

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 379 761**

51 Int. Cl.:
G10L 19/10 (2006.01)
G10L 21/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09776982 .2**
96 Fecha de presentación: **06.07.2009**
97 Número de publicación de la solicitud: **2311033**
97 Fecha de publicación de la solicitud: **20.04.2011**

54 Título: **Proporcionar una señal de activación de distorsión de tiempo y codificar una señal de audio con la misma**

30 Prioridad:
11.07.2008 US 79873 P

45 Fecha de publicación de la mención BOPI:
03.05.2012

45 Fecha de la publicación del folleto de la patente:
03.05.2012

73 Titular/es:
**Fraunhofer-Gesellschaft zur Förderung der
angewandten Forschung e.V.
Hansastraße 27c
80686 München, DE**

72 Inventor/es:
**BAYER, Stefan;
DISCH, Sascha;
GEIGER, Ralf;
FUCHS, Guillaume;
NEUENDORF, Max;
SCHULLER, Gerald y
EDLER, Bernd**

74 Agente/Representante:
Arizti Acha, Monica

ES 2 379 761 T3

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

DESCRIPCIÓN

Proporcionar una señal de activación de distorsión de tiempo y codificar una señal de audio con la misma.

5 La presente invención se refiere a la codificación y decodificación de audio y específicamente a la codificación/decodificación de una señal de audio que tiene un contenido armónico o de voz, que puede someter a un procesamiento de distorsión de tiempo.

A continuación, se dará una breve introducción al campo de la codificación de audio distorsionado en el tiempo, cuyos conceptos pueden aplicarse en combinación con algunas de las realizaciones de la invención.

10 En los últimos años, se han desarrollado técnicas para transformar una señal de audio en una representación de dominio de frecuencia, y para codificar con eficacia esta representación de dominio de frecuencia, por ejemplo teniendo en cuenta umbrales de enmascaramiento perceptivo. Este concepto de codificación de señal de audio es particularmente eficaz si la longitud del bloque, para el cual se transmite un conjunto de coeficientes espectrales codificados, es larga, y si sólo un número relativamente pequeño de coeficientes espectrales están bien por encima del umbral de enmascaramiento global mientras que un gran número de coeficientes espectrales está cerca o por debajo del umbral de enmascaramiento global y de este modo puede omitirse (o codificado con una longitud de código mínima).

15 Por ejemplo, las transformadas solapadas moduladas basadas en seno o basadas en coseno a menudo se utilizan en aplicaciones para codificación de fuente debido a sus propiedades de compactación de energía. Es decir, para tonos armónicos con frecuencias fundamentales constantes (tono), éstas concentran la energía de señal hasta un número bajo de componentes espectrales (subbandas), lo cual conduce a una representación de señal eficaz.

20 En general, el tono (fundamental) de una señal se entenderá como la frecuencia dominante más baja que se distingue del espectro de la señal. En el modelo de voz común, el tono es la frecuencia de la señal de excitación modulada por la garganta humana. Si sólo una frecuencia fundamental simple estuviera presente, el espectro sería extremadamente simple, comprendiendo únicamente la frecuencia fundamental y los sobretonos. Un espectro de este tipo podría codificarse de forma muy eficaz. Para señales con tono variable, sin embargo, la energía correspondiente a cada componente armónico se dispersa sobre varios coeficientes de la transformada, conduciendo de este modo a una
25 reducción de la eficacia de codificación.

A fin de solucionar esta reducción de la eficacia de codificación, la señal de audio a codificarse se remuestrea de forma eficaz sobre una rejilla de tiempo no uniforme. En el procesamiento posterior, las posiciones de muestra obtenidas por el remuestreo no uniforme se procesan como si representaran los valores sobre una rejilla de tiempo uniforme. Esta operación normalmente se denota por la frase "distorsión de tiempo". Los tiempos de muestra pueden elegirse de forma
30 ventajosa dependiendo de la variación de tiempo del tono, de modo que una variación de tono en la versión distorsionada en el tiempo de la señal de audio es menor que una variación de tono en la versión original de la señal de audio (antes de la distorsión de tiempo). Esta variación de tono también puede denotarse con la frase "contorno de distorsión de tiempo". Después de la distorsión en el tiempo de la señal de audio, la versión distorsionada en el tiempo de la señal de audio se convierte en el dominio de frecuencia. La distorsión en el tiempo que depende del tono tiene el efecto de que la representación de dominio de frecuencia de la señal de audio distorsionada en el tiempo exhibe típicamente una compactación de energía en un número mucho más pequeño de componentes espectrales que una representación de dominio de frecuencia de la señal de audio original (no distorsionada en el tiempo).
35

En el lado del decodificador, la representación de dominio de frecuencia de la señal de audio distorsionada en el tiempo se convierte de nuevo al dominio de tiempo, de modo que una representación de dominio de tiempo de la
40 señal de audio distorsionada en el tiempo está disponible en el lado del decodificador. Sin embargo, en la representación de dominio de tiempo de la señal de audio distorsionada en el tiempo reconstruida en el lado del decodificador, las variaciones del tono original de la señal de audio de entrada en el lado del codificador no se incluyen. En consecuencia, se aplica otra distorsión en el tiempo por el remuestreo de la representación del dominio del tiempo reconstruida en el lado del decodificador de la señal de audio distorsionada en el tiempo. Con el fin de obtener una
45 buena reconstrucción de la señal de audio de entrada en el lado del codificador en el decodificador, se desea que la distorsión en el tiempo en el lado del decodificador sea al menos aproximadamente la operación inversa con respecto a la distorsión en el tiempo en el lado del codificador. Con el fin de obtener una distorsión en el tiempo apropiada, se desea tener información disponible en el decodificador que permite un ajuste de la distorsión en el tiempo en el lado del decodificador.

50 Dado que típicamente se requiere transferir este tipo de información desde el codificador de señal de audio al decodificador de señal de audio, se desea mantener una velocidad de bits requerida para esta transmisión reducida mientras aún se permite una reconstrucción fiable de la información de distorsión de tiempo requerida en el lado del decodificador.

55 El documento de patente US 2007/0100607 A1 describe una codificación de transformada modificada distorsionada en el tiempo de señales de audio. Se deriva una representación de una señal de audio que tiene tramas de audio consecutivas, en el que se estima una distorsión de tiempo común para dos tramas vecinas cualesquiera, de forma que

5 una transformada en bloque siguiente puede adicionalmente emplear la información de distorsión. Así, pueden derivarse y aplicarse las funciones de ventana requeridas para una aplicación exitosa de un procedimiento de solapamiento y suma durante la reconstrucción, las funciones de ventana anticipándose ya al remuestreo de la señal debido a la distorsión en el tiempo. Por tanto, puede utilizarse la eficacia aumentada de la codificación de transformación basada en bloques de señales distorsionadas en el tiempo sin introducir discontinuidades audibles.

En vista de la descripción anterior, existe el deseo de crear un concepto que permita una aplicación de velocidad de bits eficaz del concepto de distorsión de tiempo en un codificador de audio.

10 Es objeto de la invención crear conceptos para mejorar la impresión de audición proporcionada por una señal de audio codificada sobre la base de información disponible in un codificador de señal de audio de distorsión en el tiempo o un decodificador de señal de audio de distorsión en el tiempo.

15 Este objeto se logra mediante un proveedor de la señal de activación de distorsión de tiempo para proporcionar una señal de activación de distorsión de tiempo sobre la base de una representación de una señal de audio según las reivindicaciones 1-3, un codificador de señal de audio para codificar una señal de audio de entrada según la reivindicación 8, un método para proporcionar una señal de activación de distorsión de tiempo según las reivindicaciones 10-12, un método para proporcionar una representación codificada de una señal de audio de entrada según la reivindicación 13, o un programa informático según la reivindicación 14.

Las realizaciones según la invención se refieren a métodos para un codificador de transformación MDCT distorsionada en el tiempo. Algunas realizaciones se refieren a herramientas sólo para el codificador. Sin embargo, otras realizaciones también se refieren a herramientas para el decodificador.

20 La invención crea un proveedor de la señal de activación de distorsión de tiempo para proporcionar una señal de activación de distorsión de tiempo sobre la base de una representación de una señal de audio. El proveedor de la señal de activación de distorsión de tiempo comprende un proveedor de información de compactación de energía configurado para proporcionar una información de compactación de energía que describe una compactación de energía en una representación espectral transformada por la distorsión de tiempo de la señal de audio. El proveedor de la señal de
25 activación de distorsión de tiempo también comprende un comparador configurado para comparar la información de compactación de energía con un valor de referencia, y para proporcionar la señal de activación de distorsión de tiempo con dependencia de un resultado de la comparación.

30 Esta realización se basa en el hallazgo de que el uso de una funcionalidad de distorsión de tiempo en un codificador de señal de audio típicamente conlleva una mejora, en el sentido de una reducción de la velocidad de bits de la señal de audio codificada, si la representación espectral transformada por la distorsión de tiempo de la señal de audio comprende una distribución de energía lo suficientemente compacta en el sentido de que la energía se concentra en una o más regiones espectrales (o líneas espectrales). Esto se debe al hecho de que una distorsión en el tiempo exitosa conlleva el efecto de disminuir la velocidad de bits al transformar un espectro emborronado, por ejemplo de una trama de audio, en el espectro que tiene uno o más picos discernibles, y en consecuencia tiene una compactación de energía más alta que
35 el espectro de la señal de audio original (no distorsionada en el tiempo).

Con respecto a este tema, debe entenderse que una trama de señal de audio, durante la cual el tono de la señal de audio varía significativamente, comprende un espectro emborronado. El tono con variación en el tiempo de la señal de audio tiene el efecto de que una transformación de dominio de tiempo a un dominio de frecuencia sobre la trama de la señal de audio da como resultado una distribución emborronada de la energía de señal sobre la frecuencia, en particular
40 en la región de frecuencia más alta. En consecuencia, una representación espectral de esta señal de audio original (no distorsionada en el tiempo) comprende una compactación de energía baja y típicamente no exhibe picos espectrales en una parte de frecuencia más alta del espectro, o sólo exhibe picos espectrales relativamente pequeños en la parte de frecuencia más alta del espectro. En contraste, si la distorsión en el tiempo tiene éxito (en cuanto a proporcionar una mejora de la eficacia de codificación) la distorsión en el tiempo de la señal de audio original produce una señal de audio distorsionada en el tiempo que tiene un espectro con picos relativamente más altos y claros (en particular en la parte de
45 frecuencia más alta del espectro). Esto se debe al hecho de que una señal de audio con un tono con variación en el tiempo se transforma en una señal de audio distorsionada en el tiempo que tiene una variación de tono más pequeña o incluso un tono aproximadamente constante. En consecuencia, la representación espectral de la señal de audio distorsionada en el tiempo (la cual puede considerarse una representación espectral transformada por la distorsión de tiempo de la señal de audio) comprende uno o más picos espectrales claros. Es decir, el emborronado del espectro de la señal de audio original (que tiene un tono variable en el tiempo) se reduce por una operación de distorsión de tiempo con éxito, de modo que la representación espectral transformada por la distorsión de tiempo de la señal de audio comprende una compactación de energía más alta que el espectro de la señal de audio original. No obstante, la distorsión en el tiempo no siempre tiene éxito en la mejora de la eficacia de la codificación. Por ejemplo, la distorsión en
50 el tiempo no mejora la eficacia de la codificación si la señal de audio de entrada comprende grandes componentes de ruido, o si el contorno de distorsión de tiempo extraído es impreciso.

En vista de esta situación, la información de compactación de energía proporcionada por el proveedor de información de compactación de energía es un indicador valioso para decidir si la distorsión de tiempo tiene éxito en cuanto a reducción

de la velocidad de bits.

Una realización de la invención crea un proveedor de la señal de activación de distorsión de tiempo para proporcionar una señal de activación de distorsión de tiempo sobre la base de una representación de una señal de audio. El proveedor de activación de la distorsión de tiempo comprende dos proveedores de representación de la distorsión de tiempo configurados para proporcionar dos representaciones de la distorsión de tiempo de la misma señal de audio que utilizan una información del contorno de distorsión de tiempo diferente. De este modo, los proveedores de representación de la distorsión de tiempo podrían estar configurados (de forma estructural y/o funcional) de la misma forma y utilizar la misma señal de audio pero una información del contorno de distorsión de tiempo diferente. El proveedor de la señal de activación de distorsión de tiempo también comprende dos proveedores de información de compactación de energía configurados para proporcionar una primera información de compactación de energía sobre la base de la primera representación de distorsión de tiempo y para proporcionar una segunda información de compactación de energía sobre la base de la segunda representación de distorsión de tiempo. Los proveedores de información de compactación de energía podrían estar configurados de la misma forma pero para utilizar las diferentes representaciones de distorsión de tiempo. Además el proveedor de la señal de activación de distorsión de tiempo comprende un comparador para comparar las dos informaciones de compactación de energía diferentes y para proporcionar la señal de activación de distorsión de tiempo con dependencia de un resultado de la comparación.

En una realización preferida, el proveedor de información de compactación de energía está configurado para proporcionar una medida de planicidad espectral que describe la representación espectral transformada por la distorsión de tiempo de la señal de audio como la información de compactación de energía. Se ha encontrado que la distorsión de tiempo tiene éxito, en cuanto a reducir una velocidad de bits, si transforma un espectro de una señal de audio de entrada en un espectro de distorsión de tiempo menos plano representando una versión distorsionada en el tiempo de la señal de audio de entrada. En consecuencia, la medida de planicidad espectral puede utilizarse para decidir, sin realizar un proceso de codificación espectral completo, si la distorsión de tiempo debe activarse o desactivarse.

En esta realización, el proveedor de información de compactación de energía está configurado para calcular un cociente de una media geométrica del espectro de potencia retransformada por la distorsión de tiempo y una media aritmética del espectro de potencia transformada por la distorsión de tiempo para obtener la medida de planicidad espectral. Se ha encontrado que este cociente es una medida de planicidad espectral la cual está bien adaptada para describir los posibles ahorros de velocidad de bits que pueden obtenerse mediante una distorsión en el tiempo.

En otra realización preferida, el proveedor de información de compactación de energía está configurado para enfatizar una parte de frecuencia más alta de la representación espectral transformada por la distorsión de tiempo cuando se compara con una parte de frecuencia más baja de la representación espectral transformada por la distorsión de tiempo, para obtener la información de compactación de energía. Este concepto se basa en el hallazgo de que la distorsión de tiempo típicamente tiene un impacto mucho mayor sobre el rango de frecuencia más alta que sobre el rango de frecuencia más baja. En consecuencia, una valoración dominante del rango de frecuencia más alta es apropiada para determinar la efectividad de la distorsión de tiempo utilizando una medida de planicidad espectral. Además, las señales típicas de audio exhiben un contenido armónico (que comprende la armonía de una frecuencia fundamental) que disminuye en intensidad con la frecuencia creciente. Un énfasis de una parte de frecuencia más alta de la representación espectral transformada por la distorsión de tiempo cuando se compara con una parte de frecuencia más baja de la representación espectral transformada por la distorsión de tiempo también ayuda a compensar esta decadencia típica de las líneas espectrales con frecuencia creciente. Para resumir, una consideración enfatizada de la parte de frecuencia más alta del espectro conlleva una fiabilidad aumentada de la información de compactación de energía y por lo tanto permite una provisión más fiable de la señal de activación de distorsión de tiempo.

En otra realización preferida, el proveedor de información de compactación de energía está configurado para proporcionar una pluralidad de medidas banda por banda de planicidad espectral, y para calcular un promedio de la pluralidad de medidas banda por banda de planicidad espectral, para obtener la información de compactación de energía. Se ha encontrado que la consideración de medidas espectrales de planicidad banda por banda otorga una información particularmente fiable con respecto a si la distorsión de tiempo es eficaz para reducir la velocidad de bits de una señal de audio codificada. En primer lugar, la codificación de la representación espectral transformada por la distorsión de tiempo se realiza típicamente en el modo banda por banda, de modo que una combinación de las medidas banda por banda de planicidad espectral se adapta bien a la codificación y por lo tanto representa una mejora que puede obtenerse de la velocidad de bits con buena precisión. Además, un cálculo banda por banda de medidas de planicidad espectral elimina sustancialmente la dependencia de la información de compactación de energía a partir de una distribución de la armonía. Por ejemplo, incluso si una banda de frecuencia más alta comprende una energía relativamente pequeña (más pequeña que las energías de bandas de frecuencia inferiores), la banda de frecuencia más alta puede ser incluso perceptualmente relevante. Sin embargo, el impacto positivo de una distorsión de tiempo (en el sentido de una reducción del emborronado de las líneas espectrales) sobre esta banda de frecuencia más alta sería considerado como bajo, simplemente debido a la baja energía de la banda de frecuencia más alta, si la medida de planicidad espectral no se computase en el modo banda por banda. A diferencia, al aplicar el cálculo banda por banda, puede tenerse en consideración un impacto positivo de la distorsión de tiempo con un peso apropiado, porque las medidas de planicidad espectral banda por banda son independientes de las energías absolutas en las bandas de

frecuencia respectivas.

5 En otra realización preferida, el proveedor de la señal de activación de distorsión de tiempo comprende un calculador de valor de referencia configurado para calcular una medida de planicidad espectral que describe una representación espectral no distorsionada en el tiempo de la señal de audio, para obtener el valor de referencia. En consecuencia, la señal de activación de distorsión de tiempo puede proporcionarse sobre la base de una comparación de la planicidad espectral de una versión no distorsionada en el tiempo (o sin distorsionar) de la señal de audio de entrada y una planicidad espectral de una versión de distorsión por el tiempo de la señal de audio de entrada.

10 En otra realización preferida, el proveedor de información de compactación de energía está configurado para proporcionar una medida de entropía perceptiva que describe la representación espectral transformada por la distorsión de tiempo de la señal de audio como la información de compactación de energía. Este concepto se basa en el hallazgo de que la entropía perceptiva de la representación espectral transformada por la distorsión de tiempo es una buena estimación de varios bits (o una velocidad) requeridos para codificar el espectro transformado por la distorsión de tiempo. En consecuencia, la medida de entropía perceptiva de la representación espectral transformada por la distorsión de tiempo es una buena medida respecto si puede esperarse una reducción de la velocidad de bits por la distorsión en el tiempo, incluso en vista del hecho de que una información de distorsión de tiempo adicional debe ser codificada si se utiliza la distorsión de tiempo.

20 En otra realización preferida, el proveedor de información de compactación de energía está configurado para proporcionar una medida de autocorrelación que describe una autocorrelación de una representación de la distorsión por el tiempo de la señal de audio como la información de compactación de energía. Este concepto se basa en el hallazgo de que la eficacia de la distorsión de tiempo (en cuanto a reducción de la velocidad de bits) puede medirse (o al menos estimar) sobre la base de una señal de dominio de tiempo distorsionada en el tiempo (o un remuestreo no uniforme). Se ha encontrado que la distorsión en el tiempo es eficaz si la señal del dominio de tiempo distorsionada en el tiempo comprende un grado relativamente alto de periodicidad, lo cual se refleja por la medida de autocorrelación. A diferencia, si la señal del dominio de tiempo distorsionada en el tiempo no comprende una periodicidad significativa, puede concluirse que la distorsión en el tiempo no es eficaz.

25 Este hallazgo se basa en el hecho de que una distorsión de tiempo eficaz transforma una parte de una señal sinusoidal de una frecuencia variable (la cual no comprende una periodicidad) en una parte de una señal sinusoidal de frecuencia aproximadamente constante (la cual comprende un grado alto de periodicidad). A diferencia, si la distorsión en el tiempo no puede proporcionar una señal de dominio de tiempo con un grado alto de periodicidad, puede esperarse que la distorsión en el tiempo tampoco proporcione un ahorro de velocidad de bits significativa, lo cual justificaría su aplicación.

30 En esta realización, el proveedor de información de compactación de energía está configurado para determinar una suma de valores absolutos de una función de autocorrelación normalizada (sobre una pluralidad de valores de retardo) de la representación de distorsión por tiempo de la señal de audio, para obtener la información de compactación de energía. Se ha encontrado que una determinación del complejo por computación de los picos de autocorrelación no es necesaria para estimar la eficacia de la distorsión en el tiempo. En cambio, Se ha encontrado que una evaluación de de la suma de la autocorrelación sobre un rango (amplio) de valores de retardo de la autocorrelación además otorga resultados muy fiables. Esto se debe al hecho de que la distorsión de tiempo realmente transforma una pluralidad de componentes de señal (por ejemplo una frecuencia fundamental y armónica de la misma) de frecuencia variable en componentes de señal periódicos. En consecuencia, la autocorrelación de una señal distorsionada en el tiempo de este tipo presenta picos a una pluralidad de valores de retardo de autocorrelación. Por tanto, una formación de la suma es una forma eficaz desde el punto de vista computacional para extraer la información de compactación de energía de la autocorrelación.

35 En otra realización preferida, el proveedor de la señal de activación de distorsión de tiempo comprende un calculador de valor de referencia configurado para calcular el valor de referencia sobre la base de una representación espectral no distorsionada en el tiempo de la señal de audio o sobre la base de una representación de dominio de tiempo no distorsionada en el tiempo de la señal de audio. En este caso, el comparador típicamente está configurado para formar un valor de relación utilizando la información de compactación de energía que describe una compactación de energía en un espectro transformado por la distorsión de tiempo de la señal de audio y el valor de referencia. El comparador también está configurado para comparar el valor de relación con uno o más valores de umbral para obtener la señal de activación de distorsión de tiempo. Se ha encontrado que la relación entre una información de compactación de energía en el caso no distorsionado por el tiempo y la información de compactación de energía en el caso de distorsionado en el tiempo permite una generación eficaz desde el punto de vista computacional a la vez de lo suficientemente fiable de la señal de activación de distorsión de tiempo.

40 Otra realización preferida de la invención crea un codificador de señal de audio para codificar una señal de audio de entrada, para obtener una representación codificada de la señal de audio de entrada. El codificador de señal de audio comprende un transformador de distorsión de tiempo configurado para proporcionar una representación espectral transformada por la distorsión de tiempo sobre la base de la señal de audio de entrada. El codificador de señal de audio también comprende un proveedor de la señal de activación de distorsión de tiempo, tal como se describió anteriormente. El proveedor de la señal de activación de distorsión de tiempo está configurado para recibir la señal de audio de entrada

5 y proporcionar la información de compactación de energía de modo que la información de compactación de energía describe una compactación de energía en la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada. El codificador de señal de audio comprende además un controlador configurado para proporcionar de forma selectiva, con dependencia de la señal de activación de distorsión de tiempo, una parte de contorno de distorsión de tiempo no constante hallada (variable) o información de la distorsión de tiempo, o una parte de contorno de distorsión de tiempo estándar constante (no-variable) o información de la distorsión de tiempo al transformador de distorsión de tiempo. De esta forma, es posible aceptar o rechazar de forma selectiva una parte de contorno de distorsión de tiempo no constante hallada en la derivación de la representación de la señal de audio codificada de la señal de audio de entrada.

10 Este concepto se basa en el hallazgo de que no siempre es eficaz introducir una información de distorsión de tiempo en una representación codificada de la señal de audio de entrada, porque un número importante de bits es necesario para codificar la información de distorsión de tiempo. Además, Se ha encontrado que la información de compactación de energía, la cual se calcula mediante el proveedor de la señal de activación de distorsión de tiempo, es una medida eficaz desde el punto de vista computacional para decidir respecto si es ventajoso proporcionar el transformador de distorsión de tiempo con la parte de contorno de distorsión de tiempo hallada variable (no-constante) o un contorno de distorsión de tiempo estándar (no-variable, constante). Debe observarse que cuando el transformador de distorsión de tiempo comprende una transformada superpuesta, una parte de contorno de distorsión de tiempo hallada puede usarse en el cálculo de dos o más bloques de transformada subsiguientes. En particular, Se ha encontrado que no es necesario codificar completamente tanto la versión de la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada utilizando la parte de contorno de distorsión de tiempo recientemente hallada variable como la versión de la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada utilizando una parte de contorno de distorsión de tiempo estándar (no-variable) con el fin de poder tomar una decisión si la distorsión en el tiempo permite un ahorro en la velocidad de bits o no. En cambio, Se ha encontrado que una evaluación de la compactación de energía de la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada forma una base fiable de la decisión. Por consiguiente, una velocidad de bits requerida puede mantenerse baja.

30 En otra realización preferida adicionalmente, el codificador de señal de audio comprende una interfaz de salida configurado para incluir de forma selectiva, con dependencia de la señal de activación de distorsión de tiempo, una información del contorno de distorsión de tiempo que representa un contorno de distorsión hallado variable de tiempo dentro de la representación codificada de la señal de audio. Por tanto, puede obtenerse una elevada eficacia de la codificación de señal de audio, sin tener en cuenta se la señal de entrada se ajusta bien para la distorsión de tiempo o no.

35 Una realización adicional según la invención crea un método para proporcionar una señal de activación de distorsión de tiempo sobre la base de una señal de audio. El método cumple la funcionalidad del proveedor de la señal de activación de distorsión de tiempo y puede complementarse por cualquiera de los rasgos y funcionalidad que se describen en el presente documento con respecto al proveedor de la señal de activación de distorsión de tiempo.

Otra realización según la invención crea un método para codificar una señal de audio de entrada, para obtener una representación codificada de la señal de audio de entrada. Este método puede complementarse por cualquiera de los rasgos y funcionalidades que se describen en el presente documento con respecto al codificador de señal de audio.

40 Otra realización según la invención crea un programa informático para realizar los métodos mencionados en el presente documento.

