

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 614 358**

51 Int. Cl.:

G10L 19/22 (2013.01)

G10L 19/02 (2013.01)

G10L 19/08 (2013.01)

G10L 19/09 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **21.07.2015 PCT/EP2015/066677**

87 Fecha y número de publicación internacional: **07.12.2016 WO2016016053**

96 Fecha de presentación y número de la solicitud europea: **21.07.2015 E 15739590 (6)**

97 Fecha y número de publicación de la concesión europea: **07.12.2016 EP 3000110**

54 Título: **Selección de uno de un primer algoritmo de codificación y un segundo algoritmo de codificación usando reducción de armónicos**

30 Prioridad:

28.07.2014 EP 14178809

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

30.05.2017

73 Titular/es:

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

72 Inventor/es:

**RAVELLI, EMMANUEL;
MULTRUS, MARKUS;
DÖHLA, STEFAN;
GRILL, BERNHARD y
JANDER, MANUEL**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 614 358 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).

Selección de uno de un primer algoritmo de codificación y un segundo algoritmo de codificación usando reducción de armónicos

DESCRIPCIÓN

5

La presente invención se refiere a la codificación de audio, y en particular, a la codificación de audio conmutada, donde, para diferentes porciones de una señal de audio, la señal codificada se genera utilizando diferentes algoritmos de codificación.

10

Los codificadores de audio conmutado que determinan diferentes algoritmos de codificación para diferentes porciones de la señal de audio son conocidos. En términos generales, los codificadores de audio conmutados proporcionan la conmutación entre dos modos diferentes, es decir, algoritmos tales como ACELP (Algebraic Code Excited Linear Prediction, Predicción Lineal Con Excitación por Código Algebraico) y TCX (Transform Coded Excitation, Excitación Codificada por Transformadas).

15

El modo LPD de MPEG USAC (MPEG Unified Speech Audio Coding, Codificación Unificada de Audio del Habla), se basa en dos modos diferentes: ACELP y TCX. El ACELP proporciona una mejor calidad para las señales similares al habla y similares a transitorios. El TCX proporciona una mejor calidad para señales similares a música y similares al ruido. El codificador decide cuál modo utilizar en una base trama a trama. La decisión tomada por el codificador es crítica para la calidad del códec. Una única decisión equivocada puede producir un fuerte artefacto, particularmente a bajas tasas de bits.

20

El enfoque más sencillo para decidir cuál modo utilizar es una selección de modo de bucle cerrado, es decir, llevar a cabo una codificación/decodificación completa de ambos modos, a continuación calcular un criterio de selección (por ejemplo, la SNR segmentaria) para ambos modos basándose en la señal de audio y en las señales de audio codificadas/decodificadas y finalmente elegir un modo basándose en el criterio de selección. Por lo general este enfoque produce una decisión estable y robusta. Sin embargo, también requiere una cantidad significativa de complejidad, cuando ambos modos han de ejecutarse en cada trama.

25

30

Para reducir la complejidad, un enfoque alternativo es la selección en modo de bucle abierto. La selección en bucle abierto consiste en no llevar a cabo una codificación/decodificación completa de ambos modos, sino en cambio, elegir un modo utilizando un criterio de selección calculado con baja complejidad. La complejidad del peor caso se reduce a continuación mediante la complejidad del modo menos complejo (normalmente TCX), menos la complejidad necesaria para calcular el criterio de selección. El ahorro en complejidad es normalmente significativo, lo que hace que este tipo de enfoque sea atractivo cuando la complejidad del peor caso del códec está restringida.

35

La norma AMR-WB+ (definida en International Standard 3GPP TS 26.290 V6.1.0 2004-12) incluye una selección en modo de bucle abierto, utilizada para decidir entre todas las combinaciones de ACELP/TCX20/TCX40/TCX80 en una trama de 80 ms. Está descrito en la Sección 5.2.4 de 3GPP TS 26.290. También se describe en el artículo de conferencia "Low Complex Audio Encoding for Mobile, Multimedia, VTC 2006, Makinen et al." y los documentos US 7.747.430 B2 y US 7.739.120 B2 que se refieren al autor de este artículo de conferencia.

40

En el documento US 7.747.430 B2 divulga una selección en modo de bucle abierto basándose en un análisis de parámetros de predicción a largo plazo. El documento US 7.739.120 B2 divulga una selección en modo de bucle abierto basándose en características de la señal que indican el tipo de contenido de audio en las respectivas secciones de una señal de audio, en el que, si una selección de este tipo no es viable, la selección se basa además en una evaluación estadística llevada a cabo para secciones respectivamente adyacentes.

45

La selección en modo de bucle abierto de AMR-WSB + puede describirse en dos etapas principales. En la primera etapa principal, se calculan varios rasgos sobre la señal de audio, tales como la desviación típica de los niveles de energía, la relación de energía de baja frecuencia/alta frecuencia, energía total, distancia de ISP (immittance spectral pair, par espectral de inmitancia), retardos y ganancias de tono, inclinación espectral. Estos rasgos se utilizan a continuación para tomar una decisión entre ACELP y TCX, utilizándose un único clasificador basado en el umbral. Si se selecciona TCX en la primera etapa principal, entonces la segunda etapa principal decide entre las posibles combinaciones de TCX20/TCX40/TCX80 de una manera en bucle cerrado.

50

55

El documento WO 2012/110448 A1 desvela un enfoque para decidir entre dos algoritmos de codificación que tienen diferentes características basándose en un resultado de detección de transitorios y un resultado de la calidad de una señal de audio. Además, se desvela la aplicación de histéresis, en el que la histéresis se basa en las selecciones realizadas en el pasado, es decir, para las porciones anteriores de la señal de audio.

En el artículo de conferencia "Low Complex Audio Encoding for Mobile, Multimedia, VTC 2006, Makinen et al.", se comparan la selección en modo de bucle cerrado y en modo de bucle abierto de AMR-WB+. Las pruebas de

escucha subjetiva indican que la selección en modo de bucle abierto tiene un rendimiento significativamente peor que la selección en modo de bucle cerrado. Pero también se muestra que la selección en modo de bucle abierto reduce la complejidad del peor caso en un 40 %.

5 El documento WO2014/118136A1 desvela un enfoque para realizar selección de bucle abierto entre TCX y ACELP usando estimaciones de SNR (por segmentos) derivadas para la aproximación respectiva de cada uno de los algoritmos de codificación de TCX y ACELP.

10 Es el objeto de la invención proporcionar un enfoque mejorado que permita la selección entre un primer algoritmo de codificación y un segundo algoritmo de codificación con un buen rendimiento y complejidad reducida.

Este objetivo se consigue mediante un aparato de acuerdo con la reivindicación 1, un método de acuerdo con la reivindicación 14 y un programa de computadora de acuerdo con la reivindicación 15.

15 Algunas formas de realización de la invención están basadas en el reconocimiento que una selección en bucle abierto con un rendimiento mejorado puede implementarse mediante la estimación de una medida de la calidad para cada uno del primer y segundo algoritmos de codificación y seleccionando uno de los algoritmos de codificación basándose en una comparación entre la primera y segunda medidas de calidad. Las medidas de la calidad se estiman, es decir, la señal de audio no se codifica y decodifica realmente para obtener las medidas de calidad. De esta manera, las medidas de calidad pueden obtenerse con una complejidad reducida. El modo de selección puede
20 llevarse a cabo a continuación utilizando las medidas de calidad estimada comparables a una selección en modo de bucle abierto. Además, la invención se basa en el reconocimiento que puede obtenerse una selección de modo mejorada si la estimación de la primera medida de la calidad utiliza una versión filtrada de la porción de la señal de audio, en la que los armónicos están reducidos en comparación con la versión no filtrada de la señal de audio.

25 En algunas formas de realización de la invención, se implementa una selección en modo de bucle abierto donde la SNR segmentaria de ACELP y TCX se estima en primer lugar con baja complejidad. Y a continuación se lleva a cabo la selección de modo utilizando estos valores de SNR estimados, como en una selección de modo de bucle cerrado.

30 Las formas de realización de la invención no emplean un enfoque de rasgos+clasificador clásico como se efectúa en la selección el modo de bucle abierto de AMR-WB+. En cambio, las formas de realización de la invención intentan estimar una medida de la calidad de cada modo y seleccionar el modo que proporcione la mejor calidad.

35 Ahora se describirán formas de realización de la presente invención en mayor detalle con referencia a los dibujos adjuntos, en los que:

La Figura 1 muestra una vista esquemática de una forma de realización de un aparato para seleccionar uno de un primer algoritmo de codificación y un segundo algoritmo de codificación;

40 La Figura 2 muestra una vista esquemática de una forma de realización de un aparato para codificar una señal de audio;

La Figura 3 muestra una vista esquemática de una forma de realización de un aparato para seleccionar uno de un primer algoritmo de codificación y un segundo algoritmo de codificación;

45 La Figura 4a y 4b posibles representaciones de SNR y SNR segmentaria.

50 En la siguiente descripción, se hará referencia a los elementos/etapas similares en los diferentes dibujos mediante los mismos signos de referencia. Cabe observar que en los dibujos los rasgos, tales como las conexiones de las señales o similares, que no sean necesarias para entender la invención, se han omitido.

La Figura 1 muestra un aparato 10 para seleccionar uno de un primer algoritmo de codificación, tal como un algoritmo TCX y un segundo algoritmo de codificación, tal como un algoritmo ACELP, como el codificador para
55 codificar una porción de una señal de audio. El aparato 10 comprende un primer estimador 12 para estimar una SNR o se proporciona una SNR segmentaria de la porción de la señal de audio como primera medida de la calidad para la porción de la señal. La primera medida de la calidad está asociada con el primer algoritmo de codificación. El aparato 10 comprende un filtro 2 configurado para recibir la señal de audio, para reducir la amplitud de los armónicos en la señal de audio y para emitir una versión filtrada de la señal de audio. El filtro 2 puede ser interno en el primer estimador 12, como se muestra en la Figura 1, o puede ser externo con respecto al primer estimador 12. El primer
60 estimador 12 utiliza la versión filtrada de la señal de audio para estimar la primera medida de la calidad. En otras palabras, el primer estimador 12 estima una primera medida de la calidad que la porción de la señal de audio tendría si se codificara y decodificara utilizando el primer algoritmo de codificación, sin realmente codificar y decodificar la porción de la señal de audio utilizando el primer algoritmo de codificación. El aparato 10 comprende un segundo estimador 14 para estimar una segunda medida de la calidad para la porción de la señal. La segunda medida de la

calidad está asociada con el segundo algoritmo de codificación. En otras palabras, el segundo estimador 14 estima la segunda medida de la calidad que tendría la porción de la señal de audio si se codificara y decodificara utilizando el segundo algoritmo de codificación, sin realmente codificar y decodificar la porción de la señal de audio utilizando el segundo algoritmo de codificación. Además, el aparato 10 comprende un controlador 16 para seleccionar el primer algoritmo de codificación o el segundo algoritmo de codificación basándose en una comparación entre la primera medida de la calidad y la segunda medida de la calidad. El controlador puede comprender una salida 18 que indica el algoritmo de codificación seleccionado.

En la siguiente memoria descriptiva, el primer estimador utiliza la versión filtrada de la señal de audio, es decir, la versión filtrada de la porción de la señal de audio al estimar la primera medida de la calidad si se proporciona el filtro 2 configurado para reducir la amplitud de los armónicos y no está desactivado, incluso si no se indica explícitamente.

En una forma de realización, la primera característica asociada con el primer algoritmo de codificación es más adecuada para señales similares a música y similares a ruido, y la segunda característica de codificación asociada con el segundo algoritmo de codificación es más adecuada para señales similares al habla y similares a transitorios. En algunas formas de realización de la invención, el primer algoritmo de codificación es un algoritmo de codificación de audio, tal como un algoritmo de codificación de transformadas, por ejemplo, un algoritmo de codificación MDCT (modified discrete cosine transform, transformada de coseno discreta modificada), tal como un algoritmo de codificación TCX (transform coding excitation, excitación codificada por transformadas). Otros algoritmos de codificación de transformadas pueden estar basados en una transformada de FFT o en cualquier otra transformada o banco de filtros. En algunas formas de realización de la invención, el segundo algoritmo de codificación es un algoritmo de codificación del habla, tal como un algoritmo de codificación CELP (code excited linear prediction, predicción lineal con excitación por código), tal como un algoritmo de codificación ACELP (algebraic code excited linear prediction, predicción lineal con excitación por código algebraico).

