

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 758 799**

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 11180983 (6)**

97 Fecha y número de publicación de la concesión europea: **04.09.2019 EP 2410519**

54 Título: **Método y aparato para codificar y decodificar 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:

06.05.2020

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 2 758 799 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

Método y aparato para codificar y decodificar 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 dentro de la longitud de código mínima).

20 Por ejemplo, las transformadas del tipo lapped 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.

25 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 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 grilla 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 grilla 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 variante temporal del tono, de modo que una variante 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 variantes 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.

60 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.

Un ejemplo de un proceso de codificación se proporciona en el documento WO 2006/113921 A1.

En vista de la descripción anterior, existe el 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.

5 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 in un codificador de señal de audio de distorsión de tiempo o un decodificador de señal de audio de distorsión de tiempo.

10 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.

Es otro objetivo 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

20 Este objetivo se obtiene mediante un codificador de audio de acuerdo con la reivindicación 1, un decodificador de audio de acuerdo con la reivindicación 3, un método de codificación de audio de acuerdo con la reivindicación 4, un método de decodificación de acuerdo con la reivindicación 5, o un programa informático de acuerdo con la reivindicación 6.

25 Todos las siguientes apariciones de la palabra "realización (es)", si se refiere a combinaciones de características que no comprenden todas las características definidas por las reivindicaciones independientes adjuntas, se refieren a ejemplos que se presentaron originalmente pero que no representan realizaciones de la invención reivindicada actualmente.

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

La figura 1 muestra un diagrama esquemático en bloque de un proveedor de la señal de activación de distorsión de tiempo, de acuerdo con un ejemplo de comparación;

35 La figura 2a muestra un diagrama esquemático en bloque de un codificador de señal de audio, de acuerdo con 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 de acuerdo con un ejemplo de comparación;

40 La figura 3a muestra una representación gráfica de un espectro de una versión sin distorsión de tiempo de una señal de audio;

45 La figura 3b muestra una representación gráfica de un espectro de una versión de distorsión de tiempo de la señal de audio;

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

50 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;

55 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;

La figura 3f muestra un diagrama esquemático en bloque de un proveedor de información de compactación de energía, de acuerdo con un ejemplo de comparación;

60 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;

- La figura 3i muestra una representación gráfica de una función de autocorrelación de la señal de audio de acuerdo con la figura 3g;
- 5 La figura 3j muestra una representación gráfica de una función de autocorrelación de la señal de audio de acuerdo con la figura 3h;
- La figura 3k muestra un diagrama esquemático en bloque de un proveedor de información de compactación de energía, de acuerdo con un ejemplo de comparación;
- 10 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, de acuerdo con un ejemplo de comparación;
- 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, de acuerdo con un ejemplo de comparación;
- 15 La figura 5a ilustra una realización preferida de un codificador de audio con aspectos de la invención;
- La figura 5b ilustra una realización preferida de un decodificador de audio con aspectos de la invención;
- 20 La figura 6a ilustra una realización preferida del aspecto de relleno de ruido de la presente invención;
- La figura 6b ilustra una tabla que define la función de control llevada a cabo por el manipulador de nivel de relleno de ruido;
- 25 La figura 7a ilustra una realización preferida para llevar a cabo un cambio de bloque basado en la distorsión de tiempo de acuerdo con un ejemplo de comparación;
- La figura 7b ilustra una realización alternativa para influir en la función de generación de ventanas;
- 30 La figura 7c ilustra otra realización alternativa para ilustrar la función de generación de ventanas en base a la información de distorsión de tiempo;
- La figura 7d ilustra una secuencia de generación de ventanas de un comportamiento de AAC normal en un inicio sonoro;
- 35 La figura 7e ilustra secuencias de ventanas alternativas de acuerdo con una realización preferida de acuerdo con un ejemplo de comparación;
- 40 La figura 8a ilustra una realización preferida de un control en base a la distorsión de tiempo de la herramienta de TNS (formato de ruido temporal) de acuerdo con 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;
- 45 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;
- 50 La figura 10a ilustra una realización preferida de un controlador para controlar el número de líneas dentro de un procesador de codificación, de acuerdo con un ejemplo de comparación;
- La figura 10b ilustra la dependencia entre el número de líneas a ser descartadas/agregadas para una velocidad de muestreo;
- 55 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
- 60 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.
- La figura 1 muestra un diagrama esquemático en bloque del proveedor de la señal de activación de distorsión de tiempo, de acuerdo con una realización de la invención. El proveedor de la señal 100 de activación de distorsión de

5 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 de la señal 100 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 de la señal 100 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.

10 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.

15 La figura 2a muestra un diagrama esquemático en bloque de un codificador 200 de señal de audio, de acuerdo con una realización de la invención. El codificador 200 de señal de audio está configurado para recibir una señal 210 de audio de entrada (además designada 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.

20 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 del contorno de distorsión de tiempo (por ej. información 286 absoluta o relativa del contorno de distorsión de tiempo).

25 El codificador 200 de señal de audio además comprende un mecanismo de cambio, por ejemplo en la forma de un cambio 240 controlado, para decidir respecto si la información 286 hallada del contorno de distorsión de tiempo o una información 288 estándar del contorno de distorsión de tiempo se utiliza para 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 242 nueva del contorno de distorsión de tiempo, 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 250 procesamiento 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 perceptual 270 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 de acuerdo con la percepción humana. El codificador 200 de señal de audio además comprende una interfaz de salida 280 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 suministrada por el cuantificador/codificador 260.

30 El codificador 200 de señal de audio además comprende un proveedor 230 de la señal de activación de distorsión de tiempo, 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 212 codificada, por ejemplo en la forma de una marca de un bit que indica la activación o desactivación de la distorsión de tiempo.

A fin de facilitar el entendimiento, debe observarse que un transformador 220 de distorsión de tiempo típicamente

comprende una ventana 220a de análisis, un remuestreador o “distorsionador 220b de tiempo” (time warper) y un transformador 220c de dominio espectral (o convertidor de tiempo/frecuencia). Dependiendo de la implementación, sin embargo, el distorsionador 220b de tiempo puede ser colocado – en una dirección de procesamiento de señal - antes de la ventana 220a de análisis. Sin embargo, la distorsión de tiempo y el dominio de tiempo para la transformación del dominio espectral pueden combinarse en una sola unidad en algunas realizaciones.

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

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.

