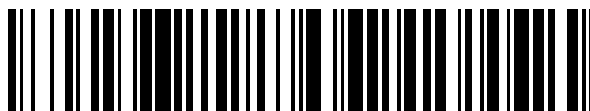


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 741 963**

51 Int. Cl.:

G10L 19/022 (2013.01)

G10L 19/025 (2013.01)

G10L 25/78 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.07.2009 E 11180988 (5)**

97 Fecha y número de publicación de la concesión europea: **26.06.2019 EP 2410520**

54 Título: **Codificadores de señal de audio, métodos para codificar una señal de audio y programas informáticos**

30 Prioridad:

11.07.2008 US 79873 P

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

12.02.2020

73 Titular/es:

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

72 Inventor/es:

**BAYER, STEFAN;
DISCH, SASCHA;
GEIGER, RALF;
NEUENDORF, MAX;
SCHULLER, GERALD;
FUCHS, GUILLAUME y
EDLER, BERND**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 741 963 T3

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

DESCRIPCIÓN

Codificadores de señal de audio, métodos para codificar una señal de audio y programas informáticos

5 La presente invención se refiere a codificación y decodificación de audio y específicamente a la codificación/decodificación de la señal de audio que tiene un contenido armónico o de expresión, el cual se puede someter a un procesamiento de distorsión (warp) de tiempo.

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

15 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 eficiencia esta representación de dominio de frecuencia, por ejemplo tomando en cuenta umbrales de enmascaramiento perceptual. Este concepto de codificación de señal de audio es particularmente eficiente 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 ser insignificante (o codificado con la longitud de código mínima).

20 Por ejemplo, las transformadas con superposición moduladas por seno o 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 (sub-bandas), lo cual conduce a una representación de señal eficiente. Se proporciona un ejemplo en "A Window Switching Algorithm for AVS Audio Coding" de Shuixian *Et al.*, NETWORKING AND MOBILE COMPUTING, WICOM 2007, INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, EE.UU, 21 de septiembre de 2007".

30 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 expresión común, el tono es la frecuencia de la señal de la excitación modulada por la garganta humana. Si solo una frecuencia fundamental simple estuviera presente, el espectro sería extremadamente simple, comprendiendo la frecuencia fundamental y los sobretonos únicamente. Un espectro de este tipo podría codificarse en forma muy eficiente. Para señales con tono variable, sin embargo, la energía correspondiente a cada componente armónico se dispersa sobre varios coeficientes de transformación, conduciendo de este modo a una reducción de la eficiencia de codificación.

35 A fin de solucionar esta reducción de la eficiencia de codificación, la señal de audio a ser codificada se re-muestrea en forma eficaz sobre una rejilla temporal no uniforme. En el procesamiento subsiguiente, las posiciones de muestra obtenidas por el re-muestreado no uniforme se procesan como si representaran los valores sobre una rejilla temporal uniforme. Esta operación normalmente se denota por la frase "distorsión de tiempo". Los tiempos de muestra se pueden elegir en forma ventajosa dependiendo de la variación temporal del tono, de modo que una variación de tono en la versión de tiempo distorsionado de la señal de audio es más pequeña que una variante de tono en la versión original de la señal de audio (antes de la distorsión de tiempo). Esta variante de tono además puede denotarse con la frase "contorno de distorsión de tiempo". Después de la distorsión de tiempo de la señal de audio, la versión distorsionada de tiempo de la señal de audio se convierte en el dominio de frecuencia. La distorsión de 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 (sin distorsión de tiempo).

50 En el lado del decodificador, la representación de dominio de frecuencia de la señal de audio distorsionada en el tiempo se retro convierte al dominio de tiempo, de modo que una representación de dominio de tiempo de la 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 por el lado del decodificador no se incluyen. En consecuencia, incluso se aplica otra distorsión de tiempo por el re-muestreado de la representación del dominio del tiempo reconstruido en el lado del decodificador de la señal de audio distorsionada en el tiempo. A fin de obtener una buena reconstrucción de la señal de audio de entrada por el lado del decodificador en el decodificador, se desea que la distorsión de tiempo del lado del decodificador sea al menos aproximadamente la operación inversa con respecto a la distorsión de tiempo del lado del decodificador. A fin de obtener una distorsión de tiempo apropiada, se desea tener información disponible en el decodificador lo cual permite un ajuste de la distorsión de tiempo en el lado del decodificador.

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 pequeña velocidad de bits requerida para esta transmisión mientras aún se permite una reconstrucción confiable de la información de distorsión de tiempo requerida en el lado del decodificador.

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

10 Es el objetivo de algunos ejemplos de comparación crear conceptos para mejorar la impresión del oído provista por una señal de audio codificada sobre la base de información disponible en un codificador de señal de audio de distorsión de tiempo o un decodificador de señal de audio de distorsión de tiempo.

15 Este objetivo se logra mediante un proveedor de la señal de activación de distorsión de tiempo para aportar una señal de activación de distorsión de tiempo sobre la base de una representación de una señal de audio tal como se describe a continuación, un codificador de señal de audio para codificar una señal de audio de entrada tal como se describe a continuación, un método para proveer una señal de activación de distorsión de tiempo tal como se describe a continuación, un método para proveer una representación codificada de una señal de audio de entrada tal como se describe a continuación, o un programa informático tal como se describe a continuación.

20 Es un objetivo adicional de la presente invención proveer un esquema de codificación/decodificación de audio mejorado, el cual provee una velocidad de bits más baja o de calidad superior

Este objetivo se obtiene mediante un codificador de audio según la reivindicación 1 ó 2, un método de codificación de audio según la reivindicación 3 ó 4 o un programa informático según la reivindicación 5.

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

30 Una realización según un ejemplo de comparación 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 proveer una información de compactación de energía que describe una compactación de energía en una representación del espectro transformado por la distorsión de tiempo de la señal de audio. El proveedor de la señal de activación de distorsión de tiempo también comprende un comparador configurado para comparar la información de compactación de energía con un valor de referencia, y proveer la señal de activación de distorsión de tiempo con dependencia de un resultado de la comparación.

40 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 otorga una mejora, en el sentido de una reducción de la velocidad de bits de la señal de audio codificada, si la representación del espectro transformado por la distorsión de tiempo de la señal de audio comprende una distribución de energía lo suficientemente compacta en que la energía se concentra en una o más regiones espectrales (o líneas espectrales). Esto se debe al hecho de que una distorsión de tiempo exitosa otorga el efecto de disminuir la velocidad de bits al transformar un espectro embarrado, por ejemplo de un marco de audio, dentro del espectro que tiene uno o más picos discernibles, y en consecuencia tiene una compactación de energía más alta que el espectro de la señal de audio original (sin distorsión de tiempo).

50 Con respecto a esta publicación, debe entenderse que un marco de la señal de audio, durante el cual el tono de la señal de audio varía significativamente, comprende un espectro embarrado. El tono con variación en el tiempo de la señal de audio tiene el efecto de que una transformación de dominio de tiempo en un dominio de frecuencia sobre el marco de la señal de audio da por resultado una distribución embarrada de la energía de señal sobre la frecuencia, en particular en la región de mayor frecuencia. En consecuencia, una representación del espectro de esta señal de audio original (sin distorsión de tiempo) comprende una compactación de energía baja y típicamente no exhibe picos espectrales en una porción de frecuencia más alta del espectro, o sólo exhibe picos espectrales relativamente pequeños en la porción de frecuencia más alta del espectro. A diferencia, si la distorsión de tiempo es con éxito (en cuanto a suministrar una mejora de la eficiencia de codificación) la distorsión de tiempo de la señal de audio original da una señal de audio distorsionada en el tiempo que tiene un espectro con picos relativamente más altos y claros (en particular en la porción de 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 del espectro de la señal de audio distorsionada en el tiempo (la cual se puede considerar como una representación del espectro transformado por la distorsión de tiempo de la señal de audio) comprende uno o más picos espectrales claros. En otras palabras, el manchado del espectro de la señal de audio original (que tiene tono temporalmente variable) es reducido por una operación de distorsión de tiempo con éxito, de modo que la representación del espectro transformado por la distorsión de tiempo de la señal de audio comprende compactación de energía más

alta que el espectro de la señal de audio original. No obstante, la distorsión de tiempo no siempre es con éxito en la mejora de la eficiencia de la codificación. Por ejemplo, la distorsión de tiempo no mejora la eficiencia 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.

5 En vista de esta situación, la información de compactación de energía provista por el proveedor de información de compactación de energía es un indicador valioso para decidir si la distorsión de tiempo es con éxito en cuanto a reducción de la velocidad.

10 Una realización crea un proveedor de la señal de activación de distorsión de tiempo para proveer 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 configurado para proveer dos representaciones de la distorsión de tiempo de la misma señal de audio que utiliza diferente información del contorno de distorsión de tiempo. De este modo, los proveedores de representación de la distorsión de tiempo pueden estar configurados (en forma estructural y/o funcional) de la misma forma y utilizar la misma señal de audio pero diferente información del contorno de distorsión de tiempo. El proveedor de la señal de activación de distorsión de tiempo además comprende dos proveedores de información de compactación de energía configurado para proveer una primera información de compactación de energía sobre la base de la primera representación de distorsión de tiempo y proveer 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 pueden estar configurados de la misma forma pero para utilizar las diferentes representaciones de distorsión de tiempo. Más aún 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 proveer la señal de activación de distorsión de tiempo con dependencia de un resultado de la comparación.

25 En una realización preferida, el proveedor de información de compactación de energía está configurado para proveer una medida de planicidad espectral que describe la representación del espectro transformado por la distorsión de tiempo de la señal de audio como la información de compactación de energía. Se ha hallado que la distorsión de tiempo es con 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 se puede utilizar para decidir, sin realizar un proceso de codificación espectral completo, si la distorsión de tiempo debe ser activada o desactivada.

35 En una realización preferida, el proveedor de información de compactación de energía está configurado para computar un cociente de un medio geométrico de un medio geométrico de espectro de potencia de transformación de la distorsión de tiempo y un medio aritmético de espectro de potencia de transformación de la distorsión de tiempo, para obtener la medida de la planicidad espectral. Se ha hallado 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 de tiempo.

45 En otra realización preferida, el proveedor de información de compactación de energía está configurado para enfatizar una porción de frecuencia más alta de la representación del espectro transformado por la distorsión de tiempo cuando se compara con una porción de frecuencia más baja de la representación del espectro transformado 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 alto que sobre el rango de frecuencia más bajo. En consecuencia, una evaluación dominante del rango de frecuencia más alto es apropiada para determinar la efectividad de la distorsión de tiempo utilizando una medida de la planicidad espectral. Además, las señales típicas de audio exhiben un contenido armónico (que comprende la armonía de una frecuencia fundamental) la cual decae en intensidad con la frecuencia creciente. Un énfasis de una porción de frecuencia superior de la representación del espectro transformado por la distorsión de tiempo cuando se compara con una porción de frecuencia inferior a la de la representación del espectro transformado por la distorsión de tiempo además ayuda a compensar esta decadencia típica de las líneas espectrales con frecuencia creciente. Para resumir, una consideración enfatizada de la porción de frecuencia más alta del espectro otorga una confiabilidad aumentada de la información de compactación de energía y por lo tanto permite una provisión más confiable de la señal de activación de distorsión de tiempo.

60 En otra realización preferida, el proveedor de información de compactación de energía está configurado para proveer una pluralidad de medidas banda por banda de planicidad espectral, y para computar un promedio de la pluralidad de medidas banda por banda de planicidad espectral, para obtener la información de compactación de energía. Se ha hallado que la consideración de medidas espectrales de planicidad banda por banda otorga una información particularmente confiable respecto a si la distorsión de tiempo es efectiva para reducir la velocidad de bits de una señal de audio codificada. En primer lugar, la codificación de la representación del espectro transformado 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, una computación de banda por banda de medidas de planicidad espectral sustancialmente elimina 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 manchado 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, se puede tomar 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.

En otra realización preferida, el proveedor de la señal de activación de distorsión de tiempo comprende un calculador de valor de referencia configurado para computar una medida de planicidad espectral que describe una representación del espectro sin distorsión de tiempo de la señal de audio, para obtener el valor de referencia. En consecuencia, la señal de activación de distorsión de tiempo puede ser provista sobre la base de una comparación de la planicidad espectral de una versión sin distorsión de tiempo de la señal de audio de entrada y una planicidad espectral de una versión de distorsión de tiempo de la señal de audio de entrada.

En otra realización preferida, el proveedor de información de compactación de energía está configurado para proveer una medida de entropía perceptual que describe la representación del espectro transformado 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 perceptual de la representación del espectro transformado por la distorsión de tiempo es un buen estimado de un número de bits (o una velocidad) requerido para codificar el espectro transformado por la distorsión de tiempo. En consecuencia, la medida de entropía perceptual de la representación del espectro transformado por la distorsión de tiempo es una buena medida respecto si puede esperarse una reducción de la velocidad de bits por la distorsión de 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.

En otra realización preferida, el proveedor de información de compactación de energía está configurado para proveer una medida de autocorrelación que describe una autocorrelación de una representación de 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 eficiencia de la distorsión de tiempo (en cuanto a reducción de la velocidad de bits) se puede medir (o al menos estimar) sobre la base de una señal de dominio de tiempo (o un remuestreado no uniforme) distorsionada en el tiempo. Se ha hallado que la distorsión de tiempo es eficiente 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 de tiempo no es eficiente.

Este hallazgo se basa en el hecho de que una distorsión de tiempo eficiente transforma una porción de una señal senoide de una frecuencia variable (la cual no comprende una periodicidad) en una porción de una señal senoide de frecuencia aproximadamente constante (la cual comprende un grado alto de periodicidad). A diferencia, si la distorsión de tiempo no es capaz de proveer una señal de dominio de tiempo con un grado alto de periodicidad, puede esperarse que la distorsión de tiempo tampoco provea un ahorro de velocidad de bits significativa, lo cual justificaría su aplicación.

En una realización preferida, 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 de tiempo de la señal de audio, para obtener la información de compactación de energía. Se ha hallado que la determinación del complejo por computación de los picos de autocorrelación no es necesaria para estimar la eficiencia de la distorsión de tiempo. En cambio, se ha hallado 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 confiables. 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 este tipo señal de distorsión de tiempo exhibe picos a una pluralidad de valores de retardo de autocorrelación. De este modo, la formación de la suma es una forma eficiente desde el punto de vista computacional para extraer la información de compactación de energía de la autocorrelación.

En otra realización preferida, el proveedor de la señal de activación de distorsión de tiempo comprende un calculador de valor de referencia configurado para computar el valor de referencia sobre la base de una representación espectral sin distorsión de tiempo de la señal de audio o sobre la base de una representación de dominio de tiempo sin distorsión de 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 además 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 hallado que la relación entre una
 5 información de compactación de energía en el caso sin distorsión de tiempo y la información de compactación de energía en el caso de distorsión de tiempo permite una generación eficiente desde el punto de vista computacional a la vez de lo suficientemente confiable de la señal de activación de distorsión de tiempo.

Otra realización preferida crea un codificador de señal de audio para codificar una señal de audio de entrada, para
 10 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 proveer una representación del espectro transformado por la distorsión de tiempo sobre la base de la señal de audio de entrada. El codificador de señal de audio además comprende un proveedor de la señal de activación de distorsión de tiempo, como se describió anteriormente. El proveedor de la señal de activación de distorsión de tiempo está configurado para recibir la señal de audio de
 15 entrada y proveer la información de compactación de energía de modo que la información de compactación de energía describe una compactación de energía en la representación del espectro transformado por la distorsión de tiempo de la señal de audio de entrada. El codificador de señal de audio además comprende un controlador configurado para proveer en forma selectiva, con dependencia de la señal de activación de distorsión de tiempo, una porción de contorno de distorsión de tiempo no constante hallada (variable) o información de la distorsión de tiempo, o una porción de contorno de distorsión de tiempo estándar constante (no-variable) o información de la distorsión de tiempo a un transformador de distorsión de tiempo. De esta forma, es posible aceptar o rechazar en forma selectiva una porción de contorno de distorsión de tiempo no constante hallada en la derivación de la representación de la
 20 señal de audio codificada de la señal de audio de entrada.

Este concepto se basa en el hallazgo de que no siempre es eficiente 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 hallado que la información de compactación de energía, la cual se computa mediante el proveedor de la señal de activación de distorsión de tiempo, es una medida eficiente desde el punto de vista computacional para decidir respecto si es ventajoso proveer un transformador de distorsión de tiempo con la porción de contorno de distorsión de tiempo hallada variable (no-
 30 constante) o un contorno de distorsión de tiempo estándar (no-variable, constante). Debe observarse que cuando un transformador de distorsión de tiempo comprende una transformada superpuesta, una porción de contorno de distorsión de tiempo hallada puede ser utilizada en el cómputo de dos o más bloques de transformada subsiguientes. En particular, se ha hallado 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 porción de contorno de distorsión de tiempo recientemente hallada variable y la versión de la representación espectral transformada por la distorsión de tiempo de la señal de audio de entrada utilizando una porción de contorno de distorsión de tiempo estándar (no variable) a fin de poder tomar la decisión si la distorsión de tiempo permite un ahorro en la velocidad de bits o no. En cambio, se ha hallado 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 confiable de la decisión. En consecuencia, una velocidad de bits requerida se puede mantener baja.
 40

En una realización preferida adicional, el codificador de señal de audio comprende una interfaz de salida configurada para incluir en 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. De este modo, se puede obtener una elevada eficiencia de la codificación de señal de audio, sin tomar en cuenta si la señal de entrada se ajusta bien para la distorsión de tiempo o no.
 45

Una realización adicional según un ejemplo de comparación crea un método para proveer 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 ser suplementado por cualquiera de los rasgos y funcionalidad que se describen en la presente con respecto al proveedor de la señal de activación de distorsión de tiempo.
 50

Otra realización según un ejemplo de comparación 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 ser suplementado por cualquier de los rasgos y funcionalidades descritos en el presente documento con respecto al codificador de señal de audio.
 55

Otra realización según un ejemplo de comparación crea un programa informático para realizar los métodos mencionados en el presente documento.
 60

Según un primer aspecto, un análisis de la señal de audio, si una señal de audio tiene una característica armónica o una característica de expresión se utiliza en forma ventajosa para controlar el procesamiento del relleno de ruido del

lado del codificador y/o del lado del decodificador. El análisis de la señal de audio se puede obtener fácilmente en un sistema, en el cual 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 expresión por un lado y la música por el otro y/o para distinguir entre la expresión sonora y la expresión sorda. Dado que esta información está disponible en un contexto de este tipo sin costos adicionales, la información disponible se utiliza en forma ventajosa para controlar el rasgo de relleno de ruido de modo que, especialmente para las señales de expresión, un relleno de ruido entre líneas armónicas se reduce, o, para señales de expresión en particular, incluso se elimina. Aún en situaciones, donde un contenido armónico se obtiene, por una expresión no se detecta directamente por parte del detector de expresiones, una reducción del relleno de ruido no obstante dará por resultado una calidad de mayor percepción. Si bien este rasgo es particularmente útil en un sistema, en el cual el análisis armónico/de expresión 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 expresión 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 en forma alternativa, la velocidad de bits es disminuida sin tener una pérdida de calidad, dado que los bits requeridos para codificar el nivel de relleno de ruido son reducidos cuando el nivel de relleno de ruido en sí mismo, el cual puede transmitirse desde un codificador a un decodificador, se reduce.

En un aspecto de la presente invención, el resultado del análisis de señal es decir, si la señal es una señal armónica o una señal de expresión se utiliza para controlar el procesamiento de la función de ventana de un codificador de audio. Se ha hallado que en una situación, en la cual una señal de expresión o una señal armónica se inicia, la posibilidad es alta de modo que un codificador sencillo cambiará de ventanas largas a ventanas cortas. Estas ventanas cortas, sin embargo, tienen una resolución de frecuencia correspondientemente reducida la cual, 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 este tipo de porción de señal. 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 expresión o armónica. En forma alternativa, las ventanas se seleccionan con una longitud aproximadamente similar a las ventanas largas, pero con una superposición más corta a fin de reducir en forma efectiva los pre-ecos. En general, la característica de la señal, si el marco de tiempo de una señal de audio tiene una característica armónica o de expresión se utiliza para seleccionar una función de generación de ventanas para este marco de tiempo.