45 Según un primer aspecto de la presente invención, un análisis de la señal de audio, si una señal de audio tiene una característica armónica o una característica de voz se utiliza de forma ventajosa para controlar el procesamiento del relleno de ruido en el lado del codificador y/o en el lado del decodificador. El análisis de la señal de audio puede obtenerse fácilmente en un sistema, en el que se utiliza una funcionalidad de distorsión de tiempo, dado que esta funcionalidad de distorsión de tiempo típicamente comprende un seguidor de tono y/o un clasificador de señal para distinguir entre la voz por un lado y la música por el otro y/o para distinguir entre la voz sonora y la voz sorda. Dado que esta información está disponible en un contexto de este tipo sin cualquier costo adicional, la información disponible se utiliza de forma ventajosa para controlar el rasgo de relleno de ruido de modo que, especialmente para las señales de voz, un relleno de ruido entre líneas armónicas se reduce, o, para señales de voz en particular, incluso se elimina. Incluso en situaciones, donde se obtiene un contenido armónico fuerte, por una voz no se detecta directamente por parte del detector de voces, una reducción del relleno de ruido no obstante dará como resultado una calidad de mayor percepción. Aunque este rasgo es particularmente útil en un sistema, en el que cual el análisis armónico/de voz se lleva a cabo de cualquier modo, y esta información, por lo tanto, está disponible sin costos adicionales, el control del esquema de relleno de ruido se basa en un análisis de la señal, si la señal tiene una característica armónica o de voz o no, también es útil, incluso cuando un analizador de señal específico debe ser inserto en el sistema, dado que la calidad se mejora sin el aumento de la velocidad de bits o, se establece de forma alternativa, la velocidad de bits se reduce sin tener una pérdida de calidad, dado que los bits requeridos para codificar el nivel de relleno de ruido se reducen cuando el nivel de relleno de ruido en si mismo, el cual puede transmitirse desde un codificador a un decodificador, se reduce.

En un aspecto adicional de la presente invención, el resultado del análisis de señal es decir, si la señal es una señal armónica o una señal de voz se utiliza para controlar el procesamiento de la función ventana de un codificador de audio. Se ha encontrado que en una situación, en la que una señal de voz o una señal armónica se inician, la posibilidad es alta de modo que un codificador sencillo cambiará desde ventanas largas hasta ventanas cortas. Estas ventanas cortas, sin embargo, tienen una resolución de frecuencia correspondientemente reducida que, por otro lado, disminuiría la ganancia de codificación para las señales fuertemente armónicas y por lo tanto aumentan el número de bits necesarios para codificar parte de señal de este tipo. En vista de esto, la presente invención que se define en este aspecto utiliza ventanas más largas que una ventana corta cuando se detecta el inicio de una señal de voz o armónica. De forma alternativa, las ventanas se seleccionan con una longitud aproximadamente similar a las ventanas largas, pero con una superposición más corto con el fin de reducir de forma efectiva los preecos. En general, la característica de la señal, si la trama de tiempo de una señal de audio tiene una característica armónica o de voz se utiliza para seleccionar una función ventana para esta trama de tiempo.

Según un aspecto adicional de la presente invención, la herramienta de TNS (conformación de ruido temporal) se controla basándose en si la señal subyacente se basa en una operación de distorsión en el tiempo o está en un dominio lineal. Típicamente, una señal que se ha procesado por una operación de distorsión en el tiempo tendrá un contenido armónico fuerte. De otro modo, un seguidor de tono asociado con una etapa de distorsión en el tiempo no dará salida a un contorno de tono válido y, en ausencia de un contorno de tono válido de este tipo, una funcionalidad de distorsión en el tiempo habría sido desactivada para esta trama de tiempo de la señal de audio. Sin embargo, las señales armónicas, normalmente serán adecuadas para ser sometidas al procesamiento de TNS. El procesamiento de TNS es particularmente útil e induce una ganancia significativa en velocidad de bits/calidad, cuando la señal procesada por la etapa de TNS tiene un espectro bastante plano. Sin embargo, cuando la apariencia de la señal es tonal es decir, no plana, como es el caso de los espectros que tienen un contenido armónico o contenido sonoro, la ganancia en calidad/velocidad de bits proporcionada por la herramienta de TNS se reducirá. Por lo tanto, sin la modificación de la invención de la herramienta de TNS, las partes distorsionadas en el tiempo típicamente no serían procesadas por la TNS, pero serían procesadas sin un filtro de TNS. Por otro lado, el rasgo de conformación de ruido de TNS no obstante proporciona una calidad mejorada específicamente en situaciones, donde la señal es variable en amplitud/potencia. En casos, donde un inicio de una señal armónica o señal de voz está presente, y donde el rasgo de cambio de bloque se implementa de modo que, en lugar de este inicio, largas ventanas o al menos ventanas más largas que las ventanas cortas se mantienen, la activación del rasgo de conformación de ruido temporal para esta trama dará como resultado una concentración del ruido alrededor del inicio de voz lo cual reduce de forma efectiva los preecos, que podrían ocurrir antes del inicio de la voz debido a una cuantificación de la trama que se produce en el procesamiento del subsiguiente codificador.

Según un aspecto adicional de la presente invención, un número variable de líneas se procesa mediante un cuantificador/codificador de entropía dentro de un aparato de codificación de audio, con el fin de representar el ancho de banda variable, el cual se introduce de trama a trama debido a la realización de la operación de distorsión en el tiempo con una característica de distorsión en el tiempo variable/contorno de distorsión. Cuando la operación de distorsión en el tiempo da como resultado la situación que el tiempo de la trama (en términos lineales) incluido en una trama de distorsión de tiempo se incrementa, el ancho de banda de una simple línea de frecuencia se disminuye, y, para un ancho de banda completo constante, el número de líneas de frecuencia a ser procesadas debe incrementarse con respecto a una situación de no distorsión de tiempo. Cuando, por otro lado, la operación de distorsión de tiempo da como resultado el hecho de que el tiempo real de la señal de audio en el dominio distorsionado en el tiempo disminuye con respecto a la longitud del bloque de la señal de audio en el dominio lineal, el ancho de banda de la frecuencia de una línea de frecuencia simple se incrementa y, por lo tanto, el número de líneas procesadas por un codificador fuente debe ser incrementado con respecto a una situación de no distorsión en el tiempo con el fin de tener una variación de ancho de banda reducida u, óptimamente, sin variación de ancho de banda.

A continuación se describen realizaciones preferidas con respecto a los dibujos adjuntos, en los que:

La figura 1 muestra un diagrama esquemático de bloques de un proveedor de la señal de activación de distorsión de tiempo, según una realización de la invención;

La figura 2a muestra un diagrama esquemático de bloques de un codificador de señal de audio, según una realización de la invención;

La figura 2b muestra otro diagrama esquemático de bloques de un proveedor de la señal de activación de distorsión de tiempo según una realización de la invención;

La figura 3a muestra una representación gráfica de un espectro de una versión no distorsionada en el tiempo de una señal de audio;

La figura 3b muestra una representación gráfica de un espectro de una versión distorsionada en el tiempo de la señal de audio;

La figura 3c muestra una representación gráfica de un cálculo individual de medidas de planicidad espectral para

diferentes bandas de frecuencia;

La figura 3d muestra una representación gráfica de un cálculo de una medida de planicidad espectral que considera únicamente la parte de frecuencia más alta del espectro;

5 La figura 3e muestra una representación gráfica de un cálculo de una medida de planicidad espectral que utiliza una representación espectral en la cual se enfatiza una parte de frecuencia más alta sobre una parte de frecuencia más baja;

La figura 3f muestra un diagrama esquemático de bloques de un proveedor de información de compactación de energía, según otra realización de la invención;

10 La figura 3g muestra una representación gráfica de una señal de audio que tiene un tono temporalmente variable en el dominio de tiempo;

La figura 3h muestra una representación gráfica de una versión de distorsión de tiempo (remuestreo no-uniforme) de la señal de audio de la figura 3g;

La figura 3i muestra una representación gráfica de una función de autocorrelación de la señal de audio según la figura 3g;

15 La figura 3j muestra una representación gráfica de una función de autocorrelación de la señal de audio según la figura 3h;

La figura 3k muestra un diagrama esquemático de bloques de un proveedor de información de compactación de energía, según otra realización de la invención;

20 La figura 4a muestra un diagrama de flujo de un método para proporcionar una señal de activación de distorsión de tiempo sobre la base de una señal de audio;

La figura 4b muestra un diagrama de flujo de un método para codificar una señal de audio de entrada para obtener una representación codificada de la señal de audio de entrada, según una realización de la invención;

La figura 5a ilustra una realización preferida de un codificador de audio con aspectos de la invención;

La figura 5b ilustra una realización preferida de un decodificador de audio con aspectos de la invención;

25 La figura 6a ilustra una realización preferida del aspecto de relleno de ruido de la presente invención;

La figura 6b ilustra una tabla que define la operación de control llevada a cabo por el manipulador de nivel de relleno de ruido;

La figura 7a ilustra una realización preferida para llevar a cabo un cambio de bloque basado en la distorsión de tiempo según la presente invención;

30 La figura 7b ilustra una realización alternativa para influir en la función ventana;

La figura 7c ilustra realización alternativa adicional para ilustrar la función ventana basándose en la información de distorsión de tiempo;

La figura 7d ilustra una secuencia de generación de ventanas de un comportamiento de AAC normal en un inicio sonoro;

La figura 7e ilustra secuencias de ventanas alternativas según una realización preferida de la presente invención;

35 La figura 8a ilustra la realización preferida de un control en base a la distorsión de tiempo de la herramienta de TNS (conformación de ruido temporal);

La figura 8b ilustra una tabla que define los procedimientos de control realizados en el generador de la señal del control de umbral en la figura 8a;

40 La figura 9a-9e ilustra diferentes características de distorsión en el tiempo y la correspondiente influencia sobre el ancho de banda de la señal de audio que se produce a continuación de una función de sin distorsión de tiempo en el lado del decodificador;

La figura 10a ilustra una realización preferida de un controlador para controlar varias líneas dentro de un procesador de codificación;

45 La figura 10b ilustra la dependencia entre varias líneas que van a descartarse/agregarse para una velocidad de muestreo;

La figura 11 ilustra una comparación entre una escala de tiempo lineal y una escala de tiempo distorsionado;

La figura 12a ilustra una implementación en el contexto de la extensión del ancho de banda; y

La figura 12b ilustra una tabla que muestra la dependencia entre la velocidad de muestreo local en el dominio distorsionado en el tiempo y el control de coeficientes espectrales.

5 La figura 1 muestra un diagrama esquemático de bloques del proveedor de la señal de activación de distorsión de tiempo, según una realización de la invención. El proveedor 100 de la señal de activación de distorsión de tiempo está configurado para recibir una representación 110 de una señal de audio y proporcionar, sobre la base del mismo, una
 10 señal 112 de activación de distorsión de tiempo. El proveedor 100 de la señal de activación de distorsión de tiempo comprende un proveedor 120 de información de compactación de energía, el cual está configurado para proporcionar una información 122 de compactación de energía, que describe una compactación de energía en una representación
 15 espectral transformada por la distorsión de tiempo de la señal de audio. El proveedor 100 de la señal de activación de distorsión de tiempo comprende además un comparador 130 configurado para comparar la información 122 de compactación de energía con un valor 132 de referencia, y proporcionar la señal 112 de activación de distorsión de tiempo con dependencia del resultado de la comparación.

20 Tal como se mencionó anteriormente, se ha encontrado que la información de compactación de energía es una información valiosa la cual permite una estimación eficaz desde el punto de vista computacional si una distorsión de tiempo otorga un ahorro de bits o no. Se ha encontrado que la presencia de ahorro de bits está íntimamente
 25 correlacionada con la cuestión si la distorsión de tiempo da como resultado una compactación de energía o no.

La figura 2a muestra un diagrama esquemático de bloques de un codificador 200 de señal de audio, según una
 30 realización de la invención. El codificador 200 de señal de audio está configurado para recibir una señal 210 de audio de entrada (también designada para $a(t)$) y proporcionar, sobre la base del mismo, una representación 212 codificada de la señal 210 de audio de entrada. El codificador 200 de señal de audio comprende un transformador 220 de distorsión de tiempo, el cual está configurado para recibir la señal 210 de audio de entrada (la cual puede estar representada en un
 35 dominio de tiempo) y proporcionar, sobre la base del mismo, una representación 222 espectral transformada por la distorsión de tiempo de la señal 210 de audio de entrada. El codificador 200 de señal de audio comprende además un analizador 284 de distorsión de tiempo, el cual está configurado para analizar la señal 210 de audio de entrada y proporcionar, sobre la base del mismo, una información 286 del contorno de distorsión de tiempo (por ejemplo información absoluta o relativa del contorno de distorsión de tiempo).

El codificador 200 de señal de audio comprende además un mecanismo de cambio, por ejemplo en forma de un cambio
 40 240 controlado, para decidir respecto si la información 286 hallada del contorno de distorsión de tiempo o una información 288 estándar del contorno de distorsión de tiempo se utiliza para un procesamiento adicional. Por tanto, el mecanismo 240 de cambio está configurado para proporcionar de forma selectiva, con dependencia de una información de activación de distorsión de tiempo, o bien la información 286 hallada del contorno de distorsión de tiempo o bien una
 45 información 288 estándar del contorno de distorsión de tiempo como información 242 nueva del contorno de distorsión de tiempo, para un procesamiento adicional, por ejemplo para un transformador 220 de distorsión de tiempo. Debe observarse, que un transformador 220 de distorsión de tiempo puede utilizar por ejemplo la información 242 nueva del contorno de distorsión de tiempo (por ejemplo una nueva parte de contorno de distorsión de tiempo) y, además, una información de distorsión de tiempo previamente obtenida (por ejemplo una o más partes de contorno de distorsión de tiempo previamente obtenidas) para la distorsión en el tiempo de una trama de audio. El pos-procesamiento del espectro opcional puede comprender por ejemplo una conformación de ruido temporal y/o un análisis de relleno de ruido. El codificador 200 de señal de audio también comprende un cuantificador/codificador 260, el cual está configurado para recibir la representación 222 espectral (procesada opcionalmente por el postprocesamiento 250 del espectro) y para cuantificar y codificar la representación 222 espectral transformada. Para este fin, el cuantificador/codificador 260 puede acoplarse con un modelo 270 perceptivo y recibir una información 272 de relevancia perceptiva desde el modelo
 50 270 perceptivo, para considerar un enmascaramiento perceptivo y ajustar las precisiones de cuantificación en diferentes cajas de frecuencia según la percepción humana. El codificador 200 de señal de audio comprende además una interfaz 280 de salida la cual está configurada para proporcionar la representación 212 codificada de la señal de audio sobre la base de la representación 262 cuantificada y codificada suministrada por el cuantificador/codificador 260.

El codificador 200 de señal de audio comprende además un proveedor 230 de la señal de activación de distorsión de tiempo, que está configurado para proporcionar una señal 232 de activación de distorsión de tiempo. La señal 232 de activación de distorsión de tiempo, por ejemplo, puede utilizarse para controlar el mecanismo 240 de cambio, para decidir respecto si la información 286 del contorno de distorsión de tiempo recientemente hallada o una información 288 estándar del contorno de distorsión de tiempo se utiliza en otros pasos de procesamiento (por ejemplo por un transformador 220 de distorsión de tiempo). Además, la información 232 de activación de distorsión de tiempo puede utilizarse en un conmutador 280 para decidir respecto si la información 242 del contorno de distorsión de tiempo nueva seleccionada (seleccionada de la información 286 del contorno de distorsión de tiempo recientemente hallada y la información estándar del contorno de distorsión de tiempo) está incluida en la representación 212 codificada de la señal 210 de audio de entrada. Típicamente, la información del contorno de distorsión de tiempo sólo se incluye en la representación 212 codificada de la señal de audio si la información del contorno de distorsión de tiempo seleccionada

describe un contorno de distorsión de tiempo no constante (variable). Además, la información 232 de activación de distorsión de tiempo puede en si misma estar incluida en la representación 212 codificada, por ejemplo en forma de una marca de un bit que indica la activación o desactivación de la distorsión de tiempo.

5 Con el fin de facilitar el entendimiento, debe observarse que un transformador 220 de distorsión de tiempo típicamente comprende un generador 220a de ventana de análisis, un remuestreador o "distorsionador 220b de tiempo" y un transformador de dominio espectral (o convertidor 220c de tiempo/frecuencia). Dependiendo de la implementación, sin embargo, el distorsionador 220b de tiempo puede ser colocado – en una dirección de procesamiento de señal – antes del generador 220a de ventana de análisis. Sin embargo, la distorsión de tiempo y el dominio de tiempo para la transformación del dominio espectral pueden combinarse en una sola unidad en algunas realizaciones.

10 A continuación se describirán los detalles con relación al funcionamiento del proveedor 230 de la señal de activación de distorsión de tiempo. Debe observarse que el proveedor 230 de la señal de activación de distorsión de tiempo puede ser equivalente al proveedor 100 de la señal de activación de distorsión de tiempo.

15 El proveedor 230 de la señal de activación de distorsión de tiempo está configurado preferiblemente para recibir la representación 210 de la señal de audio del dominio de tiempo (también designada con $a(t)$), la información 286 del contorno de distorsión de tiempo recientemente hallada, y la información 288 estándar del contorno de distorsión de tiempo. El proveedor 230 de la señal de activación de distorsión de tiempo está configurado además para obtener, utilizando la señal de audio del dominio de tiempo 210, la información 286 del contorno de distorsión de tiempo recientemente hallada y la información 288 estándar del contorno de distorsión de tiempo, una información de compactación de energía que describe una compactación de energía debido a la información 286 del contorno de distorsión de tiempo recientemente hallada, y proporcionar la señal 232 de activación de distorsión de tiempo sobre la base de esta información de compactación de energía.

20 La figura 2b muestra un diagrama esquemático de bloques de un proveedor 234 de la señal de activación de distorsión de tiempo, según una realización de la invención. El proveedor 234 de la señal de activación de distorsión de tiempo puede tomar la función del proveedor 230 de la señal de activación de distorsión de tiempo en algunas realizaciones. El proveedor 234 de la señal de activación de distorsión de tiempo está configurado para recibir una señal 210 de audio de entrada, y dos informaciones 286 y 288 del contorno de distorsión de tiempo, y proporcionar, sobre la base del mismo, una señal 234p de activación de distorsión de tiempo. La señal 234p de activación de distorsión de tiempo puede tomar la función de la señal 232 de activación de distorsión de tiempo. El proveedor de la señal de activación de distorsión de tiempo comprende dos proveedores 234a, 234g de la representación de la distorsión de tiempo idénticos, los cuales están configurados para recibir la señal 210 de audio de entrada y la información 286 y 288 del contorno de distorsión de tiempo respectivamente y proporcionar, sobre la base del mismo, dos representaciones 234e y 234k de distorsión de tiempo, respectivamente. El proveedor 234 de la señal de activación de distorsión de tiempo comprende además dos proveedores 234f y 234l de información de compactación de energía idénticos, los cuales están configurados para recibir las representaciones 234e y 234k de distorsión de tiempo, respectivamente, y, sobre la base del mismo, proporcionar la información 234m y 234n de compactación de energía, respectivamente. El proveedor de la señal de activación de distorsión de tiempo comprende además un comparador 234o, configurado para recibir la información 234m y 234n de compactación de energía, y, sobre la base del mismo proporcionar la señal 234p de activación de distorsión de tiempo.

30 Con el fin de facilitar el entendimiento, debe observarse que los proveedores 234a y 234g de representación de la distorsión de tiempo típicamente comprenden (opcional) idénticos generadores 234b y 234h de ventana de análisis, remuestreadores o distorsionadores de tiempo 234c y 234i idénticos, y (opcional) transformadores 234d y 234j del dominio espectral idénticos.

A continuación, se mencionarán diferentes conceptos para obtener la información de compactación de energía. De antemano, se dará una introducción que explica el efecto de distorsión de tiempo sobre una señal de audio típica.

45 A continuación, el efecto de distorsión en el tiempo sobre una señal de audio se describirá tomando como referencia las figuras 3a y 3b. La figura 3a muestra una representación gráfica de un espectro de una señal de audio. Una abscisa 301 describe una frecuencia y una ordenada 302 describe la intensidad de la señal de audio. Una curva 303 describe una intensidad de la señal de audio no distorsionada en el tiempo como una función de la frecuencia f

50 La figura 3b muestra una representación gráfica de un espectro de una versión de distorsionada en el tiempo de la señal de audio representada en la figura 3a. Nuevamente, una abscisa 306 describe una frecuencia y una ordenada 307 describe la intensidad de la versión distorsionada de la señal de audio. Una curva 308 describe la intensidad de la versión distorsionada en el tiempo de la señal de audio sobre la frecuencia. Tal como puede observarse a partir de una comparación de la representación gráfica de las figuras 3a y 3b, la versión no distorsionada en el tiempo (sin distorsionar" de la señal de audio comprende un espectro emborronado, en particular en una región de frecuencia más alta. A diferencia, la versión distorsionada en el tiempo de la señal de audio de entrada comprende un espectro que tiene picos espectrales claramente distinguibles, incluso en la región de frecuencia más alta. Además, un moderado afilado de los picos espectrales incluso puede observarse en la región espectral más baja de la versión distorsionada en el tiempo de la señal de audio de entrada.

5 Debe observarse que el espectro de la versión de la distorsión de tiempo de la señal de audio de entrada, el cual se muestra en la figura 3b, puede cuantificarse y codificarse, por ejemplo mediante el cuantificador/codificador 260, con una velocidad de bits más baja que el espectro de la señal de audio de entrada sin distorsionar mostrado en la figura 3a. Esto se debe al hecho de que un espectro emborronado comprende típicamente un gran número de coeficientes espectrales perceptualmente relevantes (es decir un número relativamente pequeño de coeficientes espectrales cuantificados en cero o cuantificados en valores pequeños), mientras que un espectro "menos plano" tal como se muestra en la figura 3 típicamente comprende un mayor número de coeficientes espectrales cuantificado en cero o cuantificados en valores pequeños. Los coeficientes espectrales cuantificados en cero o cuantificados en valores pequeños pueden codificarse con menos bits que los coeficientes espectrales cuantificados a valores superiores, de modo que el espectro de la figura 3b puede codificarse utilizando menos bits que el espectro de la figura 3a.

10 No obstante, debe observarse además que el uso de una distorsión de tiempo no siempre da como resultado una mejora significativa de la eficacia de codificación de la señal distorsionada en el tiempo. Por consiguiente, en algunos casos, el precio, en cuanto a velocidad de bits, requerido para la codificación de la información de distorsión de tiempo (por ejemplo contorno de distorsión de tiempo) puede exceder los ahorros, en cuanto a velocidad de bits, para codificar el espectro transformado por la distorsión de tiempo (cuando se compara con el espectro transformado sin distorsión de tiempo). En este caso, es preferible proporcionar la representación codificada de la señal de audio utilizando un contorno de distorsión de tiempo estándar (no-variable) para controlar la transformación de la distorsión de tiempo. En consecuencia, la transmisión de cualquier información de distorsión de tiempo (es decir, información del contorno de distorsión de tiempo) puede omitirse (salvo una marca que indica la desactivación de la distorsión de tiempo), manteniendo de este modo la velocidad de bits baja.

15 A continuación, diferentes conceptos para un cálculo fiable y eficaz desde el punto de vista computacional de una señal 112, 232, 234p de activación de distorsión de tiempo se describirá tomando como referencia las figuras 3c-3k. Sin embargo, antes de eso, el antecedente del concepto de la invención se resumirá brevemente.

20 La presunción básica es que la aplicación de la distorsión de tiempo sobre una señal armónica con un tono variable hace el tono constante, y al hacer el tono constante mejora la codificación de los espectros obtenidos por una siguiente transformación de la frecuencia de tiempo, porque en lugar del emborronado de las diferentes armonías sobre varias cajas espectrales (ver las figuras 3a) sólo un número limitado de líneas significativas permanecen (ver la figura 3b). Sin embargo, incluso cuando se detecta una variación de tono, la mejora en la ganancia de codificación (es decir la cantidad de bits ahorrados) puede ser insignificante (por ejemplo si uno tiene un ruido fuerte subyacente a la señal armónica, o si la variación es tan pequeña que el emborronado de las armonías más altas no es un problema), o puede ser inferior a la cantidad de bits necesarios para transferir el contorno de distorsión de tiempo al decodificador, o simplemente puede ser incorrecto. En estos casos, es preferible rechazar el contorno de distorsión de tiempo variable (por ejemplo 286) producido por un codificador del contorno de distorsión de tiempo y en su lugar utilizar una señalización de un bit eficaz, señalizando un contorno de distorsión de tiempo estándar (no-variable).

25 El alcance de la presente invención comprende la creación de un método para decidir si una parte de contorno de distorsión de tiempo obtenida proporciona suficiente ganancia de codificación (por ejemplo suficiente ganancia de codificación para compensar la sobrecarga requerida para la codificación para el contorno de distorsión de tiempo).

30 Tal como se estableció anteriormente, el aspecto más importante de la distorsión de tiempo es la compactación de la energía espectral para un número más pequeño de líneas (véanse las figuras 3a y 3b). Una mirada a ésta muestra que una compactación de energía también corresponde a un espectro más "sin planicidad" (véanse las figuras 3a y 3b), dado que la diferencia entre los picos y los valles del espectro se incrementa. La energía se concentra en menos líneas teniendo las líneas entre aquéllos menos energía que antes.

35 Las figuras 3a y 3b muestran un ejemplo esquemático con un espectro sin distorsión de una trama con fuertes armonías y variación de tono (figura 3a) y el espectro de la versión distorsionada en el tiempo de la misma trama (figura 3b).

40 En vista de esta situación, se ha encontrado que es ventajoso utilizar la medida de planicidad espectral como una medida posible para la eficacia de la distorsión de tiempo.

45 La planicidad espectral puede calcularse, por ejemplo, dividiendo la media geométrica del espectro de potencia por la media aritmética del espectro de potencia. Por ejemplo, la planicidad espectral (también designada brevemente "planicidad") puede calcularse según la siguiente ecuación:

50

$$\text{Planicidad} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\left(\frac{\sum_{n=0}^{N-1} x(n)}{N}\right)}$$

En lo anterior, $x(n)$ representa la magnitud de un número binario n . Además, en lo anterior, N representa un número total de cajas espectrales consideradas para el cálculo de la medida de planicidad espectral.