En algunas formas de realización, la medida de la calidad representa una medida de calidad perceptual. Puede calcularse un valor único que es una estimación de la calidad subjetiva del primer algoritmo de codificación y un valor único que es una estimación de la calidad subjetiva del segundo algoritmo de codificación. El algoritmo de codificación que proporciona la mejor calidad subjetiva estimada puede elegirse basándose solamente en la comparación de estos dos valores. Esto es diferente de lo que se hace en la norma AMR-WB+ donde se calculan muchos rasgos representativos de diferentes características de la señal, y a continuación se aplica un clasificador para decidir cuál algoritmo elegir.

En algunas formas de realización, la respectiva medida de la calidad se estima basándose en una porción de la señal de audio ponderada, es decir, una versión ponderada de la señal de audio. En algunas formas de realización, la señal de audio ponderada puede definirse como una señal de audio filtrada mediante una función de ponderación, donde la función de ponderación es un filtro de LPC ponderado de $A(z/g)$ siendo $A(z)$ un filtro de LPC y g una ponderación entre 0 y 1 tal como 0,68. Se ha comprobado que de esta manera pueden obtenerse buenas medidas de la calidad perceptual. Obsérvese que el filtro de LPC de $A(z)$ y el de filtro de LPC de $A(z/g)$ ponderado se determinan en una etapa de pre-procesamiento y que también se utilizan en ambos algoritmos de codificación. En otras formas de realización, la función de ponderación puede ser un filtro lineal, un filtro FIR o un filtro de predicción lineal.

En algunas formas de realización, la medida de la calidad es la SNR segmentaria (relación de señal a ruido) en el dominio de las señales ponderadas. Se ha comprobado que la SNR segmentaria en el dominio de las señales ponderadas representa una buena medida de la calidad perceptual y, por lo tanto, puede utilizarse como la medida de la calidad de una manera beneficiosa. Esta es también la medida de la calidad utilizada en los algoritmos de codificación tanto de ACELP como de TCX para estimar los parámetros de codificación.

Otra medida de la calidad puede ser la SNR en el dominio de las señales ponderadas. Otras medidas de la calidad pueden ser la SNR segmentaria, la SNR de la correspondiente porción de la señal de audio en el dominio de las señales no ponderadas, es decir, no filtradas por los coeficientes de LPC (ponderados).

En términos generales, la SNR compara las señales de audio originales y procesadas (tales como las señales del habla), muestra a muestra. Su objetivo es el de medir la distorsión de los codificadores de forma de onda que reproducen la forma de onda introducida. La SNR puede calcularse como se muestra en la Figura 4a, donde $x(i)$ e $y(i)$ son las muestras originales y procesadas indexadas por i y N es número total de muestras. La SNR segmentaria, en lugar de funcionar sobre la totalidad de la señal, calcula el promedio del valor de los valores de SNR de los segmentos cortos, tal como de 1 a 10 ms, tal como 5 ms. La SNR puede calcularse como se muestra en la Figura 4b, donde N y M son la longitud de segmento y el número de segmentos, respectivamente.

En algunas formas de realización de la invención, la porción de la señal de audio representa una trama de la señal de audio que se obtiene mediante la puesta en ventanas de la señal de audio y se lleva a cabo la selección de un algoritmo de codificación adecuado para una pluralidad de tramas sucesivas obtenidas mediante la aplicación de la

puesta en ventanas de una señal de audio. En la siguiente memoria descriptiva, y en relación con la señal de audio, los términos “porción” y “trama” se utilizan de una manera intercambiable. En algunas formas de realización, cada trama se divide en subtramas y la SNR segmentaria se estima para cada trama calculando la SNR para cada subtrama, convertida en dB y calculándose el promedio de las SNR de las subtramas en dB.

5 Por lo tanto, en algunas formas de realización, no se estima la SNR (segmentaria) entre la señal de audio introducida y la señal de audio decodificada, sino que se estima la SNR (segmentaria) entre la señal de audio introducida ponderada y la señal de audio decodificada ponderada. En lo que se refiere a esta SNR (segmentaria),
10 puede hacerse referencia al capítulo 5.2.3 de la norma AMR-WB+ (International Standard 3GPP TS 26.290 V6.1.0 2004-12).

En algunas formas de realización de la invención, la respectiva medida de la calidad se estima basándose en la energía de una porción de la señal de audio ponderada y basándose en una distorsión estimada que se introduce cuando se codifica la porción de la señal mediante el respectivo algoritmo, en el que el primer y segundo
15 estimadores están configurados para determinar las distorsiones estimadas en función de la energía de una señal de audio ponderada.

En algunas formas de realización de la invención, se determina una distorsión estimada del cuantificador introducida por un cuantificador utilizado en el primer algoritmo de codificación cuando se cuantifica la porción de la señal de audio, y se determina la primera medida de la calidad basándose en la energía de la porción de la señal de audio ponderada y de la distorsión estimada del cuantificador. En tales formas de realización, puede estimarse una ganancia global para la porción de la señal de audio de manera tal que la porción de la señal de audio produciría una tasa de bits objetivo dada cuando se codifica con un cuantificador y con un codificador de entropía utilizado en el primer algoritmo de codificación, en el que se determina la distorsión estimada del cuantificador basándose en la ganancia global estimada. En tales formas de realización, la distorsión estimada del cuantificador puede determinarse basándose en una potencia de la ganancia estimada. Cuando el cuantificador utilizado en el primer algoritmo de codificación es un cuantificador uniforme escalar, el primer estimador puede configurarse para determinar la distorsión estimada del cuantificador utilizando la fórmula $D = G \cdot G / 12$, en la que D es la distorsión estimada del cuantificador y G es la ganancia global estimada. En el caso en que el primer algoritmo de codificación utilice otro cuantificador, la distorsión del cuantificador puede determinarse a partir de la ganancia global, de una manera diferente.

Los inventores han reconocido que una medida de la calidad, tal como una SNR segmentaria, que se obtendría cuando se codifica y decodifica la porción de la señal de audio utilizando el primer algoritmo de codificación, tal como el algoritmo TCX, puede estimarse de una manera apropiada utilizando los rasgos anteriores en cualquier combinación de los mismos.

En algunas formas de realización de la invención, la primera medida de la calidad es una SNR segmentaria y la SNR segmentaria se estima calculando una SNR estimada asociada con cada una de la pluralidad de subporciones de la porción de la señal de audio basándose en una energía de la correspondiente subporción de la señal de audio ponderada y de la distorsión estimada del cuantificador y calculando un promedio de las SNR asociadas con las subporciones de la porción de la señal de audio ponderada para obtener la SNR segmentaria estimada para la porción de la señal de audio ponderada.

En algunas formas de realización de la invención, se determina una distorsión estimada de un libro de código adaptativo introducido por un libro de código adaptativo utilizado en el segundo algoritmo de codificación cuando se utiliza el libro de código adaptativo para codificar la porción de la señal de audio, y se estima la segunda medida de la calidad basándose en una energía de la porción de la señal de audio ponderada y de la distorsión estimada del libro de código adaptativo.

En tales formas de realización, para cada una de la pluralidad de subporciones de la porción de la señal de audio, el libro de código adaptativo puede aproximarse basándose en una versión de la subporción de la señal de audio ponderada desplazada hacia el pasado mediante un retardo de paso determinado en una etapa de pre-procesamiento, puede estimarse una ganancia del libro de código adaptativo de manera tal que se minimiza un error entre la subporción de la porción de la señal de audio ponderada y el libro de código adaptativo aproximado, y puede determinarse una distorsión estimada del libro de código adaptativo basándose en la energía de un error entre la subporción de la porción de la señal de audio ponderada y el libro de código adaptativo aproximado escalado por la ganancia del libro de código adaptativo.

En algunas formas de realización de la invención, la distorsión estimada del libro de código adaptativo determinada para cada subporción de la porción de la señal de audio puede reducirse en un factor constante para tener en cuenta una reducción de la distorsión que se consigue mediante un libro de código innovador en el segundo algoritmo de codificación.

En algunas formas de realización de la invención, la segunda medida de la calidad es una SNR segmentaria y la SNR segmentaria se estima calculando una SNR estimada asociada con cada subporción basándose en la energía de la correspondiente subporción de la señal de audio ponderada y de la distorsión estimada del libro de código adaptativo y calculando un promedio de las SNR asociadas con las subporciones para obtener la SNR segmentaria estimada.

En algunas formas de realización de la invención, el libro de código adaptativo se aproxima basándose en una versión de la porción de la señal de audio ponderada desplazada hacia el pasado mediante un retardo de paso determinado en una etapa de pre-procesamiento, se estima una ganancia del libro de código adaptativo de manera tal que se minimiza un error entre la porción de la señal de audio ponderada y el libro de código adaptativo aproximado, y se determina la distorsión estimada del libro de código adaptativo basándose en la energía entre la porción de la señal de audio ponderada y el libro de código adaptativo aproximado escalado por la ganancia del libro de código adaptativo. De esta manera, puede determinarse la distorsión del libro de código adaptativo estimado con una baja complejidad.

Los inventores han reconocido que la medida de la calidad, tal como una SNR segmentaria, que se obtendría cuando se codifica y decodifica la porción de la señal de audio utilizando el segundo algoritmo de codificación, tal como un algoritmo ACELP, puede estimarse de una manera apropiada utilizando los rasgos anteriores en cualquier combinación de los mismos.

En algunas formas de realización de la invención, se utiliza un mecanismo de histéresis al comparar las medidas de calidad estimadas. Esto puede hacer que la decisión de qué algoritmo utilizar sea más estable. El mecanismo de histéresis puede depender de las medidas de calidad estimadas (tales como las diferencias entre ellas) y de otros parámetros, tales como las estadísticas relacionadas con decisiones anteriores, la cantidad de tramas temporalmente estacionarias, los transitorios en las tramas. En lo que a dichos mecanismos de histéresis se refiere, se puede hacer referencia al documento WO 2012/110448 A1, por ejemplo.

En algunas formas de realización de la invención, un codificador para codificar una señal de audio comprende el aparato 10, una etapa para llevar a cabo el primer algoritmo de codificación y una etapa para llevar a cabo el segundo algoritmo de codificación, en el que el codificador está configurado para codificar la porción de la señal de audio utilizando el primer algoritmo de codificación o el segundo algoritmo de codificación en función de la selección por el controlador 16. En algunas formas de realización de la invención, un sistema para codificar y decodificar comprende el codificador y un decodificador configurado para recibir la versión codificada de la porción de la señal de audio y una indicación del algoritmo utilizado para codificar una porción de la señal de audio y para decodificar la versión codificada de la porción de la señal de audio utilizando el algoritmo indicado.

Un algoritmo de selección de modo de bucle abierto de este tipo como se muestra en la Figura 1 y se ha descrito anteriormente (con la excepción del filtro 2) se describe en una solicitud anterior PCT/EP2014/051557. Este algoritmo se utilizar para tomar una decisión entre dos modos, tales como ACELP y TCX, en una base trama a trama. La selección puede estar basada en una estimación de la SNR segmentaria tanto de ACELP como de TCX. Se selecciona el modo con la SNR segmentaria estimada más elevada. Opcionalmente, puede utilizarse un mecanismo de histéresis para proporcionar una selección más robusta. La SNR segmentaria de ACELP puede estimarse utilizando una aproximación de la distorsión del libro de código adaptativo y una aproximación de la distorsión de libro de código innovador. El libro de código adaptativo puede aproximarse en el dominio de las señales ponderadas utilizando un retardo de paso estimado mediante un algoritmo de análisis de paso. La distorsión puede calcularse en el dominio de las señales ponderadas suponiendo una ganancia óptima. La distorsión puede reducirse a continuación mediante un factor constante, aproximándose a la distorsión del libro de código innovador. La SNR segmentaria de TCX puede estimarse utilizando una versión simplificada del codificador TCX real. La señal introducida puede transformarse en primer lugar con una MDCT, y a continuación conformarse usando un filtro LPC ponderado. Finalmente, puede estimarse la distorsión en el dominio del MDCT ponderado, utilizándose una ganancia global y un estimador de la ganancia global.