La figura 2b muestra un diagrama esquemático en bloque de un proveedor 234 de la señal de activación de distorsión de tiempo, de acuerdo con 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) idénticas ventanas 234b y 234h de análisis, idénticos 234c y 234i remuestreadores o distorsionadores 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 302 y una ordenada 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.

5
10 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 términos de velocidad de bits, requerido para la codificación de la información de distorsión de tiempo (por ej. contorno de distorsión de tiempo) puede exceder los salvados, en términos de 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.

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

25 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 ej. 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 ej. 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).

30
35 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).

40 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 (ver las figuras 3a y 3b). Una mirada a ésta muestra que una compactación de energía además corresponde a un espectro más “sin planicidad” (ver 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.

45
50 Las figuras 3a y 3b muestran un ejemplo esquemático con un 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 la 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.

55 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 de acuerdo con la siguiente ecuación:

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

60 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:

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

En este caso, N puede ser igual al número de líneas espectrales provistas por el transformador 234d, 234j de domino espectral y $|X|_{t_w}(n)$ 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 (ver 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 de acuerdo con 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", además denominada como AAC.

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 primer medida de planicidad espectral puede computarse para la primera 311 banda 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 311 banda de frecuencia (N variable puede tomar el ancho en término de cajas de frecuencia de la primera banda de frecuencia). En consecuencia, se obtiene una medida de planicidad para la primera 311 banda de frecuencia. De manera similar, una medida de planicidad puede computarse para la segunda 312 banda de frecuencia, tomando en cuenta las cajas de frecuencia de las segundas 312 bandas de frecuencia 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 313 banda de frecuencia, pueden computarse de la misma forma.

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.

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 superior 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 completa de frecuencia más alta se puede considerar en su totalidad para el cálculo de la medida de planicidad espectral.

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.

Por ejemplo, el proveedor de la señal 100, 230, 234 de activación de distorsión de tiempo (o el comparador 130, 234o 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.

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 altos 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.

En términos de 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.

La figura 3f muestra un proveedor de información de compactación de energía 325, 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 de la señal 100, 290, 234 de activación de distorsión de tiempo. El proveedor de información de compactación de energía 325 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 $|X|_{t_w}$. El proveedor 325 de información de compactación de energía además está configurado para proveer una información de entropía perceptual 326, 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 del factor de forma 327, 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 del factor de forma 328, la cual puede estar asociada con una banda de frecuencia. El proveedor 325 de información de compactación de energía además comprende un calculador de energía de la banda de frecuencia 329, el cual está configurado para calcular una información de energía de la banda de frecuencia en(n) (330) 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 326 de entropía perceptual sobre la base de la información de energía de la banda de frecuencia 330 y de la información 332 estimada del número de líneas. Por ejemplo, el calculador del factor de forma 327 puede estar configurado para computar el factor forma de acuerdo con

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

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 , de acuerdo con 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, $e_n(n)$ designa una energía en la banda de frecuencia o banda de factor de escala n . $k_{\text{Offset}}(n+1) - k_{\text{Offset}}(n)$ designa un ancho de la banda de frecuencia o banda de factor de escala de índice n en términos de cajas de frecuencia.

- 5 Más aún, el calculador de entropía perceptual 332 puede estar configurado para computar la información de entropía perceptual $sfbPe$ de acuerdo con la siguiente ecuación:

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

10 (3)

En lo anterior, se pueden sostener las siguientes relaciones:

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

- 15 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.

- 20 Como se mencionó anteriormente, la información 326 de entropía de percepción 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)".

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

- 30 Otra mirada al TW-MDCT (transformada de coseno discreto modificada por la distorsión de tiempo) es la idea básica para cambiar la señal de una manera que tenga un tono constante o casi constante dentro de un bloque. Si se obtiene un tono constante, esto significa que la máxima de la autocorrelación de un bloque en proceso aumenta. Dado que no es trivial 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.

- 35 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 observarse en la figura 3g.

- 45 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 $a(t_w)$ de la señal $a(t)$. Como puede observarse en la figura 3h, la versión de la distorsión de tiempo $a(t_w)$ 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.

- 50 En otras palabras, la figura 3h ilustra el hecho de que una señal de tiempo de una frecuencia temporalmente variable se transforma en una señal de tiempo de una frecuencia temporalmente constante mediante una operación de distorsión de tiempo apropiada, 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 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 $\tau \neq 0$.

- 60 La figura 3j muestra una representación gráfica de la función de autocorrelación R_{t_w} de la señal de tiempo de distorsión de tiempo $a(t_w)$. Como puede observarse a partir de la figura 3j, la función de autocorrelación R_{t_w}

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 $a(t_w)$. Esta periodicidad se refleja mediante los picos adicionales de la función de autocorrelación $R_{t_w}(\tau)$ cuando se compara con la función de autocorrelación $R_{uw}(\tau)$. 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 términos de una reducción de la velocidad de bits) de la distorsión de tiempo.

10 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 la ventana 234b y 234h 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 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 $R_{t_w}(\tau)$ de la señal de distorsión de tiempo $a(t_w)$ 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 $R_{t_w}(\tau)$ (por ejemplo, sobre un rango predeterminado de valores 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 de audio de entrada 310 sólo si se halla, sobre la base de la información 122, 234m, 234n de compactación de energía provista por el proveedor 370 de información de compactación de energía, que la distorsión de tiempo realmente otorga una eficiencia de codificación mejorada.

30 Para resumir lo anterior, las realizaciones de acuerdo con 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 términos de 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.

55 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 de acuerdo con 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 una ventana de análisis 502, el cual está, en esta realización, conectado a un controlador de función de generación de ventanas 504. La ventana de análisis 502 está conectado a un distorsionador de tiempo 506 (time warper). Dependiendo de la implementación, sin embargo, el distorsionador de tiempo 506 puede ser colocado – en una dirección de procesamiento de señal antes de la ventana de análisis 502. 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 2005., la colocación entre el distorsionador de tiempo 506 y la ventana de análisis 502 se puede establecer según lo requerido. Además, un convertidor de tiempo/frecuencia 508 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 de TNS (formato de ruido temporal) 510, 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 bloque codificador 512 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.

Además, el codificador que se ilustra en la figura 5a comprende un analizador de distorsión de tiempo 516, 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 de señal 520 está implementado, la decisión sonora/sorda puede efectuarse además mediante el clasificador de señal 520. En este caso, el analizador de distorsión de tiempo no necesariamente debe realizar la misma funcionalidad. La salida del analizador de distorsión de tiempo 518 está conectada a al menos una y con preferencia más de una funcionalidad en el grupo de funcionalidad que comprende el controlador de la función de generación de ventanas 504, el distorsionador de tiempo 506, la etapa 510 de TNS, el cuantificador y codificador 512 y una interfaz 522 de salida.