5 En una realización de la invención, el cálculo mencionado anteriormente de la “planicidad”, que puede servir como una información de compactación de energía, puede llevarse a cabo utilizando las representaciones 234e, 234k del espectro transformado por la distorsión de tiempo, de modo que puede sostenerse la siguiente relación:

$$x(n) = |X|_{tw}(n).$$

10 En este caso, N puede ser igual al número de líneas espectrales proporcionadas por el transformador 234d, 234j de dominio espectral y $|X|_{tw}(n)$ es una representación 234e, 234k espectral transformada de la distorsión de tiempo.

15 Aunque la medida espectral es una cantidad útil para la provisión de la señal de activación de distorsión de tiempo, una desventaja de la medida de planicidad espectral, al igual que la medida de relación señal a ruido (SNR), es que si se aplica al espectro total, enfatiza partes con energía más alta. Normalmente, los espectros armónicos tienen una cierta inclinación espectral, lo que significa que la mayor parte de la energía se concentra en los primeros pocos tonos parciales y entonces disminuye con la frecuencia creciente, conduciendo a una subrepresentación de los parciales más elevados en la medida. Esto no se desea en algunas realizaciones, dado que intenciones desea mejorar la calidad de estos parciales más altos, porque se emborronan la mayoría de las veces (véase la figura 3a). A continuación, se describirán varios conceptos opcionales para la mejora de la relevancia de la medida de planicidad espectral.

20 En una realización según la invención, se elige un enfoque similar a la denominada medida de “SNR segmental”, que conduce a una medida de planicidad espectral banda por banda. Un cálculo de la medida de planicidad espectral se realiza (por ejemplo de forma separada) dentro de un número de bandas, y se toma la principal (o medio). Las diferentes bandas podrían tener igual ancho de banda. Sin embargo, preferiblemente, los anchos de banda pueden seguir una escala perceptiva, al igual que las bandas críticas, o corresponder, por ejemplo, a las bandas del factor de escala de la denominada “codificación de audio avanzada”, también conocida como AAC.

25 El concepto mencionado anteriormente se explicará brevemente a continuación, tomando como referencia la figura 3c, que muestra una representación gráfica de un cálculo individual de medidas de planicidad espectral para diferentes bandas de frecuencia. Tal como puede observarse, el espectro puede dividirse en diferentes bandas 311, 312, 313 de frecuencia, que pueden un ancho de banda igual o que pueden tener diferentes anchos de banda. Por ejemplo, una primera medida de planicidad espectral puede calcularse para la primera banda 311 de frecuencia, por ejemplo, utilizando la ecuación para la “planicidad” dada anteriormente. En este cálculo, pueden considerarse las cajas de frecuencia de la primera banda de frecuencia (n variable en funciones puede tomar los índices de caja de frecuencia de las cajas de frecuencia de la primera banda de frecuencia), y puede considerarse el ancho de la primera banda 311 de frecuencia (N variable puede tomar el ancho en término de cajas de frecuencia de la primera banda de frecuencia). Por consiguiente, se obtiene una medida de planicidad para la primera banda 311 de frecuencia. De manera similar, una medida de planicidad puede calcularse para la segunda banda 312 de frecuencia, teniendo en cuenta las cajas de frecuencia de las segundas bandas 312 de frecuencia y además el ancho de la segunda banda de frecuencia. Además, las medidas de planicidad de las bandas de frecuencia adicionales, del tipo tercera banda de frecuencia 313, pueden calcularse de la misma forma.

40 A continuación, un promedio de las medidas de plenitud para diferentes bandas 311, 312, 313 de frecuencia se puede calcular y el promedio puede ser útil como la información de compactación de energía.

45 Otro enfoque (para la mejora de la derivación de la señal de activación de distorsión de tiempo) es aplicar la medida de planicidad espectral sólo por encima de una cierta frecuencia. Un enfoque de este tipo se ilustra en la figura 3b. Tal como puede observarse, sólo cajas de frecuencia en una parte 316 de frecuencia superior de los espectros se consideran para un cálculo de la medida de planicidad espectral. Una parte de frecuencia inferior del espectro es insignificante para el cálculo de la medida de planicidad espectral. La parte 316 de frecuencia más alta puede considerarse banda de frecuencia por banda de frecuencia para el cálculo de la medida de planicidad espectral. De forma alternativa, la parte 316 completa de frecuencia más alta puede considerarse en su totalidad para el cálculo de la medida de planicidad espectral.

50 Para resumir lo anterior, puede establecerse que la disminución en la planicidad espectral (originada por la aplicación de la distorsión de tiempo) puede considerarse como una primera medida para la eficacia de la distorsión de tiempo.

55 Por ejemplo, el proveedor 100, 230, 234 de la señal de activación de distorsión de tiempo (o el comparador 130, 234° del mismo) puede comparar la medida de planicidad espectral de la representación 234e espectral transformada por la distorsión de tiempo con una medida de planicidad espectral de la representación 234k espectral transformada por la distorsión de tiempo utilizando una información estándar del contorno de distorsión de tiempo, y para decidir sobre la base de dicha comparación si la señal de activación de distorsión de tiempo debe ser activa o inactiva. Por ejemplo, la

distorsión de tiempo se activa por medio de una configuración apropiada de la señal de activación de distorsión de tiempo si la distorsión en el tiempo da como resultado una reducción suficiente de la medida de planicidad espectral cuando se compara con un caso sin distorsión en el tiempo.

5 Además de los enfoques mencionados anteriormente, la parte de frecuencia superior del espectro puede enfatizarse (por ejemplo por un escalamiento apropiado) sobre la parte de frecuencia inferior para el cálculo de la medida de planicidad espectral. La figura 3c muestra una representación gráfica de un espectro transformado por la distorsión de tiempo en que una parte de frecuencia más alta se enfatiza sobre una parte de frecuencia más baja. Por consiguiente, se compensa una subrepresentación de parciales más altos en el espectro. Por tanto, la medida de planicidad puede
10 calcularse sobre el espectro escalado completo en el que las cajas de frecuencia más alta se enfatizan sobre las cajas de frecuencia más baja, tal como se muestra en la figura 3e.

En cuanto a ahorros de bits, una medida típica para codificar la eficacia sería la entropía perceptiva, que puede definirse de un modo que se correlacione muy bien con el número real de bits necesarios para codificar un cierto espectro tal como se describe en 3GPP TS 26.403 V7.0.0: 3rd Generation Partnership Project; Technical Specification Group
15 Services and System Aspects; General audio codec audio processing functions; Enhanced aacPlus general audio codec; Encoder specification AAC part: Sección 5.6.1.1.3 Relation between bit demand and perceptual entropy. Como resultado, la reducción de la entropía perceptiva es otra medida para determinar como sería la eficacia de la distorsión en el tiempo.

La figura 3f muestra un proveedor 325 de información de compactación de energía, que puede tomar el lugar del proveedor 120, 234f, 234l de información de compactación de energía, y el cual puede utilizarse en los proveedores
20 100, 290, 234 de la señal de activación de distorsión de tiempo. El proveedor 325 de información de compactación de energía está configurado para recibir una representación de la señal de audio, por ejemplo, en forma de una representación 234e, 234k espectral transformada por la distorsión de tiempo, también designado con $|X|_{tw}$. El proveedor 325 de información de compactación de energía también está configurado para proporcionar una información 326 de entropía perceptiva, que puede tomar el lugar de la información 122, 234m, 234n de compactación de energía.

25 El proveedor 325 de información de compactación de energía comprende un calculador 327 del factor de forma, que está configurado para recibir la representación 234e, 234k espectral transformada por la distorsión de tiempo y proporcionar, sobre la base del mismo, una información 328 del factor de forma, que puede estar asociada con una banda de frecuencia. El proveedor 325 de información de compactación de energía comprende además un calculador 329 de energía de la banda de frecuencia, que está configurado para calcular una información (330) de energía de la
30 banda de frecuencia en(n) sobre la base de la representación 234e, 234k espectral de distorsión de tiempo. El proveedor 325 de información de compactación de energía comprende además un estimador 331 de número de líneas, el cual está configurado para proporcionar una información estimada del número de líneas nl (332) para una banda de frecuencia que tiene índice n. Además, el proveedor 325 de información de compactación de energía comprende un
35 calculador 333 de entropía perceptiva, que está configurado para calcular la información 326 de entropía perceptiva sobre la base de la información 330 de energía de la banda de frecuencia y de la información 332 estimada del número de líneas. Por ejemplo, el calculador 327 del factor de forma puede estar configurado para calcular el factor forma según

$$ffac(n) = \sum_{k=kOffset(n)}^{kOffset(n+1)-1} \sqrt{|X(k)|}$$

(1)

40 En la ecuación anterior, ffac(n) designa el factor forma para la banda de frecuencia que tiene un índice de banda de frecuencia n. k designa una variable en funcionamiento, que funciona sobre los índices de caja espectral de la banda de factor de escala (o banda de frecuencia) n. X(k) designa un valor espectral (por ejemplo, un valor de energía o un valor de magnitud) de la caja espectral (o caja de frecuencia) que tiene un índice de caja espectral (o un índice de caja de frecuencia) k.

45 El estimador del número de líneas puede estar configurado para estimar el número de líneas que no son cero, designadas con nl, según la siguiente ecuación:

$$nl = \frac{ffac(n)}{\left(\frac{en(n)}{kOffset(n+1)-kOffset(n)}\right)^{0.25}}$$

(2)

50 En la ecuación anterior, en(n) designa una energía en la banda de frecuencia o banda de factor de escala n. kOffset(n+1)-kOffset(n) designa un ancho de la banda de frecuencia o banda de factor de escala de índice n en cuanto

a cajas de frecuencia.

Además, el calculador 332 de entropía perceptiva puede estar configurado para calcular la información de entropía perceptiva $sfbPe$ según la siguiente ecuación:

$$sfbPe = nl \cdot \begin{cases} \log_2\left(\frac{en}{thr}\right) & \text{for } \log_2\left(\frac{en}{thr}\right) \geq c1 \\ (c2 + c3 \cdot \log_2\left(\frac{en}{thr}\right)) & \text{for } \log_2\left(\frac{en}{thr}\right) < c1 \end{cases} \quad (3)$$

En lo anterior, pueden sostenerse las siguientes relaciones:

$$c1 = \log_2(8) \quad c2 = \log_2(2.5) \quad c3 = 1 - c2 / c1 \quad (4)$$

Una entropía perceptiva total pe puede calcularse como la suma de las entropías perceptivas de múltiples bandas de frecuencia o bandas de factor de escala.

Tal como se mencionó anteriormente, la información 326 de entropía de percepción puede utilizarse como una información de compactación de energía.

Para más detalles que se refieren al cálculo de la entropía perceptiva, se hace referencia a la sección 5.6.1.1.3 del International Standard "3GPP TS 26.403 V7.0.0(2006-06)".

A continuación, se describirá un concepto para el cómputo de la información de compactación de energía en el dominio de tiempo.

Otra mirada al TW-MDCT (transformada de coseno discreto modificada por la distorsión de tiempo) es la idea básica para cambiar la señal de una manera que tenga un tono constante o casi constante dentro de un bloque. Si se obtiene un tono constante, esto significa que la máxima de la autocorrelación de un bloque en proceso aumenta. Dado que no es trivial encontrar la máxima correspondiente en la autocorrelación para el caso de distorsión en el tiempo y sin distorsión de tiempo, la suma de los valores absolutos para la autocorrelación normalizada puede utilizarse como medida para la mejora. Un aumento en esta suma corresponde a un aumento en la compactación de energía.

Este concepto se explicará en más detalles a continuación, tomando como referencia las figuras 3g, 3h, 3i, 3j y 3k.

La figura 3g muestra una representación gráfica de una señal no distorsionada en el tiempo en el dominio de tiempo. Una abscisa 350 describe el tiempo, y una ordenada 351 describe un nivel $a(t)$ de la señal de tiempo no distorsionada en el tiempo. Una curva 352 describe la evolución temporal de la señal de tiempo no distorsionada en el tiempo. Se presume que la frecuencia de la señal no distorsionada en el tiempo descrita por la curva 352 aumenta a lo largo del tiempo, tal como puede observarse en la figura 3g.

La figura 3h muestra una representación gráfica de una versión distorsionada en el tiempo de la señal de tiempo de la figura 3g. Una abscisa 355 describe el tiempo distorsionado (por ejemplo, de forma normalizada) y una ordenada 356 describe el nivel de la versión distorsionada en el tiempo $a(t_w)$ de la señal $a(t)$. Tal como puede observarse en la figura 3h, la versión distorsionada en el tiempo $a(t_w)$ de la señal de tiempo no distorsionada en el tiempo $a(t)$ comprende (al menos aproximadamente) una frecuencia temporalmente constante en el dominio de distorsión de tiempo.

En otras palabras, la figura 3h ilustra el hecho de que una señal de tiempo de una frecuencia temporalmente variable se transforma en una señal de tiempo de una frecuencia temporalmente constante mediante una operación de distorsión de tiempo apropiada, que puede comprender un remuestreo de la distorsión de tiempo.

La figura 3i muestra una representación gráfica de una función de autocorrelación de la señal de tiempo no distorsionada en el tiempo $a(t)$. Una abscisa 360 describe un retardo en la autocorrelación τ , y una ordenada 361 describe una magnitud de la función de autocorrelación. Las marcas 362 describen una evolución de la autocorrelación $R_{uw}(\tau)$ como función del retardo de la autocorrelación τ . Tal como puede observarse a partir de la figura 3i, la función de autocorrelación R_{uw} de la señal de tiempo no distorsionada en el tiempo $a(t)$ comprende un pico para $\tau = 0$ (que refleja la energía de la señal $a(t)$) y toma pequeños valores para $\tau \neq 0$.

La figura 3j muestra una representación gráfica de la función de autocorrelación R_{tw} de la señal de tiempo de distorsión de tiempo $a(t_w)$. Tal como puede observarse a partir de la figura 3j, la función de autocorrelación R_{tw} comprende un pico para $\tau = 0$, y también comprende picos para otros valores τ_1, τ_2, τ_3 del retardo de la autocorrelación τ . Estos picos

adicionales para τ_1 , τ_2 , τ_3 se obtienen por el efecto de la distorsión de tiempo para aumentar la periodicidad de la señal de tiempo de distorsión de tiempo $a(t_w)$. Esta periodicidad se refleja mediante los picos adicionales de la función de autocorrelación $R_{tw}(\tau)$ cuando se compara con la función de autocorrelación $R_{uw}(\tau)$. Por tanto, la presencia de picos adicionales (o la intensidad aumentada de picos) de la función de autocorrelación de la señal de audio distorsionada en el tiempo, cuando se compara con la función de autocorrelación de la señal de audio original puede utilizarse como una indicación de la efectividad (en cuanto a una reducción de la velocidad de bits) de la distorsión de tiempo.

La figura 3k muestra un diagrama esquemático de bloques de un proveedor 370 de información de compactación de energía configurado para recibir una representación de dominio de tiempo de distorsión de tiempo de la señal de audio, por ejemplo, la señal 234e, 234k de distorsión de tiempo (donde la transformación 234d, 234j de dominio espectral y opcionalmente el generador 234b y 234h de ventana de análisis se omite), y proporcionar, sobre la base del mismo, una información 374 de compactación de energía, que puede tomar la función de la información 372 de compactación de energía. El proveedor 370 de información de compactación de energía de la figura 3k comprende un calculador de autocorrelación 371 configurado para calcular la función de autocorrelación $R_{tw}(\tau)$ de la señal de distorsión de tiempo $a(t_w)$ a lo largo de un rango predeterminado de valores discretos de τ . El proveedor 370 de información de compactación de energía comprende además un sumador 372 de autocorrelación configurado para sumar una pluralidad de valores de la función de autocorrelación $R_{tw}(\tau)$ (por ejemplo, a lo largo de un rango predeterminado de valores discretos de τ) y proporcionar la suma obtenida como la información 122, 234m, 234n de compactación de energía.

Por tanto, el proveedor 370 de información de compactación de energía permite la provisión de información fiable que indica la eficacia de la distorsión de tiempo sin realizar realmente la transformación de dominio espectral de la versión del dominio de tiempo de distorsión de tiempo de la señal 210 de audio de entrada. Por lo tanto, es posible realizar una transformación de dominio espectral de la versión de la distorsión de tiempo de la señal 310 de audio de entrada sólo si se encuentra, sobre la base de la información 122, 234m, 234n de compactación de energía proporcionada por el proveedor 370 de información de compactación de energía, que la distorsión de tiempo realmente otorga una eficacia de codificación mejorada. Para resumir lo anterior, las realizaciones según la invención crean un concepto para una revisión final de la calidad. Un contorno de tono resultante (utilizado en un codificador de señal de audio de distorsión de tiempo) se evalúa en cuanto a su ganancia de codificación y si se acepta o se rechaza. Varias medidas referentes a la escasez de ganancia de codificación o del espectro pueden tenerse en cuenta para esta decisión, por ejemplo, una medida de planicidad espectral, una medida de plenitud espectral segmental banda por banda, y/o una entropía perceptiva.

El uso de diferente información de compactación espectral se ha discutido, por ejemplo, el uso de una medida de planicidad espectral, el uso de una medida de entropía perceptiva, y el uso de una medida de autocorrelación de dominio de tiempo. No obstante, existen otras medidas que muestran una compactación de la energía en un espectro distorsionado en el tiempo.

Todas estas medidas pueden utilizarse. Preferiblemente, para todas estas medidas, se define una relación entre la medida para espectro sin distorsión de tiempo y uno con distorsión de tiempo, y se establece un umbral para esta relación en el codificador para determinar si un el contorno de distorsión de tiempo obtenido tiene un beneficio en la codificación o no.

Todas estas medidas pueden aplicarse a una trama completa, donde sólo la tercera parte del contorno de tono es nuevo (donde, por ejemplo, tres partes del contorno de tono están asociadas con la trama completa), o preferiblemente sólo para la parte de la señal, para que se obtuvo esta nueva parte, por ejemplo, utilizando una transformada con una ventana superpuesta baja centrada sobre la parte de señal (respectiva).

Naturalmente, una sola medida o una combinación de las medidas antes mencionadas pueden utilizarse según se desee.

La figura 4a muestra un diagrama de flujo de un método para proporcionar una señal de activación de distorsión de tiempo sobre la base de una señal de audio. El método 400 de la figura 4a comprende un paso 410 para proporcionar una información de compactación de energía que describe una compactación de energía en una representación espectral transformada por la distorsión de tiempo de la señal de audio. El método 400 comprende además un paso 420 para comparar la información de compactación de energía con un valor de referencia. El método 400 comprende además un paso 430 para proporcionar la señal de activación de distorsión de tiempo con dependencia del resultado de la comparación.

El método 400 puede complementarse por cualquiera de los rasgos y funcionalidades que se describen en el presente documento con respecto a la provisión de la señal de activación de distorsión de tiempo.

La figura 4b muestra un diagrama de flujo de un método para codificar una señal de audio de entrada para obtener una representación codificada de la señal de audio de entrada. El método 450 comprende opcionalmente un paso 460 para proporcionar una representación espectral transformada por la distorsión de tiempo sobre la base de la señal de audio de entrada. El método 450 comprende además un paso 470 para proporcionar una señal de activación de distorsión de tiempo. El paso 470, por ejemplo, comprende la funcionalidad del método 400. Por tanto, la información de

compactación de energía puede proporcionarse de modo que la información de compactación de energía describa una compactación de energía en la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada. El método 450 comprende además un paso 480 para proporcionar de forma selectiva, con dependencia de la señal de activación de distorsión de tiempo, una descripción de la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada utilizando una información del contorno de distorsión de tiempo recientemente hallada o una descripción de una representación espectral transformada sin distorsión de tiempo de la señal de audio de entrada utilizando una información estándar del contorno de distorsión de tiempo (no-variable) para la inclusión en la representación codificada de la señal de audio de entrada.

El método 450 puede complementarse por cualquiera de los rasgos y funcionalidades mencionados en la presente con respecto a la codificación de la señal de audio de entrada.

La figura 5 ilustra una realización preferida de un codificador de audio según la presente invención, en la que se implementan varios aspectos de la presente invención. Se proporciona una señal de audio en una entrada del codificador 500. Esta señal de audio será típicamente una señal de audio discreta la cual se ha derivado de una señal de audio análoga utilizando una velocidad de muestreo la cual se denomina además la velocidad de muestreo normal. Esta velocidad de muestreo normal es diferente de una velocidad de muestreo local generada en una operación de la operación de distorsión de tiempo, y la velocidad de muestreo normal de la señal de audio en la entrada 500 es una velocidad de muestreo constante que de por resultado muestras de audio separadas por una parte de tiempo constante. La señal se coloca en un generador 502 de ventana de análisis, el cual está, en esta realización, conectado a un controlador de función 504 ventana. El generador 502 de ventana de análisis está conectado a un distorsionador 506 de tiempo. Dependiendo de la implementación, sin embargo, el distorsionador 506 de tiempo puede ser colocado – en una dirección de procesamiento de señal – antes del generador 502 de ventana de análisis. Se prefiere esta implementación, cuando una característica de distorsión de tiempo se requiere para la generación 502 de ventanas de análisis en bloque, y cuando la operación de distorsión de tiempo deba realizarse sobre muestras de distorsión de tiempo en vez de muestras sin distorsión de tiempo. Específicamente en el contexto de distorsión de tiempo basado en MDCT como se describe en Bernd Edler et al., “Time Warped MDCT”, Solicitud de Patente Internacional PCT/EP2009/002118. Para otras aplicaciones de distorsión de tiempo tal como las que se describen en L. Villemoes, “Time Warped Transform Coding of Audio Signals”, Solicitud de Patente Internacional PCT/EP2006/010246, noviembre de 2005., la colocación entre el distorsionador 506 de tiempo y el generador 502 de ventana de análisis se puede establecer según lo requerido. Además, un convertidor de tiempo/frecuencia 508 se proporciona para realizar una conversión de tiempo/frecuencia de una señal de audio distorsionada en el tiempo en una representación espectral. La representación espectral puede ser ingresada en una etapa de TNS (conformación de ruido temporal) 510, que proporciona, como una salida 510a, información de TNS y, como una salida 510b, valores residuales espectrales. La salida 510b está acoplada a un cuantificador y bloque 512 codificador que puede controlarse por un modelo 514 perceptivo para cuantificar una señal de modo que el ruido de cuantificación se esconda por debajo del umbral de enmascaramiento perceptivo de la señal de audio.

Adicionalmente, el codificador que se ilustra en la figura 5a comprende un analizador de distorsión de tiempo 516, que puede implementarse como un seguidor de tono, que proporciona una información de la distorsión de tiempo en la salida 518. La señal en la línea 518 puede comprender una característica de distorsión de tiempo, una característica de tono, un contorno de tono, o una información, si la señal analizada por el analizador de distorsión de tiempo es una señal armónica o una señal sin armonía. El analizador de distorsión de tiempo además puede implementar la funcionalidad de distinguir entre la voz sonora y la voz sorda. Sin embargo, dependiendo de la implementación, y si un clasificador 520 de señal está implementado, la decisión sonora/sorda puede efectuarse además mediante el clasificador 520 de señal. En este caso, el analizador de distorsión de tiempo no necesariamente debe realizar la misma funcionalidad. La salida del analizador de distorsión de tiempo 518 está conectada a al menos una y preferiblemente más de una funcionalidad en el grupo de funcionalidad que comprende el controlador de la función 504 ventana, el distorsionador 506 de tiempo, la etapa 510 de TNS, el cuantificador y codificador 512 y una interfaz 522 de salida.

De forma análoga, una salida 522 del clasificador 520 de señal puede estar conectada a una o más de las funcionalidades de un grupo de funcionalidades que comprende el controlador 504 de la función ventana, la etapa 510 de TNS, un analizador 524 de relleno de ruido o la interfaz 522 de salida. Además, una salida del analizador de distorsión de tiempo 518 puede estar conectada al analizador 524 de relleno de ruido.

Aunque la figura 5a ilustra una situación, donde la señal de audio en la entrada del generador 500 de ventana de análisis se introduce en un analizador 516 de distorsión de tiempo y el clasificador 520 de señal, las señales de entrada para estas funcionalidades además pueden tomarse de la salida del generador 502 de ventana de análisis y, con respecto al clasificador de señal, incluso se pueden tomar de la salida del distorsionador 506 de tiempo, la salida del convertidor de tiempo/frecuencia 508 o la salida de la etapa del TNS 510.

Además de la salida de una señal por el cuantificador 512 codificador que se indica en 526, la interfaz 522 de salida recibe la información en el lado del TNS 510a, una información 528 en el lado del modelo perceptivo, el cual puede incluir factores de escala de forma codificada, datos de indicación de distorsión de tiempo para más información en el lado de la distorsión de tiempo avanzada tal como el contorno de tono en la línea 518 y la información de clasificación

de señal en la línea 522. Además, el analizador 524 de relleno de ruido puede producir además datos de relleno de ruido en la salida 530 en la interfaz 522 de salida. La interfaz 522 de salida está configurada para generar datos de salida de audio codificada en la línea 532 para la transmisión a un decodificador o para almacenar en un dispositivo de almacenamiento tal como un dispositivo de memoria. Dependiendo de la implementación, los datos 532 de salida pueden incluir todos los de la entrada en la interfaz 522 de salida o pueden contener menos información, siempre que la información no sea requerida por el correspondiente decodificador, que tiene una funcionalidad reducida, siempre que la información ya esté disponible en el decodificador debido a una transmisión mediante un canal de transmisión diferente.

El codificador que se ilustra en la figura 5a puede implementarse tal como se define en detalle en MPEG-4 estándar a parte de las funcionalidades adicionales que se ilustran en el codificador de la invención en la figura 5a representada por el controlador de la función 504 ventana, el analizador 524 de relleno de ruido, el cuantificador 512 codificador y la etapa 510 de TNS, que tienen, en comparación con el MPEG-4 estándar, una funcionalidad avanzada. Otra descripción se encuentra en el AAC estándar (estándar internacional 13818-7) o 3GPP TS 26.403 V7.0.0: Third generation partnership project; technical specification group services and system aspect; general audio codec audio processing functions; enhanced AAC plus general audio codec.