Resultó que este algoritmo de selección de modo de bucle abierto descrito en la solicitud anterior proporciona la decisión esperada en la mayoría de las veces, seleccionándose ACELP en las señales similares al habla y similares a los transitorios y seleccionándose TCX en las señales similares a música y similares a ruido. Sin embargo, los inventores han reconocido que puede suceder que a veces se seleccione ACELP en algunas señales de música de armónicos. En tales señales, el libro de código adaptativo tiene por lo general una elevada ganancia de predicción, debido a la elevada predictibilidad de las señales de armónicos, produciéndose una baja distorsión y a continuación una SNR segmentaria más elevada que en TCX. Sin embargo, el TCX tiene mejor sonido en la mayoría de las señales de música de armónicos, por lo que en estos casos se preferiría TCX.

Por lo tanto, la presente invención sugiere llevar a cabo la estimación de la SNR o de la SNR segmentaria como la primera medida de la calidad utilizando una versión de la señal introducida, que se filtra para reducir los armónicos de la misma. De esta manera puede obtenerse una selección de modo mejorada sobre las señales música de

armónicos.

En términos generales podría utilizarse cualquier filtro adecuado para reducir los armónicos. En algunas formas de realización de la invención, el filtro es un filtro de predicción a largo plazo. Un ejemplo simple de un filtro de predicción a largo plazo es:

$$F(z) = 1 - g \cdot z^{-T}$$

donde los parámetros de filtro son la ganancia "g" y el retardo de paso "T", que se determinan a partir de la señal de audio.

Algunas formas de realización de la invención están basadas en un filtro de predicción a largo plazo que se aplica a la señal de audio antes del análisis de MDCT en la estimación de la SNR segmentaria de TCX. El filtro de predicción a largo plazo reduce la amplitud de los armónicos en la señal introducida antes del análisis de MDCT. La consecuencia es que la distorsión en el dominio de MDCT se reduce, se incrementa la SNR segmentaria estimada de TCX y finalmente se selecciona el TCX con mayor frecuencia en las señales de música de armónicos.

En algunas formas de realización de la invención, una función de transferencia del filtro de predicción a largo plazo comprende una parte de número entero de un retardo de paso y un filtro de múltiples derivaciones en función de una parte fraccional del retardo de paso. Esto permite una implementación eficiente ya que la parte de número entero se utiliza únicamente en la estructura de la tasa de muestreo normal ($z^{-T_{int}}$). Al mismo tiempo, puede conseguirse elevada exactitud debido al uso de la parte fraccional en el filtro de múltiples derivaciones. Considerando la parte fraccional en el filtro de múltiples derivaciones puede conseguirse una eliminación de la energía de los armónicos mientras se evita una eliminación de la energía de las porciones cercanas a los armónicos.

En algunas formas de realización de la invención, el filtro de predicción a largo plazo se describe como sigue:

$$P(z) = 1 - \beta g B(z, T_{fr}) z^{-T_{int}}$$

en el que T_{int} y T_{fr} son la parte de número entero y fraccional de un retardo de paso, g es una ganancia, β es una ponderación y $B(z, T_{fr})$ es un filtro FIR de paso bajo cuyos coeficientes dependen de la parte fraccional del retardo de paso. A continuación se proporcionan detalles adicionales acerca de formas de realización de un filtro de predicción a largo plazo de este tipo.

El retardo de paso y la ganancia pueden estimarse en una base trama a trama.

El filtro de predicción puede estar desactivado (ganancia = 0) basándose en una combinación de una o más medidas de armonicidad (por ejemplo, correlación normalizada o ganancia de predicción) y/o una o más medidas de estructura temporal (por ejemplo, una medida de uniformidad temporal o el cambio de energía).

El filtro puede aplicarse a la señal de audio introducida en una base de trama a trama. Si los parámetros del filtro cambian de una trama al siguiente, puede introducirse una discontinuidad en el límite entre dos tramas. En algunas formas de realización, el aparato comprende además una unidad para eliminar las discontinuidades en la señal de audio causadas por el filtro. Para eliminar las posibles discontinuidades, puede utilizarse cualquier técnica, tales como las técnicas comparables a las descritas en los siguientes documentos: US5012517, EP0732687A2, US5999899A o US7353168B2. A continuación se describe otra técnica para eliminar posibles discontinuidades.

Antes de describir una forma de realización del primer estimador 12 y del segundo estimador 14 en detalle con referencia a la Figura 3, se describe una forma de realización del codificador 20 con referencia a la Figura 2.

El codificador 20 comprende el primer estimador 12, el segundo estimador 14, el controlador 16, una unidad de pre-procesamiento 22, un conmutador 24, una primera etapa del codificador 26 configurada para llevar a cabo un algoritmo de TCX, una segunda etapa del codificador 28 configurada para llevar a cabo un algoritmo ACELP, y una interfaz de salida 30. La unidad de pre-procesamiento 22 puede ser parte de un codificador USAC común y puede estar configurada para emitir los coeficientes de LPC, los coeficientes de LPC ponderados, la señal de audio ponderada y un conjunto de retardos de paso. Cabe observar que todos estos parámetros se utilizan en ambos algoritmos de codificación, es decir el algoritmo de TCX y el algoritmo ACELP. Por lo tanto, tales parámetros no han de calcularse adicionalmente para la decisión de modo de bucle abierto. La ventaja de la utilización de parámetros ya calculados en la decisión de modo de bucle abierto consiste en ahorro en complejidad.

Como se muestra en la Figura 2, el aparato comprende el filtro de reducción de los armónicos 2. El aparato comprende además una unidad de desactivación opcional 4 para desactivar el filtro de reducción de los armónicos 2 basándose en una combinación de una o más medidas de armonicidad (por ejemplo, correlación normalizada o ganancia de predicción) y/o de una o más medidas de estructura temporal (por ejemplo, medida de uniformidad

temporal o cambio de energía). El aparato comprende una unidad de eliminación de discontinuidad opcional 6 para eliminar las discontinuidades de la versión filtrada de la señal de audio. Además, el aparato comprende opcionalmente una unidad 8 para estimar los parámetros de filtro del filtro de reducción de armónicos 2. En la Figura 2, estos componentes (2,4, 6 y 8) se muestran como que son parte del primer estimador 12. Se da por entendido que estos componentes pueden implementarse externos o separados desde el primer estimador y que pueden configurarse para proporcionar la versión filtrada de la señal de audio al primer estimador.

Se proporciona una señal de audio introducida 40 en una línea de entrada. La señal de audio introducida 40 se aplica al primer estimador 12, a la unidad de pre-procesamiento 22 y a ambas etapas de codificador 26, 28. En el primer estimador 12, la señal de audio introducida 40 se aplica al filtro 2 y la versión filtrada de la señal de audio introducida se utiliza para estimar la primera medida de calidad. En el caso en que el filtro esté desactivado por la unidad de desactivación 4, la señal de audio introducida 40 se utiliza para estimar la primera medida de la calidad, en lugar de la versión filtrada de la señal de audio introducida. La unidad de pre-procesamiento 22 procesa la señal de audio introducida de una manera convencional para derivar coeficientes de LPC y coeficientes de LPC ponderados 42 y para filtrar la señal de audio 40 con los coeficientes de LPC ponderados 42 para obtener la señal de audio ponderada 44. La unidad de pre-procesamiento 22 emite los coeficientes de LPC ponderados 42, la señal de audio ponderada 44 y un conjunto de retardos de paso 48. Como entenderán los expertos en la materia, los coeficientes de LPC ponderados 42 y la señal de audio ponderada 44 pueden estar segmentados en tramas o subtramas. La segmentación puede obtenerse mediante la generación de ventanas a la señal de audio de una manera apropiada.

En algunas formas de realización alternativas, es posible proporcionar un preprocesador que está configurado para generar coeficientes de LPC ponderados y una señal de audio ponderada basada en la versión filtrada de la señal de audio. Los coeficientes de LPC ponderados y la señal de audio ponderada, que están basados en la versión filtrada de la señal de audio, son seguidamente aplicados al primer estimador para estimar la primera medida de la calidad, en lugar de los coeficientes de LPC ponderados 42 y de la señal de audio ponderada 44.

En algunas formas de realización de la invención, es posible utilizar coeficientes de LPC cuantificados o coeficientes de LPC ponderados cuantificados. Por lo tanto, debe entenderse que la expresión "coeficientes de LPC" también tiene por objeto abarcar la expresión "coeficientes de LPC cuantificados" y que la expresión "coeficientes de LPC ponderados" también tiene por objeto abarcar "coeficientes de LPC ponderados cuantificados". En este aspecto, vale la pena observar que el algoritmo de TCX de USAT utiliza coeficientes de LPC ponderados cuantificados para configurar el espectro de MCDT.

El primer estimador 12 recibe la señal de audio 40, los coeficientes de LPC ponderados 42 y la señal de audio ponderada 44, estima la primera medida de calidad 46 basándose los mismos y emite la primera medida de calidad al controlador 16. El segundo estimador 16 recibe la señal de audio ponderada 44 y el conjunto de retardos de paso 48, estima la segunda medida de la calidad 50 basándose en los mismos y emite la segunda medida de calidad 50 al controlador 16. Como es conocido para los expertos en la materia, los coeficientes de LPC ponderados 42, la señal de audio ponderada 44, y el conjunto de retardos de paso 48 ya están calculados en un módulo previo (es decir, la unidad de pre-procesamiento 22) y, por lo tanto, están disponibles sin coste.

El controlador toma una decisión de seleccionar cualquiera del algoritmo de TCX o el algoritmo ACELP basándose en una comparación de las medidas de calidad recibidas. Como se ha indicado anteriormente, el controlador puede utilizar un mecanismo de histéresis para decidir qué algoritmo usarse. La selección de la primera etapa de codificador 26 o de la segunda etapa de codificador 28 se muestra esquemáticamente en la Figura 2 mediante el conmutador 24 que se controla por una señal de control 52 emitida por el controlador 16. La señal de control 52 indica si ha de utilizarse la primera etapa de codificador 26 o la segunda etapa de codificador 28. Basándose en la señal de control 52, las señales requeridas esquemáticamente indicadas mediante la flecha 54 en la Figura 2 y que incluye al menos los coeficientes de LPC, los coeficientes de LPC ponderados, la señal de audio, la señal de audio ponderada, el conjunto de retardos de paso se aplican a cualquiera de la primera etapa de codificador 26 o a la segunda etapa de codificador 28. La etapa de codificador seleccionada aplica el algoritmo de codificación asociado y emite la representación codificada 56 o 58 a la interfaz de salida 30. La interfaz de salida 30 puede estar configurada para emitir una señal de audio codificada 60 que puede comprender entre otros datos la representación codificada 56 o 58, los coeficientes de LPC o los coeficientes de LPC ponderados, los parámetros para el algoritmo de codificación seleccionado e información acerca del algoritmo de codificación seleccionado.

Algunas formas de realización específicas para estimar la primera y segunda medidas de calidad, en el que la primera y segunda medidas de calidad son SNR segmentarias en el dominio de las señales ponderadas se describen ahora con referencia la Figura 3. La Figura 3 muestra el primer estimador 12 y el segundo estimador 14 y las funcionalidades de los mismos en forma de diagramas de flujo que muestran la respectiva estimación etapa a etapa.

Estimación de SNR segmentaria de TCX

El primer estimador (de TCX) recibe la señal de audio 40 (señal introducida), los coeficientes de LPC ponderados 42 y las señales de audio ponderadas 44, como entradas. Se genera la versión filtrada de la señal de audio 40, etapa 98. En la versión filtrada de la señal de audio 40 los armónicos se reducen o son suprimidos.