Según un aspecto adicional, la herramienta de TNS (formato de ruido temporal) es controlada en base a si la señal subyacente se basa en una función de distorsión de tiempo o está en un dominio lineal. Típicamente, una señal que ha sido procesada por una función de distorsión de tiempo tendrá un contenido armónico fuerte. De otro modo, un seguidor de tono asociado con una etapa de distorsión de tiempo no dará salida a un contorno de tono válido y, en ausencia de este contorno de tono válido, una funcionalidad de distorsión de tiempo habría sido desactivada para este marco 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 tiene un contenido armónico o contenido sonoro, la ganancia en calidad/velocidad de bits provista por la herramienta de TNS será reducida. Por lo tanto, sin la modificación de la herramienta de TNS, las porciones 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 formateado de ruido de TNS no obstante provee 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 expresión 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 formato del ruido temporal para este marco dará como resultado una concentración del ruido alrededor del inicio de expresión lo cual reduce en forma efectiva los pre-ecos, que podrían ocurrir antes del inicio de la expresión debido a una cuantificación del marco que se produce en el procesamiento del subsiguiente codificador.

Según un aspecto adicional, un número variable de líneas se procesa mediante un cuantificador/codificador entrópico dentro de un aparato de codificación de audio, a fin de dar cuenta del ancho de banda variable, el cual se introduce de marco a marco debido a la realización de la función de distorsión de tiempo con una característica de distorsión de tiempo variable/contorno de distorsión. Cuando la función de distorsión de tiempo da por resultado la situación que el tiempo del marco (en términos lineales) incluido en un marco 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 sin distorsión de tiempo. Cuando, por otro lado, la función de distorsión de tiempo da por resultado el hecho de que el tiempo real de la señal de audio en el dominio de distorsión de 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 sin distorsión de tiempo a 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 cuales:

- 5 la figura 1 muestra un diagrama esquemático en bloque de un proveedor de la señal de activación de distorsión de tiempo, según una realización según un ejemplo de comparación;
- 10 la figura 2a muestra un diagrama esquemático en bloque de un codificador de señal de audio, según una realización según un ejemplo de comparación;
- la figura 2b muestra otro diagrama esquemático en bloque de un proveedor de la señal de activación de distorsión de tiempo según un ejemplo de comparación;
- 15 la figura 3a muestra una representación gráfica de un espectro de una versión sin distorsión de tiempo de una señal de audio;
- la figura 3b muestra una representación gráfica de un espectro de una versión de distorsión de tiempo de la señal de audio;
- 20 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 porción de frecuencia más alta del espectro;
- 25 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 porción de frecuencia más alta sobre una porción de frecuencia más baja;
- 30 la figura 3f muestra un diagrama esquemático en bloque de un proveedor de información de compactación de energía, según otra realización según un ejemplo de comparación;
- 35 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 (remuestreada no uniforme) de la señal de audio de la figura 3g;
- 40 la figura 3i muestra una representación gráfica de una función de autocorrelación de la señal de audio según la figura 3g;
- la figura 3j muestra una representación gráfica de una función de autocorrelación de la señal de audio según la figura 3h;
- 45 la figura 3k muestra un diagrama esquemático en bloque de un proveedor de información de compactación de energía, según otra realización según un ejemplo de comparación;
- 50 la figura 4a muestra un diagrama de flujo de un método para proveer 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 según un ejemplo de comparación;
- 55 la figura 5a ilustra una realización preferida de un codificador de audio que tiene aspectos de la invención;
- la figura 5b ilustra una realización preferida de un decodificador de audio que tiene aspectos de la invención;
- 60 la figura 6a ilustra una realización preferida del aspecto de relleno de ruido según un ejemplo de comparación;
- la figura 6b ilustra una tabla que define la función 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;

5 la figura 7b ilustra una realización alternativa para influir en la función de generación de ventanas;

la figura 7c ilustra una realización alternativa adicional para ilustrar la función de generación de ventanas en base a la información de distorsión de tiempo;

10 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;

15 la figura 8a ilustra la realización preferida de un control en base a la distorsión de tiempo de la herramienta de TNS (formato de ruido temporal) según un ejemplo de comparación;

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;

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

25 la figura 10a ilustra una realización preferida de un controlador para controlar el número de líneas dentro de un procesador de codificación según un ejemplo de comparación;

la figura 10b ilustra la dependencia entre el número de líneas a ser descartadas/agregadas para una velocidad de muestreo;

30 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

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

40 La figura 1 muestra un diagrama esquemático en bloque 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 proveer, sobre la base del mismo, una 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 proveer una información 122 de compactación de energía, que describe una compactación de energía en una representación del espectro transformado por la distorsión de tiempo de la señal de audio. El proveedor 100 de la señal de activación de distorsión de tiempo además comprende un comparador 130 configurado para comparar la información 122 de compactación de energía con un valor 132 de referencia, y proveer la señal 112 de activación de distorsión de tiempo con dependencia del resultado de la comparación.

50 Como se mencionó anteriormente, se ha hallado que la información de compactación de energía es una información valiosa la cual permite si una distorsión de tiempo otorga un almacenamiento de bits o no. Se ha hallado que la presencia de almacenamiento de bits está íntimamente correlacionada con la cuestión si la distorsión de tiempo da por resultado una compactación de energía o no.

55 La figura 2a muestra un diagrama esquemático en bloque de un codificador 200 de señal de audio, según una realización de la invención. El codificador 200 de señal de audio está configurado para recibir una señal 210 de audio de entrada (también denominada para a(t)) y proveer, 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 dominio de tiempo) y proveer, 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 además comprende un analizador 284 de distorsión de tiempo, el cual está configurado para analizar la señal 210 de audio de entrada y proveer, 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).

60 El codificador 200 de señal de audio además comprende un mecanismo de cambio, por ejemplo en la forma de un

5 cambio 240 controlado, para decidir respecto si la información 286 del contorno de distorsión de tiempo hallada o una información 288 estándar del contorno de distorsión de tiempo se utiliza para otro procesamiento. De este modo, el mecanismo 240 de cambio está configurado para proveer en forma selectiva, con dependencia de una información de activación de distorsión de tiempo, ya sea la información 286 hallada del contorno de distorsión de tiempo o una información 288 estándar del contorno de distorsión de tiempo como información nueva del contorno de distorsión de tiempo 242, para un posterior procesamiento, 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 nueva información 242 del contorno de distorsión de tiempo (por ejemplo una nueva porción de contorno de distorsión de tiempo) y, además, una información de distorsión de tiempo previamente obtenida (por ejemplo una o más porciones de contorno de distorsión de tiempo previamente obtenidas) para la distorsión de tiempo de un marco de audio. El pos-procesamiento del espectro opcional puede comprender por ejemplo un formato de ruido temporal y/o un análisis de relleno de ruido. El codificador 200 de señal de audio además comprende un cuantificador/codificador 260, el cual está configurado para recibir la representación 222 espectral (opcionalmente procesada por el pos procesamiento 250 del espectro) y para cuantificar y codificar la representación espectral transformada 222. A este fin, el cuantificador/codificador 260 puede ser acoplado con un modelo 270 perceptual y recibir una información 272 de relevancia perceptual desde el modelo 270 perceptual, para considerar un enmascaramiento perceptual y ajustar las precisiones de cuantificación en diferentes cajas de frecuencia según la percepción humana. El codificador 200 de señal de audio además comprende una interfaz 280 de salida la cual está configurada para proveer la representación 212 codificada de la señal de audio sobre la base de la representación 262 cuantificada y codificada 20 suministrada por el cuantificador/codificador 260.

25 El codificador 200 de señal de audio además comprende un proveedor de la señal de activación de distorsión de tiempo 230, el cual está configurado para proveer una señal 232 de activación de distorsión de tiempo. La señal 232 de activación de distorsión de tiempo, por ejemplo, se puede utilizar 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 se puede utilizar en un contacto 280 para decidir respecto si 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 la activación de la distorsión de tiempo puede en si misma estar incluida en la representación 35 212 codificada, por ejemplo en la forma de una marca de un bit que indica la activación o desactivación de la distorsión de tiempo.

40 A fin de facilitar el entendimiento, debe observarse que un transformador 220 de distorsión de tiempo típicamente comprende un generador de ventanas (generador de ventana) de análisis 220a, un remuestreador o "distorsionador de tiempo" 220b (*time warper*) y un transformador de dominio espectral (o convertidor de tiempo/frecuencia) 220c. Dependiendo de la implementación, sin embargo, el distorsionador de tiempo 220b puede ser colocado – en una dirección de procesamiento de señal -antes del generador de ventana de análisis 220a. 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.

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

50 El proveedor 230 de la señal de activación de distorsión de tiempo está configurado con preferencia para recibir la representación 210 de la señal de audio del dominio de tiempo (además designado 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 además configurado para obtener, utilizando la señal 210 de audio del dominio de tiempo, 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 proveer la señal 232 de activación de distorsión de tiempo sobre la base de esta información de compactación de energía.

60 La figura 2b muestra un diagrama esquemático en bloque 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 proveer, 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 proveer, 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 además comprende 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, proveer la información 234m y 234n de compactación de energía, respectivamente. El proveedor de la señal de activación de distorsión de tiempo además comprende un comparador 234o, configurado para recibir la información 234m y 234n de compactación de energía, y, sobre la base del mismo proveer la señal 234p de activación de distorsión de tiempo.

A 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 ventanas de análisis, idénticos remuestreadores o distorsionadores 234c y 234i de tiempo, y (opcional) idénticos transformadores 234d y 234j del dominio espectral.

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.

A continuación, el efecto de distorsión de 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 la intensidad de la señal de audio sin distorsión de tiempo como una función de la frecuencia f .

La figura 3b muestra una representación gráfica de un espectro de una versión de distorsión de 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 de la distorsión de tiempo de la señal de audio sobre la frecuencia. Como puede observarse a partir de una comparación de la representación gráfica de las figuras 3a y 3b, la versión sin distorsión de tiempo de la señal de audio comprende un espectro embarrado, en particular en una región de frecuencia más alta. A diferencia, la versión de la distorsión de 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 de la distorsión de tiempo de la señal de audio de entrada.

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 en la figura 3a. Esto se debe al hecho de que un espectro embarrado típicamente comprende un gran número de coeficientes espectrales perceptualmente relevantes (es decir un número relativamente pequeño de coeficientes espectrales cuantificados en cero o cuantificados a valores pequeños), mientras que un espectro "menos plano" como se muestra en la figura 3 típicamente comprende un mayor número de coeficientes espectrales cuantificado en cero o cuantificados a valores pequeños. Los coeficientes espectrales cuantificados en cero o cuantificados a valores pequeños se pueden codificar con menos bits que los coeficientes espectrales cuantificados a valores superiores, de modo que el espectro de la figura 3b puede ser codificado utilizando menos bits que el espectro de la figura 3a.

No obstante, debe observarse además que el uso de una distorsión de tiempo no siempre da por resultado una mejora significativa de la eficiencia de codificación de la señal de distorsión de tiempo. En consecuencia, 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 salvados, 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 proveer 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) se puede omitir (salvo una marca que indica la desactivación de la distorsión de tiempo), manteniendo de este modo la velocidad de bits baja.

A continuación, diferentes conceptos para un cálculo confiable y eficiente a nivel computacional de una señal 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.

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 manchado 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 variante de tono, la mejora en la ganancia de codificación (es decir la cantidad de bits almacenados) puede ser insignificante (por ejemplo si uno tiene un ruido fuerte subyacente a la señal armónica, o si la variante es tan pequeña que el manchado 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 eficiente, señalizando un contorno de distorsión de tiempo estándar (no variable).

El alcance de la presente invención comprende la creación de un método para decidir si una porción de contorno de distorsión de tiempo obtenida provee suficiente ganancia de codificación (por ejemplo suficiente ganancia de codificación para compensar el espacio libre inutilizado requerido para la codificación para el contorno de distorsión de tiempo).

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 además 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.

Las figuras 3a y 3b muestran un ejemplo esquemático con una espectro sin distorsión de un marco con fuertes armonías y variación de tono (figura 3a) y el espectro de la versión de la distorsión de tiempo del mismo marco (figura 3b).

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

La planicidad espectral puede calcularse, por ejemplo, dividiendo el medio geométrico del espectro de potencia por el medio aritmético del espectro de potencia. Por ejemplo, la planicidad espectral (además designada brevemente como "planicidad") se puede computar según la siguiente ecuación:

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

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.

En una realización de la invención, el cálculo mencionado anteriormente de la "planicidad", el cual puede ser útil como una información de compactación de energía, se puede llevar a cabo utilizando las representaciones 234e, 234k del espectro transformado por la distorsión de tiempo, de modo que se puede sostener la siguiente relación:

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

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

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 a continuación disminuye con la frecuencia creciente, conduciendo a una sub-representación de los parciales más elevados en la medida. Esto no se desea en algunas realizaciones, dado que es la intención mejorar la calidad de estos parciales más altos, porque se embarran 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.

En una realización según la invención, se elige un enfoque similar a la denominada medida de "SNR segmental", que conduce a una medida de la planicidad espectral banda por banda. Un cálculo de la medida de planicidad

espectral se realiza (por ejemplo en 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, con preferencia, los anchos de banda pueden seguir una escala perceptual, 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.

5 El concepto antes mencionado se explicará brevemente a continuación, tomando como referencia la figura 3c, la cual muestra una representación gráfica de un cálculo individual de medidas de planicidad espectral para diferentes bandas de frecuencia. Como puede observarse, el espectro puede dividirse en diferentes bandas 311, 312, 313 de frecuencia, las cuales pueden un ancho de banda igual o las cuales pueden tener diferentes anchos de banda. Por ejemplo, una primera medida de planicidad espectral puede computarse para la primera banda 311 de frecuencia, por ejemplo, utilizando la ecuación para la “planicidad” dada anteriormente. En este cálculo, se pueden considerar 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 cuanto a cajas de frecuencia de la primera banda de frecuencia). En consecuencia, se obtiene una medida de planicidad para la primera banda 311 de frecuencia. De manera similar, una medida de planicidad puede computarse para la segunda banda 312 de frecuencia, tomando en cuenta las cajas de frecuencia de las segundas bandas de frecuencia 312 y además el ancho de la segunda banda de frecuencia. Más aún, las medidas de planicidad de las bandas de frecuencia adicionales, del tipo tercera banda 313 de frecuencia, pueden computarse de la misma forma.

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

25 Otro enfoque (para la mejora de la derivación de la señal de activación de distorsión de tiempo) consiste en aplicar la medida de planicidad espectral sólo por encima de una cierta frecuencia. Este enfoque se ilustra en la figura 3b. Como puede observarse, sólo cajas de frecuencia en una porción 316 de frecuencia más alta de los espectros se consideran para un cálculo de la medida de planicidad espectral. Una porción de frecuencia inferior del espectro es insignificante para el cálculo de la medida de planicidad espectral. La porción 316 de frecuencia más alta se puede considerar banda de frecuencia por banda de frecuencia para el cálculo de la medida de planicidad espectral. En forma alternativa, la porción 316 de frecuencia más alta completa se puede considerar en su totalidad para el cálculo de la medida de planicidad espectral.

35 Para resumir lo anterior, se puede establecer que la disminución en la planicidad espectral (originada por la aplicación de la distorsión de tiempo) se puede considerar como una primera medida para la eficiencia de la distorsión de tiempo.

40 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 de tiempo da por resultado una reducción suficiente de la medida de planicidad espectral cuando se compara con un caso sin distorsión de tiempo.

45 Además de los enfoques mencionados anteriormente, la porción de frecuencia superior del espectro se puede enfatizar (por ejemplo por un escalamiento apropiado) sobre la porción 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 la cual una porción de frecuencia más alta es enfatizada sobre una porción de frecuencia más baja. En consecuencia, se compensa una sub-representación de parciales más alta en el espectro. De este modo, la medida de planicidad se puede computar sobre el espectro escalado completo en el cual las cajas de frecuencia más alta son enfatizadas sobre las cajas de frecuencia más baja, como se muestra en la figura 3e.

55 En cuanto a almacenamientos de bits, una medida típica para codificar la eficiencia sería la entropía perceptual, la cual puede definirse de un modo que se correlacione muy bien con el número real de bits necesarios para codificar un cierto espectro como se describe en 3GPP TS 26.403 V7.0.0: 3rd Generation Partnership Project; Technical Specification Group 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 de ello, la reducción de la entropía perceptual es otra medida para determinar cómo sería la eficiencia de la distorsión de tiempo.

60 La figura 3f muestra un proveedor 325 de información de compactación de energía, el cual puede tomar el lugar del proveedor 120, 234f, 234l de información de compactación de energía, y el cual se puede utilizar en los proveedores 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 la forma de una representación 234e, 234k del espectro transformado por la distorsión de tiempo, además designado con $|X|_{tw}$. El proveedor 325 de información de compactación de energía además está configurado para proveer una información 326 de entropía perceptual, la cual puede tomar el lugar de la información 122, 234m, 234n de compactación de energía.

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

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

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, la cual funciona sobre los índices de caja espectral de la banda de factor en 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 .

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:

$$nl = \frac{ffac(n)}{\left(\frac{en(n)}{kOffset(n+1)-kOffset(n)}\right)^{0.25}} \quad (2)$$

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.

Más aún, el calculador de entropía perceptual 332 puede estar configurado para computar la información de entropía perceptual $sfbPe$ 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, se pueden sostener las siguientes relaciones:

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

Una entropía perceptual total pe se puede computar como la suma de las entropías perceptuales de múltiples bandas de frecuencia o bandas de factor de escala.

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

Para otros detalles que se refieren al cómputo de la entropía perceptual, se hace referencia a la sección 5.6.1.1.3 del International Standard “3GPP TS 26.403 V7.0.0(2006-06)”.

5 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. 10 Dado que no es trivial hallar la máxima correspondiente en la autocorrelación para el caso de distorsión de tiempo y sin distorsión de tiempo, la suma de los valores absolutos para la autocorrelación normalizada se puede utilizar como medida para la mejora. Un aumento en esta suma corresponde a un aumento en la compactación de energía.

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

La figura 3g muestra una representación gráfica de una señal sin distorsión de 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 sin distorsión de tiempo. Una curva 352 describe la evolución temporal de la señal de tiempo sin distorsión de tiempo. Se presume que la frecuencia de la señal sin distorsión de tiempo descrita por la curva 352 aumenta en el tiempo, como puede 20 observarse en la figura 3g.

La figura 3h muestra una representación gráfica de una versión de distorsión de tiempo de la señal de tiempo de la figura 3g. Una abscisa 355 describe la distorsión de tiempo (por ejemplo, en forma normalizada) y una ordenada 356 describe el nivel de la versión de la distorsión de tiempo $a_{(tw)}$ de la señal $a(t)$. Como puede observarse in La figura 25 3h, la versión de la distorsión de tiempo $a_{(tw)}$ de la señal de tiempo sin distorsión de 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 30 distorsión de tiempo apropiada, la cual puede comprender un remuestreado de la distorsión de tiempo.

La figura 3i muestra una representación gráfica de una función de autocorrelación de la señal de tiempo sin distorsión de tiempo $a(t)$. Una abscisa 360 describe un retardo en la autocorrelación τ , y una ordenada 361 describe la magnitud de la función de autocorrelación. Las marcas 362 describen una evolución de la autocorrelación $R_{uw}(\tau)$ como una función del retardo de la autocorrelación τ . Como puede observarse a partir de la figura 3i, la función de 35 autocorrelación R_{uw} de la señal de tiempo sin distorsión de 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 $a_{(tw)}$. Como puede observarse a partir de la figura 3j, la función de autocorrelación R_{tw} 40 comprende un pico para $\tau = 0$, y además 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 $a_{(tw)}$. Esta periodicidad se refleja mediante los picos adicionales de la función de autocorrelación $R_{tw}(\tau)$ cuando se compara con la función de autocorrelación 45 $R_{uw}(\tau)$. De este modo, 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 ser utilizada como una indicación de la efectividad (en cuanto a una reducción de la velocidad de bits) de la distorsión de tiempo.

50 La figura 3k muestra un diagrama esquemático en bloque 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 transformada 234d, 234j de dominio espectral y opcionalmente el generador 234b y 234h de ventanas de análisis se omite), y proveer, sobre la base del mismo, una información 374 de compactación de energía, la cual puede tomar la función de la información 372 de 55 compactación de energía. El proveedor 370 de información de compactación de energía de la figura 3k comprende un calculador 371 de autocorrelación configurado para computar la función de autocorrelación $R_{tw}(\tau)$ de la señal de distorsión de tiempo $a_{(tw)}$ sobre un rango predeterminado de valores discretos de τ . El proveedor 370 de información de compactación de energía además comprende un sumador 372 de autocorrelación configurado para sumar una pluralidad de valores de la función de autocorrelación $R_{tw}(\tau)$ (por ejemplo, sobre un rango predeterminado de valores 60 discretos de τ) y proveer la suma obtenida como la información 122, 234m, 234n de compactación de energía.