A continuación, se describe la figura 5b, la cual ilustra una realización preferida de un decodificador de audio para decodificar una señal de audio codificada recibida a través de la entrada 540. La interfaz 540 de entrada es operativa para procesar la señal de audio codificada de modo que los diferentes ítems de información de la información son extraídos de la señal en la línea 540. Esta información comprende información 541 de clasificación de señal, información de distorsión de tiempo 542, datos 543 de relleno de ruido, factores 544 de escala, datos 545 de TNS e información 546 espectral codificada. La información espectral codificada se introduce en un decodificador 547 de entropía, que puede comprender un decodificador Huffman o un decodificador aritmético, siempre que la funcionalidad del codificador en el bloque 512 en la figura 5a sea implementada en forma del correspondiente codificador tal como un codificador Huffman o un codificador aritmético. La información espectral codificada se introduce al re-cuantificador 550, el cual está conectado a un relleno 552 de ruido. La salida del relleno 552 de ruido se introduce a una etapa 554 de TNS inversa, que recibe además los datos de TNS en la línea 545. Dependiendo de la implementación, el relleno 552 de ruido y la etapa 554 de TNS pueden aplicarse en diferente orden de modo que el relleno 552 de ruido opere en los datos de salida de la etapa 554 de TNS más que en los datos de entrada de TNS. Además, se provee un convertidor de frecuencia/tiempo 556, el cual alimenta un distorsionador 558 de tiempo. A la salida de la cadena de procesamiento de señal, un generador de ventana de síntesis que realiza preferiblemente un procesamiento de superposición/agregado se aplica como se indica en 560. El orden del distorsionador de tiempo 558 y la etapa de síntesis 560 se puede cambiar, pero, en la realización preferida, se prefiere realizar un algoritmo de codificado/decodificación en base a MDCT como se define en el estándar AAC (AAC = codificación de audio avanzada). A continuación, la operación de desvanecimiento cruzado inherente de un bloque al próximo debido al procedimiento de superposición/agregado se utiliza con ventaja como la última operación en las cadenas de procesamiento de modo que todos los artefactos que bloquean la distorsión en la conversión se evitan de forma eficaz.

Adicionalmente, se proporciona un analizador 562 de relleno de ruido, que está configurado para controlar el relleno 552 de ruido y que recibe como una entrada, información 542 sobre la distorsión temporal y/o información sobre la clasificación de señales 541 e información sobre el espectro recuantificado, según sea el caso.

Preferiblemente, todas las funcionalidades descritas a continuación en el presente documento se aplican juntas en un esquema de codificador/decodificador de audio mejorado. No obstante, las funcionalidades descritas a continuación en el presente documento también pueden aplicarse independientemente una a la otra, es decir, de manera que solamente se implementa una o un grupo, pero no todas las funcionalidades en un cierto esquema de codificador/decodificador.

Posteriormente, se describe en detalle el aspecto de relleno de ruido de la presente invención.

En una realización, la información adicional proporcionada por la herramienta 516 de contornos de tono/distorsión temporal en la figura 5a se utiliza de manera beneficiosa para controlar otras herramientas del codificador/decodificador y, específicamente, la herramienta de relleno de ruido implementada por el analizador 524 de relleno de ruido en el lado del codificador y/o implementada por el analizador 562 de relleno de ruido y el relleno 552 de ruido en el lado del decodificador.

Varias herramientas del codificador dentro de la trama AAC tal como una herramienta de relleno de ruido se controlan mediante la información recopilada por el análisis de contornos de tonos y/o por un conocimiento adicional de una clasificación de señales proporcionada por el clasificador 520 de señales.

Un contorno de tonos encontrado indica segmentos de señales con una estructura armónica clara, de manera que el relleno de ruido entre las líneas armónicas podría disminuir la calidad percibida, especialmente en las señales de voz, por lo tanto el nivel de ruido se reduce, cuando se encuentra un contorno de tono. De otra manera, habría ruido entre los tonos parciales, que tienen el mismo efecto que el ruido de cuantificación incrementado para un espectro emborronado. Además, la cantidad de reducción del nivel de ruido además puede mejorarse utilizando la información del clasificador de señales, de modo que por ejemplo para las señales de voz no habría ningún relleno de ruido y se aplicaría un relleno de ruido moderado a las señales genéricas con una estructura armónica fuerte.

Generalmente, el relleno 552 de ruido es útil para insertar líneas espectrales en un espectro decodificado, donde los ceros se han transmitido de un codificador a un decodificador, es decir, donde el cuantificador 512 en la figura 5a ha cuantificado líneas espectrales en cero. Naturalmente, cuantificar las líneas espectrales en cero redujo mucho la velocidad de bits de la señal transmitida, y, en teoría, la eliminación de estas líneas espectrales (pequeñas) no es audible, cuando estas líneas espectrales están por debajo del umbral de enmascaramiento perceptivo según lo determinado por el modelo 514 perceptivo. No obstante, se ha descubierto que estos “huecos espectrales”, que pueden incluir muchas líneas espectrales adyacentes dan como resultado un sonido bastante poco natural. Por lo tanto, se proporciona una herramienta de relleno de ruido para insertar líneas espectrales en las posiciones, donde las líneas se han cuantificado en cero mediante un cuantificador en el lado del codificador. Estas líneas espectrales pueden tener una amplitud o fase aleatoria, y estas líneas espectrales sintetizadas en el lado del decodificador se escalan utilizando una medida de relleno de ruido determinada en el lado del codificador según lo ilustrado en la figura 5a o dependiendo de una medida determinada en el lado del decodificador según lo ilustrado en la figura 5b por bloque 562 opcional. El analizador 524 de relleno de ruido en la figura 5a, por lo tanto, está configurado para estimar una medida de relleno de ruido de una energía de valores de audio cuantificados en cero para una trama de tiempo de la señal de audio.

En una realización de la presente invención, el codificador de audio para codificar una señal de audio en la línea 500 comprende el cuantificador 512 que se configura para cuantificar valores de audio, donde el cuantificador 512 además está configurado para cuantificar en cero valores de audio por debajo de un umbral de cuantificación. Este umbral de cuantificación puede ser el primer paso de un cuantificador basado en pasos, que se utiliza para la decisión, si un cierto valor de audio se cuantifica en cero, es decir, en un índice de cuantificación de cero, o se cuantifica en uno, es decir, un índice de cuantificación de uno que indica que el valor de audio está por encima de este primer umbral. Aunque el cuantificador en la figura 5a se ilustra como que realiza la cuantificación de valores en el dominio de la frecuencia, el cuantificador también puede utilizarse para cuantificar valores en el dominio del tiempo en una realización alternativa, en la que el relleno de ruido se realiza en el dominio del tiempo en vez del dominio de la frecuencia.

El analizador 524 de relleno de ruido se implementa como un calculador de relleno de ruido para estimar una medida de relleno de ruido de una energía de valores de audio cuantificados en cero para una trama de tiempo de la señal de audio por parte del cuantificador 512. Adicionalmente, el codificador de audio comprende un analizador 600 de señales de audio ilustrado en la figura 6a, que se configura para analizar si la trama temporal de la señal de audio tiene una característica armónica o una característica de voz. El analizador de señales 600 puede, por ejemplo, comprender el bloque 516 de la figura 5a o el bloque 520 de la figura 5a o puede comprender cualquier otro dispositivo para analizar si una señal es una señal armónica o una señal de voz. Debido a que el analizador 516 de distorsión de tiempo se implementa para buscar siempre un contorno de tonos, y debido a que la presencia de un contorno de tonos indica una estructura armónica de la señal, el analizador 600 de señales en la figura 6a puede implementarse como un rastreador de tonos o un calculador de contornos de distorsión en el tiempo de un analizador de distorsión de tiempo.

El codificador de audio adicionalmente comprende un manipulador 602 del nivel de relleno de ruido ilustrado en la figura 6a, que produce una medida/nivel de relleno de ruido manipulado que va a producirse para la interfaz 522 de salida indicada en 530 en la figura 5a. El manipulador 602 de medida de relleno de ruido está configurado para manipular la medida de relleno de ruido dependiendo de la característica armónica o de voz de la señal de audio. El codificador de audio adicionalmente comprende la interfaz 522 de salida para generar una señal codificada para transmisión o almacenamiento, comprendiendo la señal codificada la medida de relleno de ruido manipulada producida por el bloque 602 en la línea 530. Este valor corresponde al valor producido por el bloque 562 en la implementación en el lado del decodificador ilustrada en la figura 5b.

Según lo indicado en la figura 5a y la figura 5b, la manipulación del nivel de relleno de ruido puede implementarse en un codificador o puede implementarse en un decodificador o puede implementarse en ambos dispositivos juntos. En una implementación en el lado del decodificador, el decodificador para decodificar una señal de audio codificada comprende la interfaz 539 de entrada para procesar la señal codificada en la línea 540 para obtener una medida de relleno de ruido, es decir, los datos de relleno de ruido en la línea 543, y los datos de audio codificados en la línea 546. El decodificador adicionalmente comprende un decodificador 547 y un recuantificador 550 para generar datos recuantificados.

Adicionalmente, el decodificador comprende un analizador 600 de señales (figura 6a) que puede implementarse en el analizador 562 de relleno de ruido en la figura 5b para recuperar información de si una trama temporal de los datos de audio tiene una característica armónica o de voz.

Adicionalmente, se proporciona el relleno 552 de ruido para generar datos de audio de relleno de ruido, estando el relleno 552 de ruido configurado para generar los datos de relleno de ruido en respuesta a la medida de relleno de ruido transmitida a través de la señal codificada y generada por la interfaz de entrada en la línea 543 y la característica armónica o de voz de los datos de audio según lo definido por los analizadores 516 y/o 550 de señales en el lado del codificador o según lo definido por el ítem 562 en el lado del decodificador a través del procesamiento e interpretación de la información 542 de distorsión de tiempo que indica si una cierta trama temporal se ha sometido a un procesamiento de distorsión en el tiempo o no.

Adicionalmente, el decodificador comprende un procesador para procesar los datos recuantificados y los datos de audio de relleno de ruido para obtener una señal de audio decodificada. El procesador puede incluir los ítems 554, 556, 558,

560 en la figura 5b según sea el caso. Adicionalmente, dependiendo de la implementación específica del algoritmo del codificador/decodificador, el procesador puede incluir otros bloques de procesamiento, que se proporcionan, por ejemplo, en un codificador de dominio del tiempo tal como el codificador AMR WB+ u otros codificadores de voz.

5 La manipulación de relleno de ruido de la invención puede, por lo tanto, implementarse en el lado del codificador solamente calculando la sencilla medida de ruido y manipulando esta medida de ruido sobre la base de la información armónica/de voz y transmitiendo la medida de relleno de ruido manipulada ya correcta que entonces puede aplicarse por un decodificador de manera sencilla. De forma alternativa, la medida de relleno de ruido no manipulada puede transmitirse desde un codificador a un decodificador, y el decodificador entonces analizará si la trama temporal real de una señal de audio se ha distorsionado en el tiempo, es decir, tiene una característica armónica o de voz de manera que
10 la manipulación real de la medida de relleno de ruido se produce en el lado del decodificador.

Posteriormente, se comenta la figura 6b con el fin de explicar realizaciones preferibles para manipular el cálculo aproximado del nivel de ruido.

15 En la primera realización, se aplica un nivel de ruido normal, cuando la señal no tiene una característica armónica o de voz. Éste es el caso, cuando no se aplica ninguna distorsión de tiempo. Cuando, adicionalmente, se proporciona un clasificador de señales, entonces el clasificador de señales que distingue entre voz y no voz indicaría no voz para la situación en la que la distorsión de tiempo no estaba activa, es decir, cuando no se encontró ningún contorno de tonos.

20 Cuando, sin embargo, la distorsión de tiempo estaba activa, es decir, cuando se encontró un contorno de tonos, que indica un contenido armónico, entonces el nivel de relleno de ruido se manipularía para que sea más bajo que en el caso normal. Cuando se proporciona un clasificador de señales adicional, y después este clasificador de señales indica voz, y cuando simultáneamente la información de distorsión de tiempo indica un contorno de tonos, entonces se señala un nivel de relleno de ruido más bajo o incluso igual a cero. De este modo, el manipulador 602 del nivel de relleno de ruido de la figura 6a reducirá el nivel de ruido manipulado a cero o al menos a un valor menor que el valor bajo indicado en la figura 6b. Preferiblemente, el clasificador de señales adicionalmente tiene un detector sonoro/sordo según lo
25 indicado a la izquierda de la figura 6b. En el caso de voz sonora, se señala/aplica un nivel de relleno de ruido muy bajo o igual a cero. Sin embargo, en el caso de voz sorda, cuando la indicación de distorsión de tiempo no indica un procesamiento de distorsión de tiempo debido al hecho de que no se encontró ningún tono, pero cuando el clasificador de señales señala contenido de voz, la medida de relleno de ruido no se manipula, sino que se aplica un nivel de relleno de ruido normal.

30 Preferiblemente, el analizador de señales de audio comprende un rastreador de tonos para generar una indicación de tono tal como un contorno de tonos o un tono absoluto de una trama temporal de la señal de audio. Entonces, el manipulador está configurado para reducir la medida de relleno de ruido cuando se encuentra un tono, y no reducir la medida de relleno de ruido cuando no se encuentra un tono.

35 Según lo indicado en la figura 6a, un analizador 600 de señales, cuando se aplica en el lado del decodificador, no está realizando un análisis de señales real como un rastreador de tonos o un detector sonoro/sordo, sino que el analizador de señales analiza sintácticamente la señal de audio codificada con el fin de extraer una información de distorsión de tiempo o una información de clasificación de señales. Por lo tanto, el analizador 600 de señales puede implementarse dentro de la interfaz 539 de entrada en el decodificador de la figura 5b.

Una realización adicional de la presente invención se comentará posteriormente con respecto a las figuras 7a-7e.

40 Para inicios de voz donde comienza una parte de voz sonora después de una parte de señal relativamente silenciosa, el algoritmo de cambio de bloques podría clasificarla como un ataque y podría elegir bloques cortos para esta trama particular, con una pérdida de ganancia de codificación en el segmento de señal que tiene una clara estructura armónica. Por lo tanto, la clasificación sonora/sorda del rastreador de tonos se utiliza para detectar inicios sonoros y evitar que el algoritmo de cambio de bloques indique un ataque transitorio alrededor del inicio encontrado. Esta característica también puede acoplarse con el clasificador de señales para evitar el cambio de bloques en señales de
45 voz y permitirlo en todas las otras señales. Además se podría implementar un control más preciso del cambio de bloques no sólo mediante el permiso o rechazo de la detección de ataques, sino mediante el uso de un umbral variable para la detección de ataques basado en el inicio sonoro e información de clasificación de señales. Además, la información puede utilizarse para detectar ataques como los inicios sonoros mencionados anteriormente pero en vez de cambiar a bloques cortos, utiliza ventanas largas con superposiciones cortas, que siguen siendo la resolución espectral preferible pero reducen la región temporal en la que pueden surgir pre y postecos. La figura 7d muestra el comportamiento típico sin la adaptación, la figura 7e muestra dos diferentes posibilidades de adaptación (ventanas de
50 baja superposición y prevención).

55 Un codificador de audio según una realización de la presente invención funciona para generar una señal de audio tal como la señal producida por la interfaz 522 de salida de la figura 5a. El codificador de audio comprende un analizador de señales de audio tal como el analizador 516 de distorsión de tiempo o un clasificador 520 de señales de la figura 5a. Generalmente, el analizador de señales de audio analiza si una trama temporal de la señal de audio tiene una característica armónica o de voz. Para este fin, el clasificador 520 de señales de la figura 5a puede incluir un detector

520a sonoro/sordo o un detector 520b de voz /no voz. Aunque no se muestra en la figura 7a, también puede proporcionarse un analizador de distorsión de tiempo tal como el analizador 516 de distorsión de tiempo de la figura 5a, que puede incluir un rastreador de tonos en vez de los ítems 520a y 520b o además de estas funcionalidades. Adicionalmente, el codificador de audio comprende el controlador 504 de función ventana para seleccionar una función ventana dependiendo de una característica armónica o de voz de la señal de audio según lo determinado por el analizador de señales de audio. El generador 502 de ventanas entonces aplica la función ventana a la señal de audio o, dependiendo de cierta implementación, la señal de audio distorsionada en el tiempo utilizando la función ventana seleccionada para obtener una trama dividida en ventanas. Esta trama de ventana se procesa entonces además por un procesador para obtener una señal de audio codificada. El procesador puede comprender los ítems 508, 510, 512 ilustrados en la figura 5a o más o menos funcionalidades de los codificadores de audio bien conocidos tales como codificadores de audio basados en transformadas o codificadores de audio basados en dominios de tiempo que comprenden un filtro LPC tal como codificadores de voz y, específicamente, codificadores de voz implementados según la norma AMR-WB+.

En una realización preferida, el controlador 504 de función ventana comprende un detector 700 de transitorios para detectar un transitorio en la señal de audio, estando el controlador de función ventana configurado para cambiar de una función ventana para un bloque largo a una función ventana para un bloque corto, cuando se detecta un transitorio y el analizador de señales de audio no encuentra una característica armónica o de voz. Cuando, sin embargo, se detecta un transitorio y el analizador de señales de audio encuentra una característica armónica o de voz, entonces el controlador 504 de función ventana no cambia a la función ventana para el bloque corto. Las salidas de la función ventana que indican una ventana larga cuando no se obtiene ningún transitorio y una ventana corta cuando el detector de transitorios detecta un transitorio se ilustran como 701 y 702 en la figura 7a. Este procedimiento normal según se ejecuta por el codificador AAC ampliamente conocido se ilustra en la figura 7d. En la posición del inicio de voz, el detector 700 de transitorios detecta un incremento en la energía desde una trama a la siguiente trama y, por lo tanto, cambia de una ventana 710 larga a ventanas 712 cortas. Con el fin de acomodar este cambio, se utiliza una ventana 714 de detención larga, que tiene una primera parte 714a de superposición, una parte 714b de no solapamiento (*aliasing*), una segunda parte 714c de superposición más corta y una parte cero que se extiende entre el punto 716 y el punto en el eje temporal indicado por las 2048 muestras. Entonces, se ejecuta la secuencia de ventanas cortas indicada en 712 que, entonces, se termina mediante una ventana 718 de inicio larga que tiene una parte 718a larga de superposición que se superpone con la siguiente ventana larga no ilustrada en la figura 7d. Además, esta ventana tiene una parte 718b de no solapamiento, una parte 718c de superposición corta y una parte cero que se extiende entre el punto 720 en el eje temporal hasta el punto 2048. Esta parte es una parte cero.

Normalmente, el cambio a ventanas cortas es útil para evitar preecos que podrían aparecer dentro de una trama antes del evento transitorio que es la posición de inicio sonoro o, generalmente, el comienzo de la voz o el comienzo de una señal que tiene un contenido armónico. Generalmente, una señal tiene un contenido armónico, cuando un rastreador de tonos decide que la señal tiene un tono. De forma alternativa, existen otras medidas de armonicidad tales como una medida de tonalidad por encima de un cierto nivel mínimo junto con la característica de que los picos prominentes tienen una relación armónica entre sí. Existe una pluralidad de otras técnicas para determinar, si una señal es armónica o no.

Una desventaja de las ventanas cortas es que disminuye la resolución de frecuencia, ya que se incrementa la resolución temporal. Para codificación de alta calidad de voz y, específicamente, partes de voz sonora o partes que tienen un fuerte contenido armónico, se desea una buena resolución de frecuencia. Por lo tanto, el analizador de señales de audio ilustrado en 516, 520 o 520a, 520b es operativo para producir una señal desactivada para el detector 700 de transitorios de manera que se evita un cambio a ventanas cortas cuando se detecta un segmento de voz sonora o un segmento de señales que tienen una fuerte característica armónica. Esto garantiza que, para codificar dichas partes de señal, se mantiene una alta resolución de frecuencia. Esto es un equilibrio entre los preecos por un lado y la codificación de alta calidad y alta resolución del tono para la señal de voz o el tono para una señal de no voz armónica por otro lado. Se ha descubierto que es mucho más perturbador cuando el espectro armónico no está codificado con precisión en comparación con cualquier preeco que pudiera aparecer. Con el fin de disminuir además los preecos, se favorece un procesamiento TNS para dicha situación, que se comentará con relación a las figuras 8a y 8b.

En una realización alternativa ilustrada en la figura 7b, el analizador de señales de audio comprende un detector 520a, 520b de sonoro/sordo y/o de voz /no voz. Sin embargo, el detector 700 de transitorios incluido en el controlador de función ventana no está completamente activado/desactivado como en la figura 7a, sino que el umbral incluido en el detector de transitorios se controla utilizando una señal 704 de control de umbral. En esta realización, el detector 700 de transitorios está configurado para determinar una característica cuantitativa de la señal de audio y para comparar la característica cuantitativa con el umbral controlable, detectándose un transitorio cuando la característica cuantitativa tiene una relación predeterminada con respecto al umbral controlable. La característica cuantitativa puede ser un número que indica el incremento de energía de un bloque al siguiente bloque, y el umbral puede ser un cierto incremento de energía umbral. Cuando el incremento de energía de un bloque al siguiente es mayor que el incremento de energía umbral, entonces se detecta un transitorio, de manera que, en este caso, la relación predeterminada es una relación de "mayor que". En otras realizaciones, la relación predeterminada también puede ser una relación de "menor que", por ejemplo cuando la característica cuantitativa es un incremento de energía invertido. En la realización de la figura 7b, el umbral controlable se controla de manera que se reduce la probabilidad de un cambio a una función

ventana para un bloque corto, cuando el analizador de señales de audio ha encontrado una característica armónica o de voz. En la realización de incremento de energía, la señal 704 de control de umbral dará como resultado un incremento del umbral de manera que los cambios a bloques cortos suceden solamente cuando el incremento de energía de un bloque al siguiente es un incremento de energía particularmente alto.

5 En una realización alternativa, la señal de salida del detector 520a sonoro/sordo o el detector 520b de voz/no voz también puede utilizarse para controlar el controlador 504 de función ventana de tal manera que en vez de cambiar a un bloque corto en un inicio de voz, se ejecuta un cambio a una función ventana que es más larga que la función ventana para el bloque corto. Esta función ventana garantiza una resolución de frecuencia más alta que una función ventana para el bloque corto. Esta función ventana garantiza una resolución de frecuencia más alta que una función ventana para el bloque corto, pero tiene una longitud más corta que la función ventana larga de manera que se obtiene un buen compromiso
10 entre los preecos por un lado y una resolución de frecuencia suficiente por otro lado. En una realización alternativa, se puede realizar un cambio a una función ventana que tiene una superposición más pequeña según lo indicado por la línea sombreada en la figura 7e en 706. La función 706 ventana tiene una longitud de 2048 muestras como el bloque largo, pero esta ventana tiene una parte 708 cero y una parte 710 de no solapamiento de manera que se obtiene una longitud de superposición corta 712 desde la ventana 706 a una ventana 707 correspondiente. La función 707 ventana, nuevamente, tiene una parte cero a la izquierda de la región 712 y una parte de no solapamiento a la derecha de la región 712 en analogía a la función 710 ventana. Esta realización de baja superposición, efectivamente da como resultado una longitud de tiempo más corta para reducir los preecos debido a la parte cero de la ventana 706 y 707,
15 pero por otro lado tiene una longitud suficiente debido a la parte 714 de superposición y la parte 710 de no solapamiento de manera que se mantiene una resolución de frecuencia suficiente.

20 En una implementación de MDCT preferida según se implementa por el codificador AAC, mantener una cierta superposición proporciona la ventaja adicional de que, en el lado del decodificador, se puede realizar un procesamiento de superposición/suma lo que significa que se realiza una especie de desvanecimiento cruzado entre bloques. Esto evita eficazmente "artefactos". Adicionalmente, esta característica de superposición/suma proporciona la característica de desvanecimiento cruzado sin incrementar la velocidad de bits, es decir, se obtiene un desvanecimiento cruzado críticamente muestreado. En ventanas cortas o ventanas largas regulares, la parte de superposición es una superposición del 50% tal como se indica por la parte 714 de superposición. En la realización en la que la función ventana tiene una longitud de 2048 muestras, la parte de superposición es del 50%, es decir, 1024 muestras. La función ventana que tiene una superposición más corta que va a utilizarse para aplicar eficazmente una función ventana a un inicio de voz o un inicio de una señal armónica es preferiblemente menos del 50% y tiene, en la realización de la figura
25 7e, solamente una longitud de 128 muestras, que es un 1/16 del total de la longitud de ventana. Preferiblemente, se utilizan partes de superposición entre 1/4 y 1/32 de la longitud total de la función ventana.

La figura 7c ilustra esta realización, en la que un detector 520a de sonoro/sordo ejemplar controla un selector de formas de ventanas incluido en el controlador 504 de función ventana con el fin de seleccionar una forma de ventana con una superposición corta según lo indicado en 749 o una forma de ventana con una superposición larga según lo indicado en
35 750. Se implementa la selección de una de ambas formas, cuando el detector 500a de sonoro/sordo emite una señal sonora detectada en 751, donde la señal de audio utilizada para análisis puede ser la señal de audio en la entrada 500 en la figura 5a o una señal de audio preprocesada tal como una señal de audio distorsionada en el tiempo o una señal de audio que haya sido sometida a cualquier otra funcionalidad de preprocesamiento. Preferiblemente, en selector 504 de formas de ventanas en la figura 7c que está incluido en el controlador 504 de función ventana en la figura 5a solamente utiliza la señal 751, cuando un detector de transitorios incluido en el controlador de función ventana detecta un transitorio y ordena un cambio de una función ventana larga a una función ventana corta según lo comentado con relación a la figura 7a.