5 La señal de audio 40 puede ser analizada para determinar una o más medidas de armonicidad (por ejemplo, la correlación normalizada o la ganancia de predicción) y una o más medidas de estructura temporal (por ejemplo, la medida de uniformidad temporal o el cambio de energía). Basándose en una de estas medidas o de una combinación de estas medidas, el filtro 2, y por lo tanto la filtración 98 pueden desactivarse. Si se desactiva la filtración 98, la estimación de la primera medida de calidad se lleva a cabo utilizando la señal de audio 40 en lugar de la versión filtrada de la misma.

15 En algunas formas de realización de la invención, una etapa de la eliminación de discontinuidades (no se muestra en la Figura 3) puede seguir a la filtración 98 para eliminar las discontinuidades en la señal de audio, que pueden resultar de la filtración 98.

20 En la etapa 100, la versión filtrada de la señal de audio 40 se pone en ventanas. La puesta en ventanas puede tener lugar con una ventana de seno de bajo solapamiento de 10 ms. Cuando la trama anterior es ACELP; el tamaño del bloque puede incrementarse en 5 ms, el lado izquierdo de la ventana puede ser rectangular, y la respuesta de impulso cero de la ventana del filtro de síntesis de ACELP puede eliminarse de la señal introducida puesta en ventanas. Esto es similar a lo que se hace en el algoritmo de TCX. Una trama de la versión filtrada de la señal de audio 40, que representa una porción de la señal de audio, se emite desde la etapa 100.

25 En la etapa 102, la señal de audio puesta en ventanas, es decir la trama resultante, se transforma con una MDCT (transformada de coseno discreta modificada). En la etapa 104 se lleva a cabo conformación de espectro conformando el espectro de MDCT con los coeficientes de LPC ponderados.

30 En la etapa 106, se estima una ganancia global G de manera tal que el espectro ponderado cuantificado con la ganancia G produciría un objetivo R dado, cuando se codifica con un codificador de entropía, por ejemplo un codificador aritmético. La expresión "ganancia global" se utiliza ya que se determina una ganancia para la totalidad de la trama.

35 Un ejemplo de una implementación de la estimación de la ganancia global se explica ahora. Cabe observar que esta estimación de la ganancia global es apropiada para formas de realización en las que el algoritmo de codificación de TCX utiliza un cuantificador escalar con un codificador aritmético. Un cuantificador escalar de este tipo con un codificador aritmético se supone en la norma MPEG USAC.

Inicialización

40 En primer lugar, las variables utilizadas en la estimación de ganancias se inicializan como sigue:

1. Establecer $en[i] = 9,0 + 10,0 \cdot \log_{10}(c[4 \cdot i + 0] + c[4 \cdot i + 1] + c[4 \cdot i + 2] + c[4 \cdot i + 3])$, donde $0 \leq i < L/4$, $c[]$ es el vector de coeficientes para cuantificar y L es la longitud de $c[]$.

45 2. Establecer $fac = 128$, desplazamiento = fac y objetivo = cualquier valor (por ejemplo, 1000)

Iteración

50 A continuación se lleva a cabo el siguiente bloque de operaciones NITER veces (por ejemplo, aquí NITER = 10).

1. $fac = fac/2$
2. $desplazamiento = desplazamiento - fac$
3. $ener = 0$
4. para cada i donde $0 \leq i < L/4$ hacer lo siguiente:

55 si $en[i] - desplazamiento > 3,0$, entonces $ener = ener + en[i] - desplazamiento$

5. si $ener > target$, entonces $desplazamiento = desplazamiento + fac$

60 El resultado de la iteración es el valor de desplazamiento. Después de la iteración, la ganancia global se estima como $G = 10^{(offset/20)}$.

La manera específica en la que se estima la ganancia global puede variar en función del codificador y del codificador de entropía utilizados. En la norma MPEG USAC, se supone un cuantificador escalar con un codificador aritmético.

Otros enfoques de TCX pueden utilizar un cuantificador diferente, y se entiende por los expertos en la materia cómo estimar la ganancia global para tales cuantificadores diferentes. Por ejemplo, la norma AMR-WB+ supone que se utiliza un cuantificador de malla RE8. Para un cuantificador de este tipo, la estimación de la ganancia global se estimaría como se describe en el capítulo 5.3.5.7 en la página 34 de 3GPP TS 26.290 V6.1.0 2004-12, en el que se supone una tasa de bits objetivo fija.

Después de haber estimado la ganancia global en la etapa 106, la estimación de la distorsión tiene lugar en la etapa 108. Para ser más específicos, la distorsión del cuantificador se aproxima basándose en la ganancia global estimada. En la presente forma de realización se supone que se utiliza un cuantificador escalar uniforme. Por lo tanto, la distorsión del cuantificador se determina mediante la fórmula sencilla $D = G^2/12$, en la que D representa la distorsión determinada del cuantificador y G representa la ganancia global estimada. Esto corresponde a la aproximación de elevada velocidad de una distorsión del cuantificador uniforme escalar.

Basándose en la distorsión determinada del cuantificador, se lleva a cabo el cálculo de la SNR segmentaria en la etapa 110. La SNR en cada subtrama de la trama se calcula como la relación de la energía ponderada de la señal de audio y la distorsión D que se supone que es constante en las subtramas. Por ejemplo, la trama se divide en cuatro subtramas consecutivas (véase la Figura 4). La SNR segmentaria es entonces el promedio de las SNR de las cuatro subtramas y puede indicarse en dB.

Este enfoque permite la estimación de la primera SNR segmentaria que se obtendría cuando realmente se codifica y decodifica la trama objeto utilizando el algoritmo de TCX, sin embargo sin tener realmente que codificar y decodificar la señal de audio, y por lo tanto, con una complejidad fuertemente reducida y un tiempo de cálculo reducido.

Estimación de la SNR segmentaria de ACELP

El segundo estimador 14 recibe la señal de audio ponderada 44 y el conjunto de retardos de paso 48 que ya se han calculado en la unidad de pre-procesamiento 22.

Como se muestra en la etapa 112, en cada subtrama, el libro de código adaptativo se aproxima utilizando de manera simple la señal de audio ponderada y el retardo de paso T. El libro de código adaptativo se aproxima mediante:

$$xw(n-T), n = 0, \dots, N$$

donde xw es la señal de audio ponderada, T es el retardo de paso de la correspondiente subtrama y N es la longitud del subtrama. Por lo tanto, el libro de código adaptativo se aproxima utilizando una versión del subtrama desplazada al pasado en T. Por lo tanto, en algunas formas de realización de la invención, el libro de código adaptativo se aproxima de una manera muy sencilla.

En la etapa 114, se determina una ganancia de libro de código adaptativo para cada subtrama. Para ser más específicos, en cada subtrama, se estima la ganancia de libro de código G de manera tal que minimiza el error entre la señal de audio ponderada y el libro de código adaptativo aproximado. Esto puede efectuarse simplemente comparando las diferencias entre ambas señales para cada muestra y encontrando una ganancia tal que la suma de estas diferencias sea mínima.

En la etapa 116, se determina la distorsión del libro de código adaptativo para cada subtrama. En cada subtrama, la distorsión D introducida por el libro de código adaptativo es simplemente la energía del error entre la señal de audio ponderada y el libro de código adaptativo aproximado escalado por la ganancia G.

Las distorsiones determinadas en la etapa 116 pueden ajustarse en una etapa opcional 118 para tener en cuenta el libro de código innovador. La distorsión del libro de código innovador utilizado en los algoritmos ACELP puede estimarse simplemente como un valor constante. En la forma de realización descrita de la invención, se supone simplemente que el libro de código innovador reduce la distorsión D en un factor constante. Por lo tanto, las distorsiones obtenidas en la etapa 116 para cada subtrama pueden multiplicarse en la etapa 118 por un factor constante, tal como un factor constante en el orden de 0 a 1, tal como 0,055.

En la etapa 120 tiene lugar el cálculo de la SNR segmentaria. En cada subtrama, la SNR se calcula como la relación entre la energía de la señal de audio ponderada y la distorsión D. La SNR segmentaria es entonces la media de las SNR de las cuatro subtramas y puede indicarse en dB.

Este enfoque permite la estimación de la segunda SNR que se obtendría cuando realmente se codifica y decodifica la trama objeto utilizando el algoritmo ACELP, sin embargo sin tener que realmente codificar y decodificar la señal de audio, y, por lo tanto con una complejidad considerablemente reducida y un tiempo de cálculo reducido.

El primer y segundo estimadores 12 y 14 emiten las SNR segmentarias estimadas 46, 50 hacia el controlador 16 y el

controlador 16 toma una decisión acerca de qué algoritmo debe utilizarse para la porción asociada de la señal de audio basándose en las SNR segmentarias estimadas 46, 50. El controlador puede opcionalmente utilizar un mecanismo de histéresis para hacer que la decisión sea más estable. Por ejemplo, puede utilizarse el mismo mecanismo de histéresis que en la decisión de bucle cerrado, con parámetros de ajuste ligeramente diferentes. Un mecanismo de histéresis de este tipo puede calcular un valor "dsnr" que puede depender de las SNR segmentarias estimadas (tales como la diferencia entre ellas) y otros parámetros, tales como estadísticas acerca de decisiones anteriores, el número de tramas temporalmente estacionarias y los transitorios en las tramas.

Sin un mecanismo de histéresis, el controlador puede seleccionar el algoritmo de codificación que tenga la SNR estimada más elevada, es decir, se selecciona ACELP si la segunda SNR estimada es menos elevada que la primera SNR estimada, y se selecciona TCX si la primera SNR estimada es superior a la segunda SNR estimada. Con un mecanismo de histéresis, el controlador puede seleccionar el algoritmo de codificación de acuerdo con la siguiente regla de decisión, en la que `acelp_snr` es la segunda SNR estimada y `tcx_snr` es la primera SNR estimada:

15 **si `acelp_snr + dsnr > tcx_snr` entonces seleccionar ACELP, de otra manera seleccionar TCX.**

Determinación de los parámetros del filtro para reducir la amplitud de los armónicos

Ahora se describe una forma de realización para determinar los parámetros del filtro para reducir la amplitud de los armónicos. Los parámetros de filtro pueden estimarse en el lado del codificador, tal como en la unidad 8

Estimación del paso

Se estima un retardo de paso (parte de número entero + parte fraccional) por trama (tamaño de trama, por ejemplo, 20 ms). Esto se efectúa en tres etapas para reducir la complejidad y de mejorar la exactitud de la estimación

a) Primera estimación de la parte de número entero del retardo de paso

Se utiliza un algoritmo de análisis de paso que produce un contorno de evolución de paso suave (por ejemplo, el análisis de paso de bucle abierto descrito en Rec. ITU-T G.718, sec. 6.6). Este análisis es efectuado por lo general en una base de subtramas (tamaño de la subtrama, por ejemplo de 10 ms), y produce una estimación de retardo de paso por subtrama. Obsérvese que estas estimaciones de retardo de paso no tienen ninguna parte fraccional y que por lo general estiman en una señal submuestreada (tasa de muestreo, por ejemplo, 6.400 Hz). La señal utilizada puede ser cualquier señal de audio, por ejemplo, una señal de audio ponderada de LPC tal como se describe en Rec. ITU-T G.718, sec. 6.5.

b) Perfeccionamiento de la parte de número entero T_{int} del retardo de paso

La parte de número entero final del retardo de paso se estima en una señal de audio $x[n]$ que se ejecuta a la tasa de muestreo del codificador principal, que es por lo general superior que la tasa de muestreo de la señal submuestreada utilizada en a) (por ejemplo, 12,8 kHz, 16 kHz, 32 kHz...). La señal $x[n]$ puede ser cualquier señal de audio, por ejemplo una señal de audio ponderada de LPC.