De este modo, el proveedor 370 de información de compactación de energía permite la provisión de información confiable que indica la eficiencia 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 halla, sobre la base de la información 122 de compactación de energía, 234m, 234n provista por el proveedor 370 de información de compactación de energía, que la distorsión de tiempo realmente otorga una eficiencia 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 concernientes a la escasez de ganancia de codificación o del espectro se pueden tomar 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 perceptual.

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 perceptual, 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 se pueden utilizar. Con preferencia, 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 brinda un beneficio en la codificación o no.

Todas estas medidas se pueden aplicar a un marco completo, donde sólo la tercera porción del contorno de tono es nuevo (donde, por ejemplo, tres porciones del contorno de tono están asociadas con el marco completo), o con preferencia sólo para la porción de la señal, para lo cual se obtuvo esta nueva porción, por ejemplo, utilizando una transformada con una ventana superpuesta baja centrada sobre la porción de señal (respectiva).

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

La figura 4a muestra un diagrama de flujo de un método para proveer 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 suministrar 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 además comprende un paso 420 para comparar la información de compactación de energía con un valor de referencia. El método 400 además comprende un paso 430 para suministrar la señal de activación de distorsión de tiempo con dependencia del resultado de la comparación.

El método 400 puede ser suplementado por cualquiera de los rasgos y funcionalidades que se describen en la presente 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 opcionalmente comprende un paso 460 para suministrar una representación espectral transformada por la distorsión de tiempo sobre la base de la señal de audio de entrada. El método 450 además comprende un paso 470 para suministrar una señal de activación de distorsión de tiempo. El paso 470, por ejemplo, puede comprender la funcionalidad del método 400. De este modo, la información de compactación de energía puede ser provista de modo que la información de compactación de energía describa una compactación de energía en la representación del espectro transformado por la distorsión de tiempo de la señal de audio de entrada. El método 450 además comprende un paso 480 para suministrar en 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 ser suplementado 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 cual varios aspectos de la presente invención son implementados. Se provee una señal de audio en una entrada del codificador 500. Esta señal de audio será típicamente una señal de audio discreta la cual ha sido derivada de una señal de audio análoga utilizando una velocidad de muestreo la cual además se denomina la velocidad de muestreo normal. Esta velocidad de muestreo normal es diferente de una velocidad de muestreo local generada en una

operación de la función 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 porción de tiempo constante. La señal se coloca en un generador 502 de ventanas de análisis, el cual está, en esta realización, conectado un controlador de función de generación de ventanas 504. El generador 502 de ventanas de análisis está conectado a un distorsionador 506 de tiempo (time warper). 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 ventanas de análisis. Se prefiere esta implementación, cuando una característica de distorsión de tiempo se requiere para la generación de ventanas de análisis en bloque 502, y cuando la función 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 ventanas de análisis se puede establecer según lo requerido. Además, un convertidor 508 de tiempo/frecuencia se provee 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 510 de TNS (formato de ruido temporal), la cual provee, como una salida 510a, información de TNS y, como una salida 510b, valores residuales espectrales. La salida 510b está acoplada a un cuantificador y codificador 512 en bloque el cual puede ser controlado por un modelo 514 perceptual para cuantificar una señal de modo que el ruido de cuantificación se esconda por debajo del umbral de enmascaramiento perceptual de la señal de audio.

Adicionalmente, el codificador ilustrado en la figura 5a comprende un analizador 516 de distorsión de tiempo, el cual puede ser implementado como un seguidor de tono, que provee 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 expresión sonora y la expresión 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 518 del analizador de distorsión de tiempo está conectada a al menos una y con preferencia más de una funcionalidad en el grupo de funcionalidad que comprende el controlador 504 de la función de generación de ventanas, el distorsionador 506 de tiempo, la etapa 510 de TNS, el cuantificador 512 y codificador y una interfaz 522 de salida.

En 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 de generación de ventanas, la etapa 510 de TNS, un analizador 524 de relleno de ruido o la interfaz 522 de salida. Además, una salida 518 del analizador de distorsión de tiempo 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 500 del generador de ventanas de análisis es ingresada en un analizador 516 de distorsión de tiempo y el clasificador 520 de señal, las señales de entrada para estas funcionalidades además se pueden tomar de la salida del generador 502 de ventanas 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 508 de tiempo/frecuencia o la salida de la etapa 510 de TNS.

Además de la salida de una señal por el cuantificador codificador 512 que se indica en 526, la interfaz 522 de salida recibe la información 510a del lado del TNS, una información 528 del lado del modelo perceptual, el cual puede incluir factores de escala en forma codificada, datos de indicación de distorsión de tiempo para más información del 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 de salida 532 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, el cual tiene una funcionalidad reducida, o siempre que la información ya esté disponible en el decodificador debido a una transmisión mediante un canal de transmisión diferente.

El codificador ilustrado en la figura 5a puede ser implementado como se define en detalle en estándar MPEG-4 además de las funcionalidades adicionales que se ilustran en el codificador de la invención en la figura 5a representada por el controlador 504 de la función de generación de ventanas, el analizador 524 de relleno de ruido, el cuantificador codificador 512 y la etapa 510 de TNS, los cuales tienen, en comparación con el estándar MPEG-4, una funcionalidad avanzada. Otra descripción se encuentra en el estándar AAC (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 542 de distorsión de tiempo, datos 543 de relleno de ruido, factores 544 de escala, datos 545 de TNS e información 546 espectral codificada. La información espectral codificada es ingresada en un decodificador 547 entrópico, el cual 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 la forma del correspondiente codificador tal como un codificador Huffman o un codificador aritmético. La información espectral codificada es ingresada al re-cuantificador 550, el cual está conectado a un relleno 552 de ruido. La salida del relleno 552 de ruido es ingresada a una etapa 554 de TNS inversa, la cual además recibe los datos de TNS en la línea 545. Dependiendo de la implementación, el relleno 552 de ruido y la etapa 554 de TNS se pueden aplicar 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 556 de frecuencia/tiempo, el cual alimenta un regenerador 558 de tiempo (*time dewarper*). A la salida de la cadena de procesamiento de señal, un generador de ventana de síntesis que realiza con preferencia un procesamiento de superposición/agregado se aplica como se indica en 560. El orden del regenerador 558 de tiempo y la etapa 560 de síntesis 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 en forma efectiva.

Adicionalmente, se proporciona un analizador de relleno 562 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 541 de señales e información sobre el espectro recuantificado, según sea el caso.

Preferiblemente, todas las funcionalidades descritas de aquí en adelante se aplican juntas en un esquema de codificador/decodificador de audio mejorado. No obstante, las funcionalidades descritas de aquí en adelante 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 del lado del codificador y/o implementada por el analizador de relleno 562 de ruido y el relleno 552 de ruido del lado del decodificador.

Varias herramientas del codificador dentro del marco 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 expresión, 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 manchado. Además, la cantidad de reducción del nivel de ruido además puede mejorarse utilizando la información del clasificador de señales, de modo que p. ejemplo para las señales de expresión 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 perceptual según lo determinado por el modelo 514 perceptual. No obstante, se ha descubierto que estos "agujeros 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 del lado del codificador. Estas líneas espectrales

pueden tener una amplitud o fase aleatoria, y estas líneas espectrales sintetizadas del lado del decodificador se escalan utilizando una medición de relleno de ruido determinada del lado del codificador según lo ilustrado en la figura 5a o dependiendo de una medición determinada del 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 medición de relleno de ruido de una energía de valores de audio cuantificados en cero para un marco temporal de la señal de audio.

En una realización de la presente invención, el codificador de audio para codificar una señal de audio en 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 de dominio de la frecuencia, el cuantificador también puede utilizarse para cuantificar valores de 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 medición de relleno de ruido de una energía de valores de audio cuantificados en cero para un marco de tiempo de la señal de audio por parte del cuantificador 512. Adicionalmente, el codificador de audio comprende un analizador de señales de audio 600 ilustrado en la figura 6a, que se configura para analizar, si el marco temporal de la señal de audio tiene una característica armónica o una característica de expresión. El analizador 600 de señales puede, por ejemplo, comprender el bloque 516 de la figura 5a o el bloque 520 de la figura 5a o puede comprender cualquiera de los dispositivos para analizar, si una señal es una señal armónica o una señal de expresión. Debido a que el analizador 516 de distorsión temporal 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 temporal de un analizador de distorsión temporal.

El codificador de audio adicionalmente comprende un manipulador del nivel de relleno 602 de ruido ilustrado en la figura 6a, que produce una medición/nivel de relleno de ruido a ser producido en la interfaz 522 de salida indicada en 530 en la figura 5a. El manipulador de medición de relleno 602 de ruido está configurado para manipular la medición de relleno de ruido dependiendo de la característica armónica o de expresión 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, la señal codificada que comprende la medición de relleno de ruido manipulada producida por el bloque 602 en línea 530. Este valor corresponde a los valores producidos por el bloque 562 en la implementación del lado del decodificador ilustrada en la figura 5b.

Según lo indicado en la figura 5a y 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 del lado del decodificador, el decodificador para decodificar una señal codificada de audio comprende la interfaz 539 de entrada para procesar la señal codificada en línea 540 para obtener una medición de relleno de ruido, es decir, datos de relleno de ruido en línea 543, datos de audio codificados en línea 546. El decodificador adicionalmente comprende un decodificador 547 y re-cuantificador 550 para generar datos recuantificados.

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

Adicionalmente, se proporciona el relleno 552 de ruido para generar datos de audio de relleno de ruido, en donde el relleno 552 de ruido está configurado para generar los datos de relleno de ruido en respuesta a la medición 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 expresión de los datos de audio según lo definido por el analizador de señales 516 y/o 550 del lado del codificador o según lo definido por el ítem 562 del lado del decodificador a través del procesamiento e interpretación de la información sobre la distorsión temporal 542 que indica, si un cierto marco temporal ha sido sometido a un procesamiento de distorsión temporal 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 expresión.

5 La manipulación inventiva de relleno de ruido puede, por lo tanto, implementarse del lado del codificador solamente calculando la medición sencilla de ruido y manipulando esta medición de ruido sobre la base de la información armónica/de expresión y transmitiendo la medición de relleno de ruido manipulada ya correcta que entonces puede ser aplicada por un decodificador de manera sencilla. En forma alternativa, la medición de relleno de ruido no manipulada puede transmitirse desde un codificador a un decodificador, y el decodificador entonces analizará, si el marco temporal real de una señal de audio ha sufrido una distorsión temporal, es decir, tiene una característica armónica o de expresión de manera que la manipulación real de la medición de relleno de ruido se produce del lado del decodificador.

Posteriormente, la figura 6b se debate con el fin de explicar las 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 expresión. Este es el caso, cuando no se aplica ninguna distorsión temporal. Cuando, adicionalmente, se proporciona un clasificador de señales, después el clasificador de señales que distingue entre expresión y no expresión indicaría no expresión para la situación, donde la distorsión temporal no estaba activa, es decir, donde no se encontró ningún contorno de tonos.

20 Cuando, sin embargo, la distorsión temporal estaba activa, es decir, cuando se encontró un contorno de tonos, que indica un contenido armónico, entonces el nivel de 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 expresión, y cuando simultáneamente la información sobre la distorsión temporal indica un contorno de tonos, entonces se señala un nivel de relleno de ruido más bajo o aún igual a cero. De este modo, el manipulador del nivel de relleno 602 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/mudo según lo indicado a la izquierda de la figura 6b. En el caso de expresión sonora, se señala/aplica un nivel de relleno de ruido muy bajo o igual a cero. Sin embargo, en el caso de expresión sorda, donde el indicio de distorsión temporal no indica un procesamiento de distorsión temporal debido al hecho de que no se encontró ningún tono, pero donde el clasificador de señales señala contenido de expresión, la medición de relleno de ruido no se manipula, pero se aplica un nivel de relleno de ruido normal.

30 Preferiblemente, el analizador de señales de audio comprende un rastreador de tonos para generar un indicio de tono tal como un contorno de tonos o un tono absoluto de un marco temporal de la señal de audio. Entonces, el manipulador está configurado para reducir la medición de relleno de ruido cuando se encuentra un tono, y no reducir la medición 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 del lado del decodificador, no está realizando un análisis de señales real como un rastreador de tonos o un detector sonoro/mudo, sino que el analizador de señales analiza sintácticamente la señal codificada de audio con el fin de extraer una información sobre la distorsión temporal o una información sobre la 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.

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

45 Para inicios de expresión donde comienza una parte de expresión sonora después de una porción de señales silenciosas relativas, el algoritmo de cambio de bloques podría clasificar la misma como un accidente y podría elegir bloques cortos para este marco particular, con una pérdida de ganancia de codificación en el segmento de señales 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 accidente del transitorio alrededor del inicio encontrado. Esta característica también puede asociarse con el clasificador de señales para evitar el cambio de bloques en las señales de expresión y permitir las mismas 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 accidentes, sino mediante el uso de un umbral variable para la detección de accidentes basado en inicio sonoro e información sobre la clasificación de señales. Además, la información puede utilizarse para detectar accidentes como los inicios sonoros mencionados más arriba pero en vez de cambiar a bloques cortos, utiliza ventanas largas con superposiciones cortas, que siguen siendo la resolución espectral preferible pero reduce la región temporal donde pueden surgir ecos previos o posteriores. La Figura 7d muestra el comportamiento típico sin la adaptación, la figura 7e muestra dos diferentes posibilidades de adaptación (ventanas de baja superposición y prevención).

Un codificador de audio conforme a 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 temporal o un clasificador 520 de señales de la figura 5a. Generalmente, el analizador de señales de audio analiza si un marco temporal de la señal de audio tiene una característica armónica o de expresión. Para este fin, el clasificador 520 de señales de la figura 5a puede incluir un detector 520a sonoro/mudo o un detector 520b de expresión /no expresión. Aunque no se muestra en la figura 7a, también puede proporcionarse un analizador de distorsión temporal tal como el analizador 516 de distorsión temporal 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 funciones de ventanas para seleccionar una función de ventana dependiendo de una característica armónica o de expresión de la señal de audio según lo determinado por el analizador de señales de audio. El generador 502 de ventanas entonces genera ventanas de la señal de audio o, dependiendo de cierta implementación, la señal de audio distorsionada en el tiempo utilizando la función de ventana seleccionada para obtener un marco de ventana. Este marco de ventana, entonces además es procesado por un procesador para obtener una señal codificada de audio. 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 temporales que comprenden un filtro LPC tal como codificadores de expresión y, específicamente, codificadores de expresión implementados conforme la norma AMR-WB+.

En una realización preferente, el controlador 504 de funciones de ventanas comprende un detector 700 de transitorios para detectar un transitorio en la señal de audio, en donde el controlador de funciones de ventanas está configurado para cambiar de una función de ventana para un bloque largo a una función de ventana para un bloque corto, cuando se detecta un transitorio y el analizador de señales de audio no encuentra una característica armónica o de expresión. Cuando, sin embargo, se detecta un transitorio y el analizador de señales de audio encuentra una característica armónica o de expresión, entonces el controlador 504 de funciones de ventanas no cambia a la función de ventana para bloque corto. Las salidas de la función de 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 lo ejecutado por el bien conocido codificador AAC 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 un marco al siguiente marco 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 porción 714a de superposición, una porción 714b de no aliasing (generación de señal ajena), una segunda porción 714c de superposición más corta y una porción 716 cero que se extiende entre el punto y el punto en el eje temporal indicado por las muestras 2048. 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 porción 718a larga de superposición que se superpone con la siguiente ventana larga no ilustrada en la figura 7d. Además, esta ventana tiene una porción 718b de no aliasing, una porción 718c de superposición corta y una porción cero que se extiende entre el punto 720 en el eje temporal hasta el punto 2048. Esta porción es una porción cero.

Normalmente, el cambio a ventanas cortas es útil para evitar pre-ecos que aparezcan dentro de un marco antes del evento del transitorio que es la posición de inicio sonoro o, generalmente, el comienzo de la expresión 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. En forma alternativa, existen otras mediciones de armonicidad tales como una medición de tonalidad por encima de un cierto nivel mínimo junto con la característica de que los picos prominentes tienen una relación armónica uno con otro. 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 la expresión y, específicamente, porciones de expresión sonora o porciones que tienen una fuerte contenido armónico, se desea una buena resolución de frecuencia. Por lo tanto, el analizador de señales de audio ilustrado en 516, 520 ó 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 expresión sonora o un segmento de señales que tiene una fuerte característica armónica. Esto asegura que, para codificar dichas porciones de señales, se mantiene una alta resolución de frecuencia. Este es un equilibrio entre los pre-ecos por un lado y la codificación de alta calidad y alta resolución del tono para la señal de expresión o el tono para una señal de no expresión 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 pre-eco que pudiera aparecer. Con el fin de disminuir además los pre-ecos, se favorece un procesamiento TNS para dicha situación, que se debatirá 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/mudo y/o de expresión /no expresión. Sin embargo, el detector 700 de transitorios incluido en el controlador de funciones de ventanas 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, en donde se detecta un transitorio cuando la característica cuantitativa tiene una relación predeterminada con el 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 umbral de energía. Cuando el incremento de energía de un bloque al siguiente es mayor que el incremento umbral de energía, 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 invertido de energía. En la realización de la figura 7b, el umbral controlable se controla de manera que se reduce la posibilidad de un cambio a una función de ventana para un bloque corto, cuando el analizador de señales de audio ha encontrado una característica armónica o de expresión. En la realización del 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.

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

En una implementación de MDCT preferente según lo implementado por el codificador AAC, mantener una cierta superposición proporciona la ventaja adicional de que, del lado del decodificador, se puede realizar un procesamiento de superposición/agregado lo que significa que se realiza una clase de fundido cruzado entre bloques. Esto efectivamente evita "artefactos" (en este contexto se refiere a errores en la conversión) de bloqueo. Adicionalmente, esta característica de superposición/agregado proporciona la característica de fundido cruzado sin incrementar la velocidad de bits, es decir, se obtiene un fundido cruzado críticamente mostrado. En ventanas cortas o ventanas largas regulares, la porción de superposición es una superposición del 50% según lo indicado por la porción 714 de superposición. En la realización donde la función de ventana es larga como las muestras 2048, la porción de superposición es del 50%, es decir, muestras 1024. La función de ventana que tiene una superposición más corta que se utiliza para efectivamente generar ventanas de un inicio de expresión o un inicio de una señal armónica es preferiblemente menos del 50% y es, en la realización de la figura 7e, solamente las muestras 128, que es un 1/16 del total de la longitud de ventana. Preferiblemente, se utilizan las porciones de superposición entre 1/4 y 1/32 de la longitud total de la función de ventana.

La figura 7c ilustra esta realización, en la que un detector 520a sonoro/mudo ejemplar controla un selector de formas de ventanas incluido en el controlador 504 de funciones de ventanas 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 750. Se implementa la selección de una de ambas formas, cuando el detector 500a sonoro/mudo 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, el selector 504 de formas de ventanas en la figura 7c que está incluido en el controlador 504 de funciones de ventanas en la figura 5a solamente utiliza la señal 751, cuando un detector de transitorios incluido en el controlador de funciones de ventanas detecta un transitorio y ordena un cambio de una función de ventana larga a una función de ventana corta según lo debatido con relación a la figura 7a.