Preferiblemente, la realización del cambio de función ventana se combina con una realización de conformación de ruido temporal comentada con relación a las figuras 8a y 8b. Sin embargo, la realización de TNS (conformación de ruido temporal) también puede implementarse sin la realización de cambios de bloques.
45

La propiedad de compactación de energía espectral de la MDCT distorsionada en el tiempo también influye en la herramienta de conformación de ruido temporal (TNS), debido a que la ganancia de TNS tiende a disminuir para tramas distorsionadas en el tiempo especialmente para algunas señales de voz. No obstante es deseable activar TNS, por ejemplo para reducir los preecos en los inicios sonoros o desviaciones sonoras (véase la adaptación del cambio de bloques), donde no se desea ningún cambio de bloques pero aún la envolvente temporal de la señal de voz exhibe cambios rápidos. Típicamente, un codificador utiliza alguna medida para ver si la aplicación de la TNS es provechosa para una cierta trama, por ejemplo la ganancia de predicción del filtro de TNS cuando se aplica al espectro. Entonces se prefiere un umbral de ganancia de TNS variable, que sea menor para los segmentos con un contorno de tonos activo, se manera que se garantiza que la TNS esté activa más a menudo para partes de señal críticas tales como inicios
50 sonoros. Como para las otras herramientas, esto también puede complementarse teniendo en cuenta la clasificación de señales.

El codificador de audio según esta realización para generar una señal de audio comprende un distorsionador de tiempo controlable tal como el distorsionador 506 de tiempo para la distorsión en el tiempo de la señal de audio para obtener una señal de audio distorsionada en el tiempo. Adicionalmente, se proporciona un convertidor 508 de frecuencia/tiempo

- 5 para convertir al menos una parte de la señal de audio distorsionada en el tiempo en una representación espectral. El convertidor 508 de frecuencia/tiempo preferiblemente implementa una transformada MDCT tal como se conoce del codificador AAC, pero el convertidor de frecuencia/tiempo también puede realizar cualquier otra clase de transformada tal como una transformada DCT, DST, DFT, FFT o MDST o puede comprender un banco de filtros tal como un banco de filtros QMF.
- Adicionalmente, el codificador comprende una etapa 510 de conformación de ruido temporal para realizar un filtrado de predicción sobre la frecuencia de la representación espectral conforme a la orden de control de conformación de ruido temporal, en la que no se realiza el filtrado de predicción, cuando no existe la orden de control de conformación de ruido temporal.
- 10 Adicionalmente, el codificador comprende un controlador de conformación de ruido temporal para generar la orden de control de conformación de ruido temporal sobre la base de la representación espectral.
- 15 Específicamente, el controlador de conformación de ruido temporal está configurado para incrementar la probabilidad de realizar el filtrado de predicción sobre la frecuencia, cuando la representación espectral se basa en una señal de tiempo distorsionada en el tiempo o para reducir la probabilidad de realizar el filtrado de predicción sobre la frecuencia, cuando la representación espectral no está basada en una señal de tiempo distorsionada en el tiempo. Los detalles del controlador de conformación de ruido temporal se comentan con relación a la figura 8.
- El codificador de audio adicionalmente comprende un procesador para además procesar un resultado del filtrado de predicción sobre la frecuencia para obtener la señal codificada. En una realización, el procesador comprende la etapa 512 de codificador del cuantificador ilustrada en la figura 5a.
- 20 Una etapa 510 de TNS ilustrada en la figura 5a se ilustra en detalle en la figura 8. Preferiblemente, el controlador de conformación de ruido temporal incluido en la etapa 510 comprende un calculador 800 de ganancia de TNS, un tomador 802 de decisión de TNS posteriormente conectado y un generador 804 de señal de control de umbral. Dependiendo de la señal del analizador 516 de distorsión de tiempo o el clasificador 520 de señales o ambos, el generador 804 de señal de control de umbral produce una señal 806 de control de umbral para el tomador de decisión de TNS. El tomador 802 de decisión de TNS tiene un umbral controlable, que se incrementa o se reduce conforme a la señal 806 de control de umbral. El umbral en el tomador 802 de decisión de TNS es, en esta realización, un umbral de ganancia de TNS. Cuando la ganancia de TNS realmente calculada producida por el bloque 800 excede el umbral, entonces la orden de control de TNS requiere un procesamiento de TNS como salida, mientras que, en el otro caso, cuando la ganancia de TNS está por debajo del umbral de ganancia de TNS, no se produce ninguna orden de TNS o se produce una señal que ordena que el procesamiento de TNS no es útil y no debe realizarse en esta trama temporal específica.
- 25 El calculador 800 de ganancia de TNS recibe, como una entrada, la representación espectral obtenida a partir de la señal distorsionada en el tiempo. Típicamente, una señal distorsionada en el tiempo tendrá una ganancia de TNS menor, pero por otro lado, un procesamiento de TNS debido a la característica de conformación de ruido temporal en el dominio del tiempo resulta beneficiado en la situación específica, en la que existe una señal sonora/armónica que se ha sometido a una operación de distorsión en el tiempo. Por otro lado, el procesamiento de TNS no es útil en situaciones en las que el aumento de TNS es bajo, lo que significa que la señal residual de TNS en la línea 510b tiene la misma energía o mayor energía que la señal antes de la etapa 510 de TNS. En una situación en la que la energía de la señal residual de TNS en la línea 510d es ligeramente menor que la energía antes de la etapa 510 de TNS, el procesamiento de TNS podría no ser ventajoso tampoco, debido a que la reducción de bits debida a la energía ligeramente menor en la
- 30 señal que se utiliza eficazmente por la etapa 512 del codificador de entropía/cuantificador es más pequeña que el incremento de bits introducido por la transmisión necesaria de la información secundaria de TNS indicada en 510a en la figura 5a. Aunque una realización activa automáticamente el procesamiento de TNS para todas las tramas, en las que se introduce una señal distorsionada en el tiempo indicada por la información sobre tonos del bloque 516 o la información del clasificador de señales del bloque 520, una realización preferida también mantiene la posibilidad de desactivar el procesamiento de TNS, pero sólo cuando la ganancia es realmente baja o al menos menor que en el caso normal, cuando no se procesa ninguna señal armónica/de voz.
- 35 La figura 8b ilustra una implementación en la que se implementan tres ajustes de umbral diferentes por el generador 804 de señal de control de umbral/tomador 802 de decisión de TNS. Cuando no existe un contorno de tonos, y cuando un clasificador de señal indica una voz sorda o ninguna voz, entonces el umbral de decisión de TNS se ajusta para estar en un estado normal que requiere una ganancia de TNS relativamente alta para activar la TNS. Cuando, sin embargo, se detecta un contorno de tonos, pero el clasificador de señales no indica ninguna voz o el detector sonoro/sordo detecta una voz sorda, entonces el umbral de decisión de TNS se ajusta a un nivel menor, lo que significa que aún cuando se calculan ganancias de TNS comparativamente bajas por el bloque 800 en la figura 8a, aún así se activa el procesamiento de TNS.
- 40
- 45
- 50
- 55 En una situación, en la que se detecta un contorno de tonos activo y en la que se encuentra una voz sonora, entonces, el umbral de decisión de TNS se ajusta al mismo nivel más bajo o se ajusta a un estado aún menor de manera que incluso las pequeñas ganancias de TNS son suficientes para activar un procesamiento de TNS.

En una realización, el controlador 800 de ganancia de TNS está configurado para estimar una ganancia en calidad o velocidad de bits, cuando la señal de audio se somete al filtrado de predicción sobre la frecuencia. Un tomador 802 de decisión de TNS compara la ganancia estimada con un umbral de decisión, y una información de control de TNS a favor del filtrado de predicción se produce por el bloque 802, cuando la ganancia estimada está en una relación predeterminada con respecto al umbral de decisión, donde esta relación predeterminada puede ser una relación de "mayor que", pero también puede ser una relación de "menor que" para una ganancia invertida de TNS por ejemplo. Según lo comentado, el controlador de conformación de ruido temporal está configurado además para variar el umbral de decisión utilizando preferiblemente la señal 806 de control de umbral de manera que, para la misma ganancia estimada, se activa el filtrado de predicción, cuando la representación espectral se basa en la señal de audio distorsionada en el tiempo, y no se activa, cuando la representación espectral no se basa en la señal de audio distorsionada en el tiempo. Normalmente, la voz sonora mostrará un contorno de tonos, y la voz sorda tal como sonidos fricativos o sonidos sibilantes no mostrará un contorno de tonos. Sin embargo, existen señales no de voz, con fuerte contenido armónico y, por lo tanto, tienen un contorno de tonos, aunque el detector de voz no detecte voz. Adicionalmente, existe cierta voz sobre música o música sobre señales de voz, que el analizador de señales de audio (516 de la figura 5a por ejemplo) determina que tienen un contenido armónico, pero que no se detectan por el clasificador 520 de señales como señal de voz. En dicha situación, también pueden aplicarse todas las operaciones de procesamiento para señales de voz sonoras y también darán como resultado una ventaja.

Posteriormente, se describe otra realización preferida de la presente invención con respecto a un codificador de audio para codificar una señal de audio. Este codificador de audio es específicamente útil en el contexto de la extensión del ancho de banda, pero también es útil en aplicaciones de codificador autónomas, en las que el codificador de audio se ajusta para codificar un cierto número de líneas con el fin de obtener una cierta limitación de ancho de banda/operación de filtrado de paso bajo. En las aplicaciones no distorsionadas en el tiempo, esta limitación de ancho de banda seleccionando un cierto número predeterminado de líneas dará como resultado un ancho de banda constante, ya que la frecuencia de muestreo de la señal de audio es constante. En situaciones, sin embargo, en las que se realiza un procesamiento de distorsión de tiempo tal como por el bloque 506 en la figura 5a, un codificador que depende de un número fijo de líneas dará como resultado un ancho de banda variable introduciendo fuertes artefactos no sólo perceptibles para oyentes entrenados sino también perceptibles para oyentes no entrenados.

El codificador central AAC normalmente codifica un número fijo de líneas, ajustando todas las demás por encima de la línea máxima a cero. En el caso no distorsionado esto lleva a un efecto de paso bajo con una frecuencia de corte constante y por lo tanto a un ancho de banda constante de la señal AAC decodificada. En el caso distorsionado en el tiempo el ancho de banda varía debido a la variación de la frecuencia de muestreo local, una función del contorno de distorsión en el tiempo local, que lleva a artefactos audibles. Los artefactos pueden reducirse eligiendo de forma adaptativa el número de líneas, como una función del contorno de distorsión en el tiempo local y su velocidad de muestreo promedio obtenida, que van a codificarse en el codificador central dependiendo de la frecuencia de muestreo local de manera que se obtiene un ancho de banda promedio constante después de redistorsionar en el tiempo en el decodificador para todas las tramas. Un beneficio adicional es el ahorro de bits en el codificador.

El codificador de audio según esta realización comprende el distorsionador 506 de tiempo para distorsionar en el tiempo una señal de audio que utiliza una característica de distorsión en el tiempo variable. Adicionalmente, se proporciona un convertidor 508 de tiempo/frecuencia para convertir una señal de audio distorsionada en el tiempo en una representación espectral que tiene un número de coeficientes espectrales. Adicionalmente, se utiliza un procesador para procesar un número variable de coeficientes espectrales para generar la señal de audio codificada, donde este procesador que comprende el bloque 512 de cuantificador/codificador de la figura 5a está configurado para ajustar un número de coeficientes espectrales para una trama de la señal de audio basándose en la característica de distorsión en el tiempo para la trama de manera que se reduzca o se elimine una variación del ancho de banda representada por el número procesado de coeficientes de frecuencia de una trama a otra.

El procesador implementado por el bloque 512 puede comprender un controlador 1000 para controlar el número de líneas, en el que el resultado del controlador 1000 es que, con respecto a un número de líneas ajustado para el caso de que una trama de tiempo se codifique sin ninguna distorsión de tiempo, se suma o se descarta un cierto número variable de líneas en el extremo superior del espectro. Dependiendo de la implementación, el controlador 1000 puede recibir información de un contorno de tonos en una cierta trama 1001 y/o una frecuencia de muestreo promedio local en la trama indicada en 1002.

En las figuras 9(a) a 9(e), los dibujos de la derecha ilustran una cierta situación de ancho de banda para ciertos contornos de tonos sobre una trama, donde los contornos de tonos sobre la trama se ilustran en los respectivos dibujos de la izquierda para la distorsión de tiempo y se ilustran en los dibujos del medio después de la distorsión de tiempo, donde se obtiene una característica de tono sustancialmente constante. Este es el objetivo de la funcionalidad de la distorsión en el tiempo de que, después de la distorsión en el tiempo, la característica del tono sea lo más constante posible.

El ancho 900 de banda ilustra el ancho de banda que se obtiene cuando se toma un cierto número de líneas producido por un convertidor 508 de tiempo/frecuencia o producido por una etapa 510 de TNS de la figura 5a, y cuando no se

realiza una operación de distorsión en el tiempo, es decir, cuando el distorsionador 506 de tiempo se desactiva, según lo indicado por la línea 507 sombreada. Cuando, sin embargo, se obtiene un contorno de distorsión de tiempo no constante, y cuando este contorno de distorsión de tiempo se lleva a un tono más alto induciendo un incremento en la velocidad de muestreo (figura 9(a), (c)) el ancho de banda del espectro disminuye con respecto a una situación no distorsionada en el tiempo, normal. Esto significa que el número de líneas que va a transmitirse para esta trama debe incrementarse con el fin de equilibrar esta pérdida de ancho de banda.

De forma alternativa, llevar el tono a un tono constante más bajo ilustrado en la figura 9(b) o figura 9(d) da como resultado una reducción de la velocidad de muestreo. La reducción de la velocidad de muestreo da como resultado un incremento del ancho de banda del espectro de esta trama con respecto a la escala lineal, y este incremento de ancho de banda debe equilibrarse utilizando una supresión o descarte de un cierto número de líneas con respecto al valor del número de líneas para la situación normal no distorsionada en el tiempo.

La figura 9(e) ilustra un caso especial, en el que un contorno de tonos se lleva a un nivel medio de manera que la frecuencia de muestreo promedio dentro de una trama es, en vez de realizar la operación de distorsión en el tiempo, igual que la frecuencia de muestreo sin ninguna distorsión en el tiempo. De este modo, el ancho de banda de la señal no se ve afectado, y puede procesarse el número sencillo de líneas que va a utilizarse para el caso normal sin distorsión en el tiempo, aunque se tenga que realizar la operación de distorsión en el tiempo. A partir de la figura 9, se vuelve claro que la realización de una operación de distorsión en el tiempo no influye necesariamente en el ancho de banda, sino que la influencia en el ancho de banda depende del contorno de tonos y la forma en que se lleva a cabo la distorsión de tiempo en una trama. Por lo tanto, es preferible utilizar, como valor de control, una velocidad de muestreo local o promedio. La determinación de esta velocidad de muestreo local se ilustra en la figura 11. La parte superior en la figura 11 ilustra una parte de tiempo con valores de muestreo equidistantes. Una trama incluye, por ejemplo, siete valores de muestreo indicados por T_n en la representación gráfica superior. La representación gráfica inferior muestra el resultado de una operación de distorsión en el tiempo, en la que, en conjunto, ha tenido lugar un incremento de la velocidad de muestreo. Esto significa que la longitud de tiempo de la trama distorsionada en el tiempo es más pequeña que la longitud de tiempo de la trama no distorsionada en el tiempo. Sin embargo, debido a que la longitud de tiempo de la trama distorsionada en el tiempo que va a introducirse en el convertidor de tiempo/frecuencia es fija, el caso de un incremento de la velocidad de muestreo hace que una parte adicional de la señal de tiempo no correspondiente a la trama indicada por T_n se introduzca en la trama distorsionada en el tiempo tal como se indica por las líneas 1100. De este modo, una trama distorsionada en el tiempo cubre una parte de tiempo de la señal de audio indicada por T_{in} que es mayor que el tiempo T_n . En vista de esto, la distancia eficaz entre dos líneas de frecuencia o el ancho de banda de la frecuencia de una única línea en el dominio lineal (que es el valor inverso para la resolución) ha disminuido, y el número de líneas N_n ajustado para un caso no distorsionado en el tiempo cuando se multiplica por la distancia de frecuencia reducida da como resultado un ancho de banda más pequeño, es decir, una reducción del ancho de banda.

En el otro caso, no ilustrado en la figura 11, en el que el distorsionador de tiempo lleva a cabo una reducción de la velocidad de muestreo, la longitud de tiempo eficaz de una trama en el dominio distorsionado en el tiempo es más pequeña que la longitud de tiempo del dominio no distorsionado en el tiempo de manera que el ancho de banda de la frecuencia de una única línea o la distancia entre dos líneas de frecuencia ha incrementado. Ahora, la multiplicación de este Δf incrementado por el número N_n de líneas para el caso normal dará como resultado un ancho de banda incrementado debido a la resolución de frecuencia reducida/distancia de frecuencia incrementada entre dos coeficientes de frecuencias adyacentes.

La figura 11 adicionalmente ilustra cómo se calcula una velocidad de muestreo promedio f_{SR} . Para este fin, se determina la distancia de tiempo entre dos muestras distorsionadas en el tiempo y se toma el valor inverso, que se define como la velocidad de muestreo local entre dos muestras distorsionadas en el tiempo. Dicho valor puede calcularse entre cada par de muestras adyacentes, y puede calcularse el valor de la media aritmética y este valor finalmente da como resultado la velocidad de muestreo promedio local, que preferiblemente se utiliza para introducirse en el controlador 1000 de la figura 10a.

La figura 10b ilustra una representación gráfica que indica cuántas líneas deben sumarse o descartarse dependiendo de la frecuencia de muestreo local, donde la frecuencia de muestreo f_n para el caso no distorsionado junto con el número de líneas N_n para el caso no distorsionado en el tiempo define el ancho de banda previsto, que debe mantenerse constante tanto como sea posible para una secuencia de tramas distorsionadas en el tiempo o para una secuencia de tramas distorsionadas en el tiempo y no distorsionadas en el tiempo.

La figura 12b ilustra la dependencia entre los diferentes parámetros comentados con relación a la figura 9, figura 10b y figura 11. Básicamente, cuando la velocidad de muestreo, es decir, la velocidad de muestreo promedio f_{SR} disminuye con respecto al caso no distorsionado en el tiempo, deben suprimirse líneas, mientras que deben sumarse líneas, cuando la velocidad de muestreo aumenta con respecto a la velocidad de muestreo normal f_n para el caso no distorsionado en el tiempo de manera que las variaciones del ancho de banda de una trama a otra se reducen o preferiblemente incluso se eliminan tanto como sea posible.

El ancho de banda que resulta por el número de líneas N_n y la velocidad de muestreo f_n define preferiblemente la frecuencia 1200 de cruce para un codificador de audio que, además de un codificador de audio central fuente, tiene un

- codificador de extensión de ancho de banda (codificador BWE). Tal como se conoce en la técnica, un codificador de extensión de ancho de banda solamente codifica un espectro con una alta velocidad de bits hasta la frecuencia de cruce y codifica el espectro de la banda alta, es decir, entre la frecuencia 1200 de cruce y la frecuencia f_{MAX} con una velocidad de bits baja, donde esta velocidad de bits baja normalmente es incluso menor que 1/10 o menos de la velocidad de bits requerida para la banda baja entre una frecuencia de 0 y la frecuencia 1200 de cruce. La figura 12a ilustra además el ancho de banda BW_{AAC} de un codificador de audio AAC sencillo, que es mucho mayor que la frecuencia de cruce. Por lo tanto, las líneas no solamente pueden descartarse, sino que también pueden sumarse. Además, también se ilustra la variación del ancho de banda para un número de líneas constante dependiendo de la velocidad de muestreo local f_{SR} . Preferiblemente, el número de líneas que va a sumarse o que va a suprimirse con respecto al número de líneas para el caso normal se ajusta de manera que cada trama de los datos codificados de AAC tiene una frecuencia máxima lo más próxima posible a la frecuencia 1200 de cruce. De este modo, se evita cualquier hueco espectral debido a una reducción del ancho de banda por un lado o una sobrecarga transmitiendo información en una frecuencia por encima de la frecuencia de cruce en la trama codificada de banda baja. Esto, por un lado, incrementa la calidad de la señal de audio decodificada y, por otro lado, disminuye la velocidad de bits.
- La suma real de líneas con respecto a un número fijo de líneas o una supresión de líneas con respecto al número fijo de líneas puede realizarse antes de cuantificar las líneas, es decir, en la entrada del bloque 512, o puede realizarse después de cuantificar o también puede realizarse, dependiendo del código de entropía específico, después de la codificación de entropía.
- Además, se prefiere llevar las variaciones de ancho de banda a un nivel mínimo e incluso eliminar las variaciones del ancho de banda, pero, en otras implementaciones, incluso una reducción de las variaciones del ancho de banda determinando el número de líneas que dependen de la característica de distorsión en el tiempo incluso aumenta la calidad de audio y reduce la velocidad de bits requerida en comparación con una situación, en la que se aplica un número de líneas constante independientemente de una cierta característica de distorsión de tiempo.
- Aunque se han descrito algunos aspectos en el contexto de un aparato, es evidente que estos aspectos también representan una descripción del método correspondiente, en el que un bloque o dispositivo corresponde a una etapa del método o una característica de una etapa del método. De forma análoga, los aspectos descritos en el contexto de una etapa del método representan también una descripción de un bloque o ítem correspondiente o característica del aparato correspondiente.
- Dependiendo de ciertos requerimientos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse utilizando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tenga señales de control electrónicamente legibles almacenadas en los mismos, que actúen conjuntamente (o puedan actuar conjuntamente) con un sistema informático programable de manera tal se lleve a cabo el método respectivo. Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, que pueden actuar conjuntamente con un sistema informático programable, de manera que se lleve a cabo uno de los métodos descritos en el presente documento. Generalmente, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo operativo el código de programa para realizar uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse, por ejemplo, en un soporte legible por máquina. Otras realizaciones comprenden el programa informático para llevar a cabo uno de los métodos descritos en el presente documento, almacenado en un soporte legible por máquina. En otras palabras, una realización del método según la invención es, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los métodos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador. Una realización adicional de los métodos según la invención es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. Una realización adicional del método según la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales puede configurarse por ejemplo para transferirse a través de una conexión de comunicación de datos, por ejemplo a través de Internet. Una realización adicional comprende medios de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los métodos descritos en el presente documento. Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para llevar a cabo uno de los métodos descritos en el presente documento. En algunas realizaciones, puede utilizarse un dispositivo lógico programable (por ejemplo una disposición de puertas programables en campo) para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una disposición de puertas programables en campo puede actuar conjuntamente con un microprocesador con el fin de llevar a cabo uno de los métodos descritos en el presente documento.