La parte de número entero del T_{int} es entonces el retardo que maximiza la función de autocorrelación

$$C(d) = \sum_{n=0}^N x[n]x[n-d]$$

estimándose d alrededor de un retardo de paso T en a).

$$T - \delta_1 \leq d \leq T + \delta_2$$

c) Estimación de la parte fraccional T_{fr} del retardo de paso

La parte fraccional T_{fr} se encuentra interpolando la función de autocorrelación $C(d)$ calculada en la etapa b) y seleccionando el retardo de paso fraccional que maximice la función de autocorrelación interpolada. La interpolación puede llevarse a cabo mediante un filtro FIR de pase bajo tal como se describe en por ejemplo Rec. ITU-T G.718, sec. 6.6.7.

Estimación ganancia y cuantificación

Por lo general, la ganancia se estima sobre en la señal de audio introducida a la tasa de muestreo del codificador principal, pero también puede ser cualquier señal de audio como la señal de audio ponderada de LPC. Esta señal se indica $y[n]$, y puede ser igual a o diferente de, $x[n]$.

Primero se encuentra la predicción de $y_F[n]$ de $y[n]$ filtrando $y[n]$ con el siguiente filtro:

$$P(z) = B(z, T_{fr})z^{-T_{int}}$$

Siendo T_{int} la parte de número entero del retardo de paso (estimada en b)) y $B(z, T_{fr})$ un filtro FIR de paso bajo cuyos coeficientes dependen de la parte fraccional del retardo de paso T_{fr} (estimada en c)).

5

Un ejemplo de $B(z)$ cuando la resolución del retardo de paso es de $\frac{1}{4}$:

$$T_{fr} = \frac{0}{4} \quad B(z) = 0,0000z^{-2} + 0,2325z^{-1} + 0,5349z^0 + 0,2325z^1$$

$$T_{fr} = \frac{1}{4} \quad B(z) = 0,0152z^{-2} + 0,3400z^{-1} + 0,5094z^0 + 0,1353z^1$$

$$T_{fr} = \frac{2}{4} \quad B(z) = 0,0609z^{-2} + 0,4391z^{-1} + 0,4391z^0 + 0,0609z^1$$

$$T_{fr} = \frac{3}{4} \quad B(z) = 0,1353z^{-2} + 0,5094z^{-1} + 0,3400z^0 + 0,0152z^1$$

La ganancia g se calcula como sigue:

10

$$g = \frac{\sum_{n=0}^{N-1} y[n]y_P[n]}{\sum_{n=0}^{N-1} y_P[n]y_P[n]}$$

y está limitada entre 0 y 1.

15 Finalmente, la ganancia g se cuantifica, por ejemplo, en 2 bits, utilizándose por ejemplo la cuantificación uniforme.

Se utiliza β para controlar la intensidad del filtro. β igual a 1 produce efectos completos, β igual a cero desactiva el filtro. Por lo tanto, en algunas formas de realización de la invención, el filtro puede desactivarse estableciendo β a un valor de 0. En algunas formas de realización de la invención, si el filtro está activado, β puede establecerse a un valor entre 0,5 y 0,75. En algunas formas de realización de la invención, si el filtro está desactivado, β puede establecerse a un valor 0,625. Un ejemplo de $B(z, T_{fr})$ se ha proporcionado anteriormente. El orden y los coeficientes de $B(z, T_{fr})$ pueden depender también de la tasa de bits y de la tasa de muestreo de salida. Puede diseñarse una respuesta de frecuencia diferente y ajustarse para cada combinación de tasa de bits y de tasa de muestreo de salida.

20

25

Desactivación del filtro

El filtro puede desactivarse basándose en una combinación de una o más medidas de armonicidad y/o de una o más medidas de estructura temporales. A continuación se muestran ejemplos de tales medidas:

30

i) Medida de la armonicidad como la correlación normalizada del retardo de paso de número entero estimado en la etapa b).

$$\text{norm. corr.} = \frac{\sum_{n=0}^N x[n]x[n - T_{int}]}{\sqrt{\sum_{n=0}^N x[n]x[n]} \sqrt{\sum_{n=0}^N x[n - T_{int}]x[n - T_{int}]}}$$

35

La correlación normalizada es de 1 si la señal introducida es perfectamente predecible por el retardo de paso de número entero y 0 si no es predecible en absoluto. Un valor elevado (cercano a 1) indicaría a continuación una señal armónica. Para una decisión más robusta, la correlación normalizada de la trama anterior también puede utilizarse en la decisión, por ejemplo:

40

Si $(\text{norm.corr}(\text{curr.}) * \text{norm.corr}(\text{prev.})) > 0,25$, entonces el filtro no está desactivado

ii) Medidas de estructura temporal calculadas, por ejemplo, basándose en muestras de energía también utilizadas por un detector de transitorios para la detección de transitorios (por ejemplo, medida de la uniformidad temporal, cambio de energía), por ejemplo

5 si (medida de uniformidad temporal > 3,5 o cambio de energía > 3,5) entonces el filtro está desactivado.

A continuación se exponen más detalles en cuanto a la determinación de una o más medidas de la armonicidad.

10 La medida de la armonicidad se calcula, por ejemplo, mediante una correlación normalizada de la señal de audio o de una versión premodificada de la misma en o alrededor del retardo de paso. El retardo de paso aún podría determinarse en etapas que comprenden una primera etapa y una segunda etapa, en las que, dentro de la primera etapa, se determina una estimación preliminar del retardo de paso en un dominio submuestreado de una primera tasa de muestreo, y, dentro de la segunda etapa, la estimación preliminar del retardo de paso se perfecciona en una segunda tasa de muestreo, más elevada que la primera tasa de muestreo. El retardo de paso se determina, por ejemplo, usando autocorrelación. La al menos una medida de estructura temporal se determina, por ejemplo, dentro de una región temporal temporalmente colocada en dependencia de la información de paso. Un extremo temporalmente de encabezamiento del pasado de la región temporal está colocado, por ejemplo, en función de la información de paso. El extremo temporal de encabezamiento del pasado de la región temporal puede estar colocado de manera tal que el extremo temporalmente de encabezamiento del pasado de la región temporal está desplazado en la dirección del pasado en una cantidad temporal que aumenta monótonicamente con un incremento de la información de paso. El extremo temporalmente de encabezamiento futuro de la región temporal puede estar situado en función de la estructura temporal de la señal de audio dentro de una región temporal candidata que se extiende desde el extremo temporalmente de encabezamiento pasado de la región temporal, o de la región de mayor influencia en la determinación de la medida de la estructura temporal, hacia un extremo temporalmente de encabezamiento futuro de una trama actual. Para este fin puede utilizarse la amplitud o relación entre las muestras de energía máxima y mínima dentro de la región temporal candidata. Por ejemplo, la al menos una medida de la estructura temporal puede medir una variación de energía promedio o máxima de la señal de audio dentro de la región temporal y es posible satisfacer una condición de desactivación si tanto la al menos una medida de estructura temporal es más pequeña que un primer valor umbral predeterminado y la medida de la armonicidad es, para una trama actual y/o una trama precedente, superior a un segundo umbral. La condición se satisface también si, para una trama actual, la medida de la armonicidad es superior a un tercer umbral, y, para una trama actual y una trama precedente, la medida de la armonicidad es superior a un cuarto umbral que disminuye al aumentar el retardo de paso.

35 Ahora se presenta una descripción etapa a etapa de una forma de realización concreta para determinar las medidas.

Etapa 1. Detección de transitorios y medidas temporales

40 La señal introducida $S_{HP}(n)$ se introduce en el detector de transitorios en el dominio del tiempo. La señal introducida $S_{HP}(n)$ se filtra en paso alto. La función de transferencia del filtro de HP de detección de los transitorios se proporciona mediante

$$H_{TD}(z) = 0,375 - 0,5z^{-1} + 0,125z^{-2} \quad (1)$$

45 La señal, filtrada por el filtro de HP para la detección de transitorios, se indica como $s_{TD}(n)$. La señal filtrada por HP se segmenta en 8 segmentos consecutivos de la misma longitud. La energía de la señal filtrada por HP $s_{TD}(n)$ para cada segmento se calcula como:

$$E_{TD}(i) = \sum_{n=0}^{L_{\text{segmento}}-1} (s_{TD}(iL_{\text{segmento}} + n))^2, \quad i = 0, \dots, 7 \quad (2)$$

50 donde $L_{\text{segmento}} = \frac{L}{8}$ es el número de muestras en segmentos de 2,5 milisegundos a la frecuencia de muestreo introducida.

Una energía acumulada se calcula usando:

$$E_{Acc} = \max(E_{TD}(i-1), 0,8125E_{Acc}) \quad (3)$$

Se detecta un ataque si la energía de un segmento $E_{TD}(i)$ supera la energía acumulada en un factor constante

relación de ataque = 8,5 y el índice de ataque se establece a i :

$$E_{TD}(i) > \text{relación de ataque} \cdot E_{Acc} \quad (4)$$

5 Si no se detecta un ataque basándose en el criterio anterior, pero se detecta un fuerte incremento de la energía en el segmento i , el índice de ataque se establece a i sin indicar la presencia de un ataque. El índice de ataque se establece básicamente a la posición del último ataque en una trama con algunas restricciones adicionales.

10 El cambio de energía para cada segmento se calcula como:

$$E_{cambio}(i) = \begin{cases} \frac{E_{TD}(i)}{E_{TD}(i-1)}, & E_{TD}(i) > E_{TD}(i-1) \\ \frac{E_{TD}(i-1)}{E_{TD}(i)}, & E_{TD}(i-1) > E_{TD}(i) \end{cases} \quad (5)$$

La medida de la uniformidad temporal se calcula como:

$$TFM(N_{pasado}) = \frac{1}{8 + N_{pasado}} \sum_{i=-N_{pasado}}^7 E_{cambio}(i) \quad (6)$$

El máximo cambio de energía se calcula como:

$$MEC(N_{pasado}, N_{nuevo}) = \max(E_{cambio}(-N_{pasado}), E_{cambio}(-N_{pasado}+1), \dots, E_{cambio}(N_{nuevo}-1)) \quad (7)$$

20 Si el índice de $E_{cambio}(i)$ o $E_{TD}(i)$ es negativo, entonces indica un valor del segmento precedente, con una indexación de segmento con relación a la trama actual.

25 N_{pasado} es el número de los segmentos de las tramas anteriores. Es igual a 0 si la medida de uniformidad temporal se calcula para la utilización en la decisión de ACELP/TCX. Si la medida de la uniformidad temporal se calcula para la decisión de TCX LTP entonces es igual a:

$$N_{pasado} = 1 + \min\left(8, \left\lceil 8 \frac{pasado}{L} + 0,5 \right\rceil\right) \quad (8)$$

30 N_{nuevo} es el número de segmentos de la trama actual. Es igual a 8 para las tramas no transitorias. Para las tramas transitorias se hallan primero las ubicaciones de los elementos con la máxima y mínima energía:

$$i_{max} = \arg \max_{i \in \{-N_{pasado}, \dots, 7\}} E_{TD}(i) \quad (9)$$

$$i_{min} = \arg \min_{i \in \{-N_{pasado}, \dots, 7\}} E_{TD}(i) \quad (10)$$

35 Si $E_{TD}(i_{min}) > 0,375 E_{TD}(i_{max})$ entonces N_{nuevo} se establece a $i_{max} - 3$, de otra manera N_{nuevo} se establece a 8.

Etapa 2. Conmutación de la longitud de los bloques de transformada

40 La longitud de solapamiento y la longitud del bloque de transformada del TCX dependen de la existencia de un transitorio y de su ubicación.