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

La propiedad de compactación de energía espectral de la MDCT distorsionada en el tiempo también influye en la herramienta de formato de ruido temporal (TNS), debido a que la ganancia de TNS tiende a disminuir para los marcos distorsionados en el tiempo especialmente para algunas señales de expresión. No obstante es deseable

- activar TNS, p. ejemplo para reducir los pre-ecos en los inicios sonoros o desviaciones sonoras (consultar la adaptación del cambio de bloques), donde no se desea ningún cambio de bloques pero aún el envolvente temporal de la señal de expresión exhibe cambios rápidos. Típicamente, un codificador utiliza alguna medición para ver si la aplicación de la TNS es provechosa para un cierto marco, 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, de manera que se asegure que la TNS esté activa más a menudo para dichas porciones de señales críticas como inicios sonoros. Como para las otras herramientas, esto también puede complementarse teniendo en cuenta la clasificación de señales.
- 10 El codificador de audio conforme a esta realización para generar una señal de audio comprende un distorsionador temporal controlable tal como el distorsionador 506 temporal para la distorsión temporal de la señal de audio para obtener una señal de audio distorsionada en el tiempo. Adicionalmente, se proporciona un convertidor 508 de frecuencia/tiempo para convertir al menos una porción 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.
- 15 Adicionalmente, el codificador comprende una etapa 510 de formato de ruido temporal para realizar un filtrado de predicción sobre la frecuencia de la representación espectral conforme a la orden de control del formato de ruido temporal, en donde no se realiza el filtrado de predicción, cuando no existe la orden de control del formato de ruido temporal.
- 20 Adicionalmente, el codificador comprende un controlador de formato de ruido temporal para generar la orden de control del formato de ruido temporal sobre la base de la representación espectral.
- 25 Específicamente, el controlador de formato de ruido temporal está configurado para incrementar la posibilidad de realizar el filtrado de predicción sobre la frecuencia, cuando la representación espectral se basa en una señal temporal distorsionada en el tiempo o para reducir la posibilidad de realizar el filtrado de predicción sobre la frecuencia, cuando la representación espectral no está basada en una señal temporal distorsionada en el tiempo. Los detalles del controlador de formato de ruido temporal se debaten con relación a la figura 8.
- 30 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 del etapa 512 del codificador del cuantificador ilustrada en la figura 5a.
- 35 Una etapa 510 de TNS ilustrada en la figura 5a se ilustra en detalle en la figura 8. Preferiblemente, el controlador de formato 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 de señal 804 de control de umbral. Dependiendo de la señal del analizador 516 de distorsión temporal 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, 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 se va a realizar en este marco temporal específico.
- 40 El calculador 800 de ganancia de TNS recibe, como una entrada, la representación espectral obtenida de la señal distorsionada en el tiempo. Típicamente, una señal distorsionada en el tiempo tendrá una ganancia de TNS menor, pero por otro lado, u procesamiento de TNS debido a la característica de formato de ruido temporal en el dominio del tiempo es beneficiario en la situación específica, donde existe una señal sonora/armónica que haya sido sometida a una operación de distorsión temporal. Por otro lado, el procesamiento de TNS no es útil en situaciones, donde la ganancia de TNS es baja, 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, donde la energía de la señal residual de TNS en línea 510d es levemente menor que la energía antes de la etapa 510 de TNS, el procesamiento de TNS podría no ser ventajoso, debido a la reducción de bits debido a que la energía levemente menor en la señal que efectivamente es utilizada 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 510a secundaria de TNS indicada en la figura 5a. Aunque una realización cambia automáticamente en el procesamiento de TNS para todos los marcos, en el que se ingresa una señal distorsionada en el tiempo indicada por la información sobre tonos del bloque 516 o la información 520 del clasificador de señales del bloque, una realización preferente también mantienen la posibilidad del procesamiento de TNS desactivado, pero solo cuando la ganancia es realmente baja o al menos menor que en
- 50
55
60

el caso normal, cuando no se procesa ninguna señal armónica/de expresión.

La figura 8b ilustra una implementación donde tres ajustes de umbral diferentes son implementados 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 el clasificador de señales indica una expresión sorda o ninguna expresión, entonces el umbral de decisión de TNS se ajusta para que esté en un estado normal requiriendo 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 expresión o el detector sonoro/mudo detecta una expresión sorda, entonces el umbral de decisión de TNS se fija en un nivel menor, lo que significa que aún cuando las ganancias de TNS comparativamente bajas son calculadas por el bloque 800 en la figura 8a, sin embargo se activa el procesamiento de TNS.

En una situación, en la que se detecta un contorno de tonos activo y en la que se encuentra una expresión sonora, entonces, el umbral de decisión de TNS se fija en el mismo nivel más bajo o se fija en un estado aún menor de manera que aún las pequeñas ganancias de TNS sean 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 debatido, el controlador de formato de ruido temporal además está configurado 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 expresión sonora exhibirá un contorno de tonos, y la expresión sorda tal como sonidos fricativos o sonidos sibilantes no exhibirá un contorno de tonos. Sin embargo, existen señales de no expresión, con fuerte contenido armónico y, por lo tanto, tienen un contorno de tonos, aunque el detector de expresión no detecta la expresión. Adicionalmente, existe cierta expresión sobre la música o música sobre señales de expresión, que el analizador de señales de audio (516 de la figura 5a por ejemplo) determina que tienen un contenido armónico, pero que no son detectadas por el clasificador 520 de señales como señal de expresión. En dicha situación, también pueden aplicarse todas las operaciones de procesamiento para señales de expresión sonoras y también darán como resultado una ventaja.

Posteriormente, se describe otra realización preferente 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 las aplicaciones del codificador por sí solas, donde 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 temporal tal como por 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 capacitados sino también perceptibles para oyentes no capacitados.

El codificador central AAC normalmente codifica un número fijo de líneas, ajustando en cero todos los otros por encima de la línea máxima. En el caso no distorsionado esto lleva a un efecto de paso bajo con una frecuencia de corte constante y por lo tanto un ancho de banda constante de la señal AAC decodificada. En el caso del 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 temporal local, que lleva a artefactos audibles. Estos artefactos pueden reducirse eligiendo en forma adaptable el número de líneas, como una función del contorno de distorsión temporal local y su velocidad de muestreo promedio obtenida, a ser codificadas 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 todos los marcos. Un beneficio adicional es el ahorro de bits en el codificador.

El codificador de audio conforme a esta realización comprende el distorsionador 506 temporal para distorsionar en el tiempo una señal de audio que utiliza una característica de distorsión temporal variable. Adicionalmente, se proporciona un convertidor 508 de frecuencia/tiempo 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 del cuantificador/codificador de la figura 5a está configurado para fijar un número de coeficientes espectrales para un marco de la señal de audio basado en la característica de la distorsión temporal para el marco de manera que se reduzca o se elimine la variación del ancho

de banda representada por el número procesado de coeficientes de frecuencia de marco a marco.

El procesador implementado por el bloque 512 puede comprender un controlador 1000 para controlar el número de líneas, donde el resultado del controlador 1000 es aquel, con respecto a un número de líneas fiado para el caso de un marco temporal que está siendo codificado sin ninguna distorsión temporal, se agrega o se descarta un cierto número variable de líneas en el extremo superior del espectro. Dependiendo de la implementación, el controlador 1000 puede recibir una información sobre el contorno de tonos en un cierto marco 1001 y/o una frecuencia de muestreo promedio local en el marco indicado 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 un marco, donde los contornos de tonos sobre el marco se ilustran en los respectivos dibujos de la izquierda para la distorsión temporal y se ilustran en los dibujos del medio después de la distorsión temporal, donde se obtiene una característica de tono sustancialmente constante. Este es el objetivo de la funcionalidad de la distorsión temporal que, después de la distorsión temporal, la característica del tono es 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 frecuencia/tiempo o producido por una etapa 510 de TNS de la figura 5a, y cuando no se realiza una operación de distorsión temporal, es decir, cuando el distorsionador 506 temporal se desactivó, según lo indicado por la línea 507 sombreada. Cuando, sin embargo, se obtiene un contorno de distorsión temporal no constante, y cuando este contorno de distorsión temporal 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 a ser transmitidas para este marco debe incrementarse con el fin de equilibrar esta pérdida de ancho de banda.

En 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 tasa de muestreo. La reducción de la tasa de muestreo da como resultado un incremento del ancho de banda del espectro de este marco con respecto a la escala lineal, y este incremento de ancho de banda debe equilibrarse utilizando una supresión o desecho 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 un marco es, en vez de realizar la operación de distorsión temporal, igual que la frecuencia de muestreo sin ninguna distorsión temporal. De este modo, el ancho de banda de la señal no es afectada, y puede procesarse el sencillo número de líneas a ser utilizado para el caso normal sin distorsión temporal, aunque se tenga que realizar la operación de distorsión temporal. De la figura 9, se torna claro que realizar una operación de distorsión temporal no necesariamente influye en el ancho de banda, sino que la influencia en el ancho de banda depende del contorno de tonos y la forma, como se lleva a cabo la distorsión temporal en un marco. Por lo tanto, es preferible utilizar, como valor de control, una tasa de muestreo local o promedio. La determinación de esta tasa de muestreo local se ilustra en la figura 11. La porción superior en la figura 11 ilustra una porción temporal con valores de muestreo equidistantes. Un marco incluye, por ejemplo, siete valores de muestreo indicados por T_n en la línea superior. La línea inferior muestra el resultado de una operación de distorsión temporal, en la que, totalmente, se ha producido un incremento de la tasa de muestreo. Esto significa que la longitud del marco distorsionado en el tiempo es más pequeña que la longitud temporal del marco no distorsionado en el tiempo. Sin embargo, debido a que la longitud temporal del marco distorsionado en el tiempo a ser introducido en el convertidor de frecuencia/tiempo es fija, el caso de un incremento de la tasa de muestreo hace que una porción adicional de la señal temporal que no pertenece al marco indicado por T_n sea introducido en el marco distorsionado en el tiempo según lo indicado por las líneas 1100. De este modo, un marco distorsionado en el tiempo cubre una porción temporal de la señal de audio indicada por T_{in} el cual es mayor que el tiempo T_n . En vista de ello, la distancia efectiva entre dos líneas de frecuencia o el ancho de banda de la frecuencia de una única línea en el dominio lineal (el cual es el valor inverso para la resolución) ha disminuido, y el número de líneas N_n fijadas 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.

El otro caso, no ilustrado en la figura 11, donde el distorsionador temporal lleva a cabo una reducción de la tasa de muestreo, la longitud de tiempo efectivo de un marco en el dominio distorsionado en el tiempo es más pequeña que la longitud temporal 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 se ha incrementado. Ahora, multiplicar 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, como se calcula una tasa de muestreo promedio f_{SR} . Para este fin, se determina la distancia temporal entre dos muestras distorsionadas en el tiempo y se toma el valor inverso, el que se define que

es la tasa 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 promedio aritmético y este valor finalmente da como resultado la tasa de muestreo promedio local, que preferiblemente se utiliza para ser ingresada en el controlador 1000 de la figura 10a.

5 La figura 10b ilustra una línea que indica cuantas líneas deben agregarse 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, el que debe mantenerse constante tanto como sea posible para una secuencia de marcos distorsionados en el tiempo o para una secuencia de marcos distorsionados en el tiempo y no distorsionados en el tiempo.

15 La figura 12b ilustra la dependencia entre los diferentes parámetros debatidos con relación a la figura 9, Figura 10b y figura 11. Básicamente, cuando la tasa de muestreo, es decir, la tasa de muestreo promedio f_{SR} disminuye con respecto a el caso no distorsionado en el tiempo, las líneas deben suprimirse, mientras que las líneas deben agregarse, cuando la tasa de muestreo aumenta con respecto a la tasa de muestreo normal f_N para el caso no distorsionado en el tiempo de manera que las variaciones del ancho de banda de marco a marco se reducen o preferiblemente aún se eliminan tanto como sea posible.

20 El ancho de banda que resulta en el número de líneas N_N y la tasa de muestreo f_N preferiblemente define la frecuencia de cruce 1200 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 el arte, 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 de cruce 1200 y la frecuencia f_{MAX} con una velocidad de bits baja, donde esta velocidad de bits baja típicamente es aún menor que 1/10 o menos de la velocidad de bits requerida para la banda baja entre una frecuencia de 0 y la frecuencia de cruce 1200. La figura 12a además ilustra 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 agregarse. Además, también se ilustra la variación del ancho de banda para un número de líneas constante dependiendo de la tasa de muestreo local f_{SR} . Preferiblemente, el número de líneas a ser agregadas o a ser suprimidas con respecto al número de líneas para el caso normal se fija de manera que cada marco de los datos codificados de AAC tiene una frecuencia máxima tan cercana como sea posible a la frecuencia de cruce 1200. De este modo, se evita cualquier agujero espectral debido a la reducción del ancho de banda por un lado o un espacio libre inutilizado transmitiendo información en una frecuencia por encima de la frecuencia de cruce en el marco codificado 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.

35 El agregado 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.

40 Además, es preferible llevar las variaciones de ancho de banda a un nivel mínimo y aún eliminar las variaciones de ancho de banda, pero, en otras implementaciones, aún una reducción de las variaciones del ancho de banda determinando el número de líneas que dependen de la característica de distorsión temporal aún aumenta la calidad de audio y reduce la velocidad de bits requerida en comparación con una situación, donde se aplica un número de líneas constante independientemente de una cierta característica de distorsión temporal.

50 Aunque se han descrito algunos aspectos en el contexto de un equipo, está claro que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a un paso del método o una característica del paso del método. En forma análoga, los aspectos descritos en el contexto de un paso del método representan una descripción de un bloque correspondiente o ítem o característica del equipo correspondiente.

60 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 disquete, un DVD, un CD, un ROM, un PROM, un EPROM, un EEPROM o una memoria rápida, que posea señales de control electrónicamente legibles almacenadas en los mismos, que cooperen (o sean capaces de cooperar) con un sistema informático programable de manera tal que se lleve a cabo el método respectivo. Algunas realizaciones conforme a la invención comprenden un soporte de datos que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema informático programable, de manera tal que se lleve a cabo uno de los métodos descritos en la presente. Generalmente, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, el código de programa que es operativo para realizar uno de los métodos cuando el producto de programa informático funciona en la computadora. El código de programa puede por ejemplo almacenarse en un soporte legible en la máquina. Otras realizaciones comprenden el programa informático para llevar a cabo uno de los métodos descritos

en la presente, almacenado en el soporte legible en la máquina. En otras palabras, una realización del método inventivo es, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo uno de los métodos descritos en la presente, cuando el programa informático funciona en una computadora. Una realización adicional de los métodos inventivos es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible en computadora) que comprende, grabado en los mismos, el programa informático que lleva a cabo uno de los métodos descritos en la presente. Una realización adicional del método inventivo es, por lo tanto, una corriente de datos o una secuencia de señales que representan el programa informático para llevar a cabo uno de los métodos descritos en la presente. La corriente de datos o la secuencia de señales puede por ejemplo configurarse para ser transferida a través de una conexión de comunicación de datos, por ejemplo a través de Internet. Una realización adicional comprende un medio de procesamiento, por ejemplo una computadora, o un dispositivo lógico programable, configurado para o adaptado para llevar a cabo uno de los métodos descritos en la presente. Una realización adicional comprende una computadora que tiene el programa informático instalado en la misma para llevar a cabo uno de los métodos descritos en la presente. En algunas realizaciones, puede utilizarse un dispositivo lógico programable (por ejemplo una matriz de puertas programables del campo) para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en la presente. En algunas realizaciones, una matriz de puertas programables del campo puede cooperar con un microprocesador con el fin de llevar a cabo uno de los métodos descritos en el presente documento.

REIVINDICACIONES

1. Codificador de audio para generar una señal de audio codificada, que comprende:
 - 5 un analizador (516, 520) de señal de audio para analizar si un marco de tiempo de la señal de audio tiene una característica armónica o de expresión;
 - un controlador (504) de función de ventana para seleccionar una función de ventana dependiendo de una característica armónica o de expresión de la señal de audio;
 - 10 un generador (502) de ventanas para generar ventanas de la señal de audio utilizando la función de ventana seleccionada para obtener un marco con ventanas; y
 - un procesador (508, 512) para procesar adicionalmente el marco con ventanas para obtener la señal de audio codificada;
 - 15 en el que el controlador (504) de función de ventana comprende un detector (700) de transitorios para detectar un transitorio, en el que el controlador de función de ventana está configurado para conmutar de una función de ventana para un bloque grande a una función de ventana para un bloque corto, cuando un transitorio se detecta y el analizador (516, 520) de señal de audio no encuentra una característica armónica o de expresión, y para no conmutar a la función de ventana para el bloque corto, cuando se detecta un transitorio y el analizador (516, 520) de señal de audio encuentra una característica armónica o de expresión; y
 - 20 en el que el controlador (504) de función de ventana está configurado para conmutar de una función (706) de ventana que tiene una longitud de muestras 2048 como el bloque largo, pero que tiene una porción (708) cero y una porción (710) de no aliasing a una función (707) de ventana que es más larga que la función de ventana para un bloque corto y está adaptada para obtener una longitud (712) de superposición más corta con una ventana (706) previa que la función (714) de ventana para un bloque largo y que tiene una porción cero a la izquierda de una región (712) y una porción de no aliasing a la derecha de la región (712), cuando se detecta un transitorio y la señal tiene una característica armónica o de expresión, de modo que hay una porción de superposición entre 1/4 y 1/32 de la longitud de función de ventana completa.
2. Codificador de audio para generar una señal de audio codificada, que comprende:
 - 35 un analizador (516, 520) de señal de audio para analizar si un marco de tiempo de la señal de audio tiene una característica armónica o de expresión;
 - un controlador (504) de función de ventana para seleccionar una función de ventana dependiendo de una característica armónica o de expresión de la señal de audio;
 - 40 un generador (502) de ventanas para generar ventanas a la señal de audio utilizando la función de ventana seleccionada para obtener un marco con ventanas;
 - 45 un procesador (508, 512) para procesar adicionalmente el marco con ventanas para obtener la señal de audio codificada, y
 - un detector de transitorios;
 - 50 en el que el detector (700) de transitorios está configurado para detectar una característica cuantitativa de la señal de audio y para comparar la característica cuantitativa con un umbral controlable, en el que un transitorio es detectado cuando la característica cuantitativa tiene una relación predeterminada con el umbral controlable, y
 - 55 en el que el analizador de señal de audio está configurado para controlar el umbral variable de modo que se reduce una probabilidad de una conmutación a una función de ventana para un bloque corto, cuando el analizador (516, 520) de señal de audio ha encontrado una característica armónica o de expresión.
3. Método para generar una señal de audio codificada, que comprende:
 - 60 analizar (516, 520) si el marco de tiempo de la señal de audio tiene una característica armónica o de expresión;
 - seleccionar (504) una función de ventana dependiendo de una característica armónica o de expresión de la

señal de audio;

generar (502) ventanas a la señal de audio utilizando la función de ventana seleccionada para obtener un marco con ventanas; y

5

procesar (508, 512) el marco con ventanas para obtener la señal de audio codificada;

en el que se realiza una conmutación de una función de ventana para un bloque largo a una función de ventana para un bloque corto, cuando se detecta un transitorio y no se encuentra una característica armónica o de expresión mediante el análisis, y

10

en el que se realiza una conmutación de una función (706) de ventana que tiene una longitud de muestras 2048 como el bloque largo, pero que tiene una porción (708) cero y una porción (710) de no aliasing a una función (707) de ventana que es más larga que la función de ventana para un bloque corto y que tiene una superposición (712) más corta con una ventana (706) previa que la función (714) de ventana para un bloque largo y que tiene una porción cero a la izquierda de una región 712 de superposición y una porción de no aliasing a la derecha de la región (712) de superposición, cuando se detecta un transitorio y la señal tiene una característica armónica o de expresión, de modo que hay una porción de superposición entre 1/4 y 1/32 de la longitud de función de ventana completa.

15

20

4. Método para generar una señal de audio codificada, que comprende:

analizar (516, 520) si un marco de tiempo de la señal de audio tiene una característica armónica o de expresión;

25

seleccionar (504) una función de ventana dependiendo de una característica armónica o de expresión de la señal de audio;

generar (502) ventanas a la señal de audio utilizando la función de ventana seleccionada, para obtener un marco con ventanas; y

30

procesar (508, 512) el marco con ventanas para obtener la señal de audio codificada;

en el que se detecta una característica cuantitativa de la señal de audio y la característica cuantitativa es comparada con un umbral controlable, en el que un transitorio es detectado cuando la característica cuantitativa tiene una relación predeterminada con el umbral controlable, y

35

en el que el umbral variable es controlado de modo que se reduce una probabilidad de una conmutación a una función de ventana para un bloque corto, cuando se ha encontrado una característica armónica o de expresión.

40

5. Programa informático que tiene un código de programa para realizar, cuando se ejecuta en un ordenador, el método de las reivindicaciones 3 ó 4.