REIVINDICACIONES

1. Proveedor (100; 230; 234) de la señal de activación de distorsión de tiempo para proporcionar una señal (112; 232; 234p) de activación de distorsión de tiempo sobre la base de una representación (110; 234e; 234k) de una señal de audio, comprendiendo el proveedor de la señal de activación de distorsión de tiempo:
- 5 un proveedor (120; 234f; 234l; 325; 370) de información de compactación de energía configurado para proporcionar una información (122; 234m; 234n; 326; 374) de compactación de energía que describe una compactación de energía en una representación (222) espectral transformada por la distorsión de tiempo de la señal de audio; y
- un comparador (130; 234o) configurado para comparar la información (122; 234m; 234n; 326; 374) de compactación de energía con un valor de referencia, y para proporcionar la señal (112; 232; 234p) de activación de distorsión de tiempo con dependencia de un resultado de la comparación;
- 10 en el que el proveedor (120; 234f; 234l; 325) de información de compactación de energía está configurado para proporcionar una medida de entropía perceptiva (pe) que describe la representación (234e; 234k) espectral transformada por la distorsión de tiempo de la señal de audio como la información (122; 234m; 234n) de compactación de energía; y
- 15 en el que el proveedor (120; 234f; 234l; 325) de información de compactación de energía está configurado para calcular un número (nl) estimado de líneas que no son cero para una o más bandas de factor de escala de la representación (234e; 234k) espectral transformada por la distorsión de tiempo de la señal de audio sobre la base de una información de factor de forma (ffac(n)) de la banda de factor de escala, y para calcular la medida de entropía (326) perceptiva para una banda de factor de escala en consideración utilizando una multiplicación del número (nl) estimado de líneas que no son cero y una medida de energía de la banda de factor de escala en consideración.
- 20 2. Proveedor (100; 230; 234) de la señal de activación de distorsión de tiempo para proporcionar una señal (112; 232; 234p) de activación de distorsión de tiempo sobre la base de una representación (110; 234e; 234k) de una señal de audio, comprendiendo el proveedor de la señal de activación de distorsión de tiempo:
- 25 un proveedor (120; 234f; 234l; 325; 370) de información de compactación de energía configurado para proporcionar una información (122; 234m; 234n; 326; 374) de compactación de energía que describe una compactación de energía en una representación (222) espectral transformada por la distorsión de tiempo de la señal de audio; y
- un comparador (130; 234o) configurado para comparar la información (122; 234m; 234n; 326; 374) de compactación de energía con un valor de referencia, y para proporcionar la señal (112; 232; 234p) de activación de distorsión de tiempo con dependencia de un resultado de la comparación;
- 30 en el que el proveedor (120; 234f; 234l; 370) de información de compactación de energía está configurado para proporcionar una medida (374) de autocorrelación que describe una autocorrelación de una representación de dominio de tiempo distorsionado en el tiempo de la señal (234e; 234k) de audio como la información de compactación de energía; y
- en el que el proveedor (120; 234f; 234l; 370) de información de compactación de energía está configurado para determinar una suma de valores absolutos de una función de autocorrelación normalizada de la representación (234e; 234k) distorsionada en el tiempo de la señal de audio para obtener la información de compactación de energía.
- 35 3. Proveedor (100; 230; 234) de la señal de activación de distorsión de tiempo para proporcionar una señal (112; 232; 234p) de activación de distorsión de tiempo sobre la base de una representación (110; 234e; 234k) de una señal de audio, comprendiendo el proveedor de la señal de activación de distorsión de tiempo:
- 40 un proveedor (120; 234f; 234l; 325; 370) de información de compactación de energía configurado para proporcionar una información (122; 234m; 234n; 326; 374) de compactación de energía que describe una compactación de energía en una representación (222) espectral transformada por la distorsión de tiempo de la señal de audio; y
- un comparador (130; 234o) configurado para comparar la información (122; 234m; 234n; 326; 374) de compactación de energía con un valor de referencia, y para proporcionar la señal (112; 232; 234p) de activación de distorsión de tiempo con dependencia de un resultado de la comparación;
- 45 en el que el proveedor (120; 234f; 234l) de información de compactación de energía está configurado para proporcionar una medida de planicidad espectral que describe la representación (234e; 234k) espectral transformada por la distorsión de tiempo de la señal de audio como la información (122; 234m; 234n) de compactación de energía; y
- en el que el proveedor (120; 234f; 234l) de información de compactación de energía está configurado para calcular un cociente de una media geométrica del espectro (234e; 234k) de potencia transformado por la distorsión de tiempo de la señal de audio y una media aritmética del espectro (234e; 234k) de potencia transformada por la distorsión de tiempo de la señal de audio para obtener la medida de planicidad espectral.
- 50

4. Proveedor (100; 230; 234) de la señal de activación de distorsión de tiempo según una de las reivindicaciones 1 a 3, en el que el proveedor (120; 234f; 234l) de información de compactación de energía está configurado para enfatizar una parte de frecuencia más alta de la representación (234e; 234k) espectral transformada por la distorsión de tiempo cuando se compara con una parte de frecuencia más baja de la representación (234e; 234k) espectral transformada por la distorsión de tiempo para obtener la información (122; 234m; 234n) de compactación de energía.
5. Proveedor (100; 230; 234) de la señal de activación de distorsión de tiempo según una de las reivindicaciones 1 a 4, en el que el proveedor (120; 234m; 234n) de información de compactación de energía está configurado para obtener una pluralidad de medidas banda por banda de planicidad espectral, y para calcular un promedio de la pluralidad de medidas banda por banda de planicidad espectral para obtener la información (122, 234m; 234n) de compactación de energía.
6. Proveedor (100; 230) de la señal de activación de distorsión de tiempo según la reivindicación 1 a 5, en el que el proveedor de la señal de activación de distorsión de tiempo comprende un calculador de valor de referencia configurado para calcular el valor de referencia sobre la base de una representación espectral sin distorsionar la señal (210) de audio o sobre la base de una representación del dominio de tiempo sin distorsionar de la señal (210) de audio; y
- en el que el comparador está configurado para formar un valor de relación que utiliza la información (122) de compactación de energía que describe una compactación de energía en una representación espectral transformada por la distorsión de tiempo de la señal de audio y el valor de referencia, y para comparar el valor de relación con uno o más valores de umbral para obtener la señal de activación de distorsión de tiempo como resultado de la comparación.
7. Proveedor (230; 234) de la señal de activación de distorsión de tiempo según una de las reivindicaciones 1 a 6, en el que el proveedor de la señal de activación de distorsión de tiempo comprende un calculador de valor de referencia configurado para calcular el valor de referencia sobre la base de una representación distorsionada en el tiempo de la señal (210) de entrada, distorsionada en el tiempo utilizando una información (288) estándar del contorno de distorsión de tiempo; y
- en el que el comparador está configurado para formar un valor de relación que utiliza la información (234e) de compactación de energía que describe una compactación de energía en una representación distorsionada en el tiempo de la señal de audio y el valor de referencia, y para comparar el valor de relación con uno o más valores de umbral para obtener la señal de activación de distorsión de tiempo como resultado de la comparación.
8. Codificador (200) de señal de audio para codificar una señal (210) de audio de entrada para obtener una representación (212) codificada de la señal de audio de entrada, comprendiendo el codificador de señal de audio: un transformador (220) de distorsión de tiempo configurado para proporcionar una representación (222) espectral transformada por la distorsión de tiempo sobre la base de la señal (210) de audio de entrada utilizando un contorno de distorsión de tiempo; un proveedor (100; 230; 234) de la señal de activación de distorsión de tiempo según una de las reivindicaciones 1 a 7, en el que el proveedor de la señal de activación de distorsión de tiempo está configurado para recibir la señal (210) de audio de entrada y para proporcionar la señal (112; 232; 234p) de activación de distorsión de tiempo; y un controlador (240) configurado para proporcionar de forma selectiva, con dependencia de la señal (112; 232; 234p) de activación de distorsión de tiempo, una información (286) del contorno de distorsión de tiempo recientemente hallada, que describe una parte de contorno de distorsión de tiempo no constante, o una información (288) estándar del contorno de distorsión de tiempo, que describe una parte constante del contorno de distorsión de tiempo, al transformador (220) de distorsión de tiempo para describir el contorno de distorsión de tiempo utilizado por el transformador (220) de distorsión de tiempo.
9. Codificador de señal de audio según la reivindicación 8, en el que el codificador de señal de audio comprende una interfaz (280) de salida configurada para incluir la representación (222) espectral transformada por la distorsión de tiempo dentro de la representación (212) codificada de la señal de audio, y para incluir de forma selectiva, con dependencia de la señal (232) de activación de distorsión de tiempo, una información del contorno de distorsión de tiempo dentro de la representación (212) codificada de la señal de audio.
10. Método (400) para proporcionar una señal de activación de distorsión de tiempo sobre la base de una señal de audio, comprendiendo el método: proporcionar (410) una información de compactación de energía que describe una compactación de energía en una representación espectral transformada por la distorsión de tiempo de la señal de audio; comparar (420) la información de compactación de energía con un valor de referencia; y

- proporcionar (430) la señal de activación de distorsión de tiempo con dependencia del resultado de la comparación;
- en el que se proporciona una medida de entropía perceptiva (pe) que describe la representación (234e; 234k) espectral transformada por la distorsión de tiempo de la señal de audio como la información (122; 234m; 234n) de compactación de energía; y
- 5 en el que se calcula un número (nl) estimado de líneas que no son cero para una o más bandas de factor de escala de la representación (234e; 234k) espectral transformada por la distorsión de tiempo de la señal de audio sobre la base de una información de factor de forma ($ffac(n)$) de la banda de factor de escala, y en el que se calcula la medida de entropía (326) perceptiva para una banda de factor de escala en consideración utilizando una multiplicación del número (nl) estimado de líneas que no son cero y una medida de energía de la banda de factor de escala en consideración.
- 10 11. Método (400) para proporcionar una señal de activación de distorsión de tiempo sobre la base de una señal de audio, comprendiendo el método:
- proporcionar (410) una información de compactación de energía que describe una compactación de energía en una representación espectral transformada por la distorsión de tiempo de la señal de audio;
- comparar (420) la información de compactación de energía con un valor de referencia; y
- 15 proporcionar (430) la señal de activación de distorsión de tiempo con dependencia del resultado de la comparación;
- en el que se determina una suma de valores absolutos de una función de autocorrelación normalizada de la representación (234e; 234k) distorsionada en el tiempo de la señal de audio para obtener la información de compactación de energía.
- 20 12. Método (400) para proporcionar una señal de activación de distorsión de tiempo sobre la base de una señal de audio, comprendiendo el método:
- proporcionar (410) una información de compactación de energía que describe una compactación de energía en una representación espectral transformada por la distorsión de tiempo de la señal de audio;
- comparar (420) la información de compactación de energía con un valor de referencia; y
- proporcionar (430) la señal de activación de distorsión de tiempo con dependencia del resultado de la comparación;
- 25 en el que se proporciona una medida de planicidad espectral que describe la representación (234e; 234k) espectral transformada por la distorsión de tiempo de la señal de audio como la información (122; 234m; 234n) de compactación de energía; y
- en el que se calcula un cociente de una media geométrica del espectro (234e; 234k) de potencia transformado por la distorsión de tiempo de la señal de audio y una media aritmética del espectro (234e; 234k) de potencia transformado por la distorsión de tiempo de la señal de audio para obtener la medida de planicidad espectral.
- 30 13. Método (450) para codificar una señal de audio de entrada para obtener una representación codificada de la señal de audio de entrada, comprendiendo el método:
- proporcionar (470) una señal de activación de distorsión de tiempo según una de las reivindicaciones 10 a 12, en el que la información de compactación de energía describe una compactación de energía en una representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada; y
- 35 proporcionar (480) de forma selectiva, con dependencia de la señal de activación de distorsión de tiempo, una descripción de la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada o una descripción de una representación espectral no transformada por la distorsión de tiempo de la señal de audio de entrada para su inclusión en la representación codificada de la señal de audio de entrada.
- 40 14. Programa informático para realizar el método de reivindicación 10 a 13 cuando el programa informático se ejecuta en una computadora.

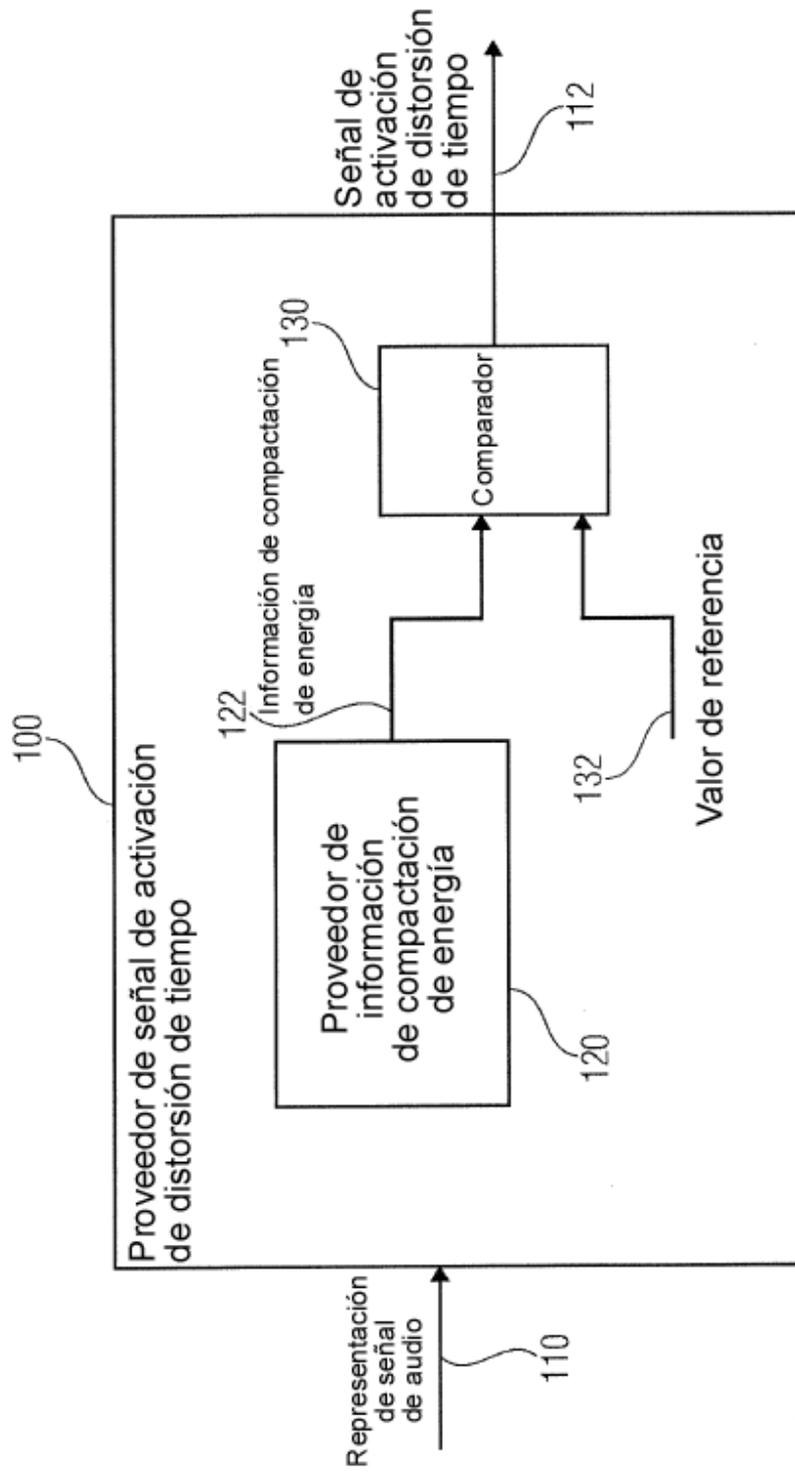


FIG 1

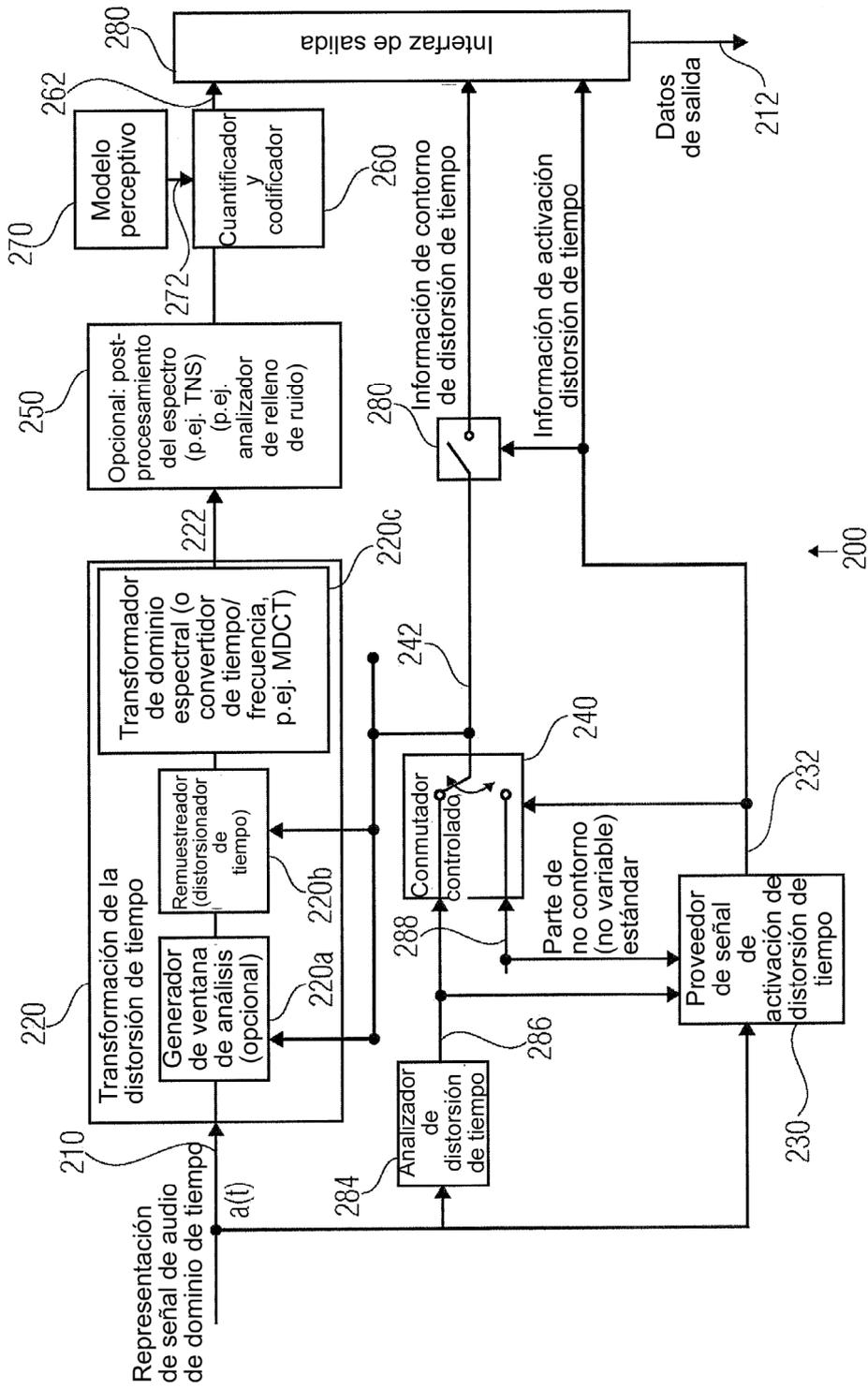


FIG 2A

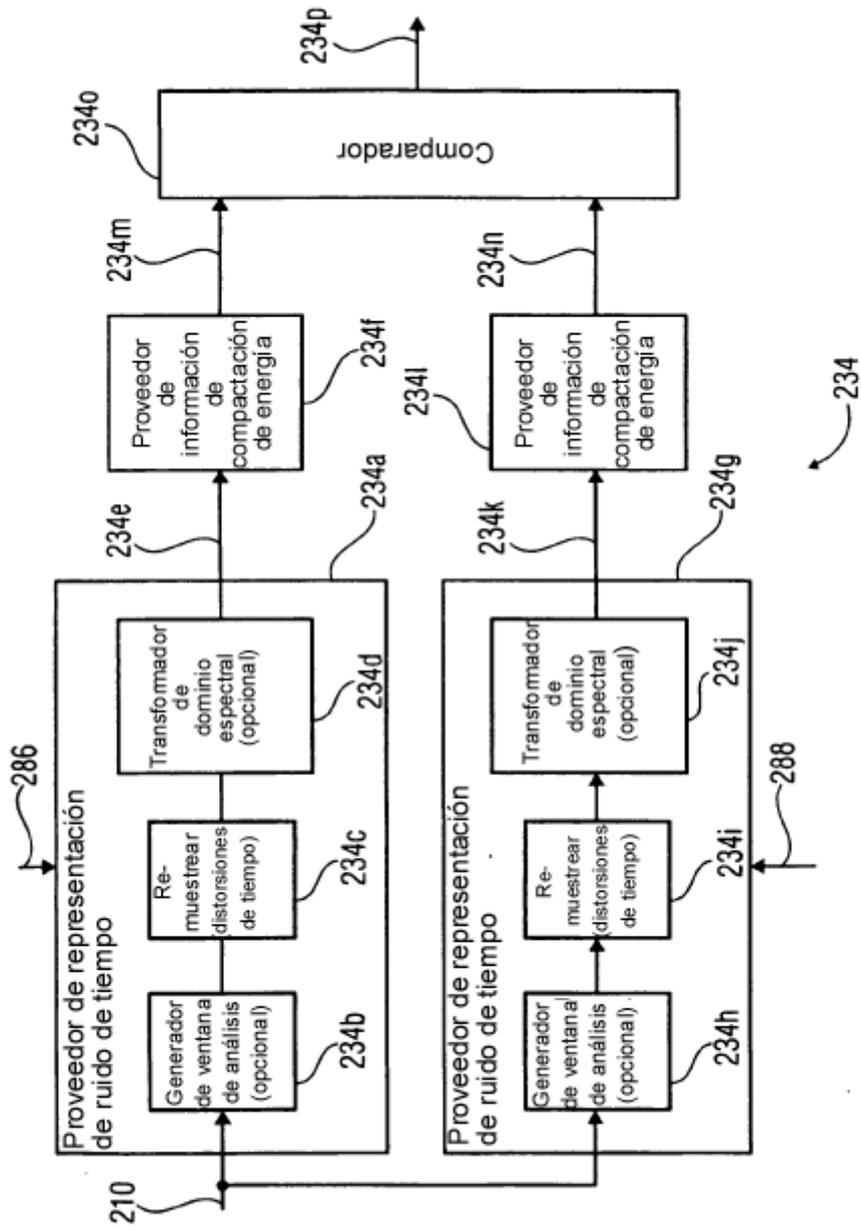
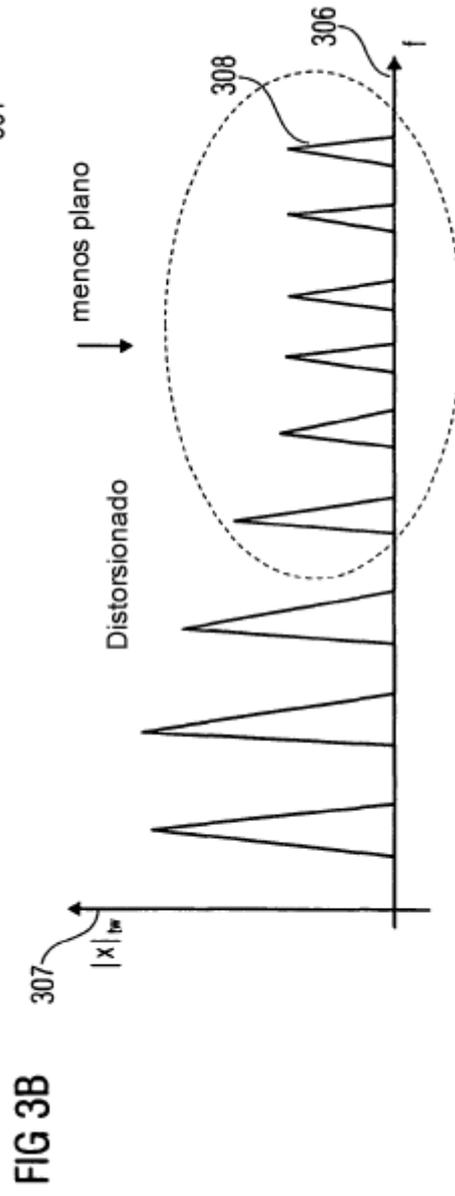
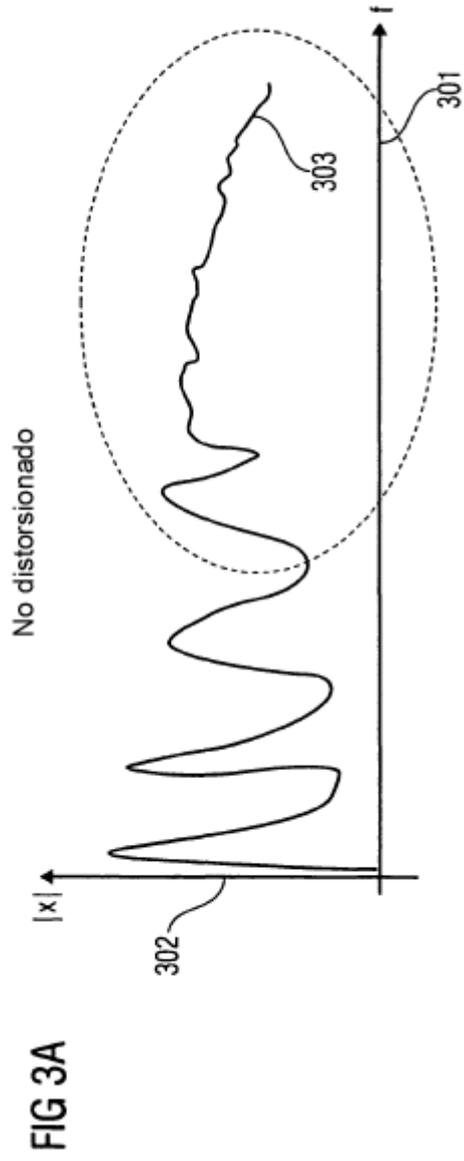


FIG 2B



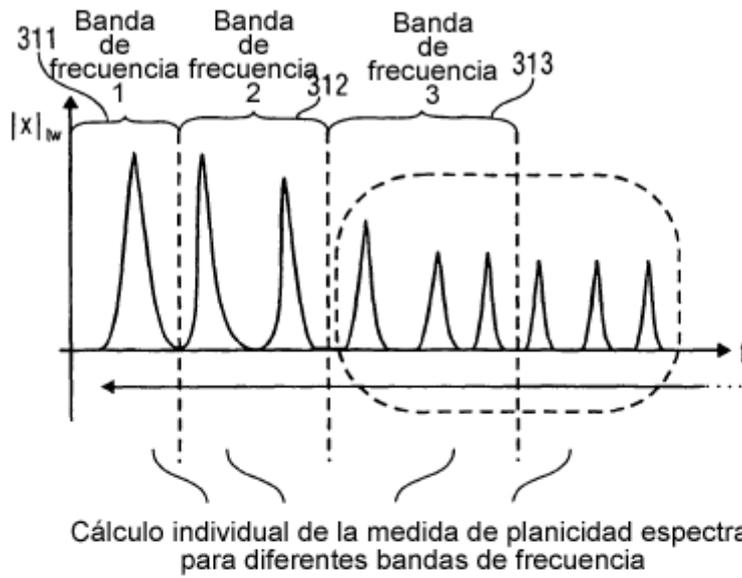
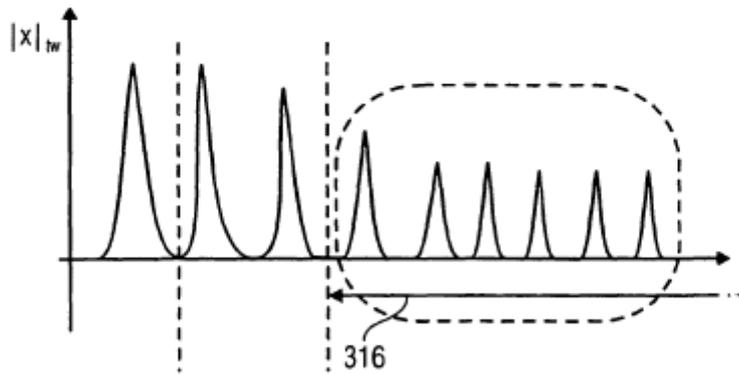


FIG 3C



316

Consideración de sólo la parte de frecuencia más alta (omitiendo la parte de frecuencia más baja) para calcular la medida de planicidad espectral (por banda de frecuencia o a lo largo de toda la parte de frecuencia más alta)