Tabla 1: codificación del solapamiento y de la longitud de la transformada basándose en la posición del transitorio

Índice de ataque	Solapamiento con la primera ventana de la siguiente trama	Decisión de transformada corta/larga (codificado en binario) 0-Larga, 1-Corta	Código binario para la anchura de solapamiento	Código de solapamiento
ninguno	ALDO	0	0	00

Índice de ataque	Solapamiento con la primera ventana de la siguiente trama	Decisión de transformada corta/larga (codificado en binario) 0-Larga, 1-Corta	Código binario para la anchura de solapamiento	Código de solapamiento
-2	COMPLETO	1	0	10
-1	COMPLETO	1	0	10
0	COMPLETO	1	0	10
1	COMPLETO	1	0	10
2	MÍNIMO	1	10	110
3	MITAD	1	11	111
4	MITAD	1	11	111
5	MÍNIMO	1	10	110
6	MÍNIMO	0	10	010
7	MITAD	0	11	011

El detector de transitorios anteriormente descrito básicamente devuelve el índice del último ataque con la restricción de que si hay múltiples transitorios entonces se prefiere el solapamiento MÍNIMO con respecto al solapamiento MITAD que se prefiere sobre solapamiento TOTAL. Si un ataque en la posición 2 o 6 no es suficientemente intenso entonces se elige el solapamiento MITAD en lugar del solapamiento MÍNIMO.

Etapa 3. Estimación de paso

Se estima un retardo de paso (parte de número entero + parte fraccional) por trama (tamaño de la trama, por ejemplo, 20 ms) como se indicó anteriormente en 3 etapas a) a c) para reducir la complejidad y mejorar la exactitud de la estimación.

Etapa 4. Bit de decisión

Si la señal de audio introducida no contiene ningún contenido de armónico o si una técnica basada en la predicción introduciría distorsiones en la estructura del tiempo (por ejemplo, la repetición de un transitorio corto), entonces se toma una decisión de desactivar el filtro.

La decisión se toma basándose en varios parámetros tales como la correlación normalizada en el retardo de paso de número entero y las medidas de estructura temporal.

La correlación normalizada en el retardo de paso de número entero $norm_corr$ se estima como se ha expuesto anteriormente. La correlación normalizada es de 1 si la señal introducida es perfectamente predecible por el retardo de paso de número entero y 0 si no es predecible en absoluto. Un valor elevado (cercano a 1) indicaría entonces una señal de armónicos. Para una decisión más robusta, además de la correlación normalizada para la trama actual ($norm_corr(curr)$) también puede usarse en la decisión la correlación normalizada de la trama pasada ($norm_corr(prev)$), por ejemplo:

$$Si (norm_corr(curr)*norm_corr(prev)) > 0,25$$

o

$$Si \max(norm_corr(curr),norm_corr(prev)) > 0,5,$$

entonces la trama actual contiene algún contenido de armónicos.

Las medidas de estructura temporal pueden calcularse mediante un detector de transitorios (por ejemplo, una medida de uniformidad temporal (ecuación (6)) y ecuación de máximo cambio de energía (7)) para evitar la activación del filtro en una señal que contenga fuertes cambios transitorios o fuertes cambios temporales. Los rasgos temporales se calculan en la señal que contiene la trama actual (N_{nuevo} segmentos) y la última trama hasta el retardo de paso (N_{pasado} segmentos). Para los transitorios similares a etapa que están decayendo lentamente, algunos de los rasgos, o todos ellos, se calculan solamente hasta la ubicación del transitorio ($i_{max}-3$) debido a que las distorsiones en la parte no armónica del espectro introducidas por la filtración de LTP se suprimirían por el enmascaramiento del transitorio fuerte de duración prolongada (por ejemplo, golpe de platillo).

Los trenes de pulsos para las señales de bajo paso pueden detectarse como un transitorio mediante un detector de transitorios. Para las señales con un bajo paso los rasgos desde el detector de transitorios se ignoran por lo tanto y en su lugar hay un umbral adicional para la correlación normalizada que depende del retardo de paso, por ejemplo:

5

Si $\text{norm_corr} \leq 1,2 \cdot T_{int}/L$, en tal caso desactivar el filtro.

10

A continuación se muestra un ejemplo de decisión en donde b1 es alguna tasa de bits, por ejemplo 48 kbps, donde TCX_20 indica que la trama se codifica utilizando un único bloque largo, donde TCX_10 indica que la trama se codifica utilizando 2, 3, 4 o más bloques cortos, donde la decisión TCX_20/TCX_10 se basa en la salida del detector de transitorios anteriormente descrito. Temp-Flatness es la Medida Temporal de Uniformidad definida en (6), maxEnergyChange es el Cambio de Energía Máximo como se define en (7). La condición $\text{norm_corr}(\text{curr}) > 1,2 \cdot T_{int}/L$ también podría escribirse como $(1,2 \cdot \text{norm_corr}(\text{curr})) \cdot L < T_{int}$.

```

enableLTP =
(tasa de bits < b1 && tcxMode == TCX_20 && (norm_corr(curr) * norm_corr(prev)) > 0,25 &&
tempFlatness < 3,5) ||
(tasa de bits >= b1 && tcxMode == TCX_10 && max(norm_corr(curr), norm_corr(prev)) > 0,5 &&
maxEnergyChange < 3,5) ||
(tasa de bits >= b1 && norm_corr(curr) > 0,44 && norm_corr(curr) > 1,2 * T_int/L) ||
(tasa de bits >= b1 && tcxMode == TCX_20 && norm_corr(curr) > 0,44 &&
(tempFlatness < 6,0 || (tempFlatness < 7,0 && maxEnergyChange < 22,0)));

(tasa de bits >= b1 && tcxMode == TCX_20 && norm_corr > 0,44 &&

```

15

Es obvio a partir de los ejemplos anteriores que la detección de un transitorio influye sobre qué mecanismo de decisión para la predicción a largo plazo se utilizará y qué parte de la señal se utilizará para las mediciones utilizadas en la decisión y que no desencadenen directamente la desactivación del filtro de predicción a largo plazo.

20

Las mediciones temporales utilizadas para la decisión de longitud de transformada pueden ser completamente diferentes de las mediciones temporales utilizadas por la decisión del filtro de LTP, o pueden solaparse o ser exactamente iguales pero calcularse en diferentes regiones. Para señales de bajo paso la detección de los transitorios puede ignorarse por completo si se alcanza el umbral para la correlación normalizada que depende del retardo de paso.

25

Técnica para eliminar posibles discontinuidades

30

Ahora se describe una posible técnica para eliminar las discontinuidades causadas por la aplicación de un filtro lineal H(z), trama a trama. El filtro lineal puede ser el filtro LTP descrito. El filtro lineal puede ser un filtro FIR (respuesta de impulso finita) o un filtro IIR (respuesta de impulso infinita). El enfoque propuesto no filtra una porción de la trama dirigente con los parámetros de filtro de la trama pasada y por lo tanto evita posibles problemas de los enfoques conocidos. El enfoque propuesto utiliza un filtro de LPC para eliminar la discontinuidad. Este filtro de LPC se estima en la señal de audio (filtrada por un filtro lineal invariable en el tiempo (Hz), o no) y es por lo tanto un buen modelo de la forma espectral de la señal de audio (filtrada por H(z), o no). El filtro de LPC se utiliza entonces de manera tal que la forma espectral de la señal de audio enmascara la discontinuidad.

35

El filtro de LPC puede estimarse de diversas maneras. Puede estimarse por ejemplo utilizando la señal de audio (trama actual y/o pasada) y el algoritmo de Levinson-Durbin. También puede calcularse sobre la señal de trama filtrada pasada, utilizándose el algoritmo de Levinson-Durbin.

40

Si se utiliza H(z) en un códec de audio y el códec del audio ya utiliza un filtro de LPC (cuantificado o no) para por ejemplo conformar el ruido de cuantificación en un códec de audio basado en transformada, entonces este filtro de LPC puede utilizarse directamente para suavizar la discontinuidad, sin la complejidad adicional necesaria para estimar un nuevo filtro de LPC.

45

A continuación se describe el procesamiento de la trama actual para el caso del filtro FIR y para el caso del filtro IIR. Se supone que la trama anterior ya se ha procesado.

50

Caso del filtro FIR:

1. Filtrar la trama actual con los parámetros de filtro de la trama actual, produciéndose una trama actual filtrada.
2. Considerar un filtro de LPC (cuantificado o no) de orden M, estimado en la señal de audio (filtrada o no).
3. Las M últimas muestras de la trama pasada se filtran con el filtro H(z) y con los coeficientes de la trama actual, produciéndose una primera porción de señal filtrada.

4. Las M últimas muestras de la trama pasada filtrada se restan a continuación de la primera porción de la señal filtrada, produciéndose una segunda porción de señal filtrada.
5. A continuación se genera una Respuesta de Impulso Cero (ZIR) del filtro de LPC filtrando una trama de cero muestras con el filtro de LPC y estados iniciales iguales a la segunda porción de señal filtrada.
6. El ZIR puede ser opcionalmente puesto en ventanas de manera tal que su amplitud vaya más rápidamente que 0.
7. Una porción de comienzo del ZIR se resta de una correspondiente porción de comienzo de la trama actual filtrada.

10 Caso del filtro IIR:

1. Considerar un filtro de LPC (cuantificado o no) de orden M, estimado en la señal de audio (filtrada o no).
2. Las M últimas muestras de la trama pasada se filtran con el filtro H(z) y con los coeficientes de la trama actual, produciéndose una primera porción de señal filtrada.
3. Las M últimas muestras de la trama pasada filtrada se restan a continuación de la primera porción de la señal filtrada, produciéndose una segunda porción de señal filtrada.
4. A continuación se genera una Respuesta de Impulso Cero (ZIR) del filtro de LPC filtrando una trama de cero muestras con el filtro de LPC y con estados iniciales iguales a la segunda porción de señal filtrada.
5. El ZIR puede ponerse en ventanas opcionalmente de manera tal que su amplitud vaya más rápido a 0.
6. A continuación se procesa una porción de comienzo de la trama actual muestra a muestra empezándose con la primera muestra de la trama actual.
7. La muestra se filtra con el filtro H(z) y con los parámetros de trama actuales, produciéndose una primera muestra filtrada.
8. La muestra correspondiente del ZIR se resta a continuación de la primera muestra filtrada, produciéndose la correspondiente muestra de la trama actual filtrada.
9. Pasar a la muestra siguiente.
10. Repetir 9 a 12 hasta que se haya procesado la última muestra de la porción de comienzo de la trama actual.
11. Filtrar las muestras restantes de la trama actual con los parámetros del filtro de la trama actual.

30 Por lo tanto, las formas de realización de la invención permiten estimar las SNR segmentarias y la selección de un algoritmo de codificación apropiado, de una manera simple y precisa. En particular, las formas de realización de la invención permiten una selección de bucle abierto, de un algoritmo de codificación apropiado, en el que se evita la selección inapropiada de un algoritmo de codificación en el caso de una señal de audio que tenga armónicos.

35 En las formas de realización anteriores, las SNR segmentarias se estiman calculando un promedio de las SNR estimadas para respectivas subtramas. En formas de realización alternativas, la SNR de una trama completa podría estimarse sin dividir la trama en subtramas.

40 Las formas de realización de la invención permiten una fuerte reducción en el tiempo de cálculo en comparación con la selección de bucle cerrado ya que se omite una cantidad de etapas requeridas en la selección de bucle cerrado.

45 Por lo tanto, una gran cantidad de etapas y el tiempo de cálculo asociado con ellas pueden ahorrarse gracias al enfoque inventivo, sin dejar de permitir la selección de un algoritmo de codificación apropiado con un buen rendimiento.

50 Aunque se han descrito algunos aspectos 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 de método o a un rasgo de una etapa de método. De manera análoga, los aspectos descritos en el contexto de una etapa de método también representan una descripción de un bloque o elemento o rasgo correspondiente de un aparato correspondiente.

55 Las formas de realización de los aparatos descritos en el presente documento y los rasgos de las mismas pueden implementarse mediante una computadora, uno o más procesadores, uno o más microprocesadores, campos de matrices de puertas programables (FPGA), circuitos integrados específicos de la aplicación (ASIC) y similares o combinaciones de los mismos, que están configurados o programados para proporcionar las funcionalidades descritas.

60 Algunas de las etapas de método, o todas ellas, pueden ejecutarse (o usarse) mediante un aparato de hardware, como por ejemplo, un microprocesador, una computadora programable o un circuito electrónico. En algunas formas de realización, algunas de las etapas de método más importantes, o varias de ellas, pueden ejecutarse mediante un aparato de este tipo.

