que comprende un codificador de audio, de acuerdo con lo descrito anteriormente, y un decodificador de audio configurado para recibir la información de audio codificada proporcionada por el codificador de audio, y para proporcionar, basándose en ésta, una información de audio decodificada. El decodificador de audio está configurado para realizar una extensión de ancho de banda basándose en la información de extensión de ancho de banda proporcionada por el codificador de audio, de tal manera que la extensión de ancho de banda se realiza con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado, y/o de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

El sistema permite una codificación y decodificación de un contenido de audio, en el que se obtiene una tasa de bits comparativamente baja utilizando una extensión de ancho de banda, y en el que se garantiza una buena reproducción de fricativos o africados mediante el uso de una resolución temporal aumentada en el entorno de un inicio de un fricativo o africado y/o en el entorno de un final de un fricativo o africado.

Otra realización de acuerdo con la invención crea un método para proporcionar una información de audio codificada basándose en una información de audio de entrada. El método comprende producir información de extensión de ancho de banda empleando una resolución temporal variable y detectar un inicio de un fricativo o africado. La resolución temporal utilizada para proporcionar la información de extensión de ancho de banda se ajusta de tal manera que se proporcione información de extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el

inicio del fricativo o africado. Este método se basa en las mismas consideraciones que el codificador de audio antes descrito.

5 Otra realización de acuerdo con la invención crea un método para proporcionar una información de audio codificada basándose en una información de audio de entrada. El método comprende proporcionar información de extensión de ancho de banda empleando una resolución temporal variable y detectar un final de un fricativo o africado. La resolución temporal utilizada para proporcionar la información de extensión de ancho de banda se ajusta de tal manera que la información de extensión de ancho de banda se proporcione con una resolución temporal aumentada en respuesta a la detección de un final de un fricativo o africado. Este método se basa en las mismas condiciones que el codificador de audio antes descrito.

15 Otra realización de acuerdo con la invención crea un método para proporcionar una información de audio decodificada basándose en una información de audio codificada. El método comprende realizar una extensión de ancho de banda basándose en una información de extensión de ancho de banda proporcionada por un codificador de audio, de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado. Este método se basa en las mismas condiciones que el decodificador de audio antes descrito.

20 Otra realización de acuerdo con la invención crea un método para proporcionar una información de audio decodificada basándose en una información de audio codificada. El método comprende realizar una extensión de ancho de banda basándose en una información de extensión de ancho de banda proporcionada por un codificador de audio, de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado. Este método se basa en las mismas condiciones que el decodificador de audio antes descrito.

30 Otra realización de acuerdo con la invención crea un programa informático para realizar uno de los métodos antes descritos.

35 Una realización de acuerdo con la invención crea una señal de audio codificada que comprende una representación codificada de una porción de baja frecuencia de un contenido de audio y una pluralidad de conjuntos de parámetros de extensión de ancho de banda. Los parámetros de extensión de ancho de banda se proporcionan con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual un inicio de un fricativo o africado está presente en el contenido de audio y durante un periodo de tiempo predeterminado posterior al momento en el cual el inicio del fricativo o africado está presente en el contenido de audio.

40 Otra realización que no forma parte de la invención tal y como se ha reivindicado crea una señal de audio codificada que comprende una representación codificada de una porción de baja frecuencia de un contenido de audio y una pluralidad de conjuntos de parámetros de extensión de ancho de banda. Los parámetros de extensión de ancho de banda se proporcionan con una resolución temporal aumentada al menos para una porción del contenido de audio en el cual está presente un final de un fricativo o africado.

Estas señales de audio codificadas se basan en las mismas condiciones que el codificador de audio antes descrito y el decodificador de audio antes descrito.

50 **Breve descripción de las figuras**

Las realizaciones de acuerdo con la presente invención se describirán posteriormente tomando referencia a las figuras adjuntas en las que:

55 Fig. 1 muestra un diagrama esquemático de bloques de un codificador de audio, de acuerdo con una realización de la presente invención;

Fig. 2 muestra un espectrograma de una señal de voz original con alineación de tramas convencional de extensión de ancho de banda (BWE) y límites de fricativos o africados detectados;

60 Fig. 3 muestra un espectrograma de una señal de voz original con una alineación de tramas de extensión de ancho de banda (BWE) de la invención;

- Fig. 4 muestra un espectrograma de voz codificada con alineación de tramas convencional de extensión de ancho de banda (BWE);
- 5 Fig. 5 muestra un espectrograma de voz codificada con una alineación de tramas de extensión de ancho de banda (BWE) de la invención;
- Fig. 6 muestra una representación esquemática de intervalos de tiempo y sub-intervalos de tiempo para los que se proporcionan conjuntos de información de extensión de ancho de banda en una realización de acuerdo con la invención;
- 10 Fig. 7 muestra una representación esquemática de intervalos de tiempo y sub-intervalos de tiempo para los que se proporcionan conjuntos de información de extensión de ancho de banda en una realización de acuerdo con la invención;
- 15 Fig. 8 muestra un diagrama esquemático de bloques de un codificador de audio de acuerdo con otra realización de la presente invención;
- Fig. 9 muestra un diagrama esquemático de bloques de un decodificador de audio de acuerdo con otra realización de la presente invención;
- 20 Fig. 10 muestra un diagrama esquemático de bloques de un decodificador de audio de acuerdo con otra realización de la presente invención;
- Fig. 11 muestra un diagrama esquemático de bloques de un sistema para la codificación de audio y la decodificación de audio, de acuerdo con una realización de la presente invención;
- 25 Fig. 12 muestra un diagrama de flujo de un método para proporcionar una información de audio codificada basándose en una información de audio de entrada de acuerdo con una realización de la presente invención; y
- 30 Fig. 13 muestra un diagrama de flujo de un método para proporcionar una información de audio decodificada basándose en una información de audio de entrada, de acuerdo con una realización de la presente invención.

Descripción detallada de las realizaciones

- 35 1. Codificador de audio de acuerdo con la Fig. 1
- La Fig. 1 muestra un diagrama esquemático de bloques de un codificador de audio de acuerdo con una realización de la invención.
- 40 El codificador de audio 100 está configurado para recibir una información de audio de entrada 110 y proporcionar, basándose en la misma, una información de audio codificada 112.
- El codificador de audio 100 comprende un detector 120 que, por ejemplo, puede recibir la información de audio de entrada 110. El detector 120 está configurado para detectar un inicio de un fricativo o africado, por ejemplo, basándose en la información de audio de entrada 110. El detector 120 puede proporcionar una información de ajuste de la resolución temporal 122.
- 45 El codificador de audio 100 comprende además un proveedor de información de extensión de ancho de banda 130 que está configurado para proporcionar una información de extensión de ancho de banda 132 empleando una resolución temporal variable. Por ejemplo, el proveedor de información de extensión de ancho de banda 130 puede estar configurado para recibir la información de audio de entrada (y posiblemente información de audio pre-procesada adicional). Además, el proveedor de información de extensión de ancho de banda 130 también puede estar configurado para recibir la información de ajuste de la resolución temporal 122 del detector 120.
- 50 El codificador de audio 100 puede comprender además una codificación de baja frecuencia 140 que, por ejemplo, puede codificar una porción de baja frecuencia de un contenido de audio representado por la información de audio de entrada 110, para proporcionar de esta manera una representación codificada 142 de una porción de baja frecuencia del contenido de audio representado por la información de audio de entrada 110. En consecuencia, la información de audio codificada 112 puede comprender la información de extensión de ancho de banda 132 y la representación codificada 142 de la porción de baja frecuencia del contenido de audio. Sin embargo, los detalles con respecto de la codificación de baja frecuencia no son esenciales para la presente invención.
- 55
- 60

A continuación, se describirá en mayor detalle la funcionalidad del codificador de audio 100.

La codificación de baja frecuencia 140 puede codificar una porción de baja frecuencia del contenido de audio representado por la información de audio de entrada 110. Por ejemplo, una porción del contenido de audio que tiene frecuencias inferiores a aproximadamente 6 kHz o inferiores a aproximadamente 7 kHz (o inferiores a cualquier otro límite de frecuencia predeterminado) puede codificarse utilizando la codificación de baja frecuencia 140. La codificación de baja frecuencia 140 puede usar, por ejemplo, cualquiera de las técnicas de codificación de audio bien conocidas, como codificación en el dominio de la transformada o codificación en el dominio de predicción lineal. En otras palabras, la codificación de baja frecuencia 140 puede usar, por ejemplo, un concepto de codificación de audio que se puede basar en la bien conocida "codificación de audio avanzada" (AAC) o que se puede basar en la bien conocida "codificación de predicción lineal". Por ejemplo, la codificación de baja frecuencia 140 puede comprender (o utilizar) una "codificación avanzada de audio" modificada como se describe en la Norma Internacional ISO/IEC 23003-3. Como alternativa, o además, la codificación de baja frecuencia 140 puede comprender (o utilizar) una codificación de predicción lineal descrita, por ejemplo, en la Norma Internacional ISO/IEC 23003-3. Sin embargo, la codificación de baja frecuencia 140 puede comprender asimismo una conmutación entre una "codificación avanzada de audio" (modificada o no modificada) y una codificación de audio en el dominio de la predicción lineal. Sin embargo, se debe tener presente que, en principio, se puede utilizar cualquier concepto conocido para la codificación de una señal de audio en la codificación de baja frecuencia 140, para proporcionar la representación codificada 142 de la porción de baja frecuencia del contenido de audio representado por la información de audio de entrada.

No obstante, el proveedor de información de extensión de ancho de banda 130 puede proporcionar información de extensión de ancho de banda (por ejemplo, en forma de parámetros de extensión de ancho de banda), que permite reconstruir una porción de alta frecuencia del contenido de audio representado por la información de audio de entrada 110, porción de alta frecuencia que no está representada por la representación codificada 142 proporcionada por la codificación de baja frecuencia 140. Por ejemplo, el proveedor de información de extensión de ancho de banda 130 puede estar configurado para presentar algunos o todos los parámetros de replicación de la banda espectral que se describen en la Norma Internacional ISO/IEC 14496-3 (o cualquier otra norma que se refiera a ISO/IEC 14496-3).

Por ejemplo, el proveedor de información de extensión de ancho de banda puede estar configurado para proporcionar algunos o todos los parámetros descritos en una sección "herramienta de SBR " y/o "SBR con bajo retardo" de la Norma Internacional ISO/IEC 14496-3. Por ejemplo, el proveedor de información de extensión de ancho de banda 130 puede estar configurado para proporcionar algunos o todos los parámetros del elemento de sintaxis `sbr_extension_data()`, `sbr_header()`, `sbr_data()`, `sbr_single_channel_element()`, `sbr_channel_pair_element()` o cualquiera de los elementos de flujo de bits a los que se hace referencia en el mismo, como se define, por ejemplo, en la Norma Internacional ISO/IEC 14496-3. En otras palabras, el proveedor de información de extensión de ancho de banda 130 puede proporcionar parámetros de replicación de ancho de banda espectral que, por ejemplo, describen de manera basta una envolvente espectral de una porción de alta frecuencia del contenido de audio representado por la información de audio de entrada 110. Sin embargo, el proveedor de información de extensión de ancho de banda 130 puede comprender además parámetros que describen un ruido en una porción de alta frecuencia del contenido de audio representado por la información de audio de entrada 110, y/o puede comprender parámetros que describen una o más señales sinusoides incluidas en la porción de alta frecuencia del contenido de audio representado por la información de audio de entrada 110. Además, el proveedor de información de extensión de ancho de banda 130 puede proporcionar, por ejemplo, un número de parámetros de configuración, como se describe también en la Norma Internacional ISO/IEC 14496-3 con respecto a la herramienta de replicación de la banda espectral. Por ejemplo, el proveedor de información de extensión de ancho de banda 130 puede proporcionar uno o más parámetros que representan una resolución temporal que se utiliza para la producción de conjuntos de información de extensión de ancho de banda, por ejemplo una resolución temporal utilizando la cual se obtienen conjuntos de parámetros actualizados que representan una envolvente espectral de la porción de alta frecuencia del contenido de audio representado por la información de audio de entrada. Por ejemplo, el proveedor de extensión de ancho de banda 130 puede proporcionar un parámetro de control que indica si se presenta uno o cuatro conjuntos de parámetros de la envolvente espectral por trama de audio. Por ejemplo, los parámetros de control proporcionados por el proveedor de información de extensión de ancho de banda 130 pueden ser similares, o incluso iguales, a los parámetros proporcionados para el caso "FIXFIX" en el elemento de sintaxis `sbr_grid()`, como se describe en la Norma Internacional ISO/IEC 14496-3.

Sin embargo, el proveedor de extensión de ancho de banda 130 puede estar configurado, como alternativa, para proporcionar una información de control que es similar, o incluso igual, a la información de control incluida en el elemento del flujo de bits `sbr_ld_grid()`, que se ha descrito, por ejemplo, en la sección 4.6.19.3.2 de la Norma Internacional ISO/IEC 14496-3.

Por ejemplo, se puede utilizar un valor de 2 bits para codificar tantos conjuntos de parámetros de forma de

envolvente proporcionados por el proveedor de información de extensión de ancho de banda 130 por trama de audio (consúltense el elemento del flujo de bits "bs_num__env" descrito en la sección 4.6.19.3.2 de ISO/IEC 14496-3).

5 Preferentemente, la señalización puede realizarse como se indica para el caso "FIXFIX", que se describe en la sección 4.6.19 "SBR de bajo retardo" de ISO/IEC 14496-3.

10 Para concluir, el proveedor de información de extensión de ancho de banda 130 proporciona información de extensión de ancho de banda 132, en el que la resolución temporal (por ejemplo, el periodo de tiempo entre las actualizaciones de los parámetros que representan una envolvente espectral de una porción de alta frecuencia del contenido de audio representado por la información de audio de entrada 110) se ajusta en dependencia a la información de ajuste de la resolución temporal 122, que se proporciona por el detector 120. Por consiguiente, la resolución temporal usada por el proveedor de información de extensión de ancho de banda 130 (por ejemplo, para proporcionar conjuntos de parámetros actualizados que describen una envolvente espectral de una porción de alta frecuencia de un contenido de audio representado por la información de audio de entrada 110) se adapta a la información de audio de entrada 110.