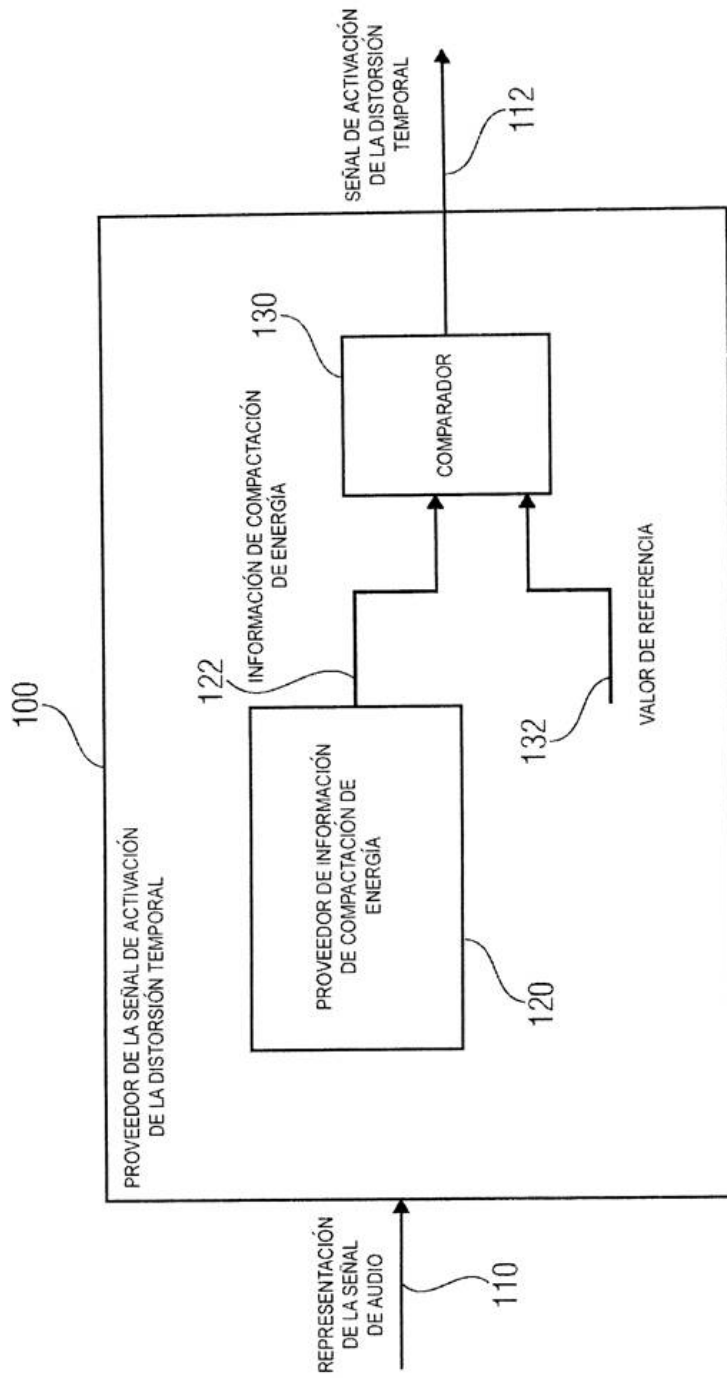


FIG 1

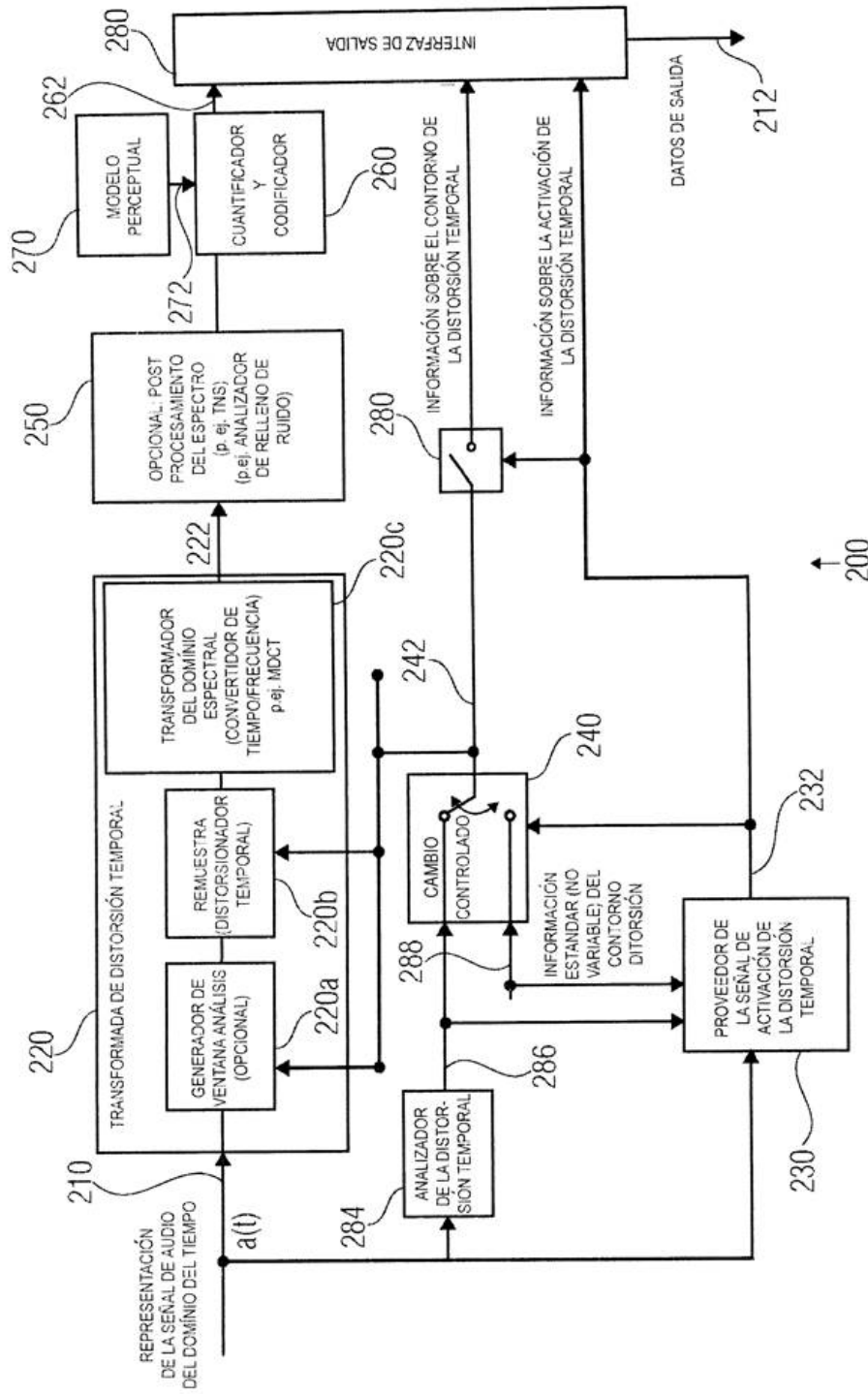


FIG 2A

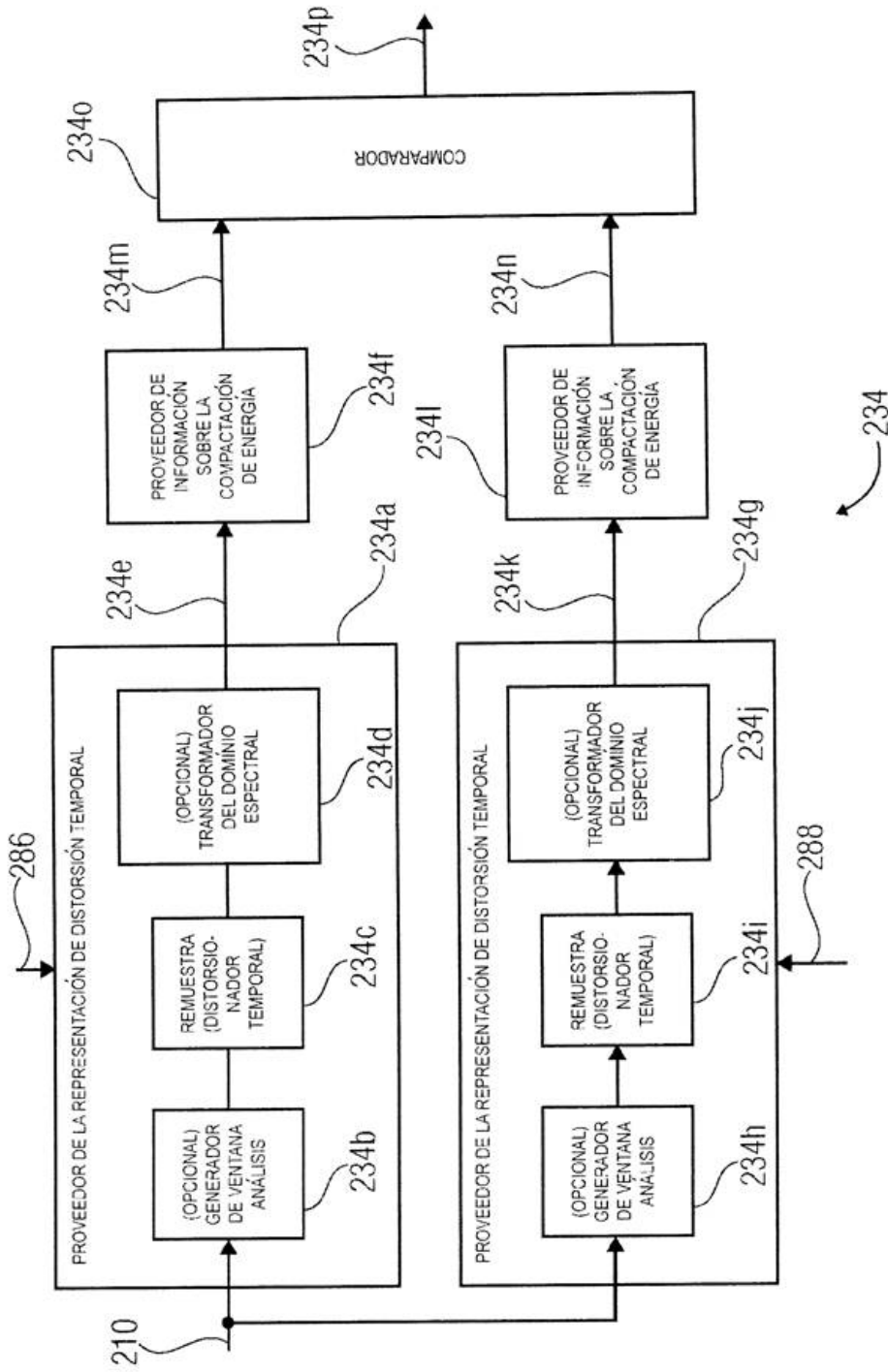
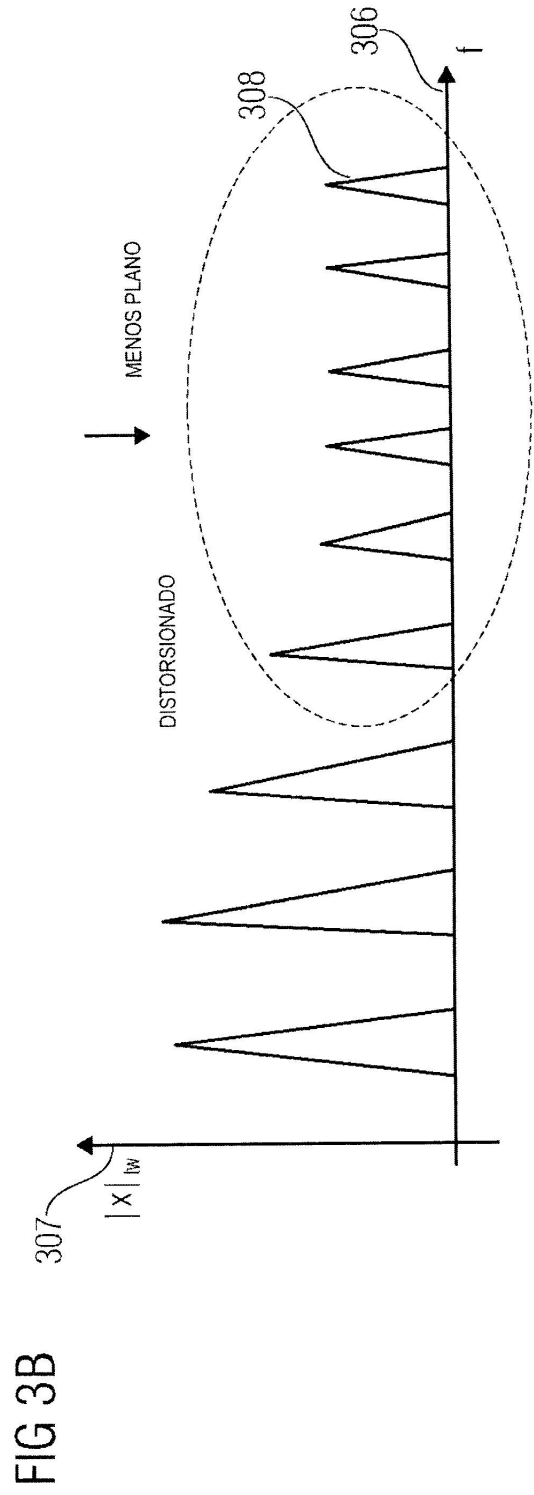
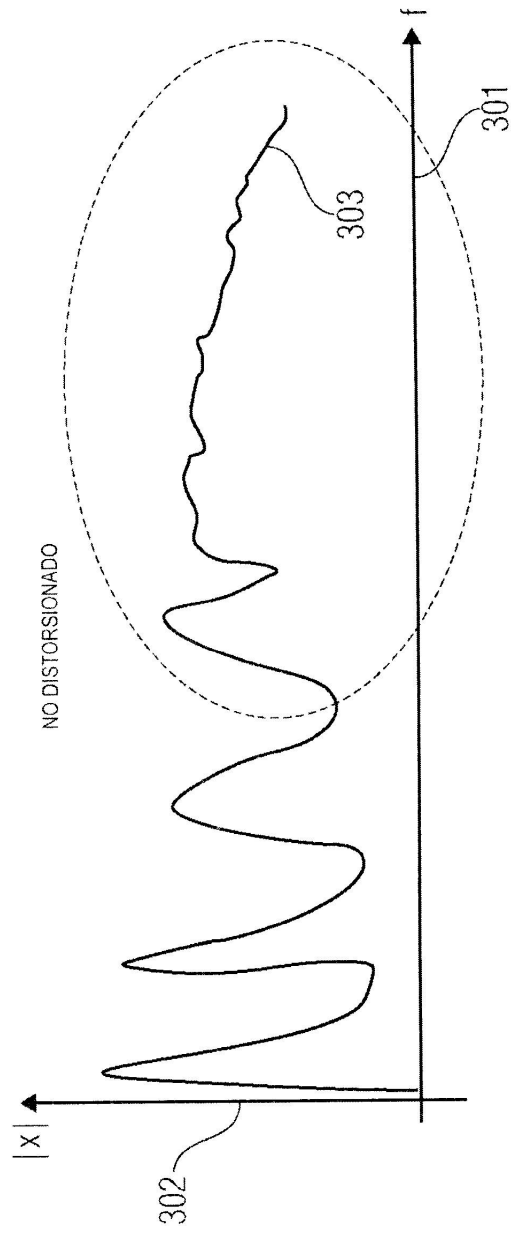


FIG 2B



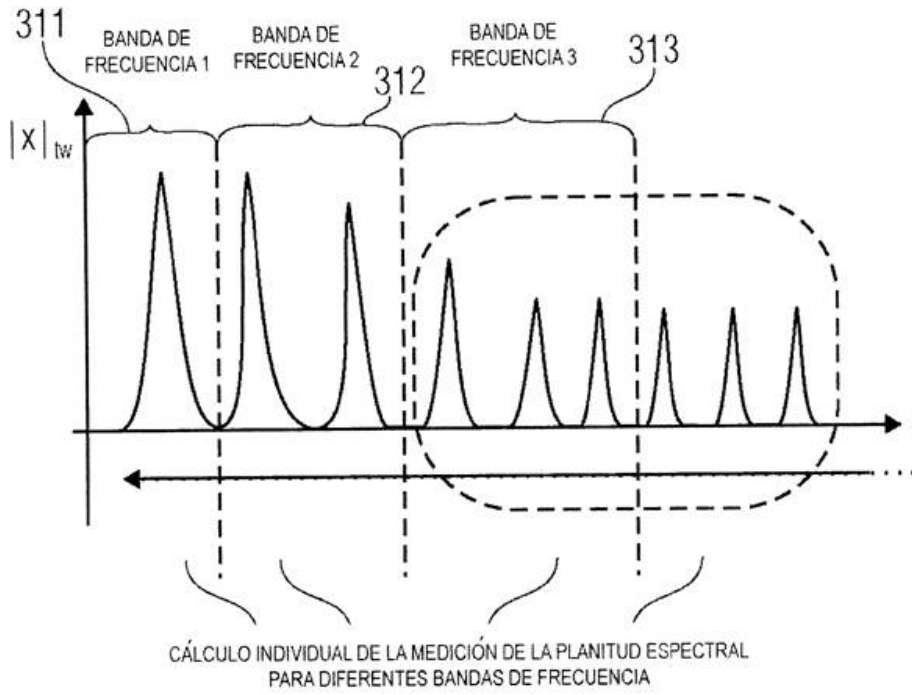
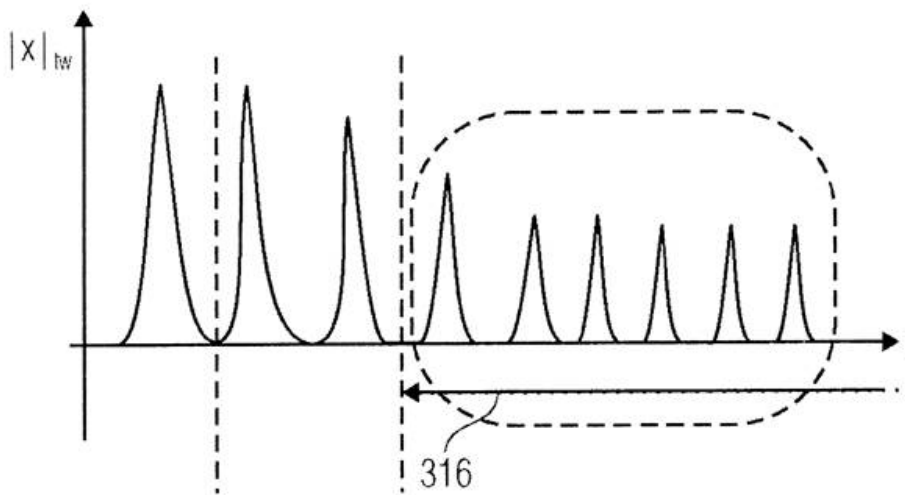


FIG 3C



CONSIDERACIÓN DE LA PORCIÓN DE FRECUENCIA MÁS ALTA SOLAMENTE (MIENTRAS SE DESATIENDE LA PORCIÓN DE FRECUENCIA MÁS BAJA) PARA EL CÁLCULO DE LA MEDICIÓN DE PLANITUD ESPECTRAL (ANCHO DE BANDA DE FRECUENCIA O POR ENCIMA DE LA PORCIÓN DE FRECUENCIA MÁS ALTA ENTERA)