Dependiendo de ciertos requisitos de implementación, las formas de realización de la invención pueden implementarse en hardware o en software. La implementación puede llevarse a cabo utilizando un medio de

almacenamiento digital no transitorio, por ejemplo un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en el mismo señales de control electrónicamente legibles, que cooperan (o son capaces de cooperar) con un sistema de computadora programable de manera tal que se lleve a cabo el respectivo método. Para ello, el medio de almacenamiento digital puede ser legible por computadora.

Algunas formas de realización de acuerdo con la invención comprenden un portador de datos que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema de computadora programable, de manera tal que se lleve a cabo uno de los métodos descritos en el presente documento.

En términos generales, las formas de realización de la presente invención pueden implementarse como un producto programa de computadora con un código de programa, siendo el código del programa operativo para llevar a cabo uno de los métodos cuando el producto programa de computadora se ejecuta en una computadora. Por ejemplo, el código del programa puede almacenarse en un portador legible por máquina.

Otras formas de realización comprenden el programa de computadora para llevar a cabo uno de los métodos descritos en el presente documento, almacenado en un portador legible por máquina.

En otras palabras, por lo tanto una forma de realización del método inventivo es un programa de computadora que tiene un código de programa para llevar a cabo uno de los métodos descritos en el presente documento, cuando el programa de computadora se ejecuta en una computadora.

Por lo tanto, una forma de realización del método inventivo consiste en un portador de datos (o un medio de almacenamiento digital, o un medio legible por computadora), que comprende, grabado en el mismo, el programa de computadora para llevar a cabo uno de los métodos descritos en el presente documento. El portador de datos, el medio de almacenamiento digital o el medio grabado son típicamente tangibles y/o no transitorios.

Una forma de realización adicional del método inventivo es, por lo tanto, una corriente de datos o una secuencia de señales que representan el programa de computadora para llevar a cabo uno de los métodos descritos en el presente documento. La corriente de datos o la secuencia de señales pueden estar configuradas, por ejemplo, para transferirse mediante una conexión de comunicación de datos, por ejemplo, mediante Internet.

Una forma de realización adicional comprende un medio de procesamiento, por ejemplo una computadora o un dispositivo lógico programable, configurado para, o programado para, llevar a cabo uno de los métodos descritos en el presente documento.

Una forma de realización adicional comprende una computadora que tiene instalado en la misma el programa de computadora para llevar a cabo uno de los métodos descritos en el presente documento.

Una forma de realización adicional de acuerdo con la invención se refiere a un aparato o sistema configurados para transferir (por ejemplo, electrónicamente u ópticamente) un programa de computadora para llevar a cabo uno de los métodos descritos en el presente documento, a un receptor. El receptor puede ser, por ejemplo, una computadora, un dispositivo móvil, un dispositivo de memoria, o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa de computadora al receptor.

En algunas formas de realización puede utilizarse un dispositivo lógico programable (por ejemplo, un campo de matrices de puertas programables), para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas formas de realización, un campo de matrices de puertas programables puede cooperar con un microprocesador para llevar a cabo uno de los métodos descritos en el presente documento. En términos generales, los métodos se llevan a cabo preferentemente mediante un aparato de hardware.

Las formas de realización anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que será evidente para los expertos en la materia modificaciones y variaciones en las disposiciones y detalles descritos en el presente documento. Por lo tanto, se pretende que la invención quede limitada solamente por el alcance de las reivindicaciones de patente siguientes y no por los detalles específicos presentados a modo de descripción y explicación de las formas de realización presentadas en el presente documento.

REIVINDICACIONES

1. Aparato (10) para seleccionar uno de un primer algoritmo de codificación que tiene una primera característica y un segundo algoritmo de codificación que tiene una segunda característica para codificar una porción de una señal de audio (40) para obtener una versión codificada de la porción de la señal de audio (40), que comprende:

- un filtro de predicción a largo plazo configurado para recibir la señal de audio, para reducir la amplitud de armónicos en la señal de audio y para emitir una versión filtrada de la señal de audio;
- un primer estimador (12) para usar la versión filtrada de la señal de audio en la estimación de una relación de señal a ruido, SNR, o relación de señal a ruido de SNR segmentaria de la porción de la señal de audio como una primera medida de calidad para la porción de la señal de audio, estando asociada la primera medida de calidad con el primer algoritmo de codificación, en el que estimar dicha primera medida de calidad comprende realizar una aproximación del primer algoritmo de codificación para obtener una estimación de distorsión del primer algoritmo de codificación y para estimar la primera medida de calidad basándose en la porción de la señal de audio y la estimación de distorsión del primer algoritmo de codificación sin codificar y decodificar realmente la porción de la señal de audio usando el primer algoritmo de codificación;
- un segundo estimador (14) para estimar una SNR o una SNR segmentaria como una segunda medida de calidad para la porción de la señal de audio, estando asociada la segunda medida de calidad con el segundo algoritmo de codificación, en el que estimar dicha segunda medida de calidad comprende realizar una aproximación del segundo algoritmo de codificación para obtener una estimación de distorsión del segundo algoritmo de codificación y para estimar la segunda medida de calidad usando la porción de la señal de audio y la estimación de distorsión del segundo algoritmo de codificación sin codificar y decodificar realmente la porción de la señal de audio usando el segundo algoritmo de codificación; y
- un controlador (16) para seleccionar el primer algoritmo de codificación o el segundo algoritmo de codificación basándose en una comparación entre la primera medida de calidad y la segunda medida de calidad,

en el que el primer algoritmo de codificación es un algoritmo de codificación por transformada, un algoritmo de codificación basado en transformada de coseno discreta modificada, MDCT, o un algoritmo de codificación de excitación codificada por transformada, TCX, y en el que el segundo algoritmo de codificación es un algoritmo de codificación de predicción lineal con excitación por código, CELP, o un algoritmo de codificación de predicción lineal con excitación por código algebraico, ACELP.

2. Aparato (10) de la reivindicación 1, en el que una función de transferencia del filtro de predicción a largo plazo comprende una parte de número entero de un retardo de paso y un filtro de múltiples derivaciones que depende de una parte fraccional del retardo de paso.

3. Aparato (10) de la reivindicación 1, en el que el filtro de predicción a largo plazo tiene la función de transferencia:

$$P(z) = 1 - \beta g B(z, T_{fr}) z^{-T_{int}}$$

T_{int} y T_{fr} son la parte de número entero y fraccional de un retardo de paso, g es una ganancia, β es una ponderación y $B(z, T_{fr})$ es un filtro FIR de paso bajo cuyos coeficientes dependen de la parte fraccional del paso.

4. Aparato de una de las reivindicaciones 1 a 3, que además comprende una unidad de desactivación para desactivar el filtro basándose en una combinación de una o más medidas de armonicidad y/o una o más medidas de estructura temporal.

5. Aparato de la reivindicación 4, en el que la una o más medidas de armonicidad comprenden al menos una de una correlación normalizada o una ganancia de predicción y en el que la una o más medidas de estructura temporal comprenden al menos una de una medida de uniformidad temporal y un cambio de energía.

6. Aparato de una de las reivindicaciones 1 a 5, en el que el filtro se aplica a la señal de audio en una base trama a trama, comprendiendo dicho aparato además una unidad para eliminar las discontinuidades en la señal de audio provocadas por el filtro.

7. Aparato (10) de una de las reivindicaciones 1 a 6, en el que el primer y segundo estimadores están configurados para estimar una SNR o SNR segmentaria de una porción de una versión ponderada de la señal de audio.

8. Aparato (10) de una de las reivindicaciones 1 a 7, en el que el primer estimador (12) está configurado para determinar una distorsión estimada de cuantificador que un cuantificador utilizado en el primer algoritmo de codificación introduciría cuando cuantifica la porción de la señal de audio y para estimar la primera medida de calidad basándose en una energía de una porción de una versión ponderada de la señal de audio y de la distorsión estimada del cuantificador, en el que el primer estimador (12) está configurado para estimar una ganancia global para la porción de la señal de audio de manera tal que la porción de la señal de audio produciría una tasa de bits

objetivo dada cuando se codifica con un cuantificador y un codificador de entropía utilizado en el primer algoritmo de codificación, en el que el primer estimador (12) está además configurado para determinar la distorsión estimada del cuantificador basándose en la ganancia global estimada.

5 9. Aparato (10) de una de las reivindicaciones 1 a 8, en el que el segundo estimador (14) está configurado para determinar una distorsión estimada del libro de código adaptativo que un libro de código adaptativo utilizado en el segundo algoritmo de codificación introduciría cuando se utiliza el libro de código adaptativo para codificar la porción de la señal de audio y en el que el segundo estimador (14) está configurado para estimar la segunda medida de calidad basándose en una energía de una porción de una versión ponderada de la señal de audio y la distorsión
10 estimada del libro de código adaptativo, en el que, para cada una de una pluralidad de subporciones de la porción de la señal de audio, el segundo estimador (14) está configurado para aproximar el libro de código adaptativo basándose en una versión de la subporción de la señal de audio ponderada desplazada hacia el pasado mediante un retardo de paso determinado en una etapa de pre-procesamiento, para estimar una ganancia del libro de código adaptativo de manera tal que se minimiza un error entre la subporción de la porción de la señal de audio ponderada
15 y el libro de código adaptativo aproximado, y para determinar la distorsión estimada del libro de código adaptativo basándose en la energía de un error entre la subporción de la porción de la señal de audio ponderada y el libro de código adaptativo aproximado escalado por la ganancia del libro de código adaptativo.

20 10. Aparato (10) de la reivindicación 9, en el que el segundo estimador (14) está además configurado para reducir la distorsión estimada del libro de código adaptativo determinada para cada subporción de la porción de la señal de audio en un factor constante.

25 11. Aparato (10) de una de las reivindicaciones 1 a 8, en el que el segundo estimador (14) está configurado para determinar una distorsión estimada del libro de código adaptativo que un libro de código adaptativo utilizado en el segundo algoritmo de codificación introduciría cuando se utiliza el libro de código adaptativo para codificar la porción de la señal de audio, y en el que el segundo estimador (14) está configurado para estimar la segunda medida de calidad basándose en una energía de una porción de una versión ponderada de la señal de audio y la distorsión estimada del libro de código adaptativo, en el que el segundo estimador (14) está configurado para aproximar el libro de código adaptativo basándose en una versión de la porción de la señal de audio ponderada desplazada hacia el pasado mediante un retardo de paso determinado en una etapa de pre-procesamiento, para estimar una ganancia del libro de código adaptativo de manera tal que se minimiza un error entre la porción de la señal de audio ponderada y el libro de código adaptativo aproximado, y para determinar la distorsión estimada del libro de código adaptativo basándose en la energía de un error entre la porción de la señal de audio ponderada y el libro de código adaptativo aproximado escalado en la ganancia del libro de código adaptativo.
30
35

40 12. Aparato (20) para codificar una porción de una señal de audio, que comprende el aparato (10) de acuerdo con una de las reivindicaciones 1 a 11, una primera etapa de codificador (26) para llevar a cabo el primer algoritmo de codificación y una segunda etapa de codificador (28) para llevar a cabo el segundo algoritmo de codificación, en el que el aparato para codificar (20) está configurado para codificar la porción de la señal de audio utilizando el primer algoritmo de codificación o el segundo algoritmo de codificación en función de la selección por el controlador (16).

45 13. Sistema para codificar y decodificar que comprende un aparato (20) para codificar de acuerdo con la reivindicación 12 y un decodificador configurado para recibir la versión codificada de la porción de la señal de audio y una indicación del algoritmo utilizado para codificar la porción de la señal de audio y para decodificar la versión codificada de la porción de la señal de audio utilizando el algoritmo indicado.