15 Por ejemplo, el codificador de audio 100 está configurado de tal manera que la resolución temporal usada por el proveedor de información de extensión de ancho de banda 130 se aumente (en comparación con una resolución temporal normal) en respuesta a una detección de un inicio de un fricativo o africado por el detector 120. Sin embargo, la resolución temporal empleada por el proveedor de información de extensión de ancho de banda se aumenta de tal manera que la información de extensión de ancho de banda (por ejemplo, los parámetros de envolvente espectral de la misma) se proporcione con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio de un fricativo o africado. En consecuencia, se codifica un inicio "completo" de un fricativo o africado (o al menos una porción suficientemente grande de un inicio de un fricativo o africado) con una resolución temporal aumentada de la información de extensión de ancho de banda. En consecuencia, se pueden codificar (y decodificar) los inicios de un fricativo o africado con suficiente precisión, por lo que se evitan las alteraciones audibles y también se evita la degradación de la calidad de audio.

20 En consecuencia, la información de audio codificada 112, que comprende la información de extensión de ancho de banda 132 y que típicamente comprende además la representación codificada 142 de la porción de baja frecuencia del contenido de audio representado por la información de audio de entrada 110, permite la decodificación del contenido de audio representado por la información de audio de entrada 110 con buena calidad, y a la vez se puede mantener razonablemente baja la tasa de bits necesaria.

25 Además, se debe tener presente que también se puede implementar cualquiera de las demás características y funcionalidades descritas en el presente documento en el codificador de audio 100. En particular, el codificador de audio 100 puede estar configurado además para ajustar la resolución temporal empleada por el proveedor de información de extensión de ancho de banda de tal manera que se proporcione información de extensión de ancho de banda con una resolución temporal aumentada en respuesta a una detección de un final de un fricativo o africado (en el que el detector 110 también puede estar configurado para detectar un final de un fricativo o africado).

30 A continuación, se describen algunos detalles adicionales con respecto a la funcionalidad del codificador de audio 100 con referencia a las Figs. 2-7.

La Fig. 2 muestra un espectrograma de una señal de voz original con alineación de tramas convencional de extensión de ancho de banda y límites de fricativos o africados detectados.

35 Una abscisa 210 describe el tiempo (en términos de bloques de tiempo) y una ordenada 212 designa subbandas de QMF. En consecuencia, la representación 200 de acuerdo con Fig. 2 representa una distribución de la energía de una señal de audio en diferentes subbandas de QMF en el tiempo.

40 Como puede observarse, las líneas verticales de guiones color magenta designan los límites temporales 220a, 220b, de una alineación de tramas de extensión de ancho de banda convencional. Además, las líneas verticales de guiones negros designan los límites de los fricativos o africados detectados 230a, 230b, 230c, 230d, ... Los límites de los fricativos o africados detectados 230a, 230b, 230c, 230d, ... pueden detectarse utilizando un detector basado en distorsión. Como se puede apreciar, los intervalos de tiempo de igual longitud, que se pueden considerar tramas de extensión de ancho de banda o, en general, tramas, están definidos por los límites 220a, ..., 220u de la alineación de tramas (convencional) de extensión de ancho de banda. En otras palabras, en el concepto convencional de acuerdo con el documento D1, la información de extensión de ancho de banda puede estar asociada con intervalos de tiempo temporalmente regulares (separados por los límites de la alineación de tramas de extensión de ancho de banda convencional) de igual longitud temporal.

Como se puede apreciar, los límites de fricativos o africados detectados pueden estar en algún punto dentro de un intervalo de tiempo definido por dos límites subsiguientes de la alineación de tramas de extensión de ancho de banda convencional.

5 Sin embargo, el esquema de tramas de extensión de ancho de banda convencional expuesto en la Fig. 2 no permite una reproducción particularmente buena de una porción de alta frecuencia de un contenido de audio, como se describirá más adelante.

10 La Fig. 3 muestra un espectrograma de la señal de voz original con la alineación de tramas de extensión de ancho de banda de la invención (en el que la alineación de tramas de extensión de ancho de banda de la invención está indicada por líneas verticales continuas negras). Una abscisa 310 describe un tiempo, en términos de bloques de tiempo, y una ordenada 312 describe una frecuencia en términos de subbandas de QMF. El espectrograma 300 de la Fig. 3 muestra una distribución de las energías (o, en general, las intensidades) de un contenido de audio (o señal de audio) en la frecuencia (o en subbandas de QMF) y en el tiempo. Como se puede apreciar, sigue existiendo una alineación de tramas regular (básica o fundamental), que está indicada por líneas verticales 330a-330u, en el que las tramas entre dos límites de tramas subsiguientes (por ejemplo, entre los límites de tramas 330a y 330b, o entre los límites de tramas 330b y 330c) se pueden considerar intervalos de tiempo de igual longitud. Sin embargo, se debe tener presente que una resolución temporal se aumenta en respuesta a una detección del inicio de un fricativo o africado y también en respuesta a la detección de un final de un fricativo o africado. Por ejemplo, una detección del inicio de un fricativo o africado en un intervalo de tiempo entre los límites de tramas 330b y 330c tiene el efecto de que la trama (o intervalo de tiempo) entre los límites de tramas 330b y 330c se subdivide en cuatro subtramas (o sub-intervalos de tiempo) 340a, 340b, 340c, 340d. Además, se debe tener presente que, en respuesta a la detección de un inicio de un fricativo o africado entre los límites de tramas 330b y 330c, se aumenta la resolución temporal no solo en la trama comprendida entre los límites de tramas 330b y 330c, sino también en dos tramas subsiguientes limitadas por los límites de tramas 330c y 330d y por los límites de tramas 330d y 330e. Por consiguiente, en respuesta a la detección de un inicio de un fricativo o africado en una única trama (o intervalo de tiempo), es decir el intervalo de tiempo delimitado por los límites de tramas 330b y 330c, se aplica una resolución temporal aumentada correspondiente a dos tramas adicionales (es decir las tramas delimitadas por los límites de tramas 330c y 330d y por los límites de tiempo 330d y 330e). En consecuencia, se puede garantizar el uso de una resolución temporal aumentada (en comparación con una resolución temporal normal) para la provisión de información de extensión de ancho de banda (o parámetros de extensión de ancho de banda) durante la duración de un inicio completo de un fricativo o africado (o al menos en toda una gran porción del inicio del fricativo o africado). De esa manera, la extensión de ancho de banda en el lado del decodificador se puede realizar con una resolución temporal aumentada durante el inicio completo del fricativo o africado, puesto que se pueden obtener conjuntos de parámetros de extensión de ancho de banda individuales (por ejemplo, los parámetros que describen una envolvente de una porción de alta frecuencia de un contenido de audio) por cada uno de los sub-intervalos de tiempo (por ejemplo, por cada uno de los sub-intervalos de tiempo 340a-340d). Además, se puede apreciar que, en respuesta a la detección de un final de un fricativo o africado en una trama comprendida entre los límites de tramas 330e y 330f, se aplica una resolución temporal aumentada a las tres tramas subsiguientes, es decir las tramas delimitadas por los límites de tramas 330e y 330f, por los límites de tramas 330f y 343g y por los límites de tramas 330g y 330h. En otras palabras, todas las tramas comprendidas entre los límites de tramas 330e y 330h se subdividen en cuatro subtramas (o sub-intervalos de tiempo), en el que se proporciona un conjunto de parámetros de extensión de ancho de banda individual por cada una de las subtramas (o sub-intervalos de tiempo). De esa manera se pueden proporcionar parámetros de extensión de ancho de banda con una resolución temporal aumentada durante la totalidad de una desactivación del fricativo o africado detectado en el intervalo de tiempo delimitado por los límites de tramas 330e y 330f.

50 Sin embargo, entre los límites de tramas 330h y 330p, se utiliza una resolución temporal "normal" (en lugar de una resolución temporal "aumentada"). Además, se utiliza una resolución temporal aumentada para la provisión de la información de extensión de ancho de banda correspondiente a las tramas comprendidas entre los límites de tramas 330p y 330s, en respuesta a una detección del inicio de un fricativo o africado en una trama (o intervalo de tiempo) delimitada por los límites de tramas 330p y 330q.

55 De modo similar, se utiliza una resolución temporal aumentada para la provisión de información de extensión de ancho de banda correspondiente a las tramas (o intervalos de tiempo) comprendidas entre los límites de tramas 330t y 330w en respuesta a la detección de un final de un fricativo o africado en una trama (o intervalo de tiempo) entre los límites de tramas 330t y 330u.

60 En conclusión, se utiliza una alineación de tramas uniforme (básico) para proporcionar información de extensión de ancho de banda en el codificador de audio 100, en el que la información de extensión de ancho de banda está asociada a tramas temporalmente regulares (intervalos de tiempo) de igual longitud temporal.

Sin embargo, el proveedor de información de extensión de ancho de banda está configurado para proporcionar un único conjunto de información de extensión de ancho de banda correspondiente a una trama (es decir, un intervalo de tiempo de una longitud temporal dada) si se utiliza una primera resolución temporal ("normal"). Por ejemplo, se proporciona un único conjunto de información de extensión de ancho de banda para una trama comprendida entre los límites de tramas 330a y 330b, y se proporciona un único conjunto de información de extensión de ancho de banda por cada una de las ocho tramas comprendidas entre los límites de tiempo 330h y 330p. Sin embargo, el proveedor de información de extensión de ancho de banda también está configurado para proporcionar una pluralidad de conjuntos de información de extensión de ancho de banda asociados a sub-intervalos de tiempo correspondientes a una trama (intervalo de tiempo) de la duración temporal dada si se utiliza una segunda resolución temporal (aumentada). Por ejemplo, se proporcionan cuatro conjuntos de información de extensión de ancho de banda por cada una de las seis tramas comprendidas entre el límite de trama 330b y el límite de trama 330h, por cada una de las tres tramas comprendidas entre los límites de tramas 330p y 330s, y por cada una de las tres tramas comprendidas entre los límites de tramas 330t y 330w. Como se puede apreciar, cada una de las tramas para las que se proporciona la información de extensión de ancho de banda con elevada resolución temporal se subdivide en cuatro subtramas (o sub-intervalos de tiempo) (por ejemplo, los sub-intervalos de tiempo 340a a 340d) de igual longitud, en el que se proporciona un conjunto de parámetros de extensión de ancho de banda por cada uno de los sub-intervalos de tiempo. Además, se debe tener presente que típicamente hay al menos una subtrama de tiempo, para la que se proporciona un conjunto de parámetros de extensión de ancho de banda, inmediatamente antes de una subtrama de tiempo durante la cual se detecta un inicio de un fricativo o africado o antes de una subtrama de tiempo durante la cual se detecta un final de un fricativo o africado. Por ejemplo, si se presume que se detecta un fricativo o africado en una segunda mitad de la trama comprendida entre los límites de tramas 330b y 330c, hay al menos dos subtramas de tiempo (que radican en una primera mitad de la trama comprendida entre los límites de tramas 330b y 330c) que preceden inmediatamente a una subtrama de tiempo durante la cual se detecta el fricativo o africado. En consecuencia, se utiliza una resolución temporal aumentada para la provisión de los parámetros de extensión de ancho de banda incluso antes del momento en el cual se detecta en realidad el inicio del fricativo o africado o antes del momento en el cual se detecta en realidad el final del fricativo o africado. En consecuencia, se puede procesar el inicio "completo" de un fricativo o africado o una desactivación "completa" de un fricativo o africado con elevada resolución temporal (por el hecho de que los parámetros de extensión de ancho de banda se proporcionan con elevada resolución temporal). En consecuencia, es posible una buena reproducción en el lado de un decodificador de audio que recibe la información de audio codificada proporcionada por el codificador de audio 100.

Tomando ahora como referencia las Figs. 4 y 5, se describirán algunas ventajas del codificador de audio 100 con respecto a los codificadores de audio convencionales.

La Fig. 4 muestra un espectrograma de voz codificada con una alineación de tramas de extensión de ancho de banda convencional. Una abscisa 410 describe un tiempo, y una ordenada 412 describe una frecuencia. Además, las elipses amarillas indican alteraciones típicas causadas por la alineación de tramas de extensión de ancho de banda convencional. El espectrograma 400 de la Fig. 4 describe por tanto una energía de una señal de voz en la frecuencia y en el tiempo.

Una primera elipse 430 describe un pre-eco que se provocaría por una alineación de tramas de extensión de ancho de banda convencional. Además, la alineación de tramas de extensión de ancho de banda convencional tiene el efecto de que se percibe el inicio expuesto en la elipse 430 como un inicio muy abrupto.

Además, una segunda elipse 440 señala un post eco, que también se provocaría por una alineación de tramas de extensión de ancho de banda convencional. Además, el final en la región indicada por la elipse 440 se percibiría típicamente como un final muy abrupto, que no sonaría natural.

Una elipse 450 muestra una fuga de vocales de una banda base, que también se provocaría por una alineación de tramas de extensión de ancho de banda convencional.

En consecuencia, se puede apreciar que surge un número de alteraciones de la alineación de tramas de extensión de ancho de banda convencional (por ejemplo, la alineación de tramas de extensión de ancho de banda mostrada en la Fig. 2).

La Fig. 5 muestra un espectrograma de voz codificada con una alineación de tramas de extensión de ancho de banda convencional (para comparar con el espectrograma de la Fig. 4). Una vez más, una abscisa 510 describe un tiempo y una ordenada 512 describe una frecuencia, por lo que el espectrograma 500 representa una energía de la señal de voz codificada (o de una señal de voz decodificada derivada de una señal de voz codificada) como una función de la frecuencia y como una función del tiempo. Como se puede apreciar, las áreas problemáticas resaltadas por las elipses 430, 440, 450, indicadas en la Fig. 4, mejoran sustancialmente. En otras palabras, el uso de una resolución temporal elevada para la provisión de la información de extensión de ancho de banda contribuye a

reducir, o incluso a evitar los pre-ecos, una percepción inadecuadamente abrupta del inicio de un fricativo o africado, post-ecos en el final de un fricativo o africado y una percepción inadecuadamente abrupta del final de un fricativo o africado. Además, el uso de una resolución temporal aumentada de acuerdo con la invención también contribuye a evitar la fuga de vocales de una banda base, como se indica en la elipse 450 en la Fig. 4.

5 A continuación, se explican algunos detalles con respecto a la provisión de la información de extensión de ancho de banda haciendo referencia a las Figs. 6 y 7.

10 La Fig. 6 muestra una representación esquemática de intervalos de tiempo y sub-intervalos de tiempo que se utilizan para la provisión de una información de extensión de ancho de banda.