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

Aunque la figura 5a ilustra una situación, donde la señal de audio en la entrada de la ventana de análisis 500 es

ingresada en un analizador de distorsión de tiempo 516 y el clasificador de señal 520, las señales de entrada para estas funcionalidades además se pueden tomar de la salida de la ventana de análisis 502 y, con respecto al clasificador de señal, incluso se pueden tomar de la salida del distorsionador de tiempo 506, la salida del convertidor de tiempo/frecuencia 508 o la salida de la etapa del TNS 510.

5 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 del lado del TNS 510a, una información del lado del modelo perceptual 528, 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.

20 El codificador que se ilustra 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 de la función de generación de ventanas 504, 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.

30 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 de entrada 540 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 de clasificación de señal 541, información de distorsión de tiempo 542, datos de relleno de ruido 543, factores de escala 544, datos de TNS 545 e información espectral codificada 546. La información espectral codificada es ingresada en un decodificador entrópico 547, el cual puede comprender un decodificador Huffman o un decodificador aritmético, siempre que la funcionalidad del codificador en el bloque 512 in 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 de TNS inversa 554, la cual además recibe los datos de TNS en la línea 545. Dependiendo de la implementación, el relleno 552 de ruido y la etapa de TNS 554 se pueden aplicar en diferente orden de modo que el relleno 552 de ruido opere en los datos de salida de la etapa de TNS 554 más que en los datos de entrada de TNS. Además, se provee un convertidor de frecuencia/tiempo 556, el cual alimenta un regenerador de tiempo 558 (time dewarper). A la salida de la cadena de procesamiento de señal, una ventana de síntesis que realiza con preferencia un procesamiento de superposición/agregado se aplica como se indica en 560. El orden del regenerador de tiempo 558 y la etapa de síntesis 560 se puede cambiar, pero, en la realización preferida, se prefiere realizar un algoritmo de codificado/decodificación en base a MDCT como se define en el estándar AAC (AAC = codificación de audio avanzada). A continuación, la operación de desvanecimiento cruzado inherente de un bloque al próximo debido al procedimiento de superposición/agregado se utiliza con ventaja como la última operación en las cadenas de procesamiento de modo que todos los artefactos que bloquean la distorsión en la conversión se evitan en forma efectiva.

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

55 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.

60 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 562 de relleno 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 poseen 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. ej. 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 500 de audio en línea 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 600 de señales de audio ilustrado en la figura 6a, que se configura para analizar, si el marco temporal de la señal de audio posee 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 602 del nivel de relleno 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 602 de medición de relleno 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 543 de relleno de ruido en línea, 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 562 de relleno 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 543 de entrada en la línea y la característica armónica o de expresión de los datos de audio según lo definido por el analizador 516 de señales 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 542 sobre la distorsión temporal 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.

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, posee 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.

En la primera realización, se aplica un nivel de ruido normal, cuando la señal no posee 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.

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 602 del nivel de relleno de ruido de la figura 6a reducirá el nivel de ruido manipulado a cero o al menos a un valor menor que el valor bajo indicado en la figura 6b. Preferiblemente, el clasificador de señales adicionalmente posee 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.

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.

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.

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

5 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 posee 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 circuito de tránsito 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 posee 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 de expresión /no expresión 520b. 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 de ventanas 502 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 circuitos de tránsito para detectar un circuito de tránsito 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 circuito de tránsito y el analizador de señales de audio no encuentra una característica armónica o de expresión. Cuando, sin embargo, se detecta un circuito de tránsito 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 circuito de tránsito y una ventana corta cuando el detector de circuitos de tránsito detecta un circuito de tránsito 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 circuitos de tránsito 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 posee una primera porción 714 de superposición, una porción 714b de no aliasing (generación de señal ajena), una segunda 714c porción de superposición más corta y una porción cero que se extiende entre el punto 716 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 de inicio larga 718 que posee una porción 718a larga de superposición que se superpone con la siguiente ventana larga no ilustrada en la figura 7d. Además, esta ventana posee 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 preecos que aparezcan dentro de un marco antes del evento del circuito de tránsito que es la posición de inicio sonoro o, generalmente, el comienzo de la expresión o el

comienzo de una señal que posee un contenido armónico. Generalmente, una señal posee un contenido armónico, cuando un rastreador de tonos decide que la señal posee 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 poseen un fuerte contenido armónico, se desea una buena resolución de frecuencia. Por lo tanto, el analizador de señales de audio ilustrado en 516, 520 o 520a, 520b es operativo para producir una señal desactivada para el detector 700 de circuitos de tránsito de manera que se evita un cambio a ventanas cortas cuando se detecta un segmento de expresión sonora o un segmento de señales que posee 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 preecos 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 preeco que pudiera aparecer. Con el fin de disminuir además los preecos, 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 circuitos de tránsito 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 circuitos de tránsito se controla utilizando una señal 704 de control de umbral. En esta realización, el detector 700 de circuitos de tránsito 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 circuito de tránsito cuando la característica cuantitativa posee 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 circuito de tránsito, 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 posee una longitud más corta que la función de ventana larga de manera que se obtiene una buena composición entre los preecos por un lado y una resolución de frecuencia suficiente por otro lado. En una realización alternativa, se puede realizar un cambio a una función de ventana que posee una superposición más pequeña según lo indicado por la línea sombreada en la figura 7e en 706. La función 706 de ventana posee una longitud 2048 de muestras como el bloque largo, pero esta ventana posee 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 707 de ventana, nuevamente, posee una porción 712 cero a la izquierda de la región y una porción 712 de no aliasing a la derecha de la región 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 preecos debido a la porción 706 y 707 cero de ventana, pero por otro lado posee 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 posee 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.

5 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
10 una señal sonora detectada en 751, donde la señal de audio utilizada para análisis puede ser la señal de audio en la entrada 500 en la figura 5a o una señal de audio preprocesada tal como una señal de audio distorsionada en el tiempo o una señal de audio que haya sido sometida a cualquier otra funcionalidad de preprocesamiento. Preferiblemente, en selector 504 de formas de ventanas en la figura 7c que está incluido en el controlador 504 de funciones de ventanas en la figura 5a solamente utiliza la señal 751, cuando un detector de circuitos de tránsito
15 incluido en el controlador de funciones de ventanas detecta un circuito de tránsito 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)
20 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
25 activar TNS, p. ej. para reducir los preecos 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, p. ej. 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
30 contorno de tonos activo, se 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.