FIG 3D

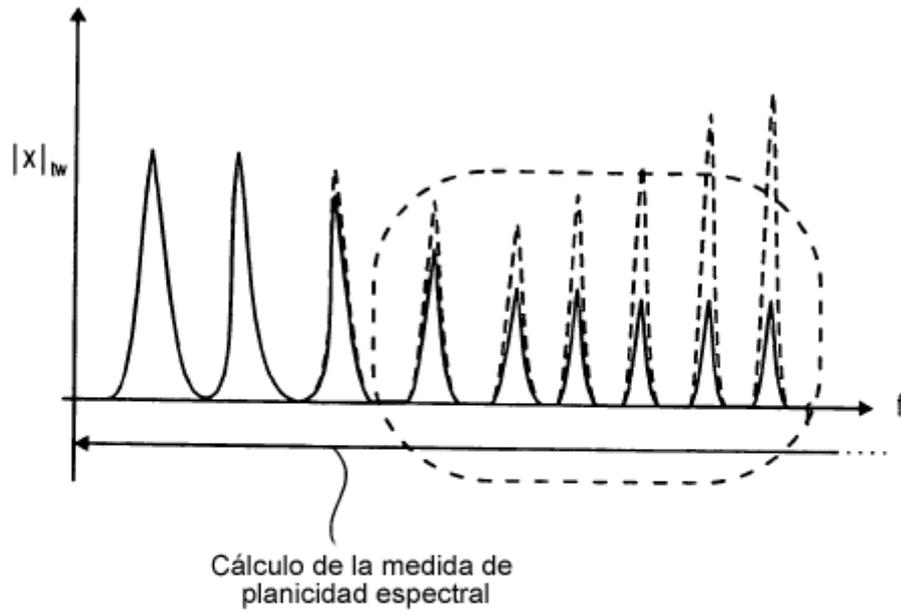


FIG 3E

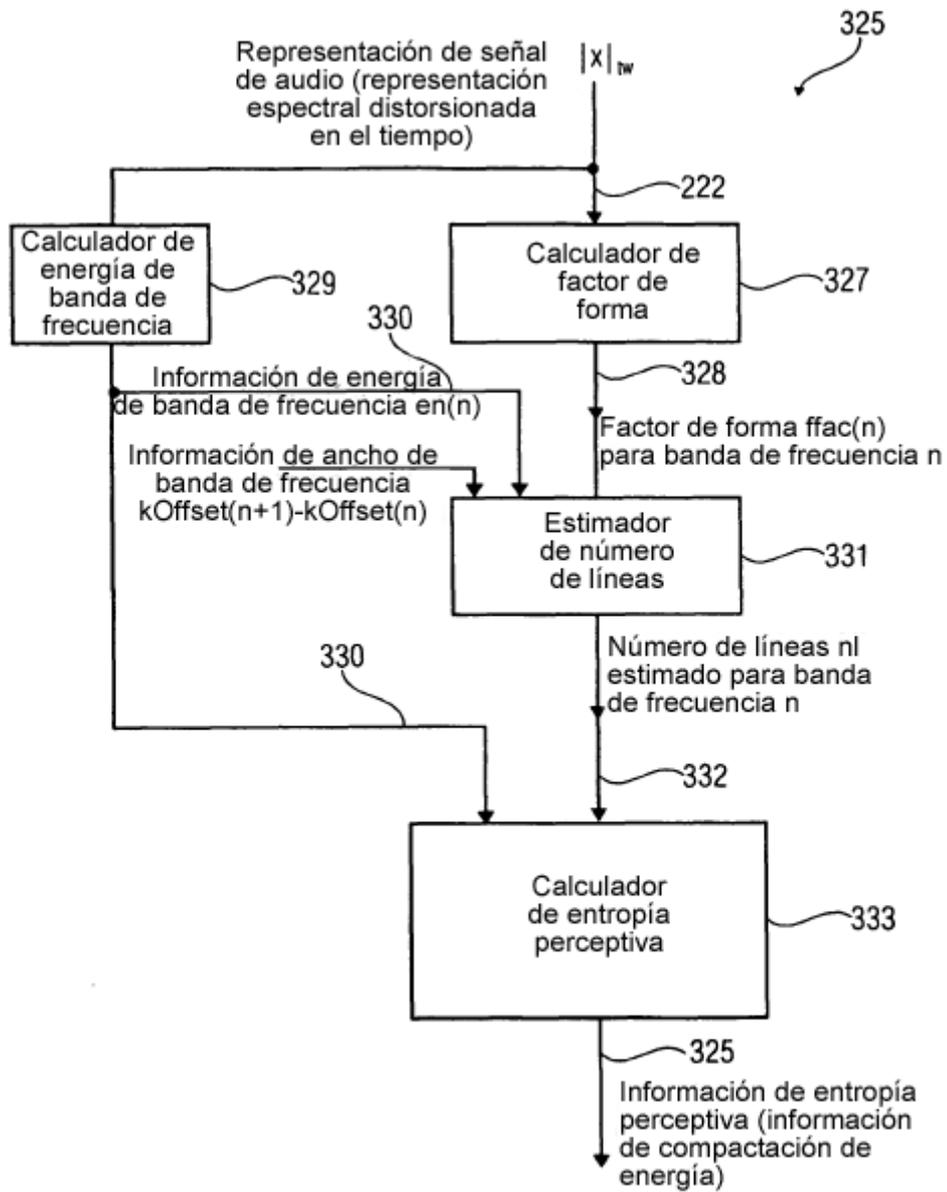


FIG 3F

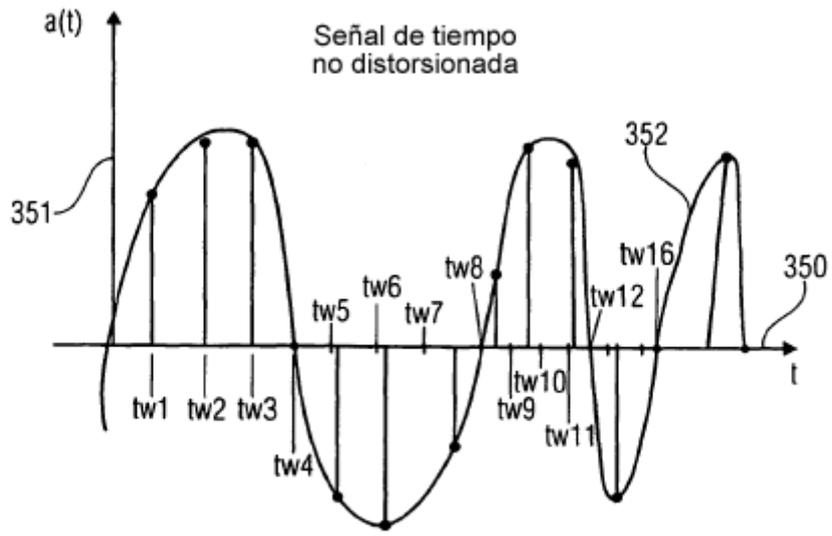


FIG 3G

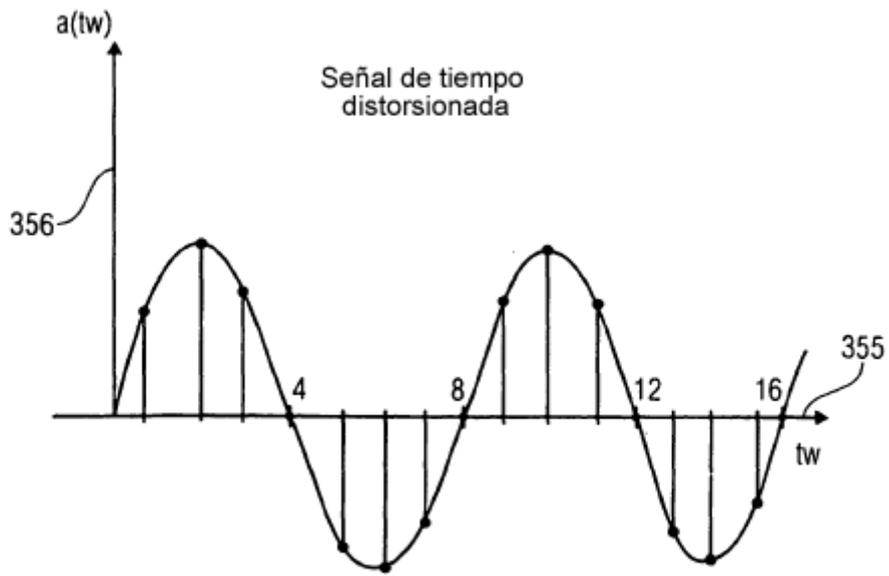


FIG 3H

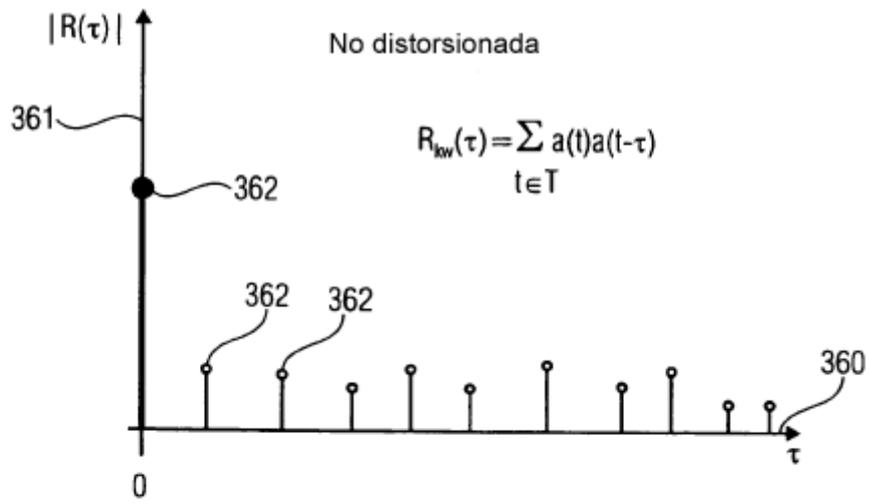


FIG 3I

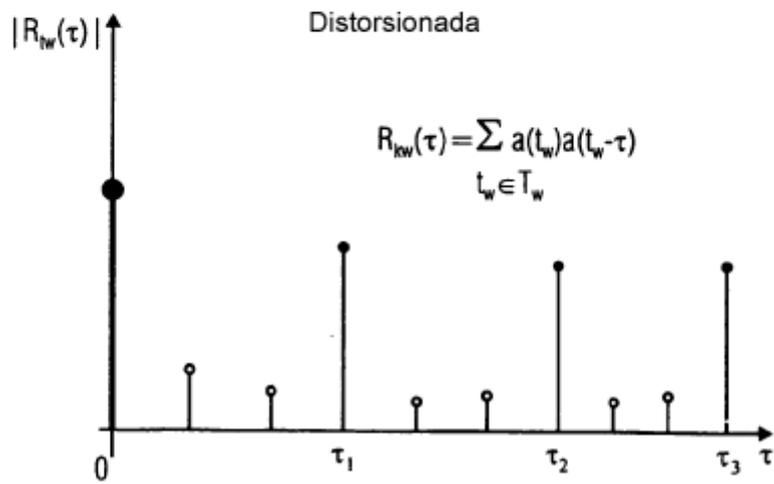


FIG 3J

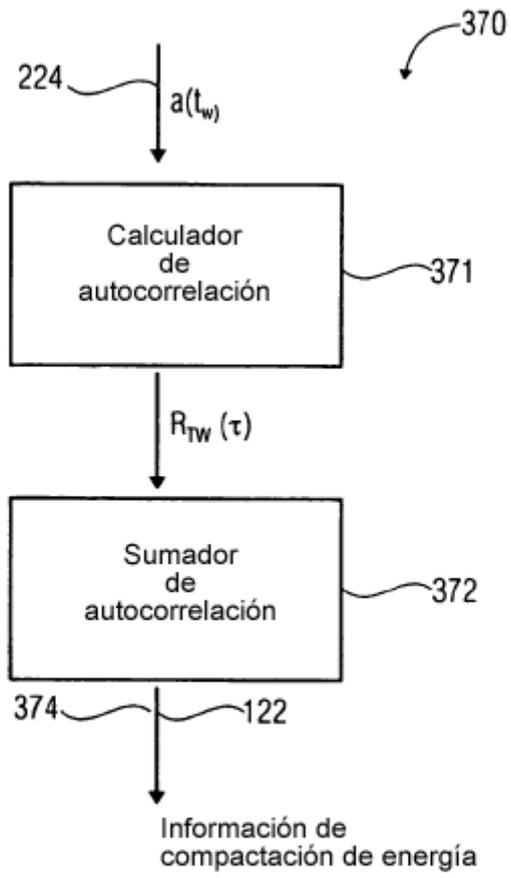


FIG 3K

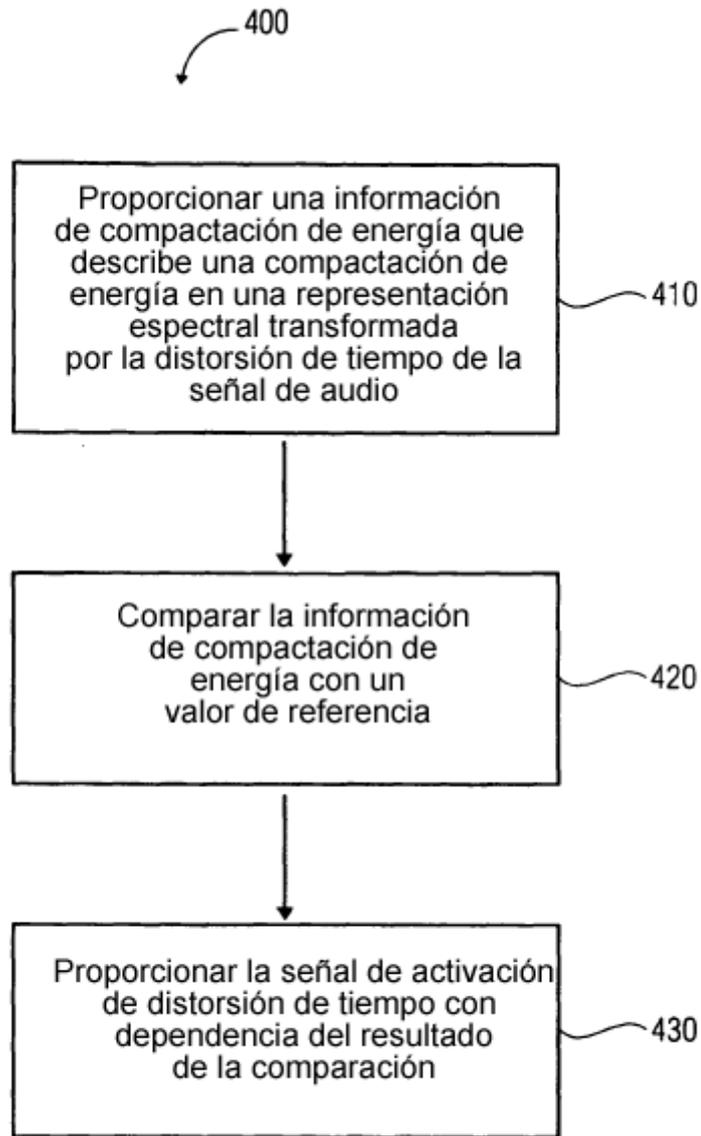


FIG 4A

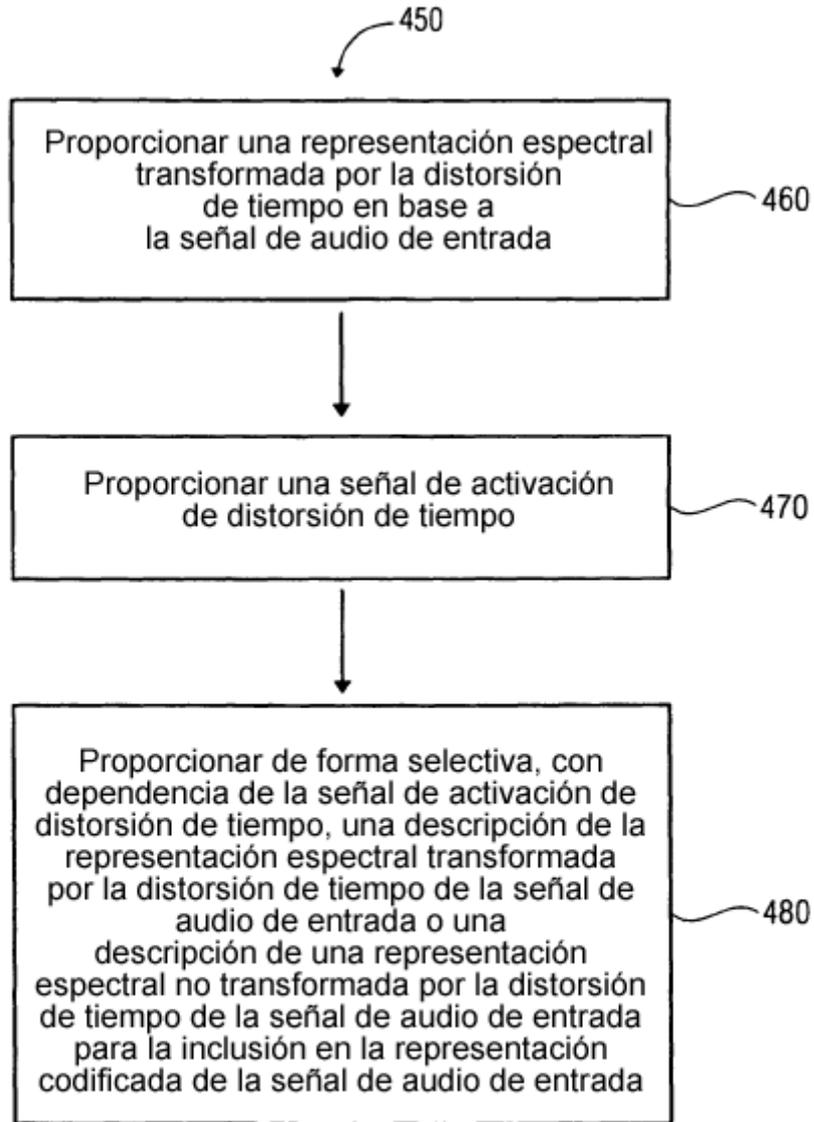


FIG 4B

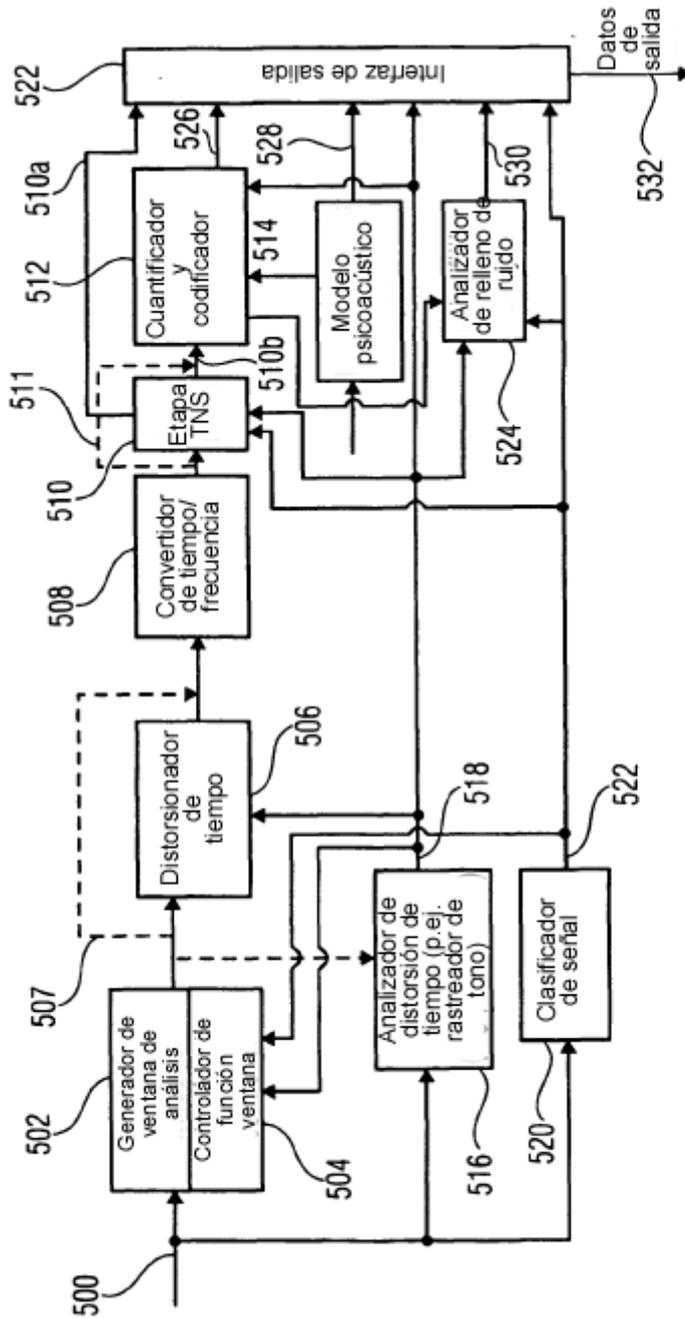


FIG 5A
(CODIFICADOR)

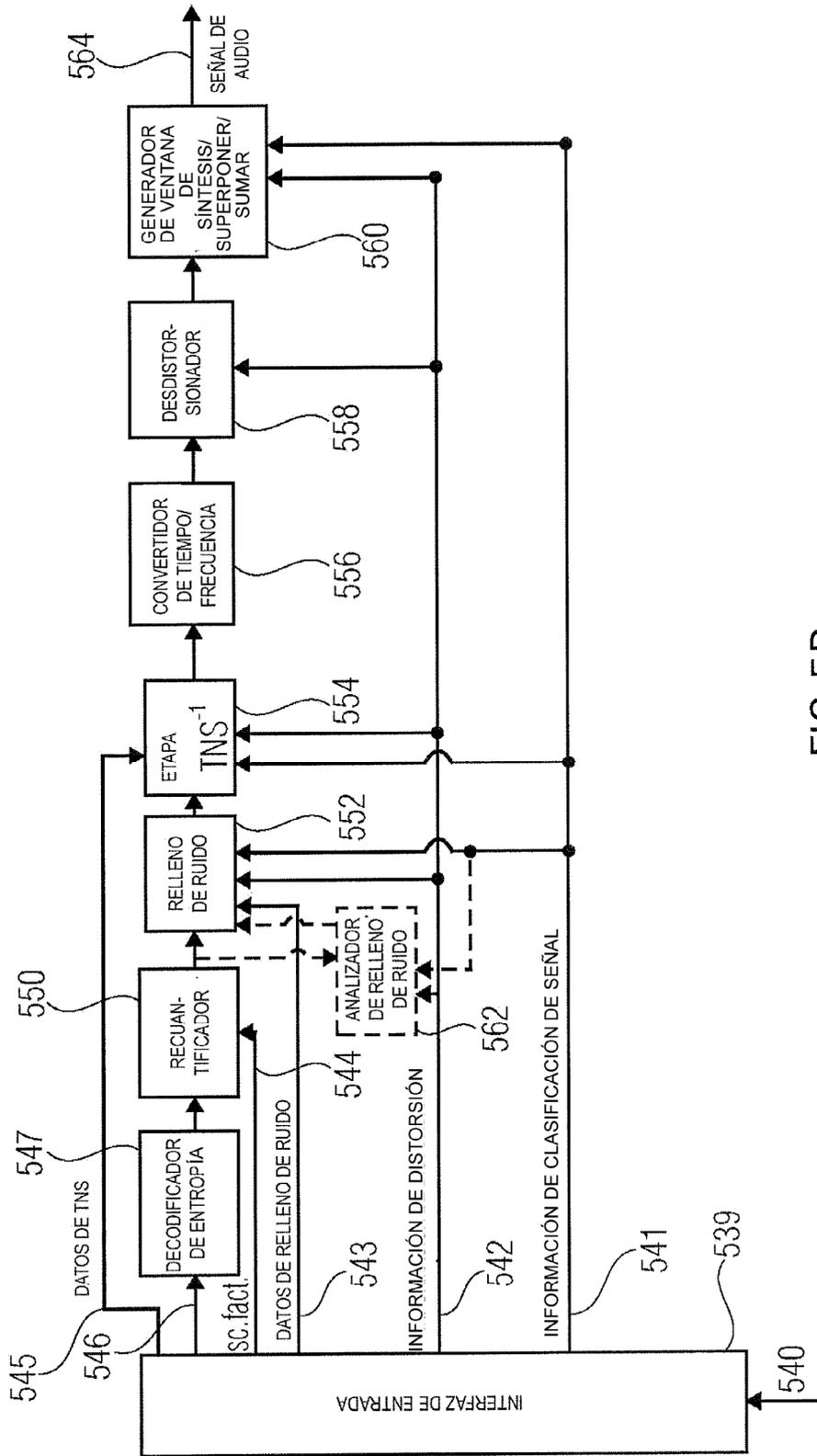


FIG 5B
(DECODIFICADOR)

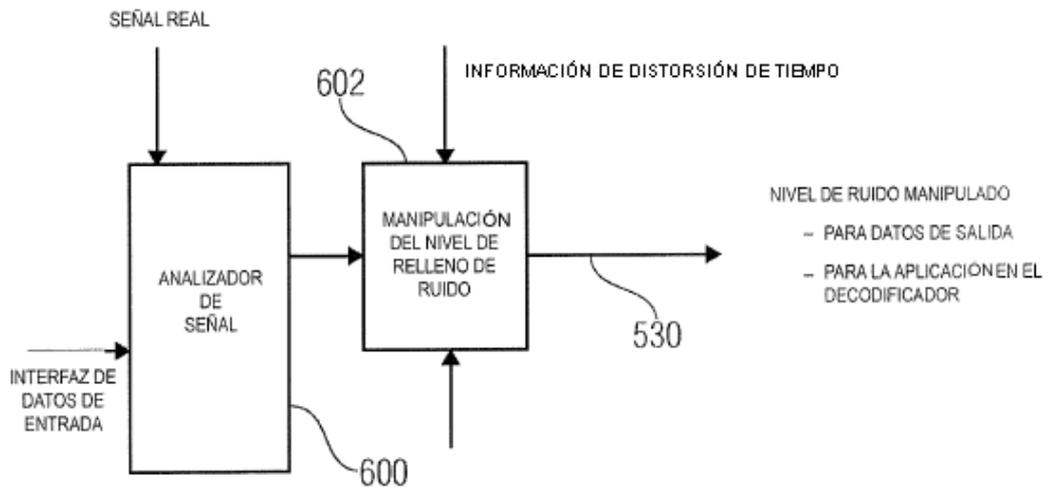


FIG 6A

CLASIFICACIÓN DE VOZ		DISTORSIÓN DE TIEMPO		CLASIFICACIÓN DE SEÑAL		NIVEL DE RELLENO DE RUIDO
V	UV	sí	NO	VOZ	SIN VOZ	
			X		X	NORMAL
		X			X	BAJO
X		X		X		MUY BAJO / CERO
	X		X	X		NORMAL