Un eje de tiempo tiene la designación 610. Como se puede apreciar, el tiempo (representado por el eje de tiempo 610) se divide en los intervalos de tiempo 620a, 620b, 620c, 620d, 620e, 620f que, por ejemplo, pueden comprender igual longitud. Los intervalos de tiempo se pueden considerar como tramas.

15 Además, el tiempo en el cual se detecta un inicio (o final) de un fricativo o africado tiene la designación t_i . El tiempo t_i radica dentro del intervalo de tiempo (o trama) 620e. Se debe tener presente que el momento en el cual se detecta el inicio (o final) del fricativo o africado puede determinarse, por ejemplo, por el detector 120 y que el momento en el cual se detecta el inicio (o final) del fricativo o africado puede estar, típicamente, algo después del comienzo mismo de un inicio del fricativo o africado o después del comienzo mismo del final del fricativo o africado.

20 Como se puede ver en la Fig. 6, la información de extensión de ancho de banda se proporciona con una resolución "normal" (comparativamente baja) durante los intervalos de tiempo 620a a 620d y 620f. Por ejemplo, se proporciona un conjunto de información de extensión de ancho de banda por cada uno de los intervalos de tiempo 620a a 620d y 620f. Por ejemplo, una forma espectral común (o modelado espectral) está representada por un conjunto de parámetros de extensión de ancho de banda por cada uno de los intervalos de tiempo 620a a 620d y 620f, de manera tal que la información de extensión de ancho de banda no representa un cambio de la forma espectral (o modelado espectral) dentro de un intervalo de tiempo individual 620a a 620d y 620f. Por el contrario, el decodificador de audio 100 está configurado para ajustar la resolución temporal empleada por el proveedor de información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda está provista de una resolución temporal aumentada en el intervalo de tiempo (o trama) 620e. En consecuencia, el proveedor de información de extensión de ancho de banda 130 puede subdividir el intervalo de tiempo 620e en cuatro sub-intervalos de tiempo 630a a 630d en respuesta al tiempo de detección del inicio (o final) de un fricativo o africado t_i dentro del intervalo de tiempo 620e. En consecuencia, el proveedor de información de extensión de ancho de banda puede proporcionar un conjunto de información de extensión de ancho de banda por cada uno de los sub-intervalos de tiempo 630a a 630d. En consecuencia, un primer conjunto de información de extensión de ancho de banda (por ej. parámetros) provisto para el sub-intervalo de tiempo 630a puede describir una forma espectral (o modelado espectral) que se ha de aplicar en la extensión de ancho de banda del sub-intervalo de tiempo 630a, un segundo conjunto de información de extensión de ancho de banda puede describir una forma espectral o modelado espectral que se ha de aplicar a una extensión de ancho de banda del sub-intervalo de tiempo 630b, un tercer conjunto de información de extensión de ancho de banda puede describir una forma espectral o modelado espectral que se ha de aplicar en la extensión de ancho de banda del sub-intervalo de tiempo 630c y un cuarto conjunto de información de extensión de ancho de banda puede describir una forma espectral o modelado espectral que se ha de aplicar en una extensión de ancho de banda del sub-intervalo de tiempo 630d. En consecuencia, los conjuntos de información de extensión de ancho de banda (o parámetros de extensión de ancho de banda) individuales se proporcionan por el proveedor de información de extensión de ancho de banda 130, de tal manera que la forma espectral o modelado espectral que se ha de aplicar en una extensión de ancho de banda de los intervalos de tiempo 630a a 630d se señalice de modo independiente. En consecuencia, una forma espectral o modelado espectral se codifica con resolución temporal aumentada (que es más elevada que la resolución temporal "normal" o "baja") para el intervalo de tiempo 620e en respuesta a la detección del inicio o final de un fricativo o africado dentro del intervalo de tiempo 620e. Sin embargo, se debe tener presente que el intervalo de tiempo 630a a 630d puede ser de igual longitud (por ejemplo en términos de tiempo o en términos de un número de muestras). Además, se debe tener presente que la resolución temporal aumentada para la provisión de la información de extensión de ancho de banda ya se utilizó en el sub-intervalo de tiempo 630a, es decir, antes del momento t_i en el cual se detecta el inicio o final del fricativo o africado. Además, también se utiliza la resolución temporal aumentada en el sub-intervalo de tiempo 630c, es decir, después del intervalo de tiempo 630b durante el cual se detecta el inicio o final del fricativo o africado. En consecuencia, el inicio o final del fricativo o africado puede codificarse con buena calidad de audio.

60 La Fig. 7 muestra otra representación esquemática de la resolución temporal utilizada para la provisión de información de extensión de ancho de banda. Se designa un eje de tiempo con 710. Como se puede apreciar, hay intervalos de tiempo 720a a 720f. Como se puede apreciar también, un momento en el cual se detecta un inicio (o final) de un fricativo o africado se designa t_i y radica en un primer cuarto del intervalo de tiempo 720e. Como se puede apreciar, una información de extensión de ancho de banda está provista de resolución temporal "normal" o

"baja" (por ejemplo, un conjunto de información de extensión de ancho de banda o un conjunto de parámetros de extensión de ancho de banda por intervalo de tiempo) para los intervalos de tiempo 720a, 720b, 720c y 720f. Sin embargo, en respuesta a la detección de que hay un inicio de un fricativo o africado en el momento t_i , el codificador de audio 100 ajusta la resolución temporal empleada por el proveedor de información de extensión de ancho de banda, de manera tal que se utiliza una resolución temporal "aumentada" (o "elevada") durante los intervalos de tiempo 720d y 720e. En consecuencia, se proporcionan conjuntos de información de extensión de ancho de banda (o parámetros de extensión de ancho de banda) individuales para los cuatro sub-intervalos de tiempo del intervalo de tiempo 720 y para los cuatro sub-intervalos de tiempo del intervalo de tiempo 720e. Por consiguiente, una envolvente espectral o forma de la envolvente espectral que se ha de utilizar para una extensión de ancho de banda (en el lado de un decodificador de audio), se representa (o codifica) con una resolución espectral aumentada durante los intervalos de tiempo 720d y 720e.

Por ejemplo, se puede proporcionar un conjunto de parámetros de extensión de ancho de banda individual por cada sub-intervalo de tiempo de los intervalos de tiempo 720d y 720e.

Sin embargo, se debe tener presente que también se utiliza la resolución temporal aumentada para el intervalo de tiempo 720d que precede (precede inmediatamente) al intervalo de tiempo 720e, en el cual se incluye el momento en el cual se detecta el inicio (o final) del fricativo o africado. Sin embargo, como es conveniente, de acuerdo con la presente invención, que al menos otro intervalo de tiempo (o sub-intervalo de tiempo), anterior (o inmediatamente anterior) al intervalo de tiempo (o sub-intervalo de tiempo) en el cual se detecta el inicio (o final) del fricativo o africado, se codifique con una resolución temporal aumentada, el codificador de audio 100 opta por la resolución temporal aumentada para la provisión (y codificación) de la información de extensión de ancho de banda del intervalo de tiempo 720d. Por consiguiente, dado que el momento en el cual se detecta el inicio del fricativo o africado radica dentro de un primer sub-intervalo de tiempo del intervalo de tiempo 720e, el decodificador de audio decide que también se debe procesar el intervalo de tiempo 720d (precedente) con elevada resolución temporal, de tal manera que la elevada resolución temporal ya se aplica a un intervalo de tiempo (o sub-intervalo de tiempo) anterior al sub-intervalo de tiempo en el cual se detecta el inicio (o final) del fricativo o africado.

Por el contrario, si solo se detectara el inicio (o final) del fricativo o africado en un segundo sub-intervalo del intervalo de tiempo 720e, el codificador de audio seleccionaría (posiblemente) una baja resolución temporal para la provisión de la información de extensión de ancho de banda para el intervalo de tiempo 720d (que es la situación expuesta en la Fig. 6). En consecuencia, es evidente, a partir de la Fig. 7 que se ejecuta una cierta "anticipación temporal" por el hecho de que se elige una resolución temporal aumentada para la provisión de la información de extensión de ancho de banda aunque esto no sea requerido por la alineación de tramas.

En consecuencia, incluso aunque el comienzo de un inicio de un fricativo o africado se procesa con elevada resolución temporal, en el que el comienzo del inicio del fricativo o africado se encuentra típicamente antes de un momento en el cual el inicio de un fricativo o africado se detecta en realidad por el detector 120. En consecuencia, se puede obtener una reproducción de audio con buena calidad perceptual sin grandes alteraciones.

En resumen, las Figs. 3, 5, 6 y 7 ilustran conceptos operativos que se pueden aplicar al codificador de audio 100 de acuerdo con la presente invención. Sin embargo, se pueden emplear en realidad diferentes conceptos de alineación de tramas siempre que se asegure que la información de extensión de ancho de banda está proporcionada con una resolución temporal aumentada (en comparación con una resolución temporal normal) al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un inicio de un fricativo o africado (o un final de un fricativo o africado) y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado (o el final del fricativo o africado).

Se debe tener presente que las Figs. 6 y 7 representan, por ejemplo, una estructura de una señal de audio codificada. Por ejemplo, la señal de audio codificada puede comprender una representación codificada de una porción de baja frecuencia de un contenido de audio. Además, la representación de audio codificada puede comprender una pluralidad de conjuntos de parámetros de extensión de ancho de banda.

Por ejemplo, se puede proporcionar un conjunto de parámetros de extensión de ancho de banda por cada una de las tramas 620a a 620d y 620f. Además, se puede proporcionar un conjunto de información de extensión de ancho de banda por cada una de las tramas 720a, 720b, 720c, 720f. Sin embargo, los conjuntos de parámetros de extensión de ancho de banda pueden proporcionarse de una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado. Por ejemplo, los conjuntos de parámetros de extensión de ancho de banda se proporcionan de una resolución temporal aumentada para la trama 620e. Por ejemplo, se puede proporcionar un total de cuatro conjuntos de parámetros de extensión de ancho de banda para la trama 620e de tal manera que la resolución temporal se aumenta en la subtrama 630a que precede a la subtrama 630b en la cual se detecta el inicio o final del fricativo o africado. Además,

se pueden proporcionar dos conjuntos de parámetros de extensión de ancho de banda más para las subtramas 630c y 630d.

5 Es evidente un concepto similar en la Fig. 7, en la que se proporcionan conjuntos de parámetros de extensión de ancho de banda con una resolución temporal aumentada para las tramas 620d y 620e.

10 En conclusión, se pueden proporcionar parámetros de extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado. Además, los parámetros de extensión de ancho de banda se proporcionan con una resolución temporal aumentada para una porción del contenido de audio en el cual se detecta la desactivación de un fricativo o africado.

15 2. Codificador de audio de acuerdo con la fig. 8

La Fig. 8 muestra un diagrama esquemático de bloques de un codificador de audio de acuerdo con una realización de la presente invención.

20 El codificador de audio 800 está configurado para recibir una información de audio de entrada 810 y para proporcionar, basándose en la misma, una información de audio codificada 812.

25 El codificador de audio 800 comprende un detector 820 configurado para detectar un final de un fricativo o africado. El detector 820 proporciona, por ejemplo, una información de ajuste de la resolución temporal 822. Además, el codificador de audio 800 comprende un proveedor de información de extensión de ancho de banda 830 que está configurado para proporcionar información de extensión de ancho de banda 832 empleando una resolución temporal variable. El codificador de audio está configurado para ajustar la resolución temporal empleada por el proveedor de información de extensión de ancho de banda 830 de tal manera que la información de extensión de ancho de banda 832 se proporcione con una resolución temporal aumentada (en comparación con una resolución temporal "normal") en respuesta a una detección de un final de un fricativo o africado. En otras palabras, la resolución temporal utilizada por el proveedor de información de extensión de ancho de banda 830 se aumenta si el detector 820 detecta un final de un fricativo o africado, de tal manera que el final del fricativo o africado se codifica con una resolución temporal comparativamente alta (superior a la normal) de la información de extensión de ancho de banda (o parámetros de extensión de ancho de banda) 832. Además, el codificador de audio 800 comprende una codificación de baja frecuencia 840 que puede proporcionar una representación codificada 842 de una porción de baja frecuencia de un contenido de audio representado por la información de audio de entrada 810.

35 Además, se debe tener presente que el detector 820 puede ser similar al detector 120 antes descrito, y que el proveedor de información de extensión de ancho de banda 130 puede ser similar (o incluso igual) al proveedor de información de extensión de ancho de banda 130 antes descrito. Además, la codificación de baja frecuencia 840 puede ser similar, o incluso igual, a la codificación de baja frecuencia 140 antes descrita.

45 Además, el codificador de audio 800 está configurado para ajustar la resolución temporal empleada por el proveedor de información de extensión de ancho de banda 830 de tal manera que la información de extensión de ancho de banda 832 se proporcione con una resolución temporal aumentada en respuesta a la detección de un final de un fricativo o africado. En consecuencia, se codifica un final de un fricativo o africado con elevada resolución temporal (al menos de la información de extensión de ancho de banda) lo que contribuye a evitar alteraciones y trae aparejada una impresión auditiva natural.

50 Sin embargo, se debe tener presente que el codificador de audio 800 puede proporcionarse, además, de cualquiera de las otras características antes descritas con respecto al codificador de audio 100, y también con respecto a las Figs. 3, 5, 6 y 7. Además, las ventajas que surgen del uso de una resolución temporal aumentada en respuesta a la detección de un final de un fricativo o africado se pueden apreciar, por ejemplo, en la Fig. 5.

55 Además, se debe tener presente que los conceptos de acuerdo con las Figs. 6 y 7 son aplicables tanto en respuesta a una detección del inicio de un fricativo o africado como en respuesta a la detección de un final de un fricativo o africado, y por lo tanto se aplican también al codificador de audio de acuerdo con la Fig. 8.