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
35 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
40 cualquier otra clase de transformada tal como una transformada DCT, DST, DFT, FFT o MDST o puede comprender un banco de filtros tal como un banco de filtros QMF.

Adicionalmente, el codificador comprende una etapa 510 de 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
45 temporal, en donde no se realiza el filtrado de predicción, cuando no existe la orden de control del formato de ruido temporal.

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.
50

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.
55 Los detalles del controlador de formato de ruido temporal se debaten con relación a la figura 8.

El codificador de audio adicionalmente comprende un procesador para además procesar un resultado del filtrado de predicción sobre la frecuencia para obtener la señal codificada. En una realización, el procesador comprende la etapa 512 del codificador del cuantificador ilustrada en la figura 5a.
60

La 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 804 de señal de control de umbral. Dependiendo de la señal 516 del analizador 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 posee 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 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.

5
10
15
20
25

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 secundaria de TNS indicada en 510a 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 del clasificador de señales del bloque 520, 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 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 la 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.

30
35

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.

40
45
50

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 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, poseen 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 poseen 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.

60

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 posee 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 del cuantificador/codificador 512 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 de banda 900 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 sombreada 507. 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.

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.

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.

El ancho de banda que resulta en el número de líneas N_N y la tasa de muestreo f_N preferiblemente define la frecuencia 1200 de cruce para un codificador de audio que, además de un codificador de audio central fuente, posee 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 1200 de cruce 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 1200 de cruce. 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 posee una frecuencia máxima tan cercana como sea posible a la frecuencia 1200 de cruce. 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.

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 512 del bloque, o puede realizarse después de cuantificar o también puede realizarse, dependiendo del código de entropía específico,

después de la codificación de entropía.

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
 5 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.

Aunque se han descrito algunos aspectos en el contexto de un equipo, está claro que estos aspectos también
 10 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.

Dependiendo de ciertos requerimientos de implementación, las realizaciones de la invención pueden implementarse
 15 en hardware o en software. La implementación puede realizarse utilizando un medio de almacenamiento digital, por ejemplo un floppy disk, un DVD, un CD, un ROM, un PROM, un EPROM, un EEPROM o una memoria FLASH, 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
 20 realizaciones conforme a la invención comprenden un soporte de datos que posee 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
 25 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 posee un código de programa para llevar a cabo uno de los métodos descritos en la presente, cuando el programa informático funciona en un ordenador. Una realización
 30 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
 35 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 un ordenador, 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 un ordenador que posee el programa informático instalado en la mismo para llevar a cabo uno de los métodos descritos en la presente. En algunas realizaciones, puede utilizarse un
 40 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 la presente.

45

REIVINDICACIONES

1. Codificador de audio para codificar una señal de audio, que comprende:
 - 5 un distorsionador (506) de tiempo;
 - un convertidor (508) de tiempo-frecuencia para llevar a cabo una conversión de tiempo/frecuencia de una señal de audio distorsionada en tiempo en una representación espectral;
 - 10 un cuantificador (512) para cuantificar valores de audio, en donde el cuantificador está configurado para cuantificar a valores de audio cero por debajo de un umbral de cuantificación;
 - un calculador (524) de relleno de ruido para estimar una medida de una energía de valores de audio cuantificados a cero durante un marco de tiempo de la señal de audio para obtener una medida de relleno de ruido;
 - 15 un analizador (516) de señal de audio para analizar si el marco de tiempo de la señal de audio tiene una característica armónica o de expresión;
 - 20 en donde el analizador (516) de señal de audio provee una información de la distorsión de tiempo en la salida (518), que está conectada al distorsionador (506) de tiempo;
 - un manipulador (602) para manipular la medida de relleno de ruido dependiendo de una característica armónica o de expresión de la señal de audio para obtener una medida de relleno de ruido manipulada; y
 - 25 una interfaz (522) de salida para generar una señal codificada para transmisión o almacenamiento, la señal codificada comprendiendo la medida (530) de relleno de ruido manipulada;
 - 30 en donde el manipulador (602) está configurado para aplicar un nivel de ruido normal cuando la señal no tiene una característica armónica o de expresión y cuando no se aplica distorsión de tiempo, y para manipular el nivel de relleno de ruido para que sea inferior que en el caso normal cuando un contorno de tono se encontró por el analizador de señal de audio, que indica un contenido armónico, y la distorsión de tiempo está activa.
- 35 2. Codificador de audio de acuerdo con la reivindicación 1,
 - en el cual el analizador (516) de señal de audio comprende un accionador de tono para generar una indicación de un tono, donde un tono se encuentra en el marco de tiempo de la señal de audio, y
 - 40 en el cual el manipulador (602) está configurado para reducir la medida de relleno de ruido, cuando se encuentra un tono.
3. Decodificador para decodificar una señal de audio codificada que comprende:
 - 45 una interfaz (539) de entrada para procesar la señal de audio codificada para obtener una medida (543) de relleno de ruido y datos (546) de audio codificados;
 - un decodificador/re-cuantificador (547, 550) para generar datos re-cuantificados;
 - 50 una analizador (600) de señal para recuperar información, si un marco de tiempo de los datos de audio tiene característica armónica o de expresión; y
 - un relleno (552) de ruido para generar datos de audio de relleno de ruido,
 - 55 donde el relleno (552) de ruido está configurado para generar datos de relleno de ruido en respuesta a la medida de relleno de ruido y a la característica armónica o de expresión de los datos de audio; y
 - un procesador (556, 558, 560) para procesar los datos re-cuantificados y los datos de audio de relleno de ruido para obtener una señal (564) de audio decodificada;
 - 60 en donde la señal de audio codificada comprende datos (542, 541) que indican, si el marco de tiempo de los datos de audio data tiene una característica armónica o de expresión, y
 - en donde el analizador (600) de señal está configurado para analizar la señal de audio codificada para