FIG 3D

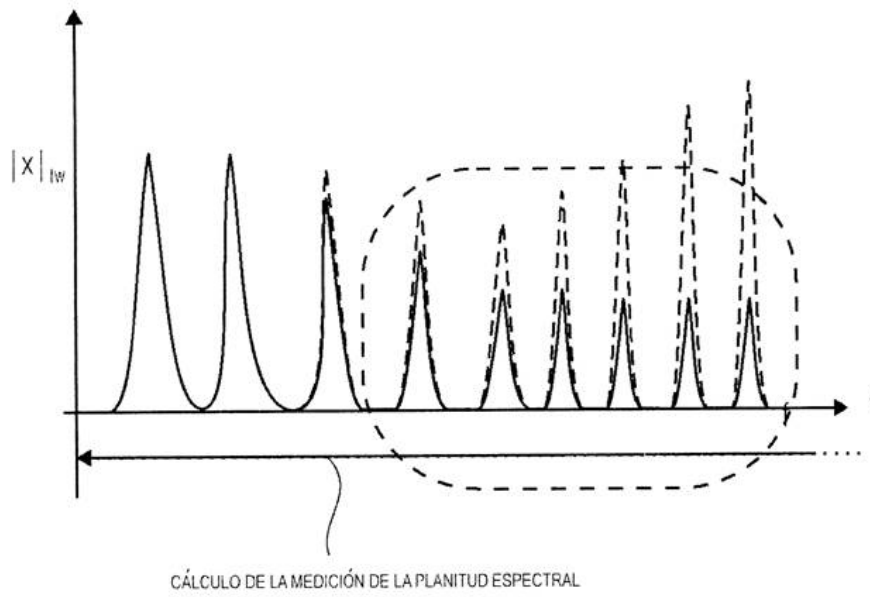


FIG 3E

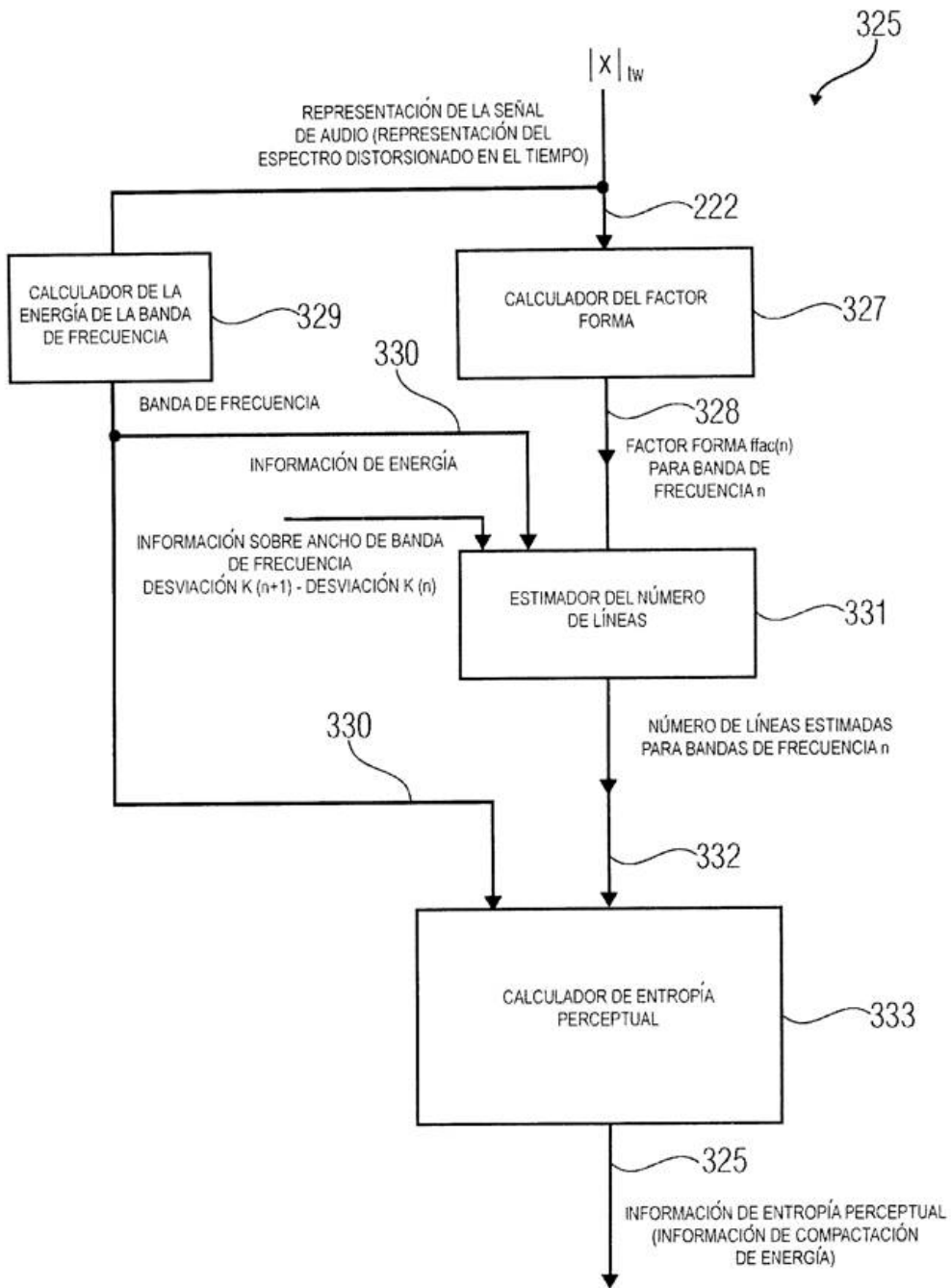


FIG 3F

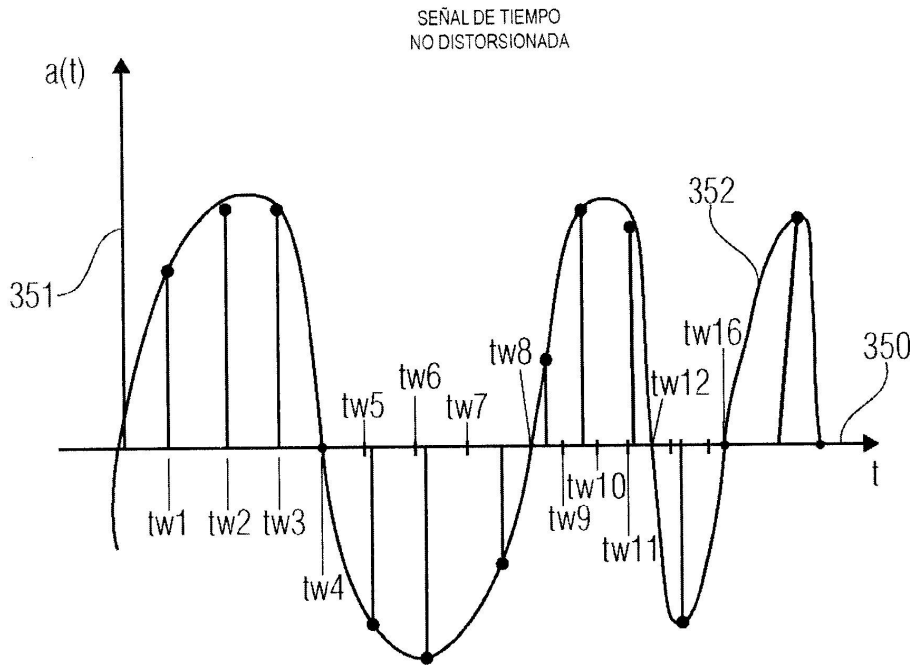


FIG 3G

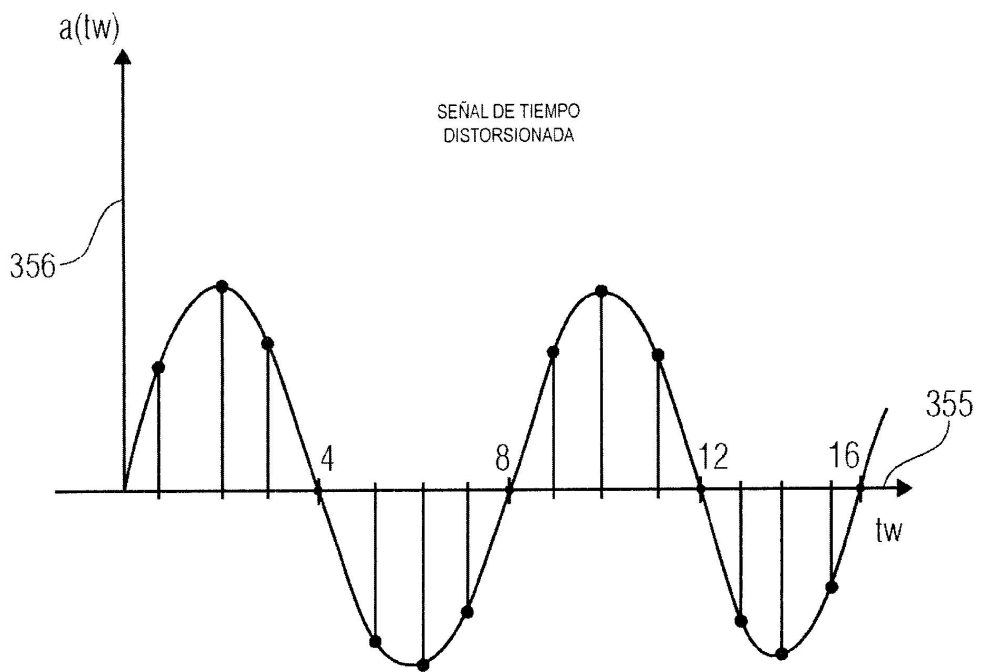


FIG 3H

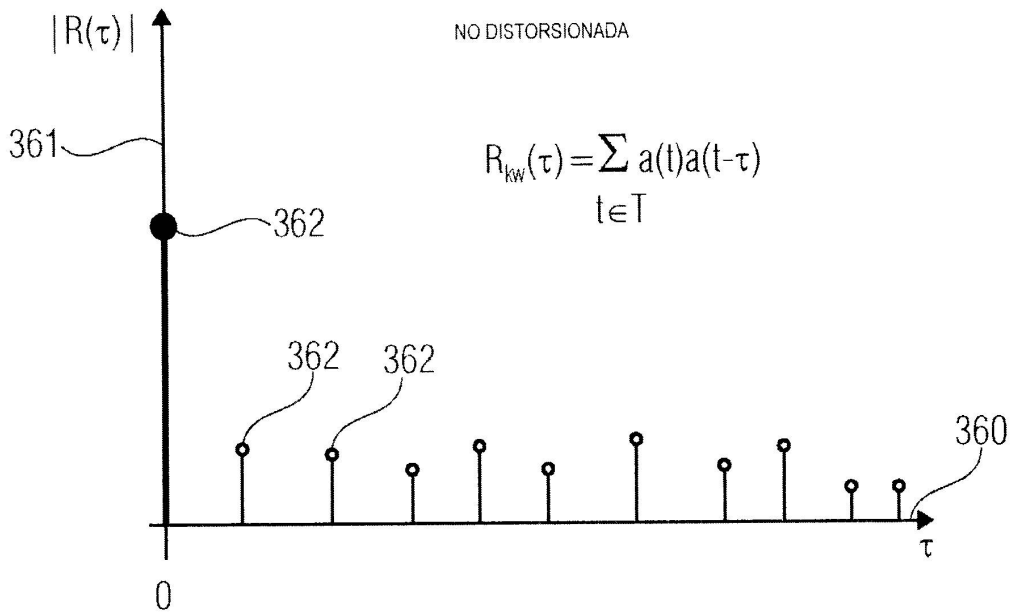


FIG 3I

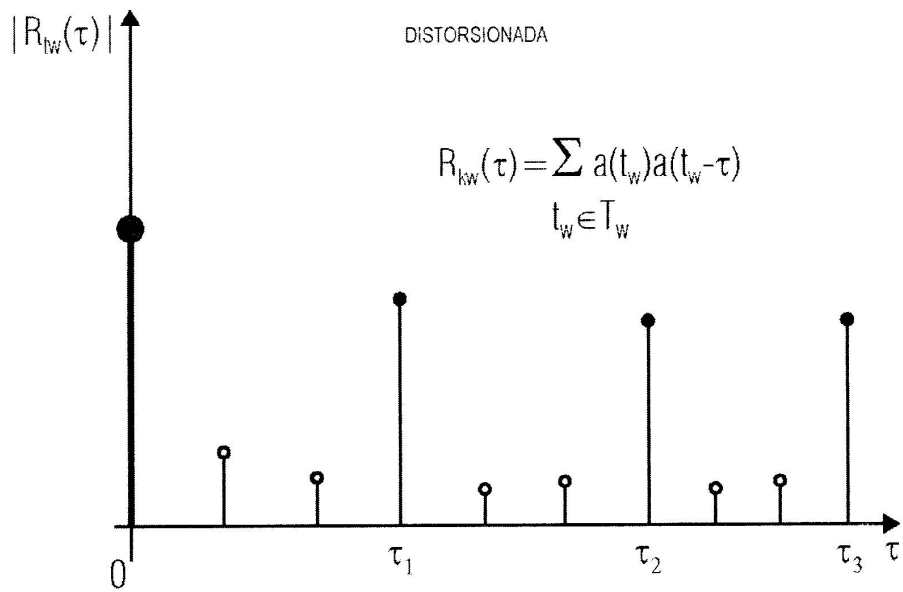


FIG 3J

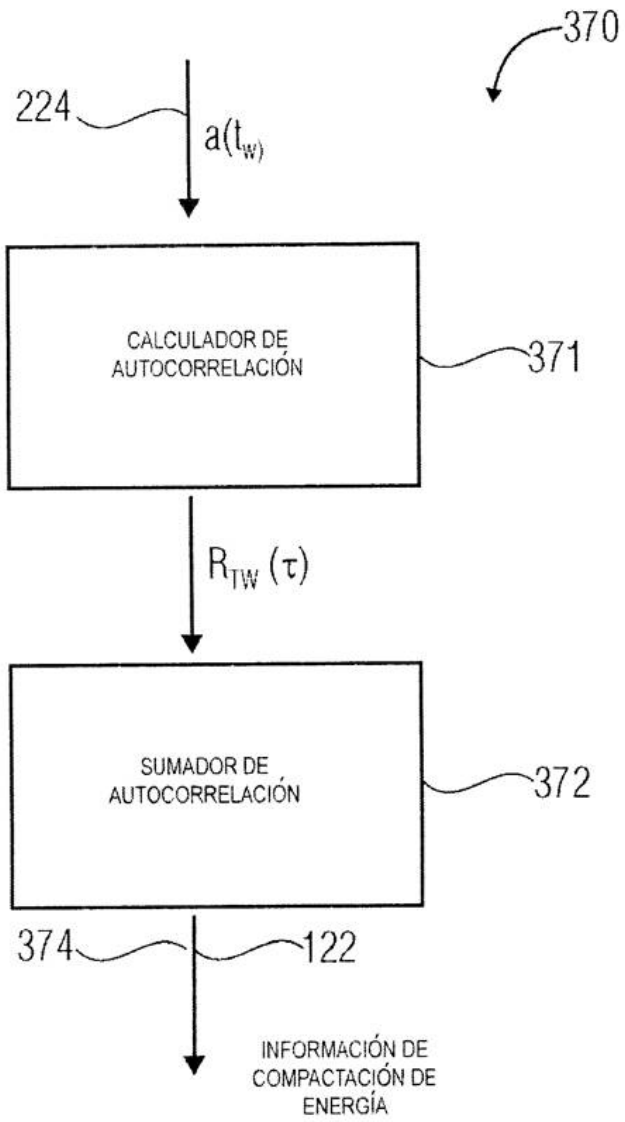


FIG 3K

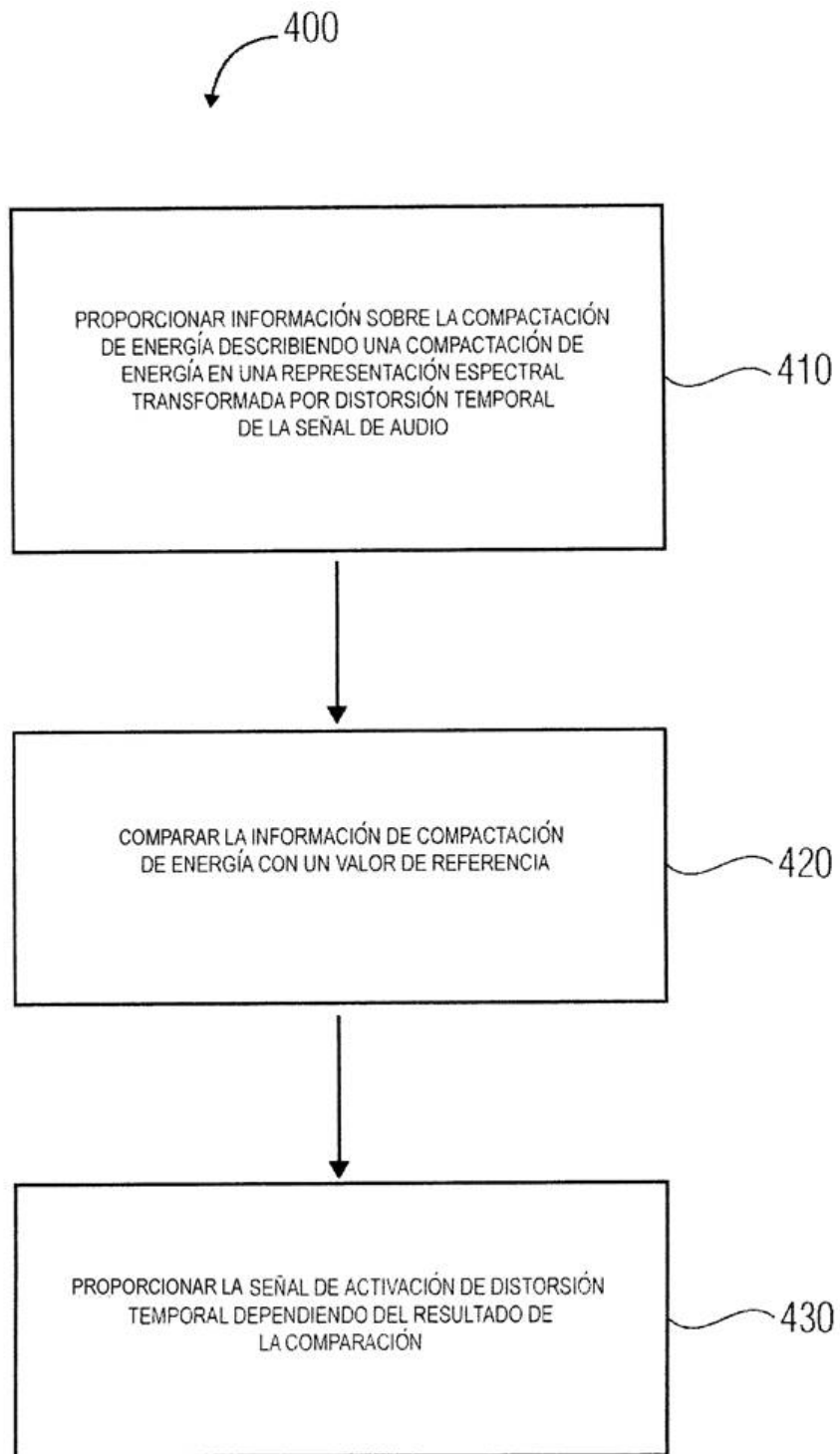


FIG 4A