3. Decodificador de audio de acuerdo con la fig. 9

60 La Fig. 9 muestra un diagrama esquemático de bloques de un decodificador de audio, de acuerdo con una realización de la invención. El decodificador de audio 900 está configurado para recibir una información de audio codificada 910 y para proporcionar, basándose en la misma, una información de audio decodificada 912. El decodificador de audio comprende una decodificación de baja frecuencia 920, que puede estar configurado para

proporcionar una representación decodificada de una porción de baja frecuencia de un contenido de audio representado por la información de audio codificada 910. Por ejemplo, la decodificación de baja frecuencia 920 puede comprender una decodificación general de audio, por ejemplo, como se describe en la Norma Internacional ISO/IEC 14496-3. En otras palabras, la decodificación de baja frecuencia 920 puede comprender, por ejemplo, una bien conocida "codificación avanzada de audio" (AAC) de MPEG-2 y puede decodificar, por ejemplo, una porción de baja frecuencia de un contenido de audio hasta una frecuencia de aproximadamente 6 kHz o 7 kHz. Sin embargo, la decodificación de baja frecuencia 920 puede usar cualquier otro concepto de decodificación tal como, por ejemplo, el muy conocido concepto de decodificación CELP o la conocidísima decodificación por excitación codificada por transformada (TCX). En términos generales, la decodificación de baja frecuencia 920 puede usar cualquier concepto de decodificación general de audio o cualquier concepto de decodificación de voz. El decodificador de audio 900 comprende además una extensión de ancho de banda 930 que está configurado para realizar una extensión de ancho de banda basándose en una información de extensión de ancho de banda 932 que se proporciona por un codificador de audio y que típicamente está incluida en la información de audio codificada 910. La extensión de ancho de banda 930 puede utilizar, típicamente, la información proporcionada por la decodificación de baja frecuencia 920. Por ejemplo, la extensión de ancho de banda 930 puede estar configurada para ejecutar una replicación de la banda espectral (SBR) basándose en una porción de baja frecuencia decodificada del contenido de audio (donde la porción de baja frecuencia decodificada del contenido de audio se proporciona por la decodificación de baja frecuencia 920). Por ejemplo, la extensión de ancho de banda 930 puede realizar la funcionalidad de la denominada "herramienta de SBR" o la de la denominada "SBR de bajo retardo" que se describe, por ejemplo, en la Norma Internacional ISO/IEC 14496-3.

Sin embargo, el decodificador de audio 900 puede estar configurado para realizar la extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado. En consecuencia, se puede obtener una buena calidad de audio incluso en el caso del inicio de un fricativo o africado o en el de la desactivación de un fricativo o africado.

Se debe tener presente que la resolución temporal, que se utiliza para la extensión de ancho de banda, puede señalizarse utilizando una información complementaria que está incluida en la información de extensión de ancho de banda 932. Por ejemplo, la señalización puede realizarse como se describiera en la Sección 4.6.19 de la Norma Internacional ISO/IEC 14496-3. En particular, la señalización de la resolución temporal puede realizarse de acuerdo con lo descrito en la Sección 4.6.19.3.2 de ISO/IEC 14496-3, subparte 4. Por consiguiente, la extensión de ancho de banda 930 puede evaluar dicha señalización para decidir qué resolución temporal se debe emplear para la extensión de ancho de banda.

Sin embargo, como alternativa, el decodificador de audio puede estar configurado para detectar un inicio de un fricativo o africado o un final de un fricativo o africado basándose en la porción decodificada de baja frecuencia del contenido de audio, que puede proporcionarse por la decodificación de baja frecuencia 920. En consecuencia, el decodificador de audio 900 puede decidir sobre la resolución temporal que se ha de utilizar para la extensión de ancho de banda de manera similar al codificador de audio antes descrito. En ese caso, puede no ser necesario incluso utilizar ninguna información complementaria adicional para señalar la resolución temporal a utilizarse para la extensión de ancho de banda, que ayuda a reducir la tasa de bits.

Con respecto a la funcionalidad del decodificador de audio 900, se debe tener presente que la funcionalidad corresponde a la funcionalidad del codificador de audio 100 de acuerdo con la Fig. 1 y del codificador de audio 800 de acuerdo con a Fig. 8. En otras palabras, la extensión de ancho de banda se realiza con una resolución "normal" o comparativamente "baja" en ausencia del inicio de un fricativo o africado o de un final de un fricativo o africado, y la extensión de ancho de banda se realiza con una resolución temporal "aumentada" o comparativamente "alta" en presencia de un inicio de un fricativo o africado o un final de un fricativo o africado. Sin embargo, la resolución temporal aumentada también se utiliza para la extensión de ancho de banda al menos durante un periodo de tiempo predeterminado anterior al momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado, de tal manera que se procesa un inicio completo de un fricativo o africado con elevada resolución temporal de la extensión de ancho de banda. En consecuencia, se evitan las alteraciones.

4. Decodificador de audio de acuerdo con la Fig. 10

La Fig. 10 muestra un diagrama esquemático de bloques de un decodificador de audio, de acuerdo con otra realización de la presente invención.

El decodificador de audio 1000 está configurado para recibir una información de audio codificada 1010 y para proporcionar, basándose en la misma, una información de audio decodificada 1012. El decodificador de audio

comprende una decodificación de baja frecuencia 1020 que puede ser sustancialmente igual a la decodificación de baja frecuencia 920 antes descrita. Además, el decodificador de audio 1000 comprende una extensión de ancho de banda 1030, que puede ser sustancialmente igual a la extensión de ancho de banda 930 antes descrita. Sin embargo, el decodificador de audio 1000 está configurado para ejecutar la extensión de ancho de banda basándose en una información de extensión de ancho de banda 1032 proporcionada por un codificador de audio, de manera que la extensión de ancho de banda se ejecuta con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado. En consecuencia, el decodificador de audio 1000 proporciona una información de audio decodificada en la cual los finales de fricativos o africados están representados con buena precisión. En consecuencia, se evitan las alteraciones.

Además, se debe tener presente que las explicaciones antes presentadas con respecto al decodificador de audio 900 también se aplican al decodificador de audio 1000. Además, se debe tener presente que el decodificador de audio 1000 puede complementarse con cualquiera de las características y funcionalidades descritas con respecto al codificador de audio 900. Además, el codificador de audio 1000 (como así también el codificador de audio 900) pueden complementarse por cualquiera de las características y funcionalidades descritas en el presente documento con respecto al decodificador de audio, puesto que la decodificación de audio corresponde a la codificación de audio antes descrita.

5. Sistema de acuerdo con la reivindicación 11

La Fig. 11 muestra un diagrama esquemático de bloques de un sistema de acuerdo con una realización de la presente invención. El sistema 1100 comprende un codificador de audio 1120, que está configurado para recibir una información de audio de entrada 1110 y para suministrar, basándose en la misma, una información de audio codificada 1130 a un decodificador de audio 1140. El decodificador de audio 1140 está configurado para proporcionar una información de audio decodificada 1150 basándose en la información de audio codificada 1130.

Sin embargo, se debe tener presente que el codificador de audio 1120 puede ser igual al codificador de audio 100 descrito con respecto a la Fig. 1 o al codificador de audio 800 descrito con respecto a la Fig. 8. Además, el decodificador de audio 1140 puede ser igual al decodificador de audio 900 descrito con respecto a la Fig. 9 o al decodificador de audio 1000 descrito con respecto a la Fig. 10. En consecuencia, el decodificador de audio puede estar configurado para recibir la información de audio codificada proporcionada por el codificador de audio, y para proporcionar, basándose en la misma, la información de audio decodificada 1150, de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado y/o de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado. En consecuencia, se puede obtener una buena reproducción de fricativos o africados.

Se debe tener presente que el sistema puede complementarse con cualquiera de las características y funcionalidades descritas anteriormente con respecto a los codificadores de audio y los decodificadores de audio.

6. Método para proporcionar una información de audio codificada basándose en una información de audio de entrada de acuerdo con la Fig. 12

La Fig. 12 muestra un diagrama de flujo de un método para proporcionar una información de audio codificada basándose en una información de audio de entrada. El método 1200 de acuerdo con la Fig. 12 comprende detectar un inicio de un fricativo o africado y/o un final de un fricativo o africado (etapa 1210). El método comprende además proporcionar 1220 información de extensión de ancho de banda empleando una resolución temporal variable. La resolución temporal utilizada para proporcionar la información de extensión de ancho de banda puede ajustarse, por ejemplo, de tal manera que la información de extensión de ancho de banda se produzca con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado. Como alternativa, la resolución temporal para proporcionar la información de extensión de ancho de banda puede ajustarse de tal manera que la información de extensión de ancho de banda se produzca con una resolución temporal aumentada en respuesta a la detección de un final de un fricativo o africado.

El método 1200 de acuerdo con la Fig. 12 se basa en las mismas condiciones que los codificadores de audio antes descritos. Además, el método 1200 puede complementarse con cualquiera de las características y funcionalidades

descritas anteriormente con respecto al codificador de audio (y también con respecto al decodificador de audio).

7. Método para proporcionar una información de audio decodificada de acuerdo con la reivindicación 13

5 La Fig. 13 muestra un diagrama de flujo de un método para proporcionar una información de audio decodificada de acuerdo con una realización de la invención. El método 1300 comprende decodificar 1310 una porción de baja frecuencia de una información de audio que, sin embargo, no es una etapa esencial del método.

10 El método 1300 comprende además ejecutar 1320 una extensión de ancho de banda basándose en una información de extensión de ancho de banda proporcionada por un codificador de audio, de tal manera que se realice una extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes del momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado y/o de tal manera que la extensión de ancho de banda se realice con una resolución temporal aumentada al menos durante un periodo
15 de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

20 El método 1300 se basa en las mismas condiciones que el codificador de audio antes descrito y el decodificador de audio antes descrito. Además, se debe tener presente que el método 1300 puede complementarse con cualquiera de las características y funcionalidades descritas en la presente con respecto al decodificador de audio. Además, el método 1300 también puede complementarse con cualquiera de las características y funcionalidades descritas en la presente con respecto al codificador de audio, teniendo en cuenta que el proceso de decodificación es sustancialmente inverso al proceso de codificación.

25 8. Conclusiones

Para concluir las explicaciones expuestas, se debe tener presente que las realizaciones de acuerdo con la invención están relacionadas con la codificación de voz y, en especial, con la codificación de voz que utiliza técnicas de extensión de ancho de banda (BWE). Las realizaciones de acuerdo con la invención apuntan a mejorar la calidad
30 perceptual de la señal decodificada mediante la detección de fricativos o africanos dentro de la señal de voz y en consecuencia la adaptación de la resolución temporal del parámetro de extensión de ancho de banda realizada con posterioridad al procesamiento (por ejemplo, mediante la adaptación de una resolución temporal que se utiliza para la provisión de conjuntos de información de extensión de ancho de banda). Las realizaciones de acuerdo con la invención comprenden la detección de inicios y finales de porciones de señales con fricativos o africanos de una
35 señal de voz y proporcionar un post-procesamiento de extensión de ancho de banda de resolución temporalmente precisa durante el periodo total de inicio y final de estas porciones con fricativos o africanos de la señal (en el que el procesamiento de extensión de ancho de banda puede comprender, por ejemplo, una provisión de dicha información de extensión de ancho de banda en el lado de un codificador de audio y puede comprender realizar una extensión de ancho de banda en el lado del decodificador de audio). Por este medio se puede reducir la aparición de alteraciones de pre- y post-eco y se puede modelar el inicio y final suficientemente suave de porciones de señal con fricativos o africanos mediante los parámetros de extensión de ancho de banda de resolución precisa. Por la presente, se evita la agudeza auditiva desagradable de los fricativos o africanos y la aparición de molestos pre- y
40 post-ecos dentro de la señal codificada.

45 Las realizaciones de acuerdo con la invención superan en eficiencia las soluciones convencionales. Por ejemplo, en [1] se propone alinear un instante de tiempo de inicio de una trama de parámetro de extensión de ancho de banda con el punto de tiempo de un cambio de distorsión espectral. El cambio de distorsión espectral podría indicar un inicio o un final repentino de una porción de señal con fricativo o africado. La técnica de alineación propuesta en [1] previene la aparición de pre-ecos de fricativos o africanos dentro de los métodos de extensión de ancho de banda.
50 Sin embargo, solo se detectan los inicios de fricativos o africanos y se pasan por alto los finales. Además, la técnica mencionada no tiene en cuenta el modelado fino de las características espectro-temporales de las inicios y finales de los fricativos o africanos individuales. Por tanto, el sonido de estos puede ser brusco y demasiado agudo.

A continuación se describen algunas realizaciones y aspectos de acuerdo con la invención.

55 Por ejemplo, un codificador de extensión de ancho de banda de la invención comprende un detector de fricativos o africanos y un conmutador de resolución espectro-temporal de extensión de ancho de banda.

60 El detector de fricativos o africanos tiene capacidad preferentemente para detectar tanto inicios como finales de fricativos o africanos. Una realización adecuada con baja complejidad informática de ese tipo de detector se puede basar, por ejemplo, en la evaluación de una tasa de cruce en cero (ZCR) y una relación de energía (para detalles consúltese, por ejemplo, las referencias [2] y [3]). El detector puede estar conectado además a un discriminador de voz/música para restringir el procesamiento subsiguiente de la invención solo a las señales de voz.

En algunas realizaciones, es conveniente o incluso necesaria una determinada anticipación del detector para poder conmutar a tiempo la resolución de la extensión de ancho de banda de tal manera que durante la totalidad de la duración de la porción de señal con inicio y final, se emplee una resolución temporal de buena precisión dentro de la estimación/síntesis de los parámetros de extensión de ancho de banda. La duración de las porciones de señal con inicio o final se pueden medir por adaptación de la señal o suponerla fija en un valor determinado en forma empírica. Por ejemplo, un número de intervalos de tiempo o sub-intervalos de tiempo que se procesan con elevada resolución temporal en respuesta a la detección del inicio de un fricativo o africado o al final de un fricativo o africado pueden predeterminarse se pueden ajustar de conformidad con las características de la señal. Por ejemplo, un fricativo o africado detectado podría activar una resolución temporal cuatro veces mayor durante un grupo de varias tramas de señal consecutivas (por ej., dos o tres tramas) que abarcan la totalidad del inicio o final del fricativo o africado detectado. Preferentemente, aunque no necesariamente, el grupo de tramas de señal con elevada resolución temporal está aproximadamente centrado con respecto al inicio o final del fricativo o africado detectado, cubriendo de esta manera toda la duración del inicio o final. En el caso de una alineación de tramas transitoria adaptativa de extensión de ancho de banda, la activación de una resolución temporal más elevada durante todo un grupo de tramas de la señal desencadenada por la detección de fricativos o africados cancela la alineación de tramas adaptativa transitoria.

A continuación se describen algunos detalles con respecto a las figuras.

La Fig. 2 muestra un espectrograma de una señal de voz original con barras verticales de guiones de color magenta que muestra una alineación de tramas de extensión de ancho de banda convencional. Las barras de guiones negros indican los límites de los fricativos o africados.