- recuperar datos que indican si el marco de tiempo de los datos de audio tiene una característica armónica o de expresión,
- 5 en donde los datos son una indicación de que la porción de tiempo ha sido sometida a un procesamiento de distorsión de tiempo, y
- en donde el procesador comprende un regenerador (558) de tiempo para regenerar en tiempo una señal de audio derivada de los datos de relleno de ruido y los datos recuantificados.
- 10 4. Método para codificar una señal de audio, que comprende:
- distorsionar (506) en tiempo una señal de audio;
- 15 realizar (508) una conversión de tiempo/frecuencia de una señal de audio distorsionada en tiempo en una representación espectral;
- cuantificar (512) valores de audio, donde los valores debajo de un umbral de cuantificación son cuantificados a cero;
- 20 estimar (524) una medida de una energía de valores de audio cuantificados a cero para un marco de tiempo de la señal de audio para obtener una medida de relleno de ruido;
- analizar (516) si el marco de tiempo de la señal de audio tiene una característica armónica o de expresión;
- 25 proveer una información de distorsión del tiempo que se usa por la distorsión del tiempo;
- manipular (602) la medida de relleno de ruido dependiendo de una característica armónica o de expresión de la señal de audio para obtener una medida de relleno de ruido manipulada de modo que se aplique un nivel de ruido normal cuando la señal no tenga una característica armónica o de expresión y cuando no se aplique distorsión de tiempo, y de modo que el nivel de relleno de ruido se manipule para que sea inferior que en el caso normal cuando se encuentre un contorno de tono, lo que indica un contenido armónico, y la distorsión de tiempo esté activa; y
- 30 generar (522) una señal codificada para transmisión o almacenamiento, la señal codificada comprendiendo la medida (530) de relleno de ruido manipulada.
- 35 5. Método para decodificar una señal de audio codificada, en donde la señal de audio codificada comprende datos (542, 541) que indican si el marco de tiempo de los datos de audio tiene una característica armónica o de expresión, que comprende:
- 40 procesar (539) la señal de audio codificada para obtener una medida (543) de relleno de ruido y datos (546) de audio codificados;
- 45 analizar la señal de audio codificada para recuperar datos que indiquen si el marco de tiempo de los datos de audio tiene una característica armónica o de expresión, en donde los datos son una indicación de que la porción de tiempo ha sido sometida a un procesamiento de distorsión de tiempo;
- generar (547, 550) datos recuantificados;
- 50 recuperar (600) información sobre si un marco de tiempo de los datos de audio tiene una característica armónica o de expresión; y
- 55 generar (552) datos de audio de relleno de tiempo en respuesta a la medida de relleno de ruido y la característica armónica o de expresión de los datos de audio; y
- procesar (556, 558, 560) los datos recuantificados y los datos de audio de relleno de ruido para obtener una señal (564) de audio decodificada, en donde el procesamiento comprende regenerar en tiempo una señal de audio derivada de los datos de relleno de ruido y datos recuantificados.
- 60 6. Programa informático que tiene un código de programa para llevar a cabo, al ponerse en marcha en un ordenador, el método de la reivindicación 4 o el método de la reivindicación 5.