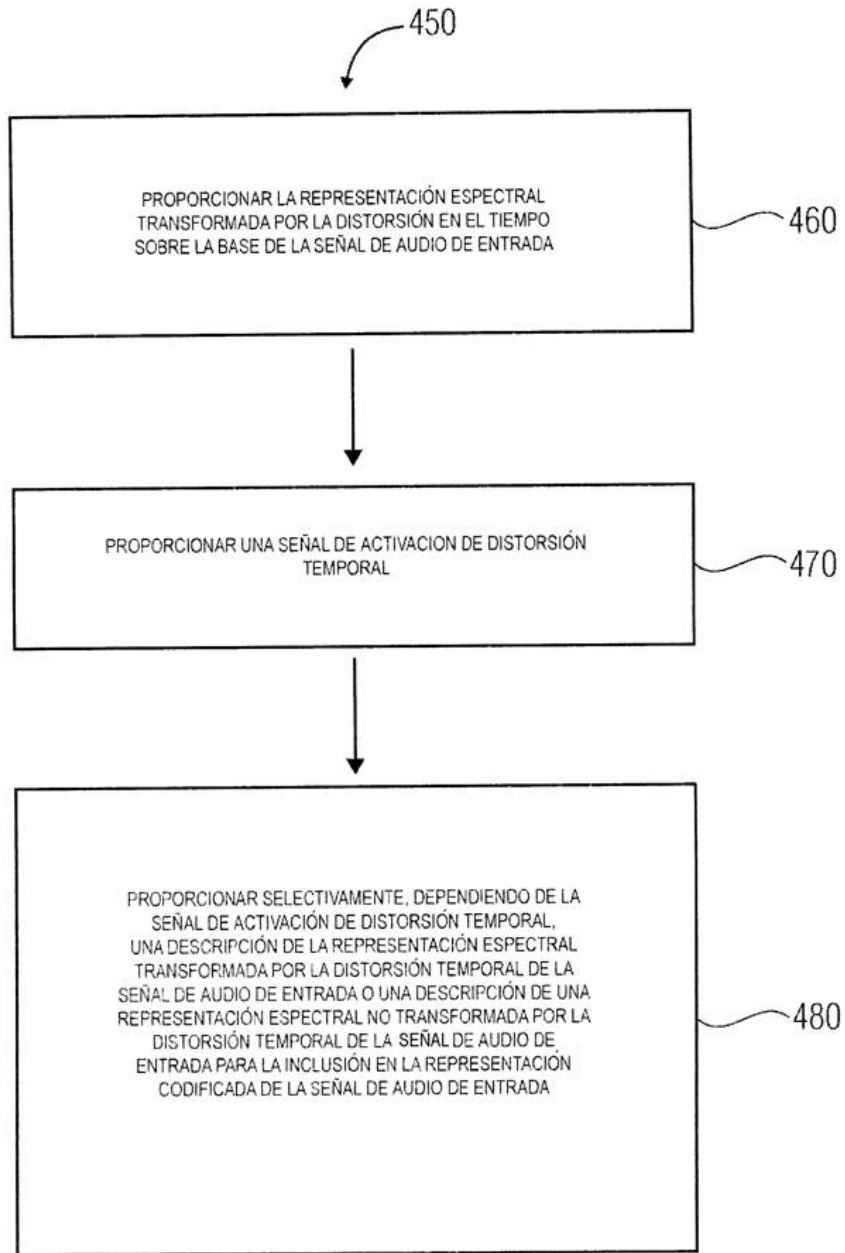


FIG 4B

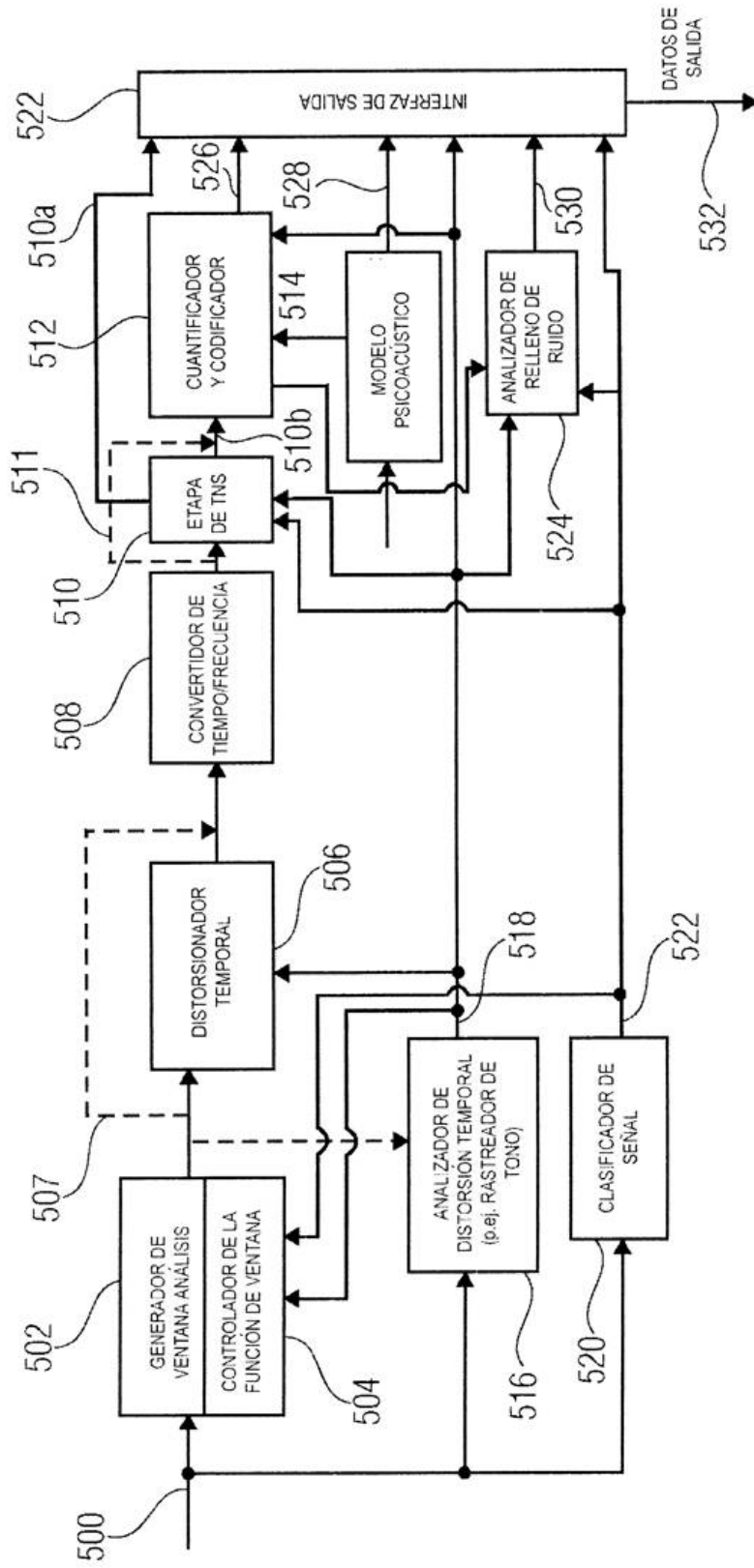


FIG 5A
(CODIFICADOR)

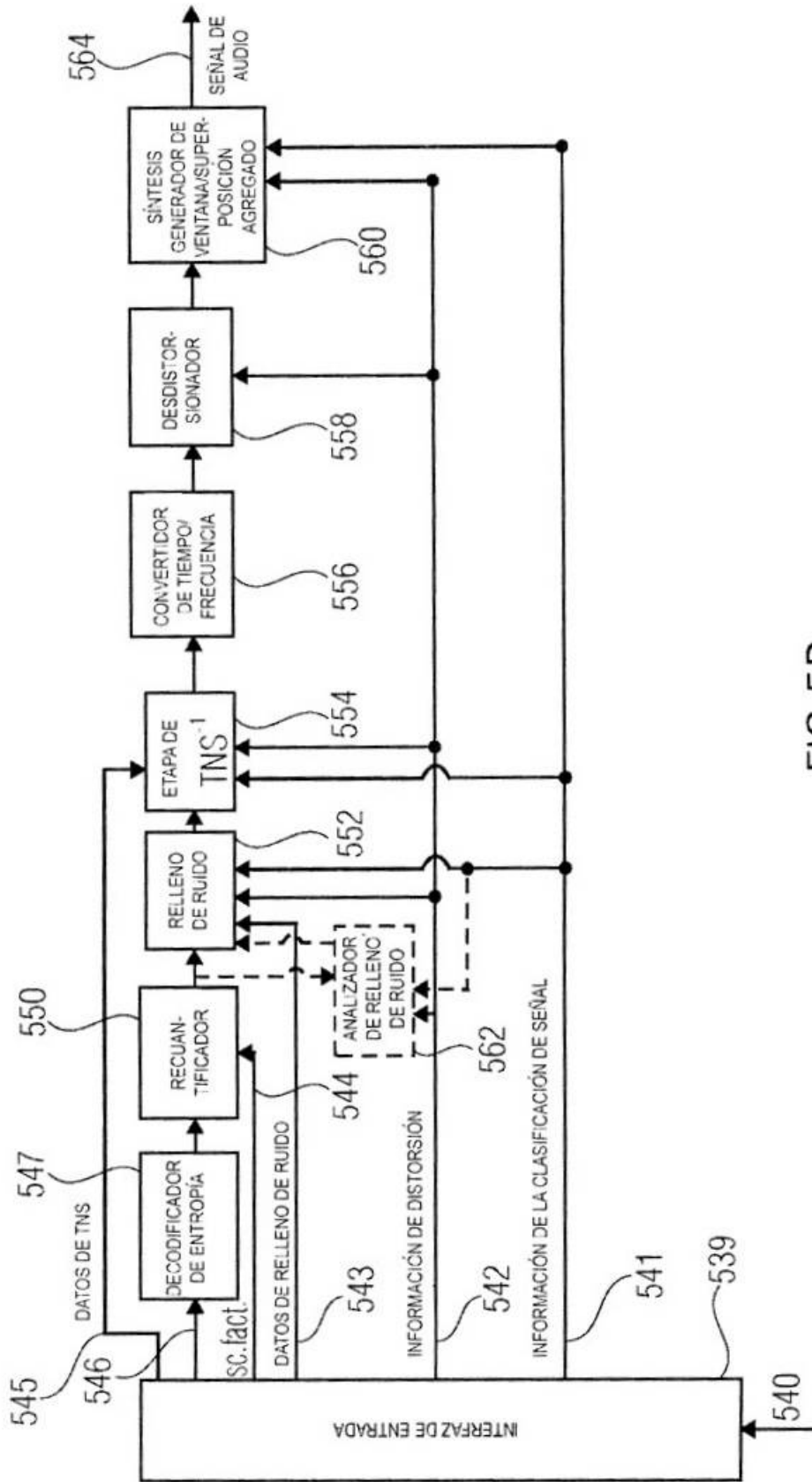


FIG 5B
(DECODIFICADOR)

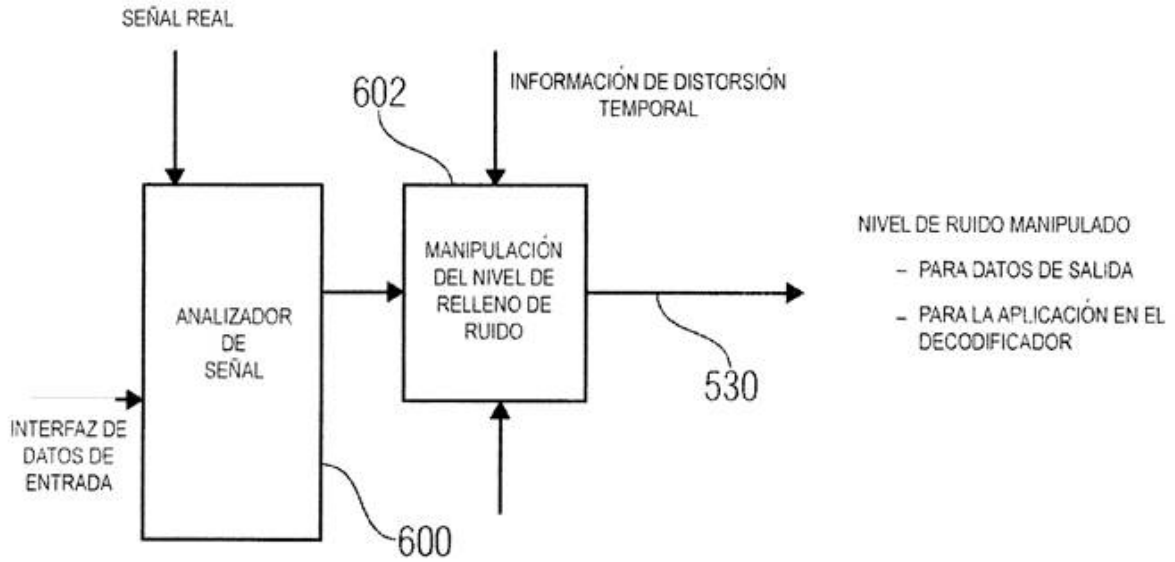


FIG 6A

CLASIFICACIÓN DE EXPRESIÓN		DISTORSIÓN TEMPORAL		CLASIFICACIÓN DE SEÑAL		NIVEL DE RELLENO DE RUIDO
V	UV	SI	NO	EXPRESIÓN	NO EXPRESIÓN	
			X		X	NORMAL
		X			X	BAJO
X		X		X		MUY BAJO / CERO
	X		X	X		NORMAL