La Fig. 3 muestra un espectrograma de una señal de voz original con una alineación de tramas de extensión de ancho de banda de la invención adaptado a los límites de los fricativos o africados, que está indicado por las líneas verticales negras continuas. En un punto de tiempo en el que se ha detectado (el inicio o final) de un límite de fricativo o africado, se perfecciona la resolución del post-procesamiento de extensión de ancho de banda conmutando a una resolución cuatro veces más elevada durante un grupo de tres tramas consecutivas.

La Fig. 4 muestra un espectrograma obtenido de la misma señal de voz codificada utilizando la alineación de tramas de extensión de ancho de banda convencional. Las elipses amarillas indican alteraciones causadas por la alineación de tramas de extensión de ancho de banda convencional (de izquierda a derecha): A: pre-eco e inicio brusco; B: post-eco y final brusco; C: fuga de energía de la vocal anterior al fricativo o africado modelado debido a una alineación de tramas demasiado basta.

La Fig. 5 muestra el espectrograma resultante de la misma señal de voz codificada utilizando la alineación de tramas de extensión de ancho de banda de la invención. Las áreas problemáticas indicadas en la Fig. 4 han mejorado sustancialmente.

En conclusión, los espectrogramas analizados en este punto indican que se puede mejorar sustancialmente la calidad del audio aplicando el concepto de acuerdo con la presente invención.

También para concluir, las realizaciones de acuerdo con la invención crean un codificador de audio o un método de codificación de audio o un programa informático relacionado, según lo descrito anteriormente.

Otras realizaciones de acuerdo con la invención crean un decodificador de audio o un método de decodificación de audio o un programa informático relacionado, según lo descrito anteriormente.

Además, las realizaciones que no forman parte de la invención tal y como se ha reivindicado crean una señal de audio codificada o un medio de almacenamiento que tiene guardada la señal de audio codificada antes descrita.

9. Alternativas de implementación

Si bien algunos aspectos se han descrito en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a una etapa del método o a una característica de una etapa del método. De manera análoga, los aspectos descritos en el contexto de una etapa del método también representan una descripción de un bloque o elemento correspondiente o de una característica de un aparato correspondiente. Algunas o todas las etapas del método pueden realizarse por medio de (o utilizando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, una cualquiera o más de las etapas más importantes del método pueden ejecutarse por un aparato de este tipo.

La señal de audio codificada de la invención puede almacenarse en un medio de almacenamiento digital o puede transmitirse por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión cableado tal como internet.

5 Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un Blue-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales control legibles electrónicamente, que cooperan (o
10 tienen capacidad para cooperar) con un sistema informático programable de tal manera que se realice el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

Algunas realizaciones que no forman parte de la invención tal y como se ha reivindicado comprenden un soporte de datos que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se realice uno de los métodos descritos en la presente.
15

En general, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo código de programa operativo para realizar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse, por ejemplo, en un soporte legible por máquina.
20

Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenado en un soporte legible por máquina.

En otras palabras, una realización del método de la invención consiste, por lo tanto, en un programa informático que tiene código de programa para realizar uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador.
25

El aparato descrito en el presente documento puede implementarse utilizando un aparato de hardware o utilizando un ordenador, o utilizando una combinación de un aparato de hardware y un ordenador.
30

Los métodos descritos en el presente documento pueden realizarse utilizando un aparato de hardware, o utilizando un ordenador, o utilizando una combinación de un aparato de hardware y un ordenador.

Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento serán evidentes para los expertos en la materia. Por lo tanto, solo es intención limitarse al alcance de las siguientes reivindicaciones de patente y no a los detalles específicos presentados a manera de descripción y explicación de las realizaciones presentadas en el presente documento.
35

Una realización proporciona un codificador de audio 100 para proporcionar una información de audio codificada 112 basándose en una información de audio de entrada 112, comprendiendo el codificador de audio un proveedor de información de extensión de ancho de banda 130 configurado para proporcionar información de extensión de ancho de banda 132 utilizando una resolución temporal variable; un detector 120 configurado para detectar un inicio de un fricativo o africado; en el que codificador de audio está configurado para ajustar una resolución temporal utilizada por
40 el proveedor de información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda se proporciona con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado 630a anterior al momento t_i en el cual se detecta un inicio de un fricativo o africado y durante un periodo de tiempo predeterminado 630c posterior al momento en el cual se detecta el inicio del fricativo o africado.
45

De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo al primer aspecto, está configurado para conmutar de una primera resolución temporal para la provisión de la información de extensión de ancho de banda a una segunda resolución temporal para la provisión de la información de extensión de ancho de banda en respuesta a la detección del inicio de un fricativo o africado, en el que la segunda resolución temporal es más elevada que la primera resolución temporal.
50

De acuerdo con un aspecto, el proveedor de información de extensión de ancho de banda del codificador de audio 100, con referencia de nuevo al primer o al segundo aspecto, está configurado para proporcionar la información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda esté asociada a intervalos de tiempo temporalmente regulares 620a, 620b, 620c, 620d, 620e, 620f; 720a-720f de igual longitud temporal, en el que el proveedor de información de extensión de ancho de banda está configurado para proporcionar un único conjunto de información de extensión de ancho de banda para un intervalo de tiempo 620a, 620b, 620c, 620d, 620f; 720a, 720b, 720c, 720f de una longitud temporal dada si se utiliza una primera resolución temporal, y en el que el proveedor de información de extensión de ancho de banda está configurado para proporcionar una
60

pluralidad de conjuntos de información de extensión de ancho de banda asociados a sub-intervalos de tiempo 630a, 630b, 630c, 630d durante un intervalo de tiempo 620e; 720d, 720e de la duración temporal dada si se utiliza una segunda resolución temporal;

5 De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo al tercer aspecto, está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que al menos un sub-intervalo de tiempo 630a; 730d, al cual está asociado un conjunto de información de extensión de ancho de banda precede inmediatamente a otro sub-intervalo de tiempo 630b; 730e al cual está asociado otro conjunto de información de extensión de ancho de banda y durante el cual se detecta un inicio de un fricativo o africado, de tal manera que la resolución temporal aumentada se utiliza en al menos un sub-intervalo de tiempo 630a; 730d anterior al sub-intervalo de tiempo 630b; 730e en el cual se detecta el inicio de un fricativo o africado.

15 De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo al tercer o cuarto aspecto, está configurado para subdividir un intervalo de tiempo 620e; 720d, 720e dado de la duración temporal dada en cuatro sub-intervalos 630a-630d; 730a-730h de igual longitud, si se utiliza una resolución temporal aumentada para proporcionar la información de extensión de ancho de banda para el intervalo de tiempo dado 620e; 720d, 720e de la duración temporal dada, de tal manera que se proporcionan cuatro conjuntos de información de extensión de ancho de banda para el intervalo de tiempo dado de la duración temporal dada.

20 De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo a uno del primer al quinto aspectos, está configurado para utilizar selectivamente una resolución temporal aumentada para proporcionar información de extensión de ancho de banda para un primer intervalo de tiempo 720d de una longitud temporal dada anterior a un segundo intervalo de tiempo 720e de la duración temporal dada, si se detecta un inicio de un fricativo o africado dentro del segundo intervalo de tiempo 720e y si una distancia temporal entre un momento en el cual se detecta el inicio del fricativo o africado y un límite entre el primer intervalo de tiempo 720d y el segundo intervalo de tiempo 720e es menor que una distancia temporal predeterminada.

30 De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo a uno del primer al sexto aspectos, está configurado para realizar una anticipación temporal, de manera que se utiliza una resolución temporal aumentada para proporcionar información de extensión de ancho de banda para un primer intervalo de tiempo 720d de una longitud temporal dada anterior a un segundo intervalo de tiempo 720e de la duración temporal dada en respuesta a una detección de un inicio de un fricativo o africado en el segundo intervalo de tiempo 720e.

35 De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo a uno del primer al séptimo aspectos, está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que se proporciona información de extensión de ancho de banda con una misma resolución temporal aumentada al menos durante un periodo de tiempo predeterminado 630a; 730d anterior a un momento t_t en el cual se detecta un inicio de un fricativo o africado y durante un periodo de tiempo predeterminado 630c; 730f posterior al momento en el cual se detecta el inicio del fricativo o africado.

45 De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo a uno del primer al octavo aspectos, está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que se proporcionan conjuntos de información de extensión de ancho de banda con la misma resolución temporal aumentada al menos para un primer sub-intervalo de tiempo 630a; 730d, un segundo sub-intervalo de tiempo 630b; 730e y un tercer sub-intervalo de tiempo 630c; 730f, en el que el primer sub-intervalo de tiempo precede inmediatamente al segundo sub-intervalo de tiempo; en el que se detecta un inicio de un fricativo o africado en el segundo sub-intervalo de tiempo; y en el que el tercer sub-intervalo de tiempo sigue inmediatamente al segundo sub-intervalo de tiempo.

50 De acuerdo con un aspecto, el detector del codificador de audio 100, con referencia de nuevo a uno del primer al noveno aspectos, está configurado para detectar un final de un fricativo o africado; y el codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que se proporciona información de extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

60 De acuerdo con un aspecto, el detector del codificador de audio 100, con referencia de nuevo a uno del primer al décimo aspectos, está configurado para evaluar una tasa de cruce en cero y/o una relación de energía y/o una distorsión espectral para detectar un inicio de un fricativo o africado.

De acuerdo con un aspecto, el detector del codificador de audio 100, con referencia de nuevo a uno del primer al undécimo aspectos, está configurado para evaluar una tasa de cruce en cero y/o una relación de energía y/o una distorsión espectral para detectar un final de un fricativo o africado.

- 5 De acuerdo con un aspecto, el codificador de audio 100, con referencia de nuevo a uno del primer al duodécimo aspectos, está configurado para ajustar selectivamente una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que se proporciona información de extensión de ancho de banda con una resolución temporal aumentada en respuesta a una detección de un inicio de un fricativo o africado solo para una porción de señal de voz, pero no para una porción de señal de música.

10

Referencias:

[1] Patente de Estados Unidos número US 20110099018, "Apparatus and Method for Calculating Bandwidth Extension Data Using a Spectral Tilt Controlled Framing"

15

[2] D. Ruinskiy y N. Dadush e Y. Lavner, "Spectral and textural feature-based system for automatic detection of fricatives and affricates", IEEE 26a Convención de Ingenieros Eléctricos y Electrónicos de Israel (IEEEI), pp. 771-775, 2010.

20

[3] H. Fujihara y M. Goto, "Three techniques for improving automatic synchronization between music and lyrics: Fricative detection, filler model, and novel feature vectors for vocal activity detection", IEEE International Conference on Audio, Speech and Signal Processing, Las Vegas, Estados Unidos, 2008.

REIVINDICACIONES

1. Un codificador de audio (800) para proporcionar una información de audio codificada (812) basándose en una información de audio de entrada (810), comprendiendo el codificador de audio:

5 un proveedor de información de extensión de ancho de banda (830) configurado para proporcionar información de extensión de ancho de banda (832) utilizando una resolución temporal variable;
 un detector (820) configurado para detectar un final de un fricativo o africado;
 10 en el que el codificador de audio está configurado para ajustar una resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que la información de extensión de ancho de banda se proporciona con una resolución temporal aumentada en respuesta a una detección de un final de un fricativo o africado,

caracterizado porque

15 el codificador de audio está configurado para ajustar la resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que se proporciona información de extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

20 2. Un decodificador de audio (1000) para proporcionar una información de audio decodificada (1012) basándose en una información de audio codificada (1010),

en el que el decodificador de audio está configurado para realizar una extensión de ancho de banda (1030) basándose en una información de extensión de ancho de banda (1032) proporcionada por un codificador de audio,

25 **caracterizado porque** la extensión de ancho de banda se ejecuta con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

30 3. Un sistema (1100), que comprende:

un codificador de audio (1120) de acuerdo con la reivindicación 1; y
 un decodificador de audio (1140) configurado para recibir la información de audio codificada (1130) proporcionada por el codificador de audio y para proporcionar, basándose en la misma, una información de audio decodificada (1150),

35 en el que el decodificador de audio está configurado para realizar una extensión de ancho de banda basándose en la información de extensión de ancho de banda proporcionada por el codificador de audio,
 de tal manera que la extensión de ancho de banda se realiza con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta el inicio de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el inicio del fricativo o africado, o

40 de tal manera que la extensión de ancho de banda se realiza con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

45 4. Un método (1200) para proporcionar una información de audio codificada basándose en una información de audio de entrada, comprendiendo el método:

proporcionar (1220) información de extensión de ancho de banda utilizando una resolución temporal variable; y
 50 detectar (1210) un final de un fricativo o africado;
 en el que una resolución temporal utilizada para proporcionar la información de extensión de ancho de banda se ajusta de tal manera que se proporciona información de extensión de ancho de banda con una resolución temporal aumentada en respuesta a una detección de un final de un fricativo o africado;

caracterizado porque

55 el método comprende ajustar la resolución temporal utilizada por el proveedor de información de extensión de ancho de banda de tal manera que se proporciona información de extensión de ancho de banda con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado anterior a un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado.

60 5. Un método (1300) para proporcionar una información de audio decodificada basándose en una información de audio codificada,

en el que el método comprende ejecutar (1320) una extensión de ancho de banda basándose en una información de

extensión de ancho de banda proporcionada por un codificador de audio,
caracterizado porque la extensión de ancho de banda se ejecuta con una resolución temporal aumentada al menos durante un periodo de tiempo predeterminado antes de un momento en el cual se detecta un final de un fricativo o africado y durante un periodo de tiempo predeterminado posterior al momento en el cual se detecta el final del fricativo o africado

5

6. Un producto de programa informático que comprende instrucciones que, cuando se ejecutan en un ordenador, harán que dicho ordenador lleve a cabo un método de acuerdo con una de las reivindicaciones 4 a 5.