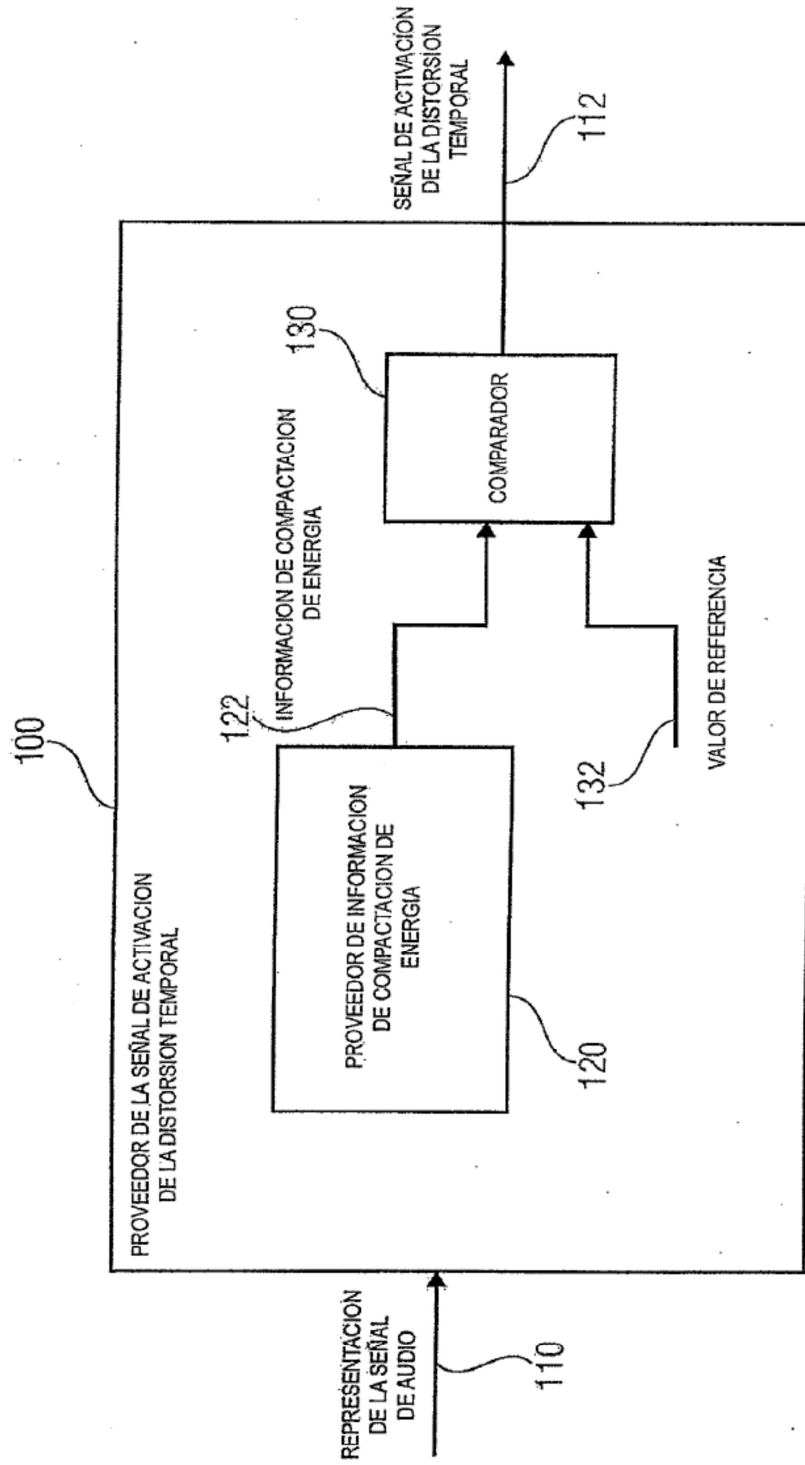


FIG 1

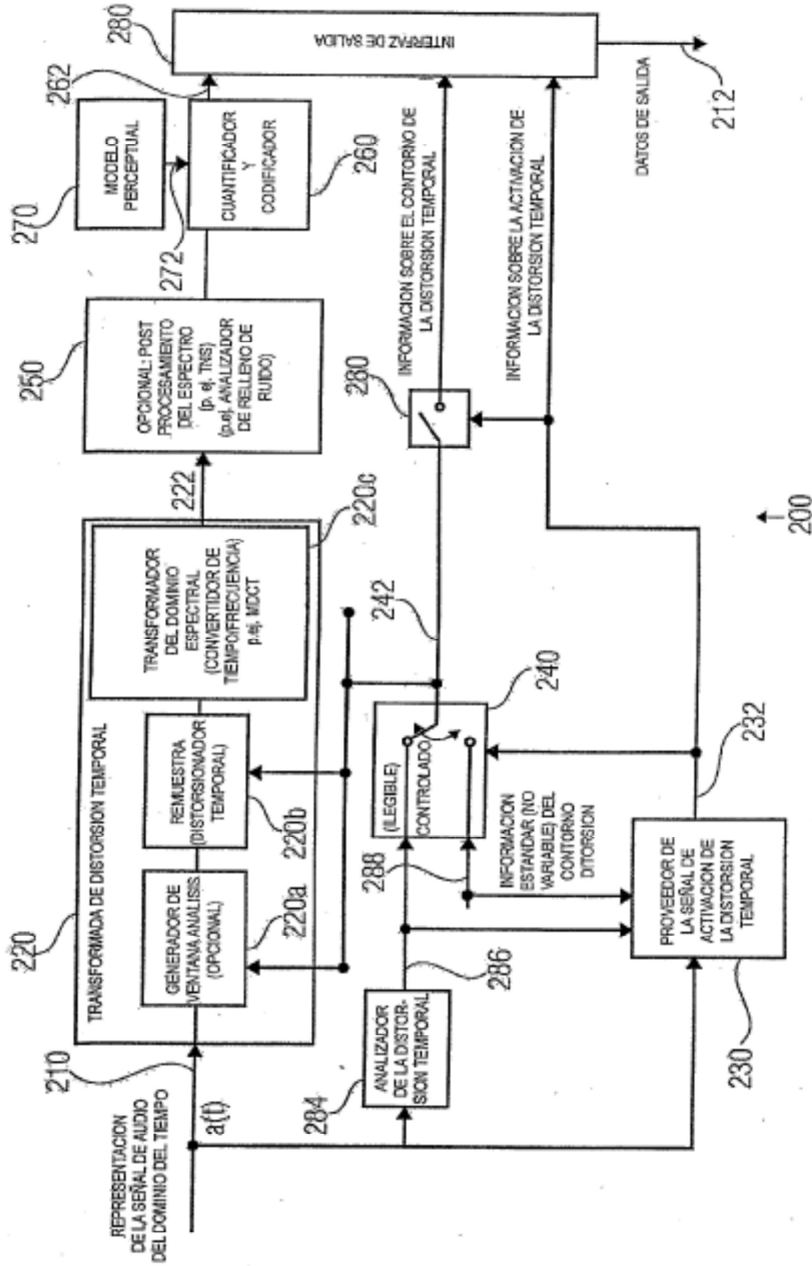


FIG 2A

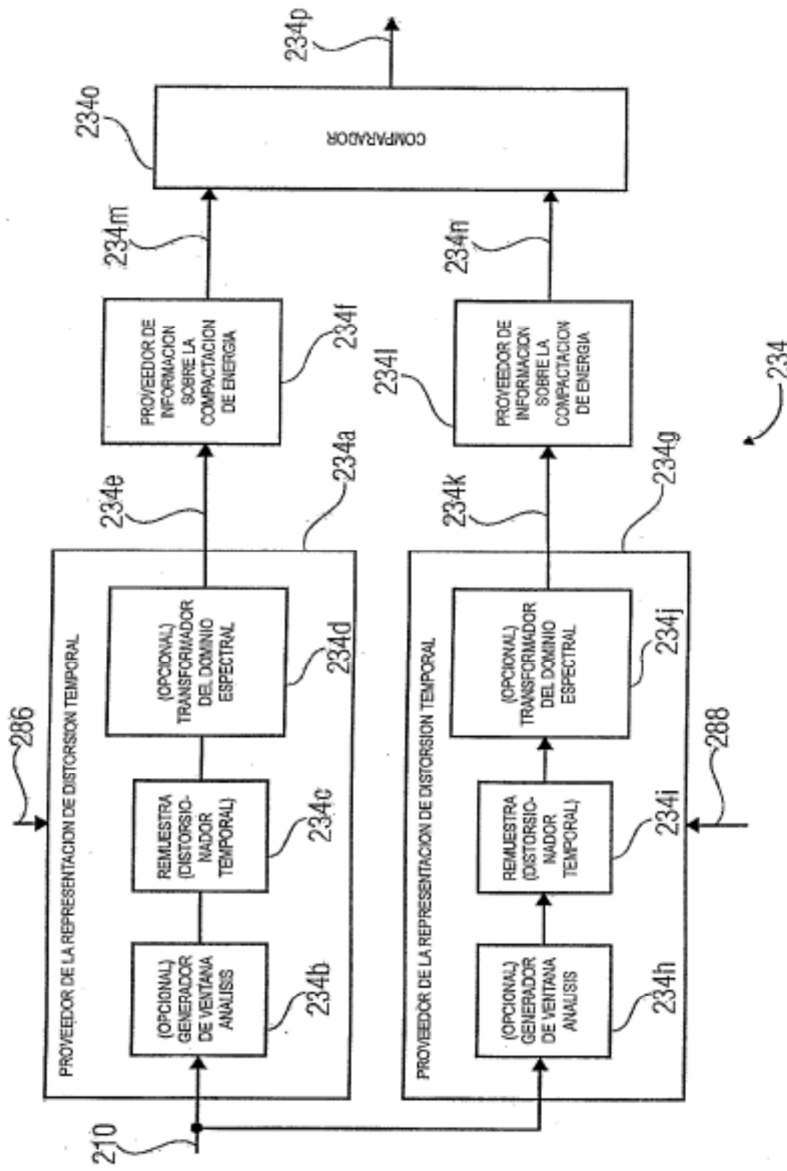