50 14. Método para seleccionar uno de un primer algoritmo de codificación que tiene una primera característica y un segundo algoritmo de codificación que tiene una segunda característica para codificar una porción de una señal de audio para obtener una versión codificada de la porción de la señal de audio, que comprende:

55 filtrar la señal de audio utilizando un filtro de predicción a largo plazo para reducir la amplitud de los armónicos en la señal de audio y para emitir una versión filtrada de la señal de audio;
utilizar la versión filtrada de la señal de audio al estimar una relación de señal a ruido, SNR, o una SNR segmentaria de la porción de la señal de audio como una primera medida de la calidad para la porción de la señal de audio, estando asociada la primera medida de la calidad con el primer algoritmo de codificación, en el que estimar dicha primera medida de la calidad comprende llevar a cabo una aproximación del primer algoritmo de codificación para obtener una estimación de la distorsión del primer algoritmo de codificación y para estimar la primera medida de la calidad basándose en la porción de la primera señal de audio y de la estimación de la distorsión del primer algoritmo de codificación sin realmente codificar y decodificar la porción de la señal de audio
60 utilizando el primer algoritmo de codificación.

estimar una SNR o SNR segmentaria como una segunda medida de la calidad para la porción de la señal de audio, estando asociada la segunda medida de la calidad con el segundo algoritmo de codificación, en el que estimar dicha segunda medida de la calidad comprende llevar a cabo una aproximación del segundo algoritmo de codificación para obtener una estimación de la distorsión del segundo algoritmo de codificación y para estimar

la segunda medida de la calidad utilizando la porción de la señal de audio y la estimación de la distorsión del segundo algoritmo de codificación sin realmente codificar y decodificar la porción de la señal de audio utilizando el segundo algoritmo de codificación; y

5 seleccionar el primer algoritmo de codificación o el segundo algoritmo de codificación basándose en una comparación entre la primera medida de la calidad y la segunda medida de la calidad,

en el que el primer algoritmo de codificación es un algoritmo de codificación por transformadas, un algoritmo de codificación basado en transformada de coseno discreta modificada, MDCT, o un algoritmo de codificación de excitación codificada por transformada, TCS, y en el que el segundo algoritmo de codificación es un algoritmo de codificación de predicción lineal con excitación por código, CELP, o un algoritmo de codificación de predicción lineal con excitación por código algebraico, ACELP.

10

15. Programa de computadora que tiene un código de programa para llevar a cabo, cuando se ejecuta en una computadora, el método de la reivindicación 14.

15

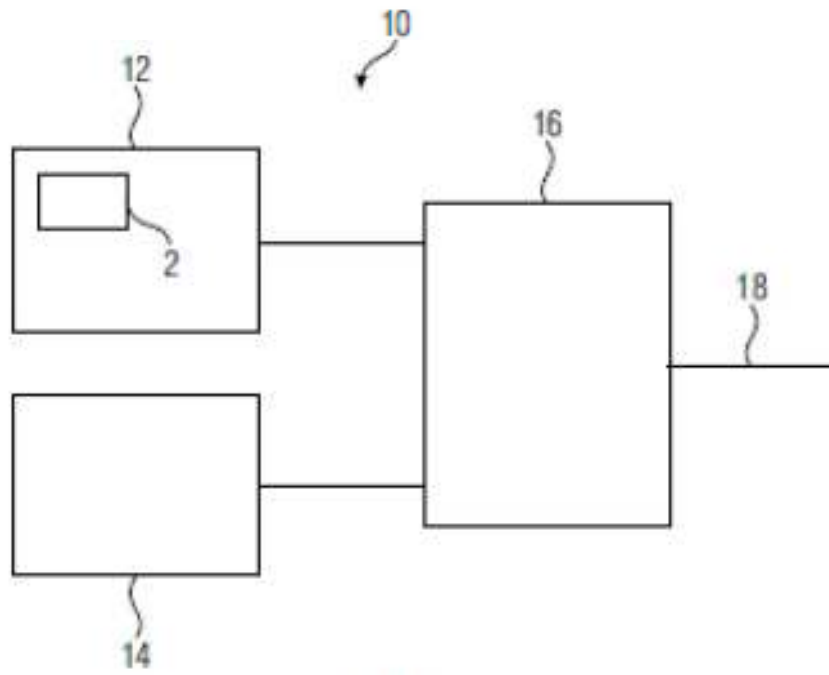


FIG 1

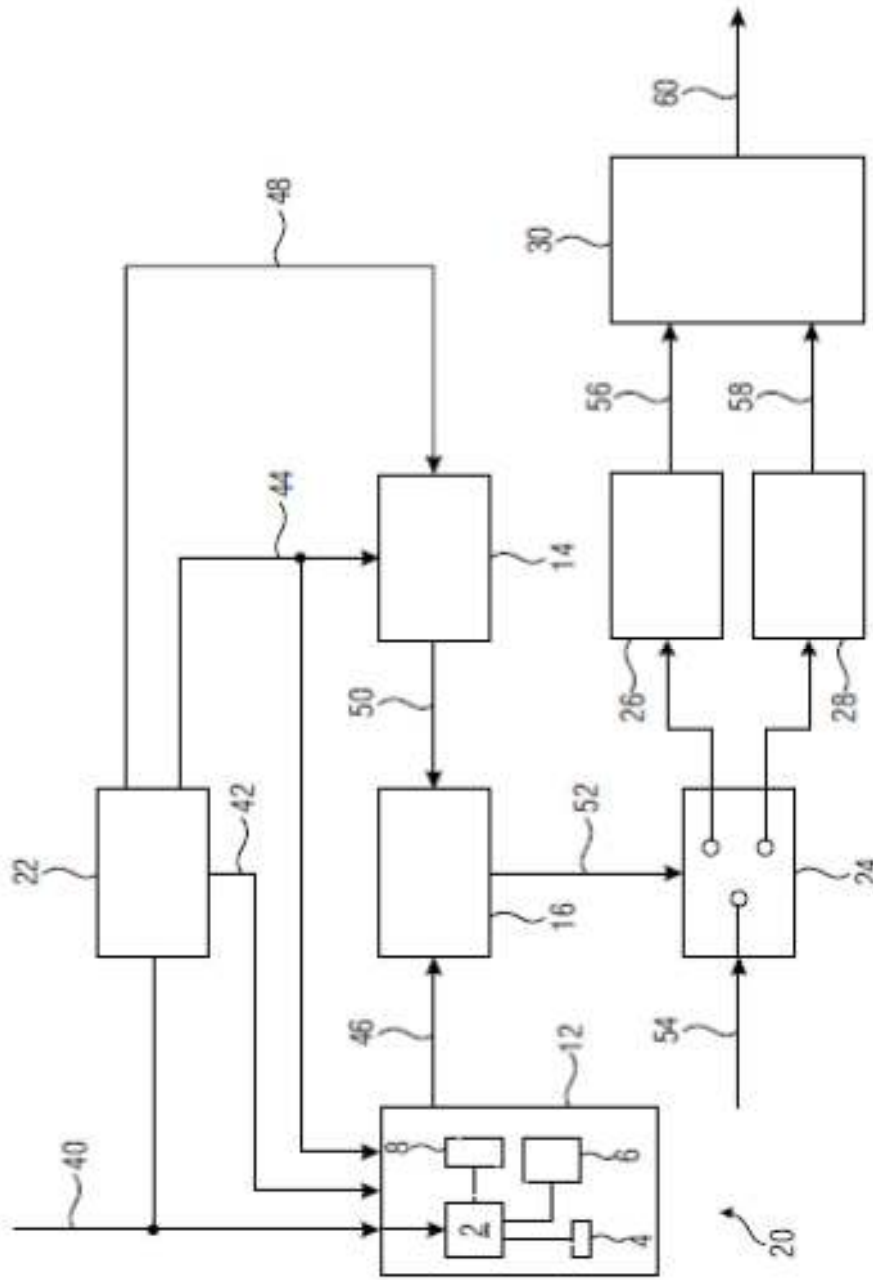


FIG 2

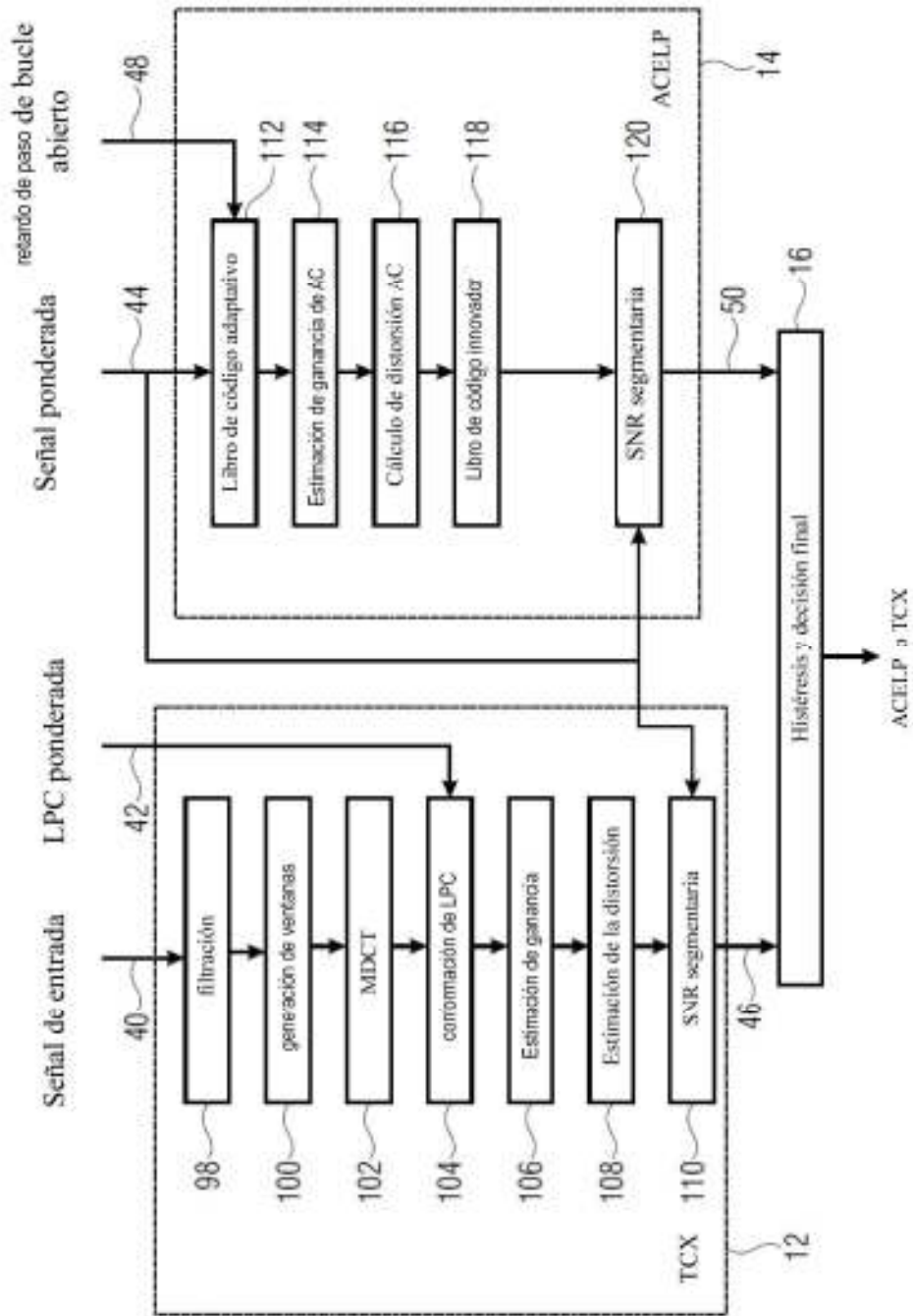


FIG 3

$$\text{SNR} = 10 \log_{10} \frac{\sum_{i=1}^N x^2(i)}{\sum_{i=1}^N (x(i) - y(i))^2}$$

FIG 4A

$$\text{SNR}_{\text{seg}} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \sum_{i=Nm}^{Nm+N-1} \left(\frac{\sum_{i=1}^N x^2(i)}{\sum_{i=1}^N (x(i) - y(i))^2} \right)$$

FIG 4B