10

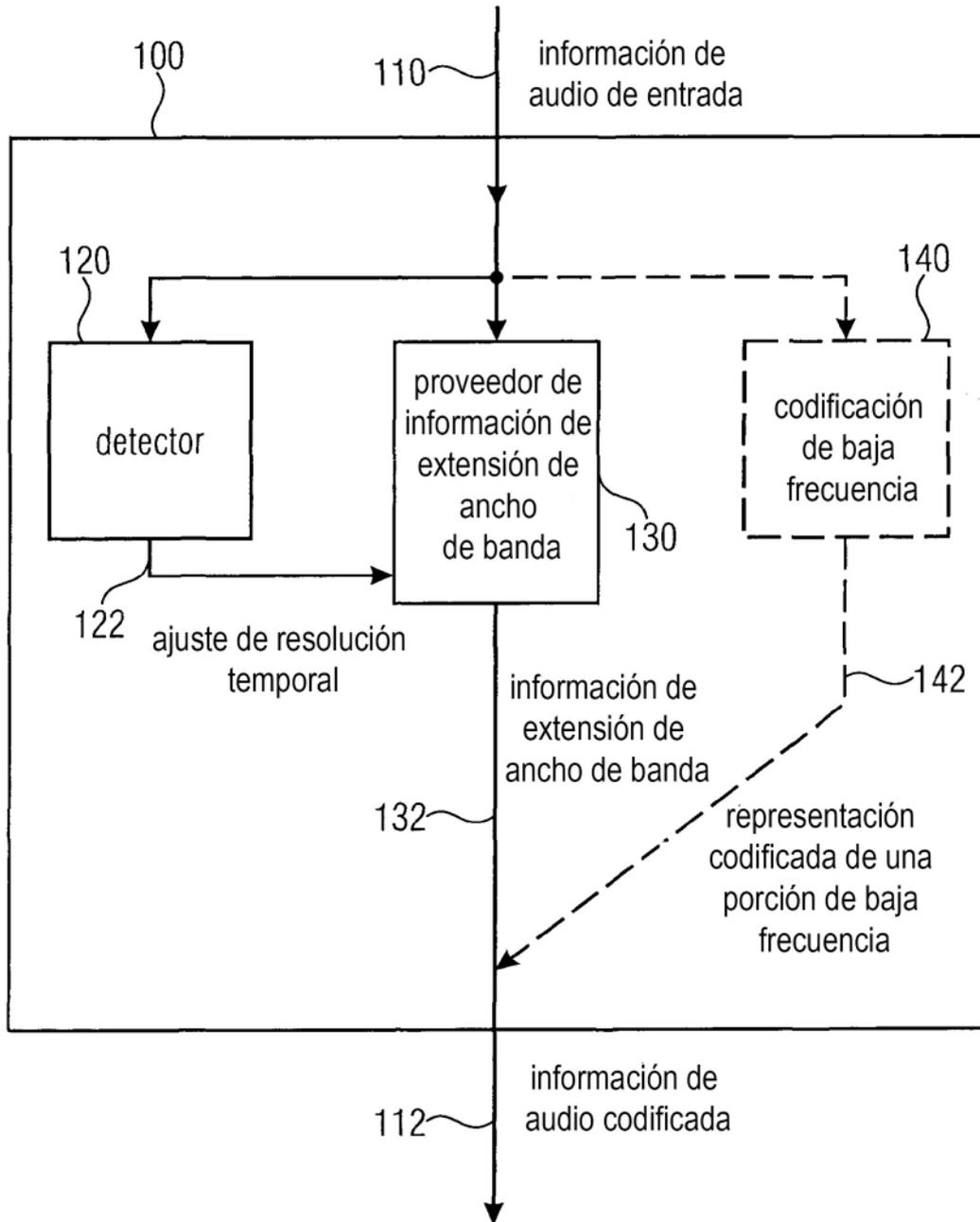
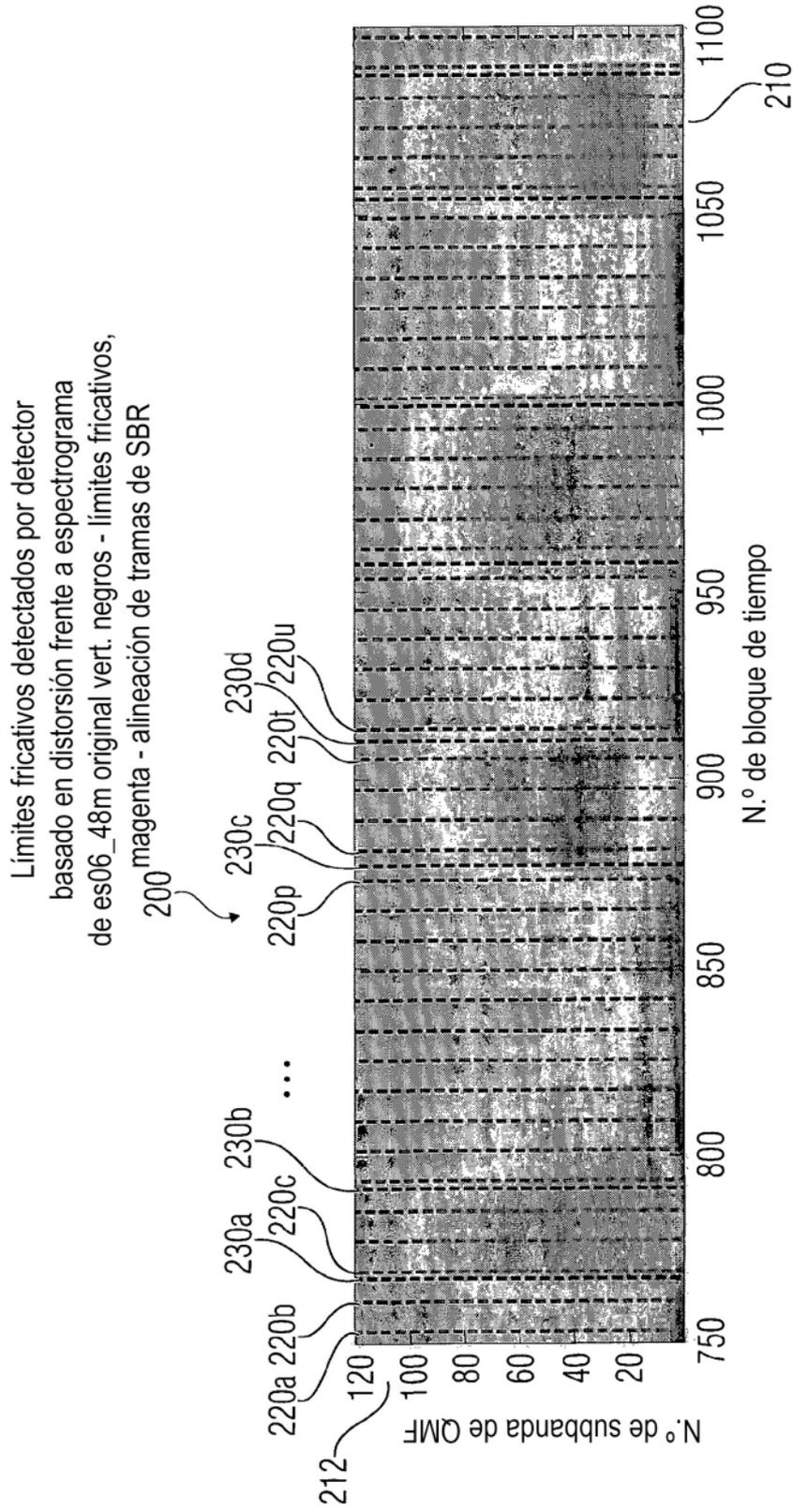


FIGURA 1

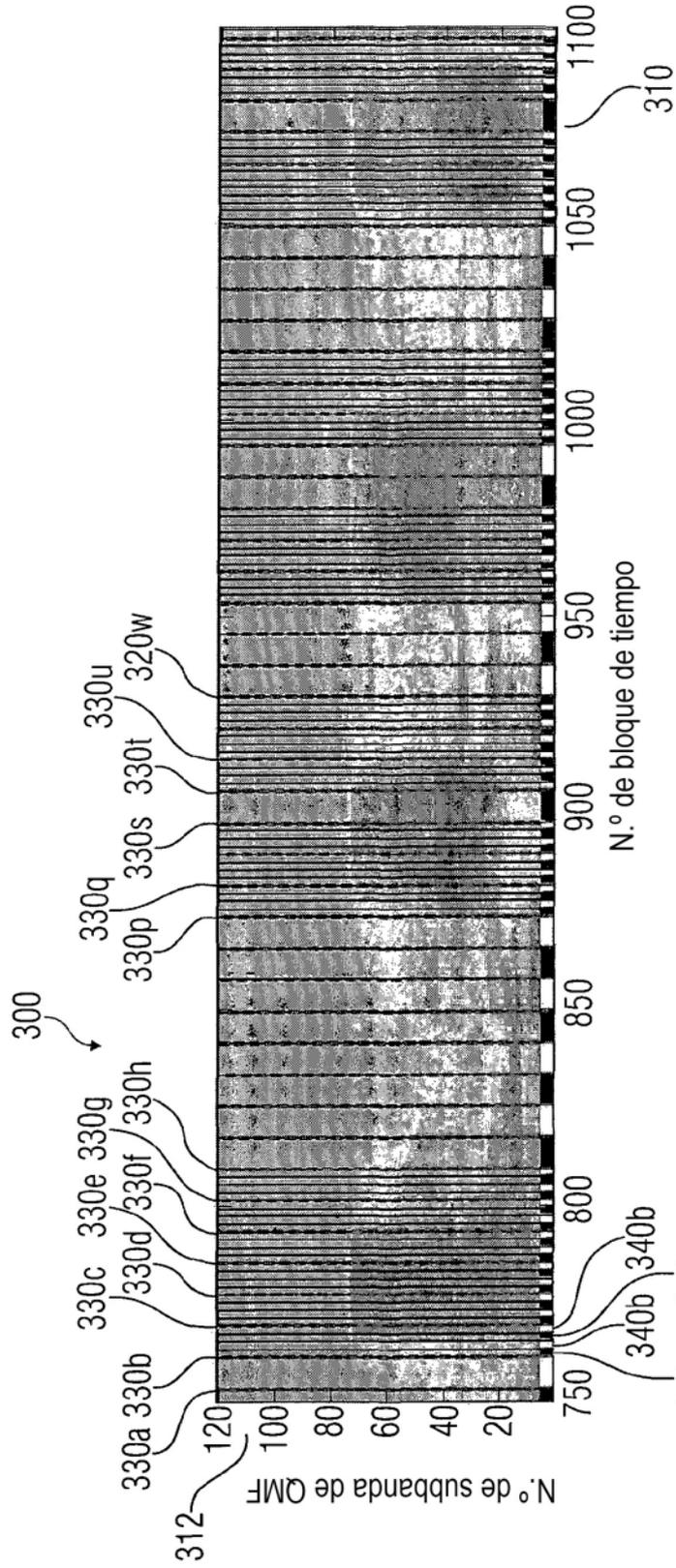


Espectrograma de la señal de voz original con alineación de tramas de BWE convencional (líneas verticales discontinuas magentas) y límites fricativos detectados (líneas verticales discontinuas negras).

FIGURA 2

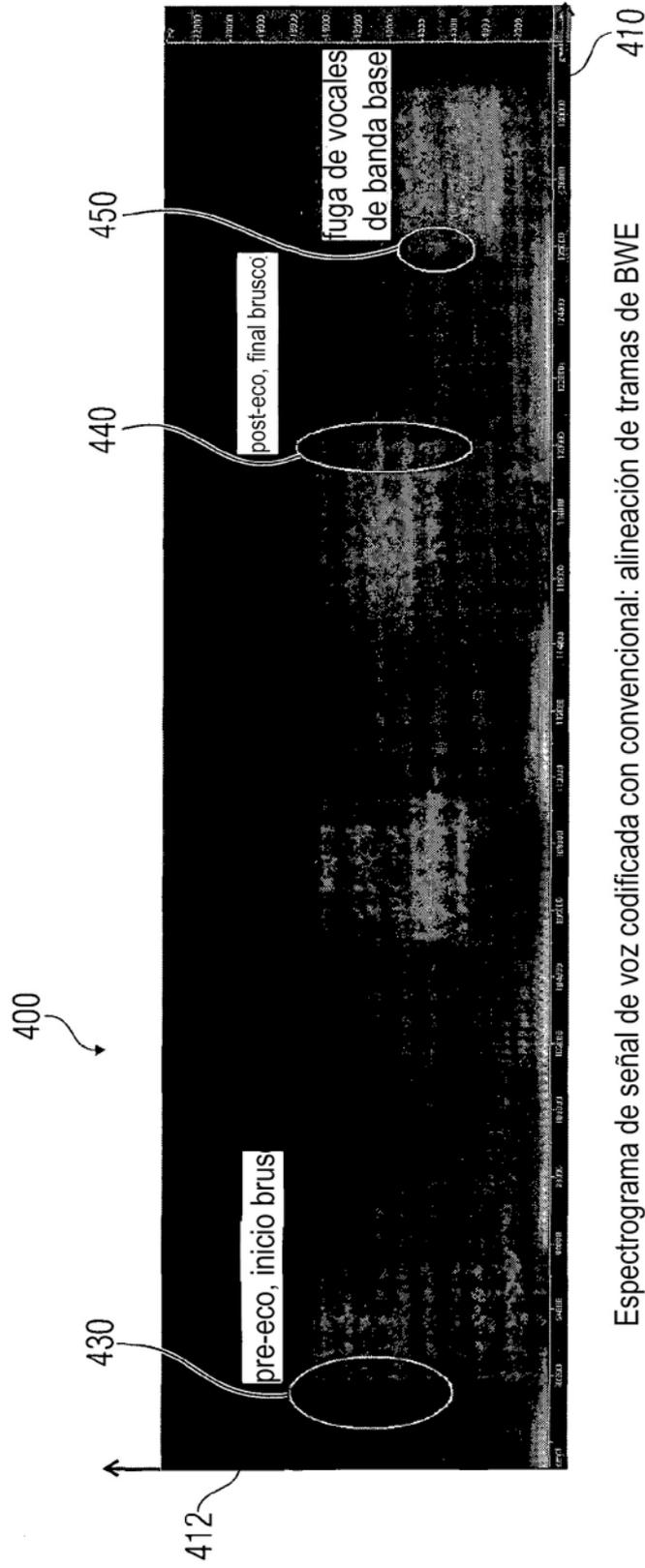
Envolventes de parámetros generados por codificador de SBR
 frente a espectrograma de USAC2 es06_48m_autoSibs4

Negro/blanco alternando - envolventes normales, verde - envolventes alienadas
 transitorias vert. negros - límites de envolvente, magenta - alineación de tramas de SBR