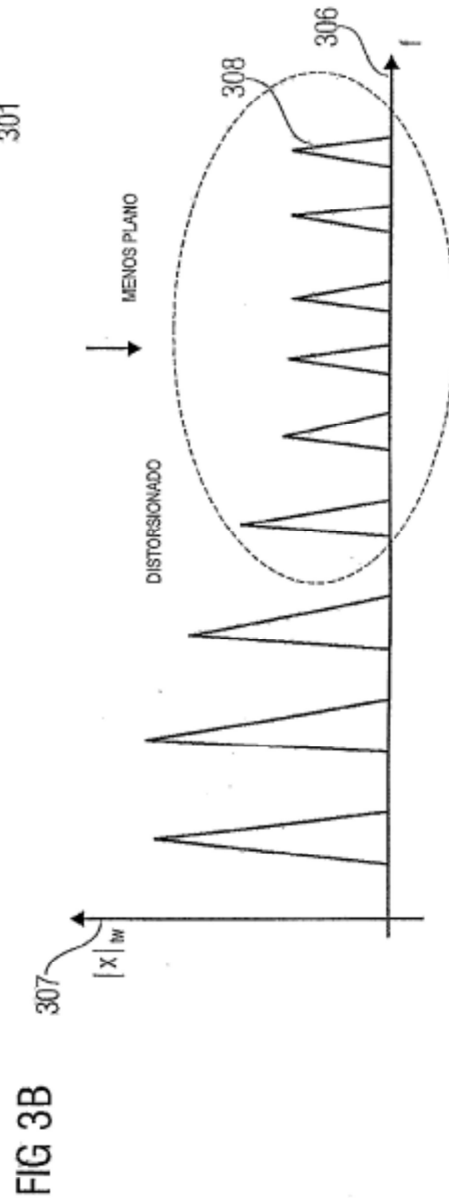
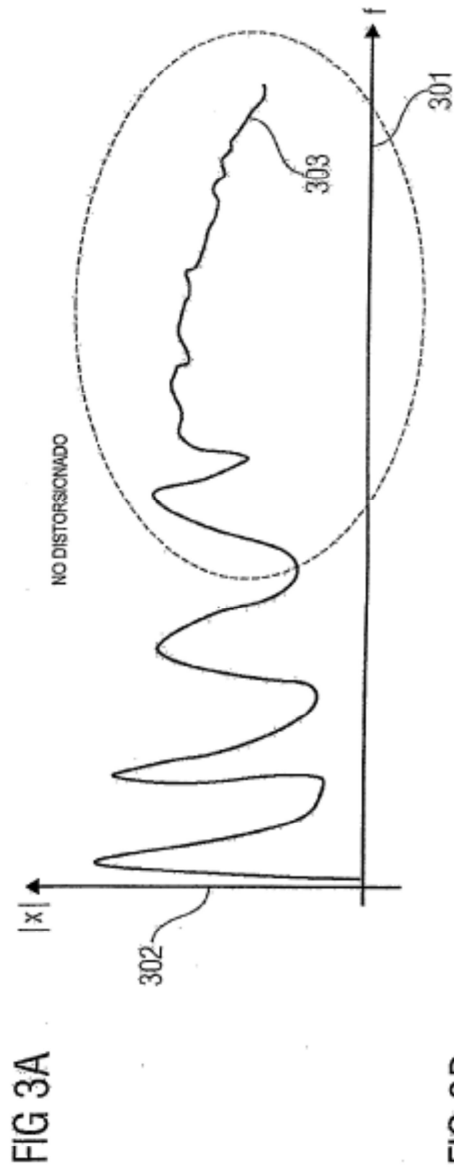


FIG 2B



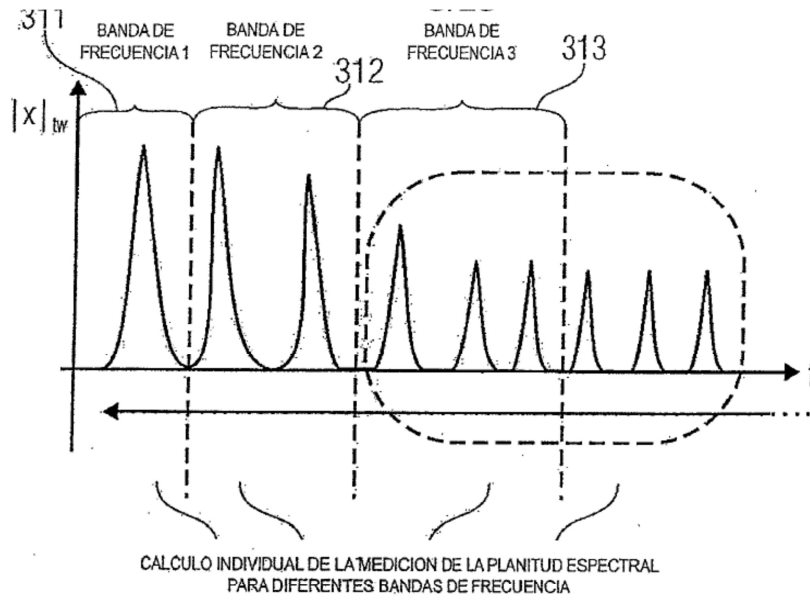
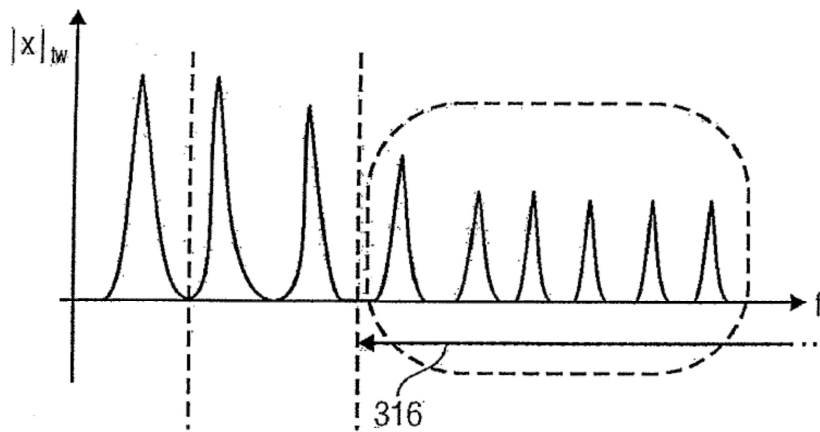