FIG 6B

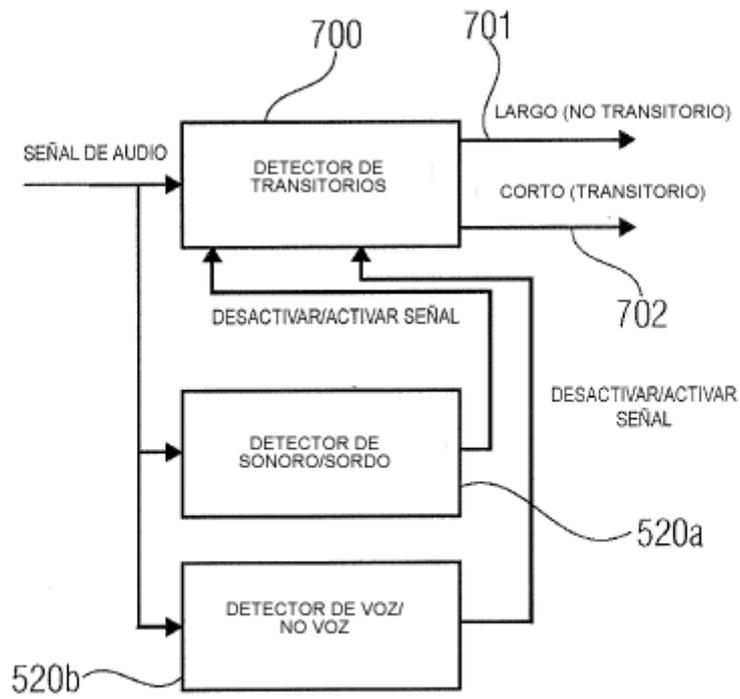


FIG 7A

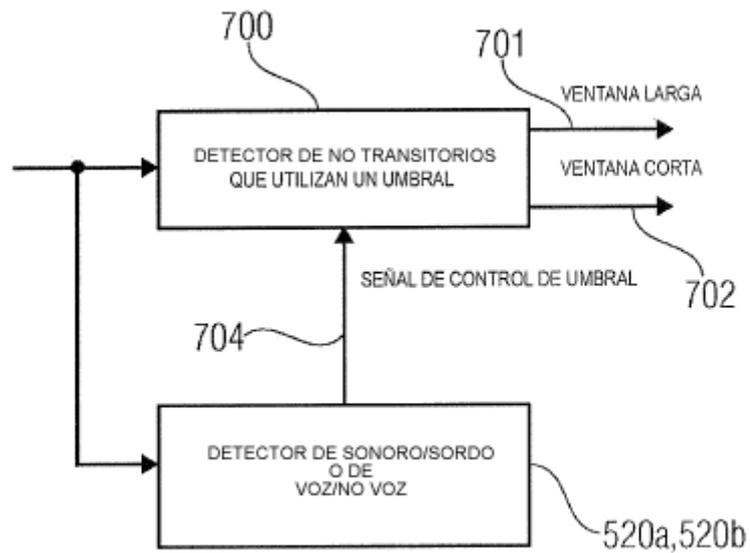


FIG 7B

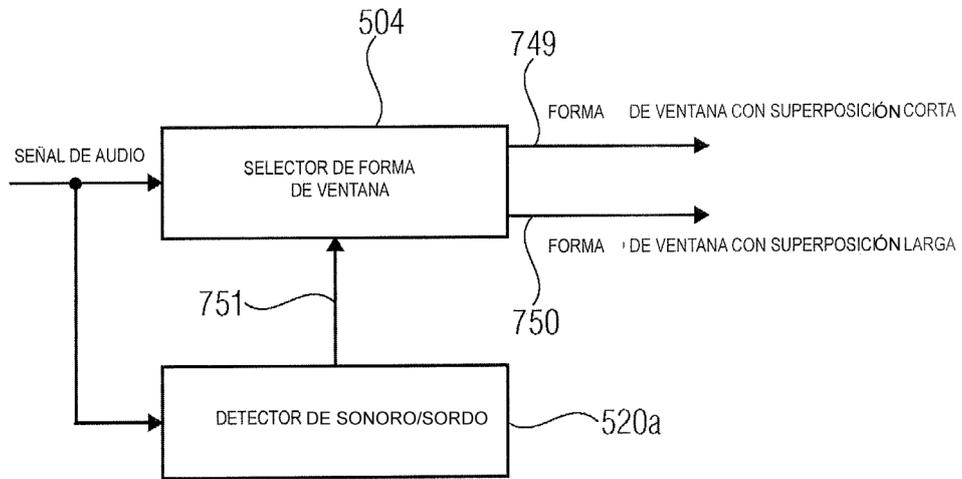


FIG 7C

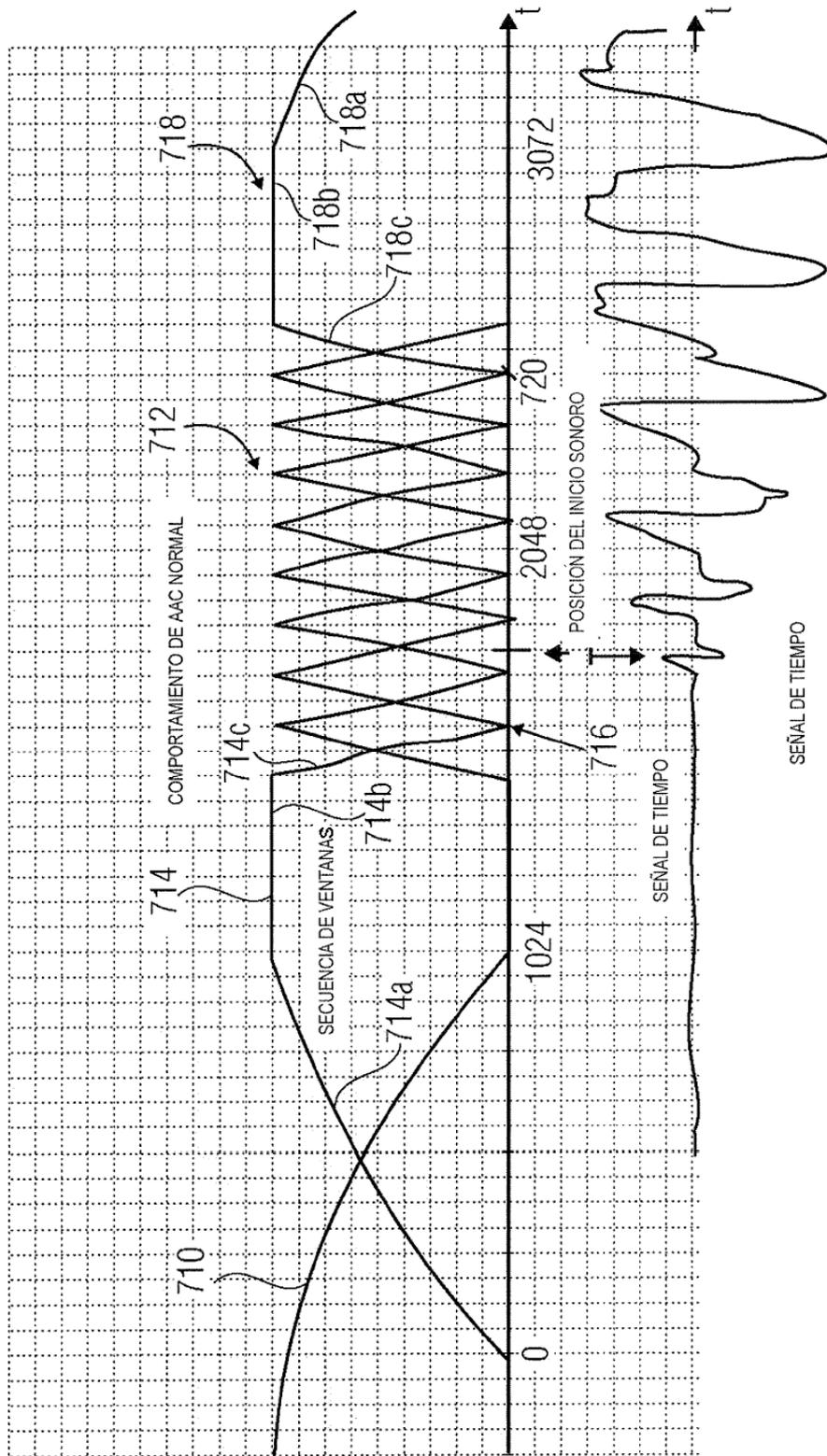


FIG 7D

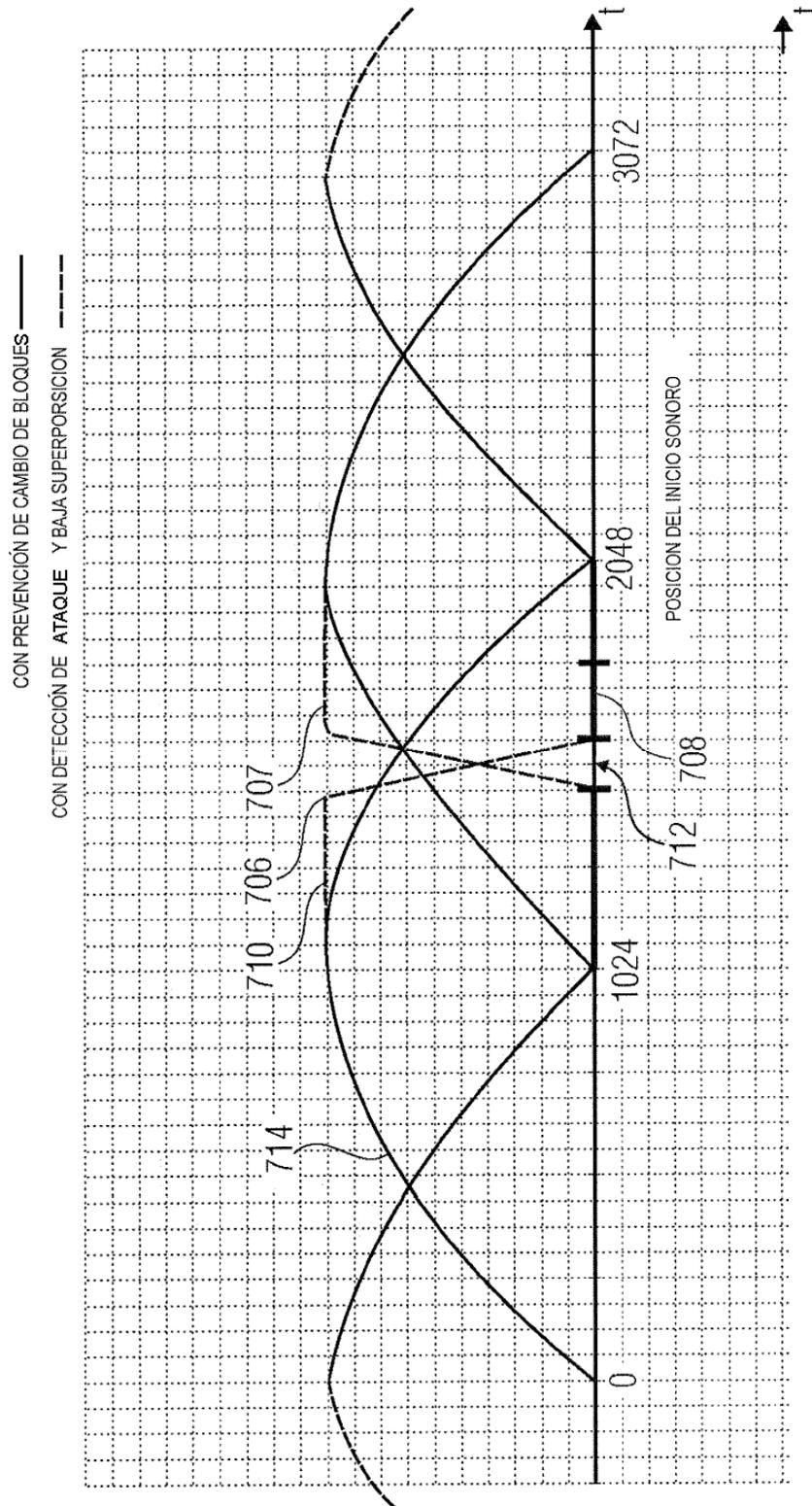


FIG 7E

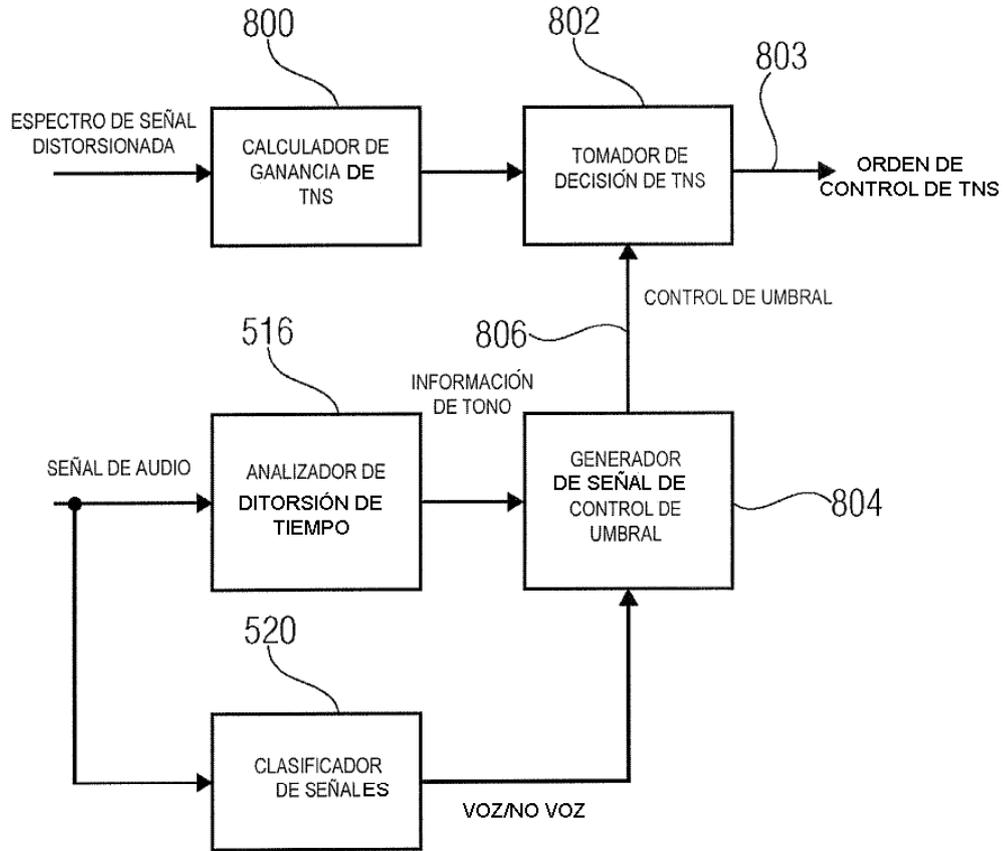


FIG 8A

CONTORNO DE TONOS		SONORA	SORDA	UMBRAL DE DECISIÓN DE TNS
		CLASIFICACIÓN DE SEÑALES		
SÍ	NO	VOZ	NO VOZ	
	X		X	NORMAL
X			X	MÁS BAJO
X		X		(AÚN) MÁS BAJO

FIG 8B

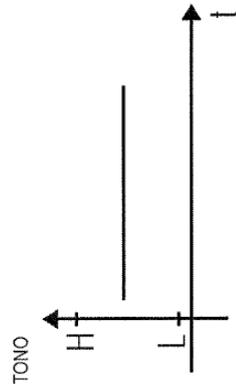
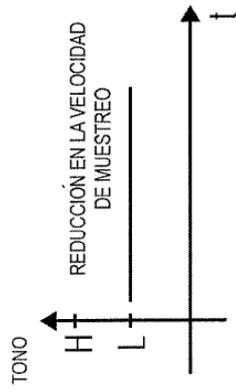
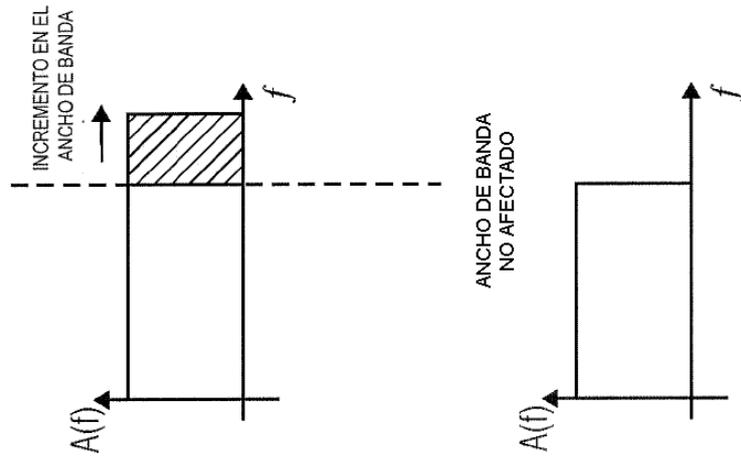


FIG 9D

FIG 9E

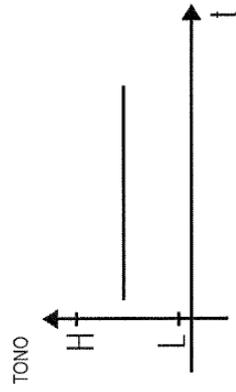
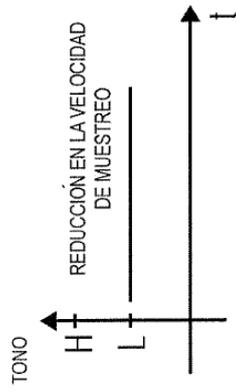
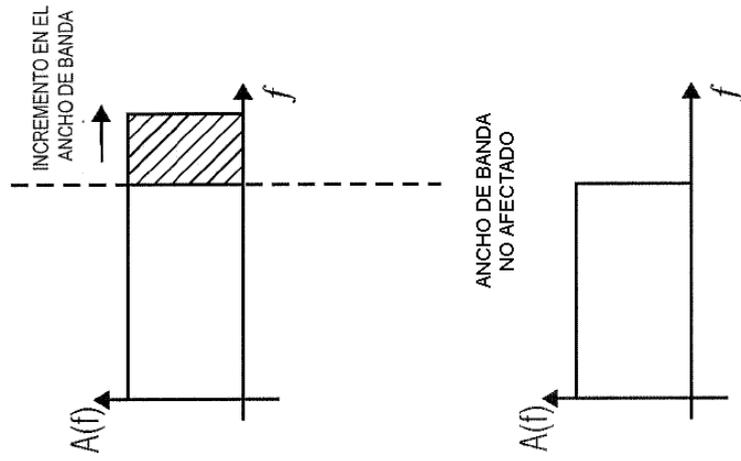


FIG 9D

FIG 9E

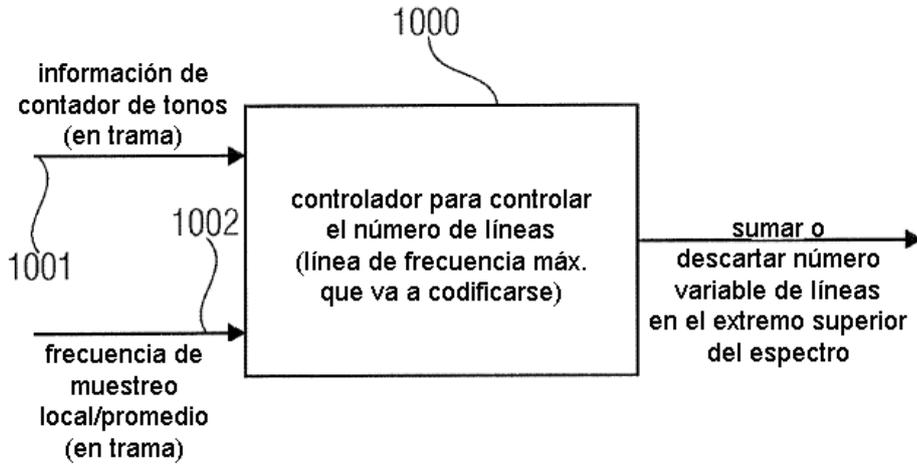


FIG 10A

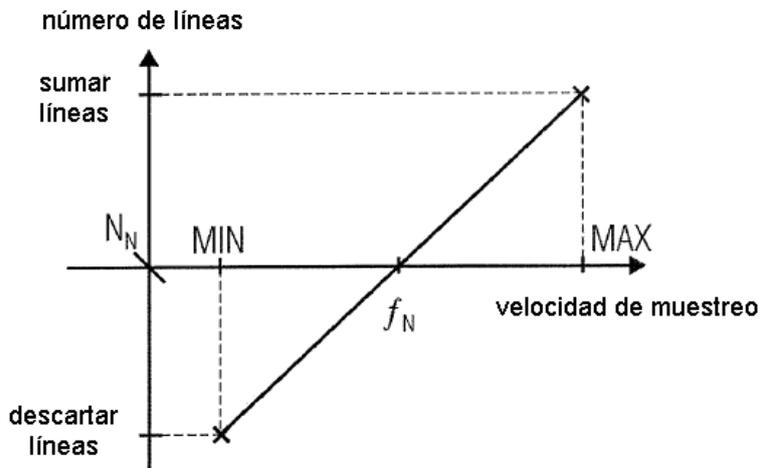


FIG 10B

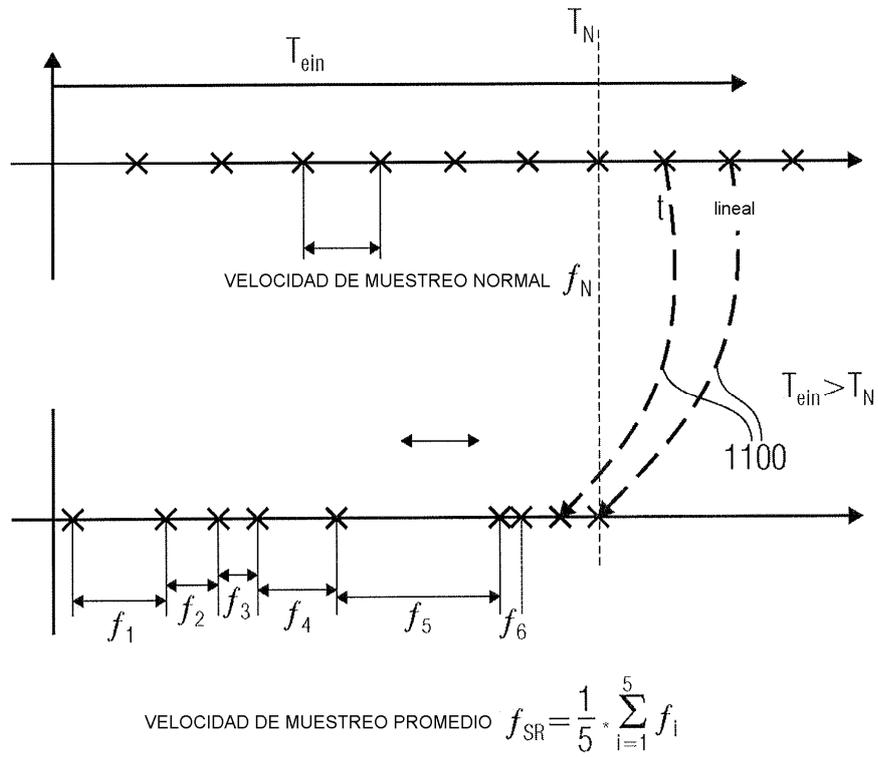


FIG 11

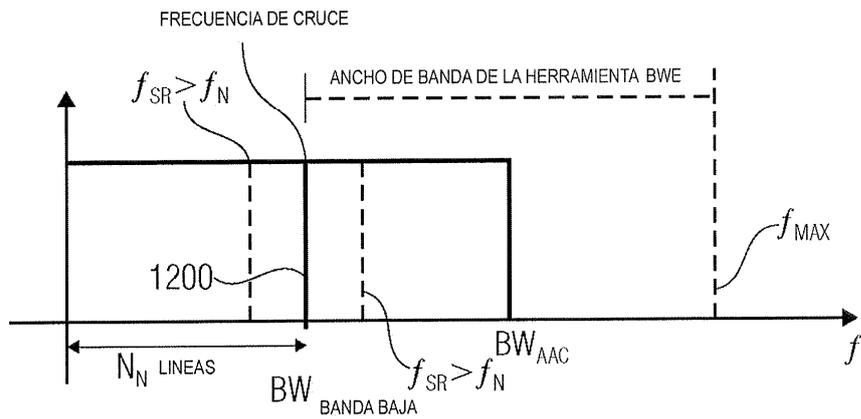


FIG 12A

f_{SR} : VELOCIDAD DE MUESTREO LOCAL
 Δf : ANCHO DE UNA LÍNEA ESPECTRAL/
 DISTANCIA ENTRE LÍNEAS ESPECTRALES

N : NÚMERO DE LÍNEAS SEGÚN LO DETERMINADO
 POR EL PROCESADOR

T_{ein} : LONGITUD DE TRAMA NO DISTORSIONADA
 (DEPENDIENDO DE LA CARACTERÍSTICA
 DE DISTORSIÓN)

T_{ein}	f_{SR}	Δf	N
$= T_N$	$= f_N$	$= \Delta f_N$	$= N_N$
$> T_N$	$> f_N$	$> \Delta f_N$	$< N_N$ ELIMINAR LÍNEAS
$< T_N$	$< f_N$	$< \Delta f_N$	$> N_N$ SUMAR LÍNEAS

FIG 12B