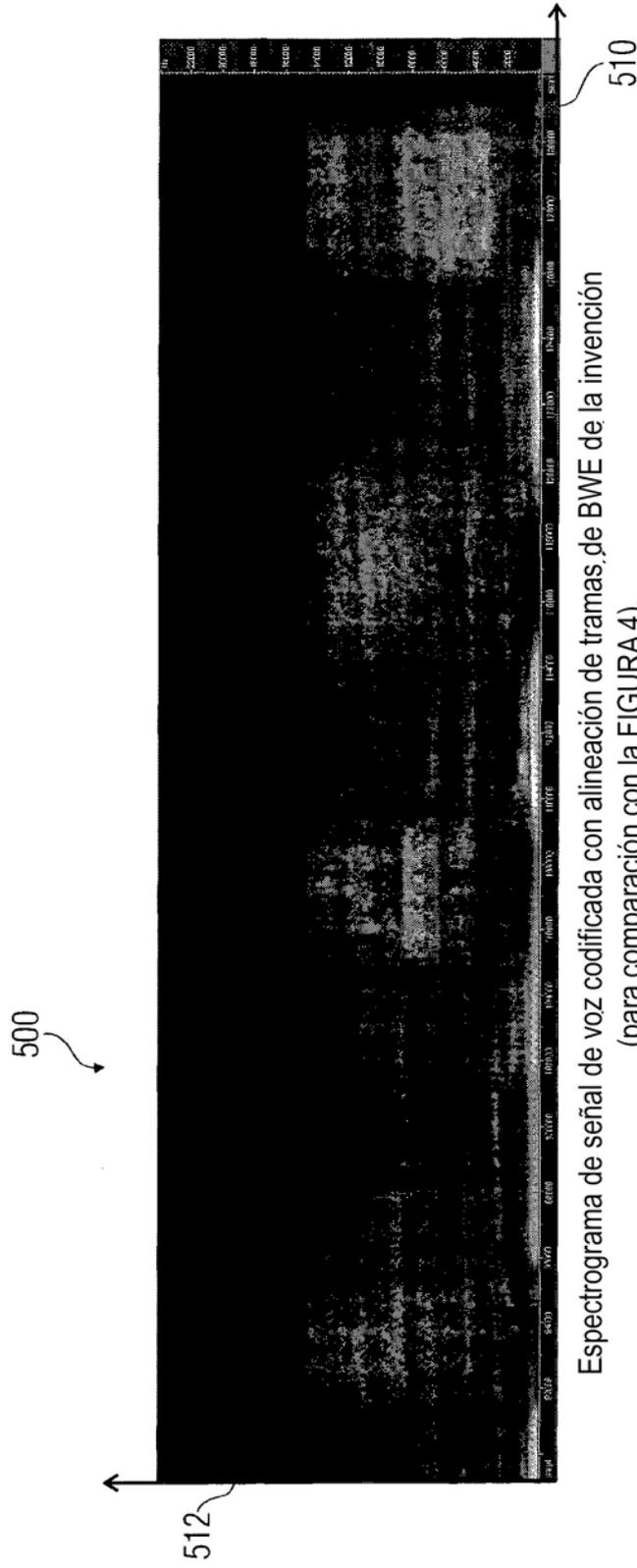
Espectrograma de la señal de voz original con alineación de tramas de BWE de la invención (líneas verticales continuas negras).

FIGURA 3



Espectrograma de señal de voz codificada con convencional: alineación de tramas de BWE (elipses amarillas indican artefactos producidos por la alineación de tramas de BWE convencional).

FIGURA 4



Espectrograma de señal de voz codificada con alineación de tramas de BWE de la invención (para comparación con la FIGURA 4).

Las áreas problemáticas como se indican en la FIGURA 4 se mejoran sustancialmente