FIG 6B

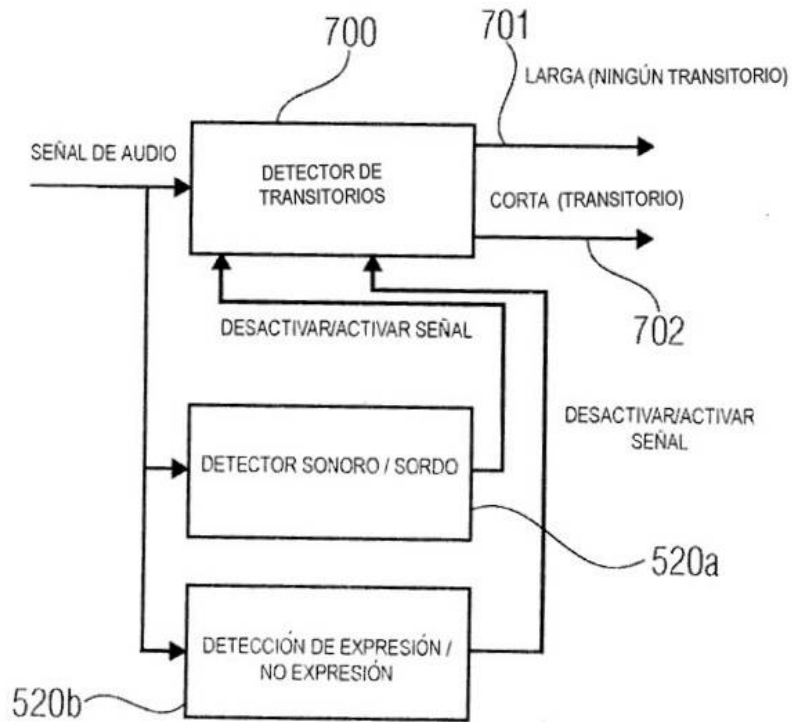


FIG 7A

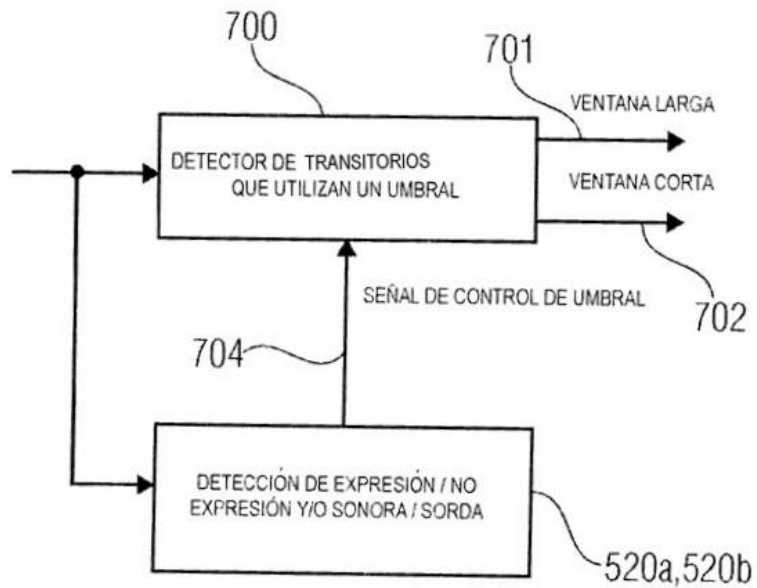


FIG 7B

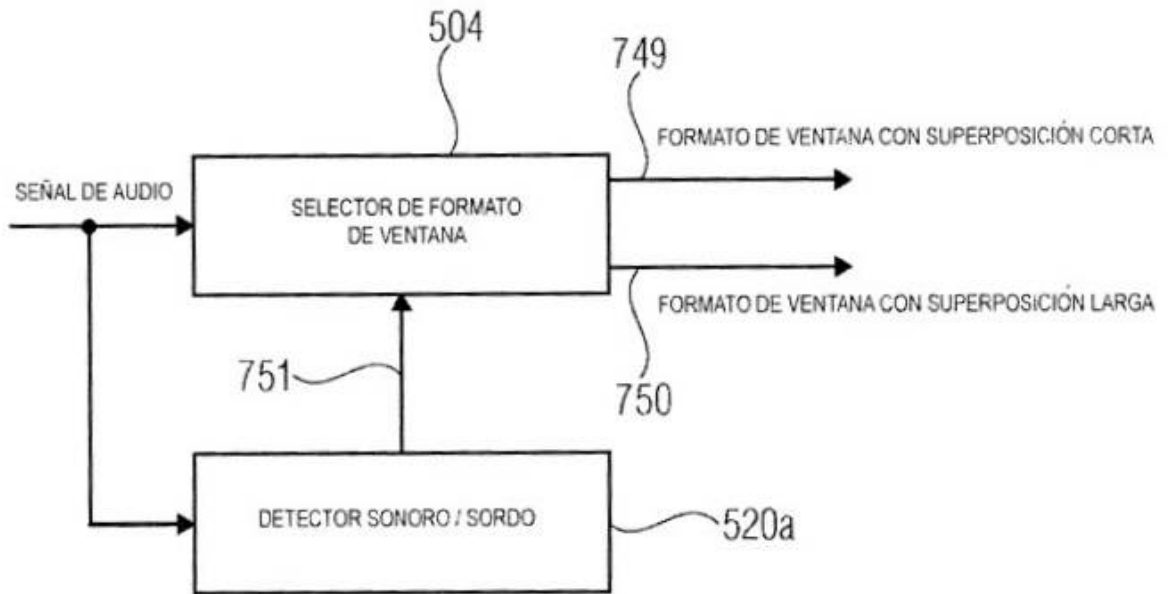


FIG 7C

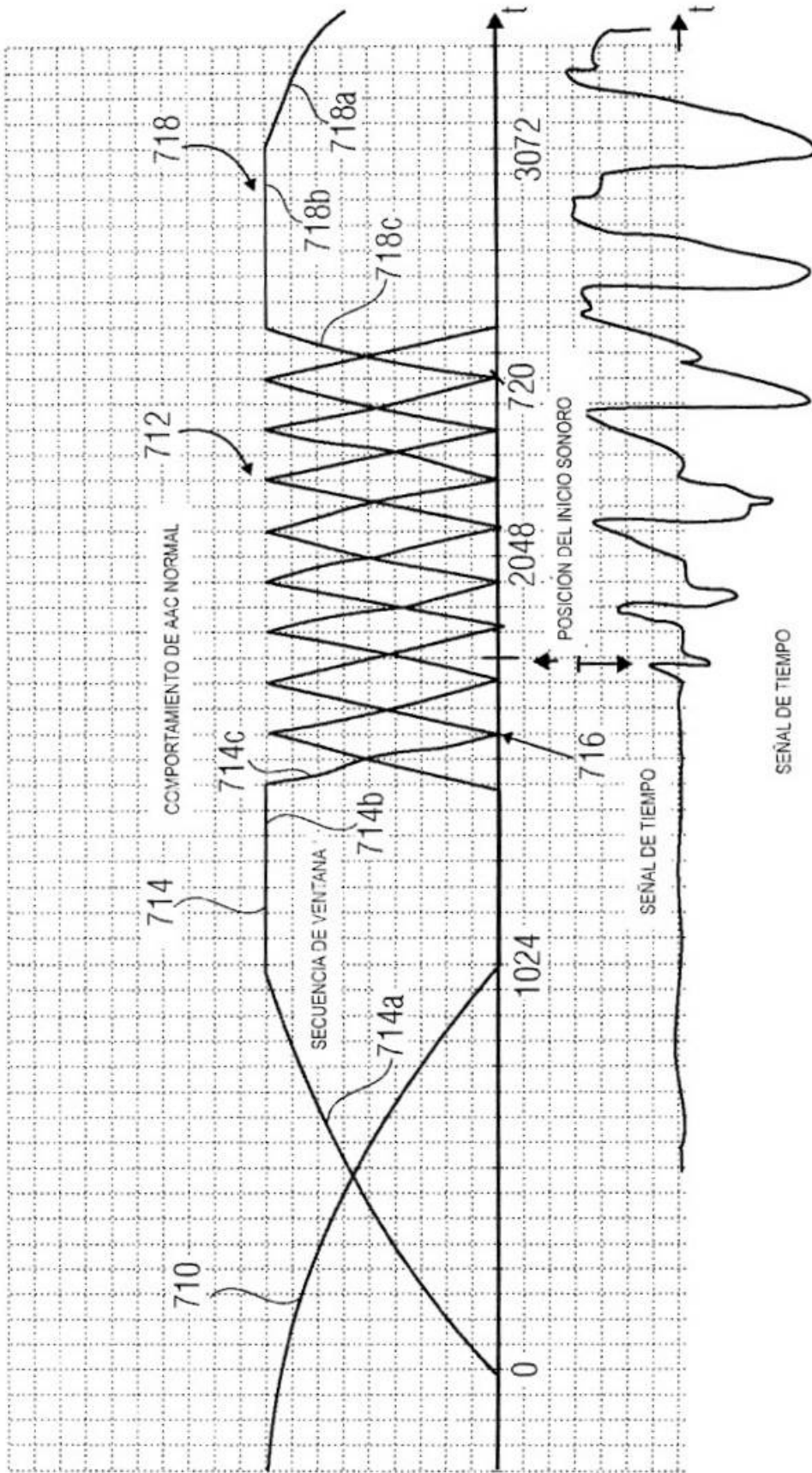


FIG 7D

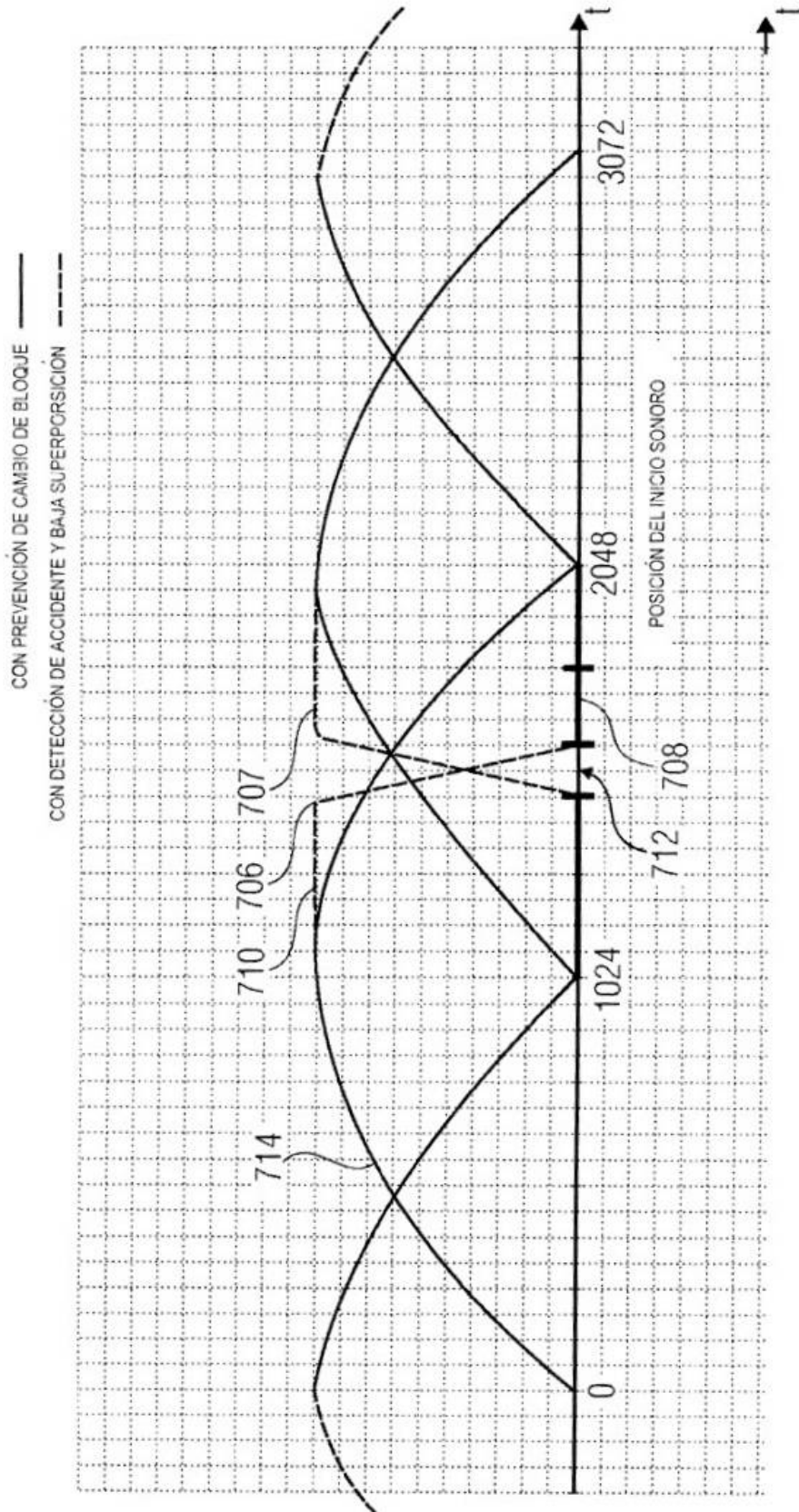


FIG 7E

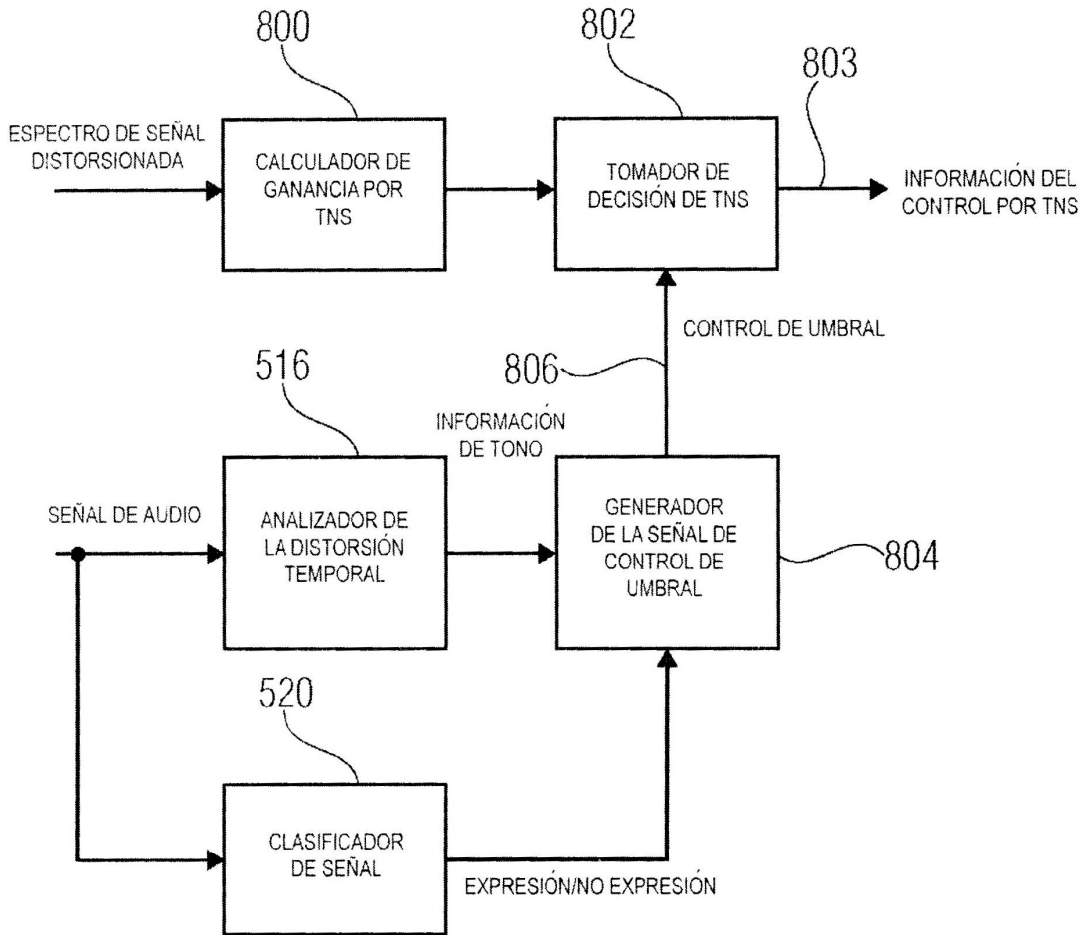


FIG 8A

		SONORA	SORDA	UMBRAL DE DECISION DE TNS
CONTORNO DE TONOS		CLASIFICACION DE SEÑAL		
sí	NO	EXPRESION	NO EXPRESION	
	X		X	NORMAL
X			X	MÁS BAJO
X		X		(AUN) MÁS BAJO

FIG 8B

FIG 9A

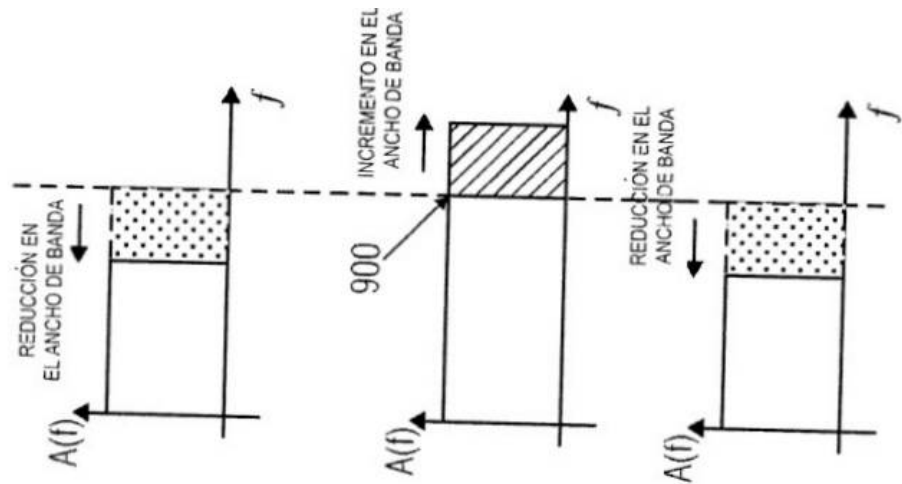


FIG 9B

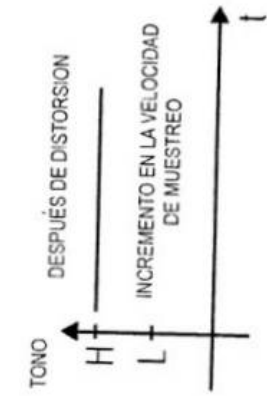


FIG 9C

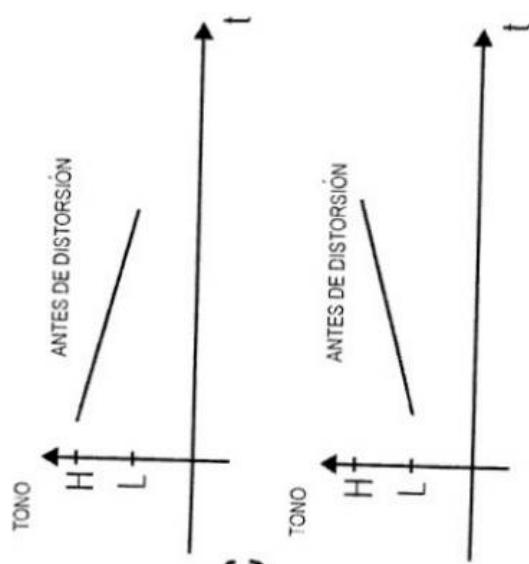


FIG 9D

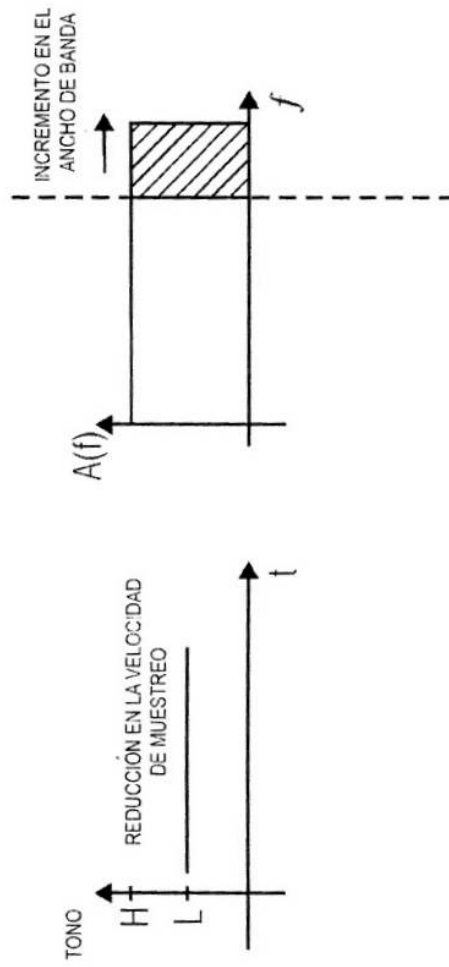
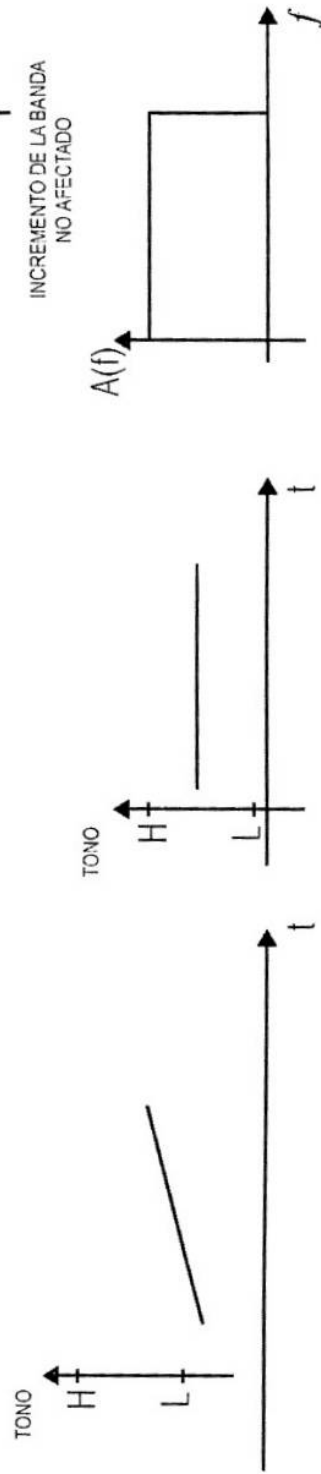


FIG 9E



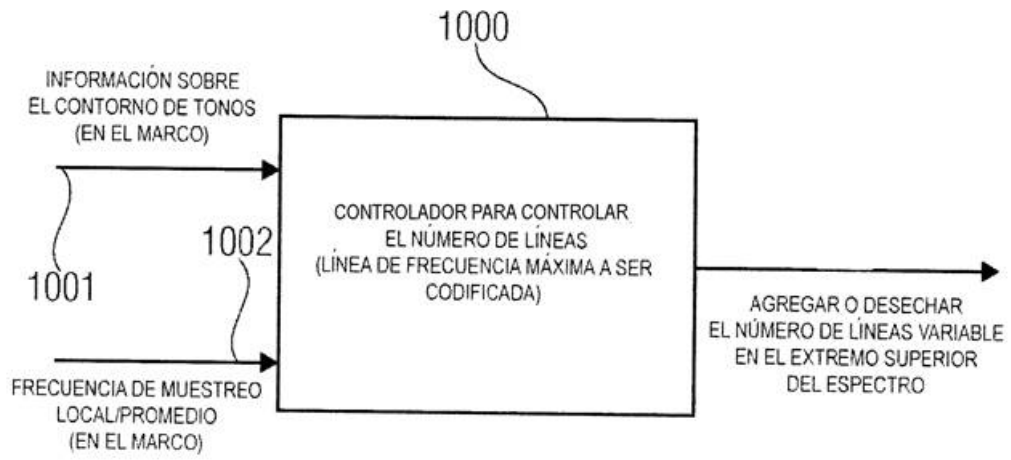


FIG 10A

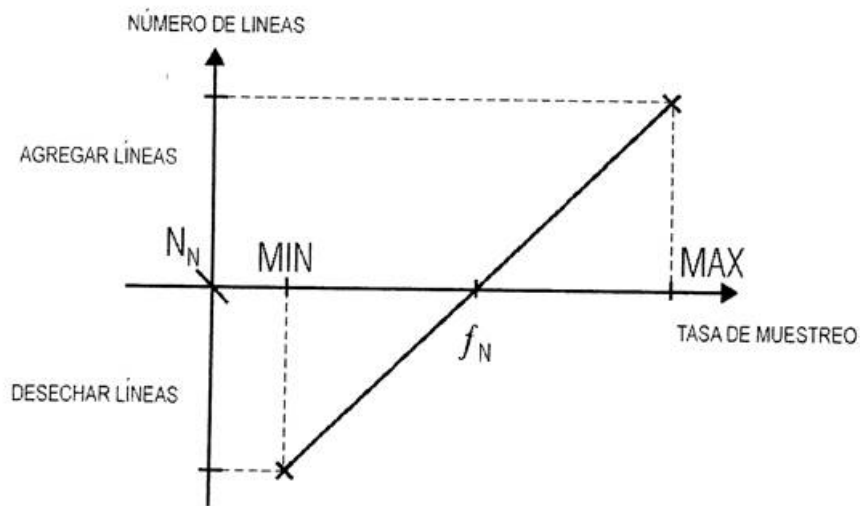


FIG 10B

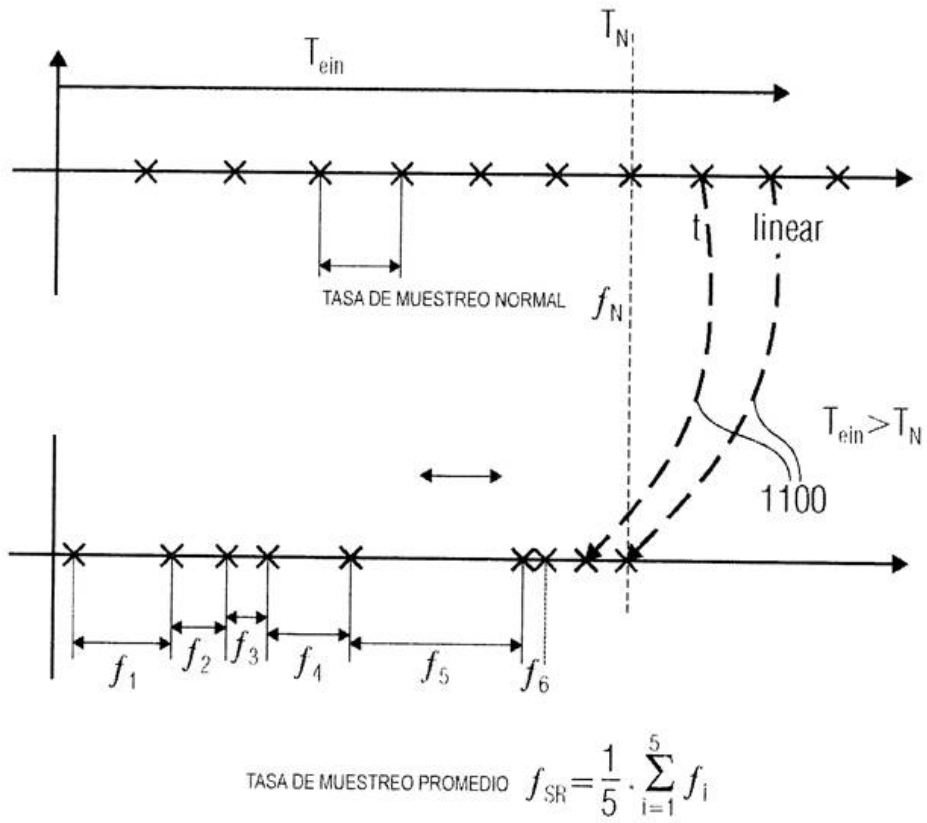


FIG 11

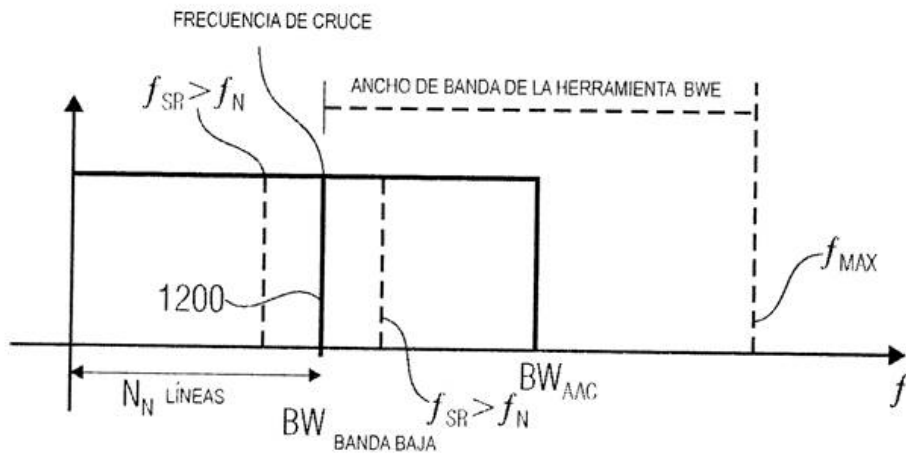


FIG 12A

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

N : NÚMERO DE LÍNEAS SEGÚN LO DETERMINADO
 POR EL PROCESADOR

T_{ein} : LONGITUD DE MARCO NO DISTORSIONADO
 (DEPENDE DE LAS CARACTERÍSTICAS 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
$< T_N$	$< f_N$	$< \Delta f_N$	$> N_N$ AGREGAR LÍNEAS

FIG 12B