FIGURA 5

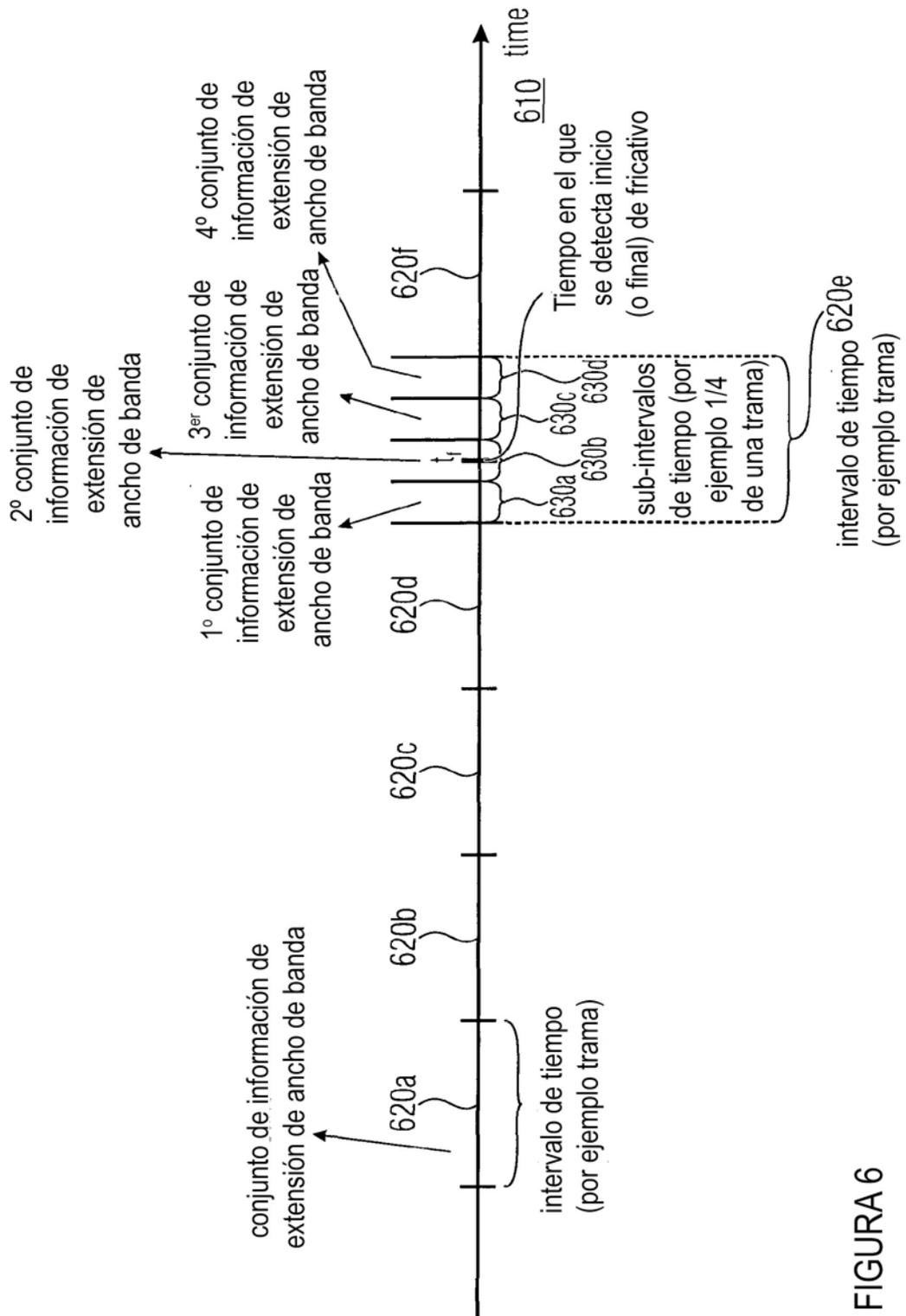


FIGURA 6

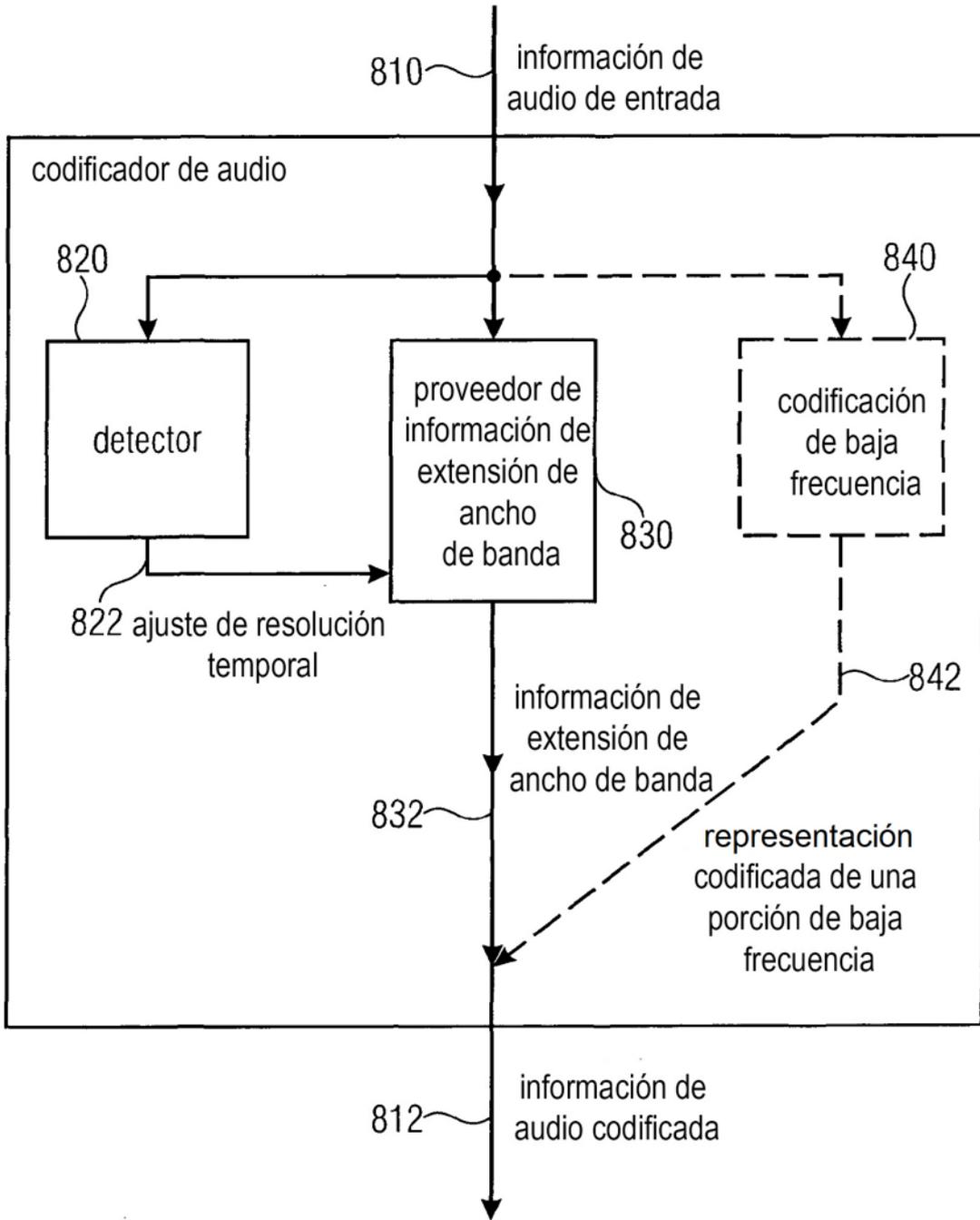


FIGURA 8

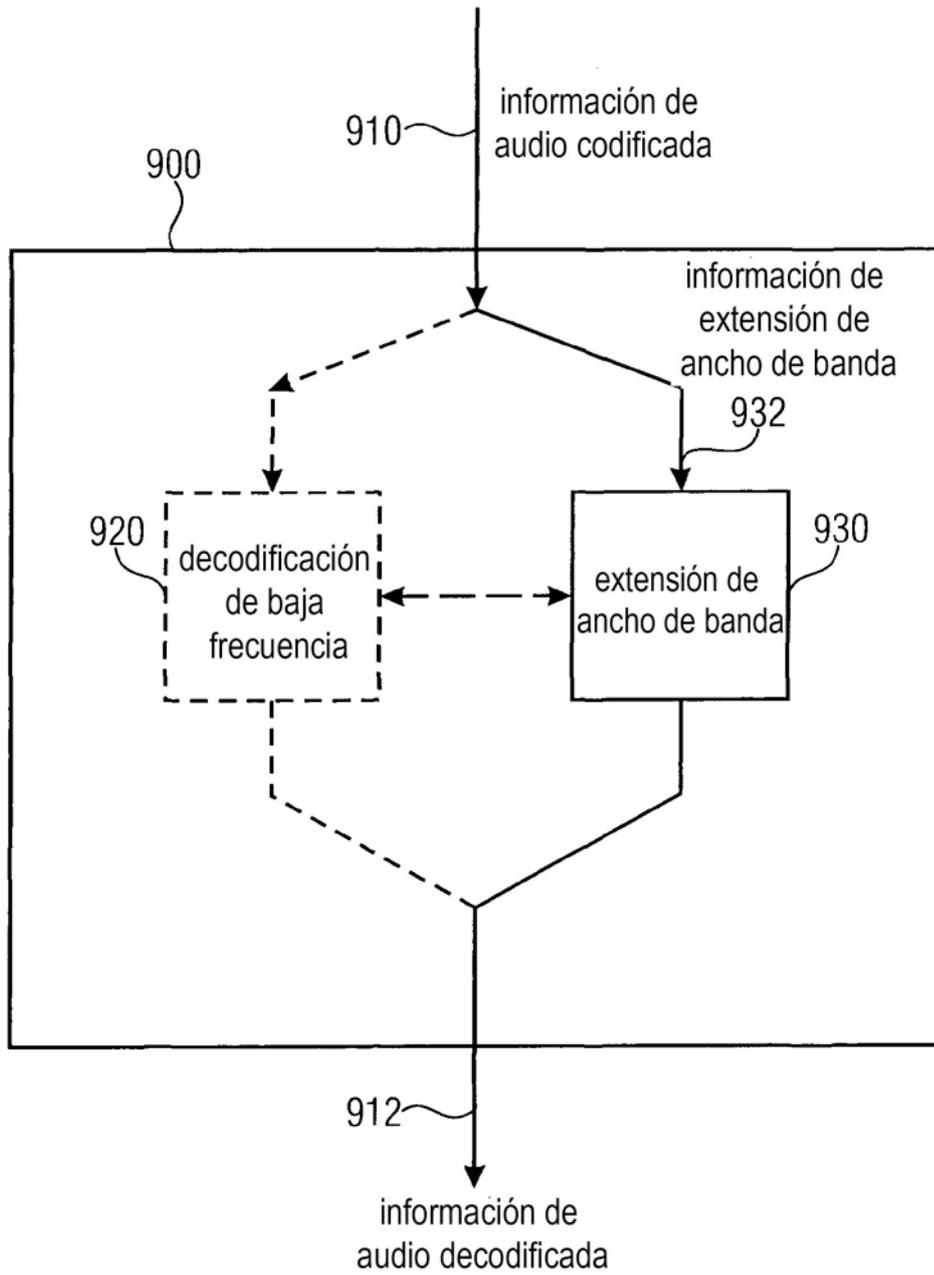


FIGURA 9

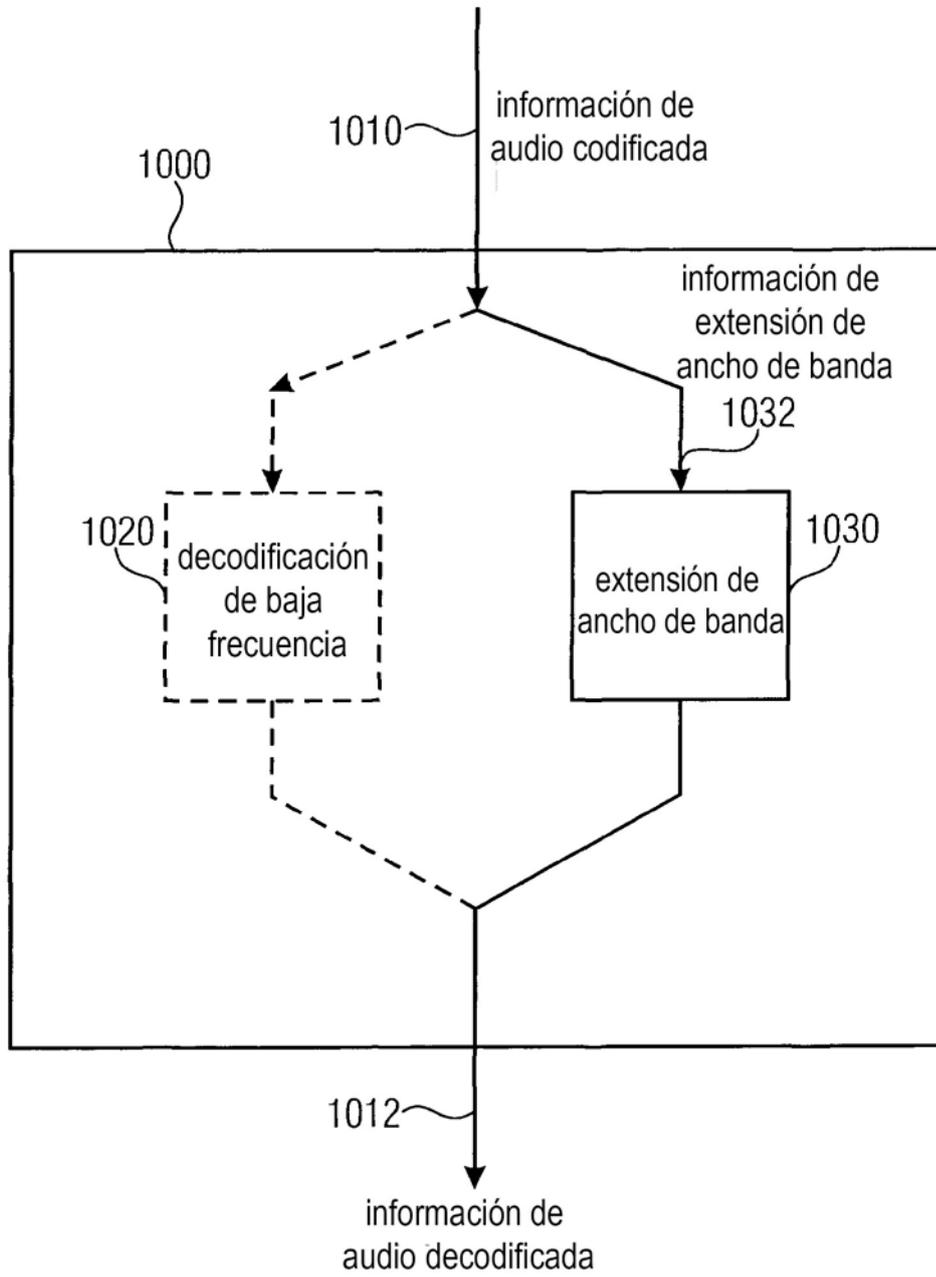


FIGURA 10

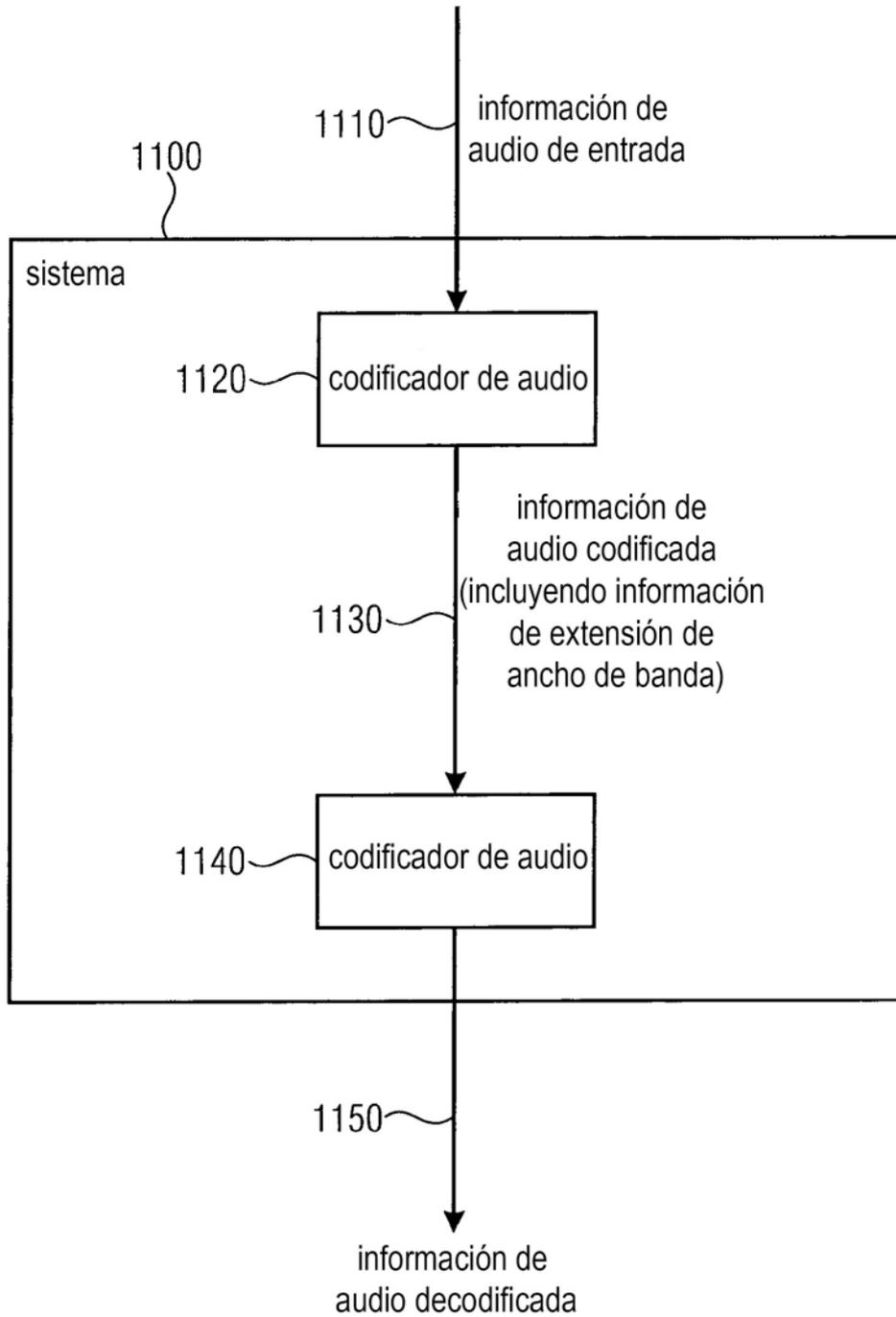


FIGURA 11

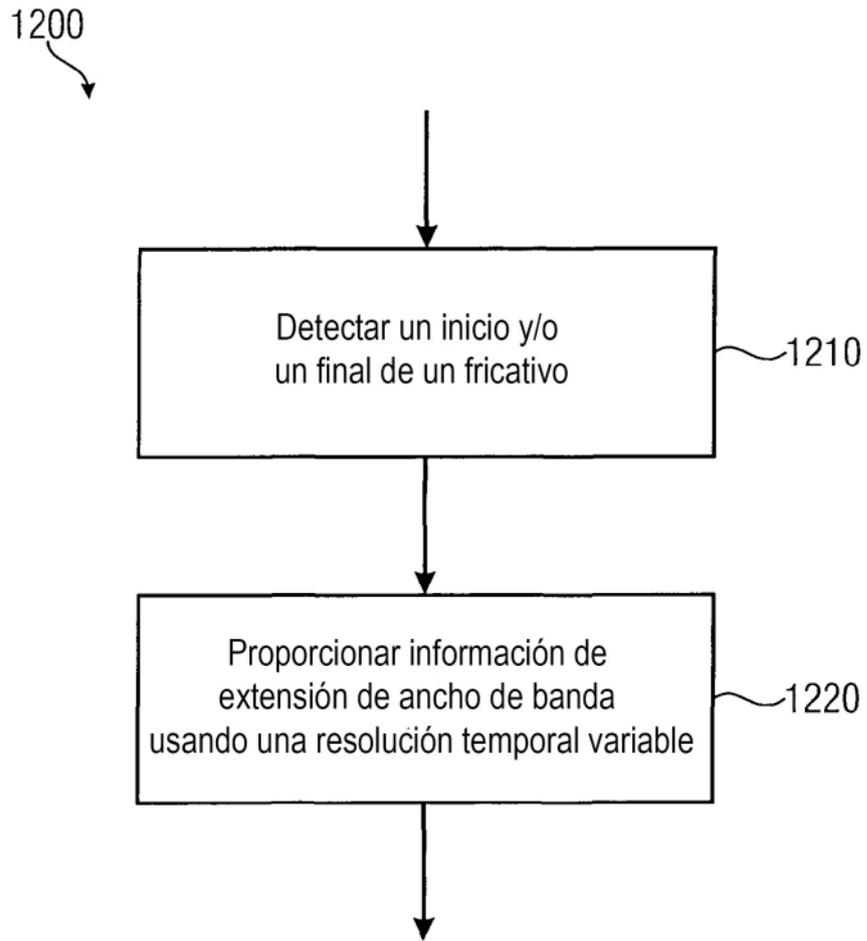


FIGURA 12

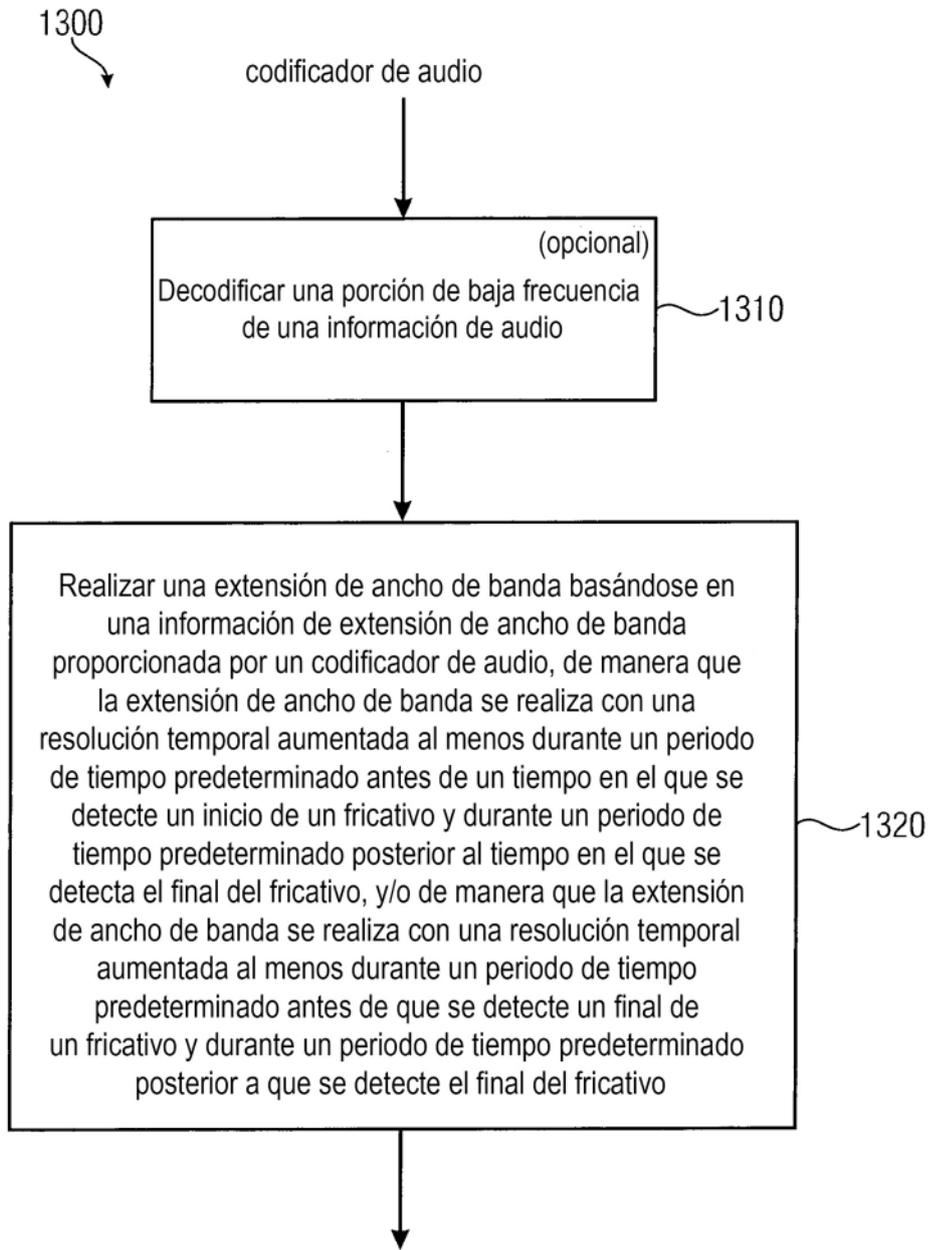


FIGURA 13