FIG 3C



CONSIDERACION DE LA PORCION DE FRECUENCIA MAS ALTA SOLAMENTE (MIENTRAS SE DESATIENDE LA PORCION DE FRECUENCIA MAS BAJA) PARA EL CALCULO DE LA MEDICION DE PLANITUD ESPECTRAL (ANCHO DE BANDA DE FRECUENCIA O POR ENCIMA DE LA PORCION DE FRECUENCIA MAS ALTA ENTERA)

FIG 3D

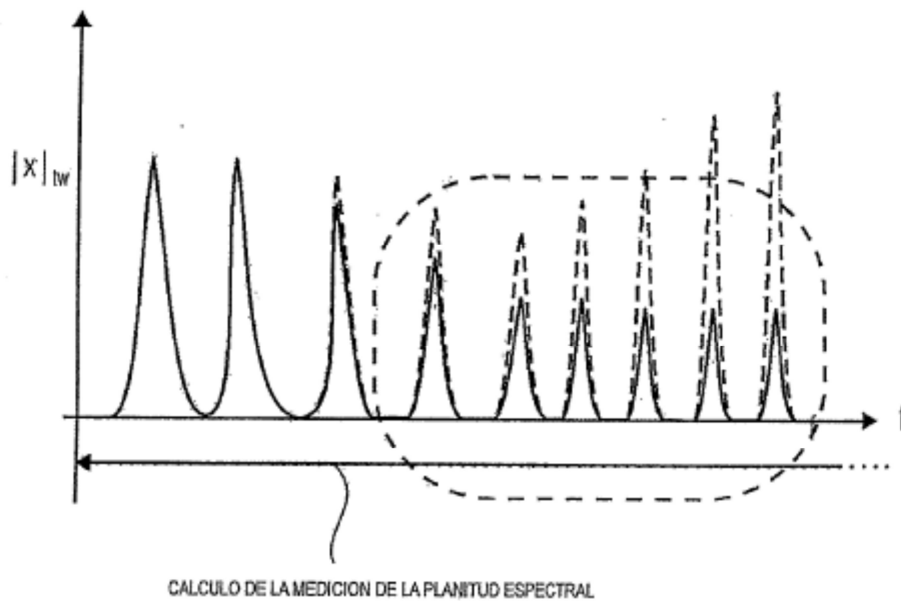


FIG 3E

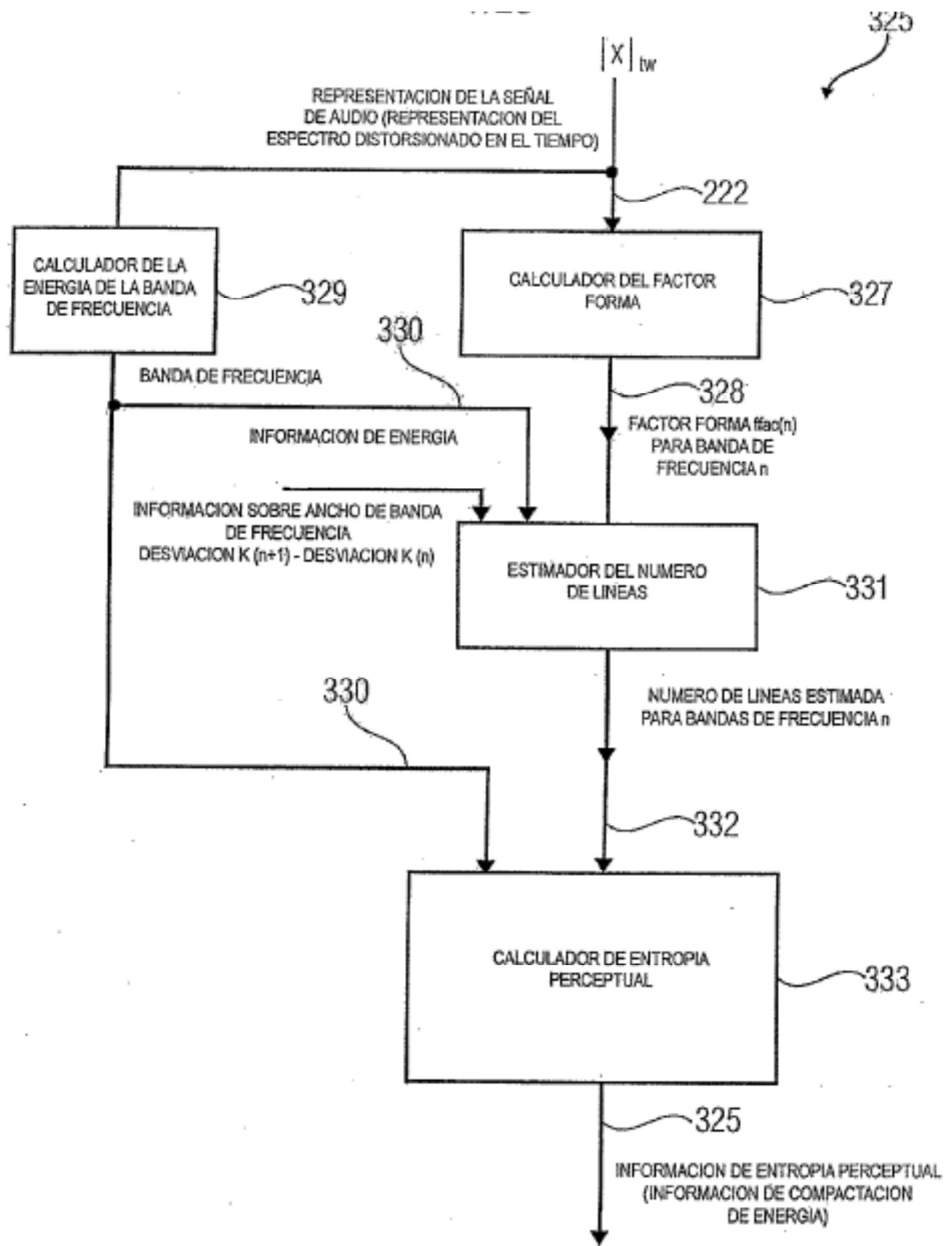


FIG 3F

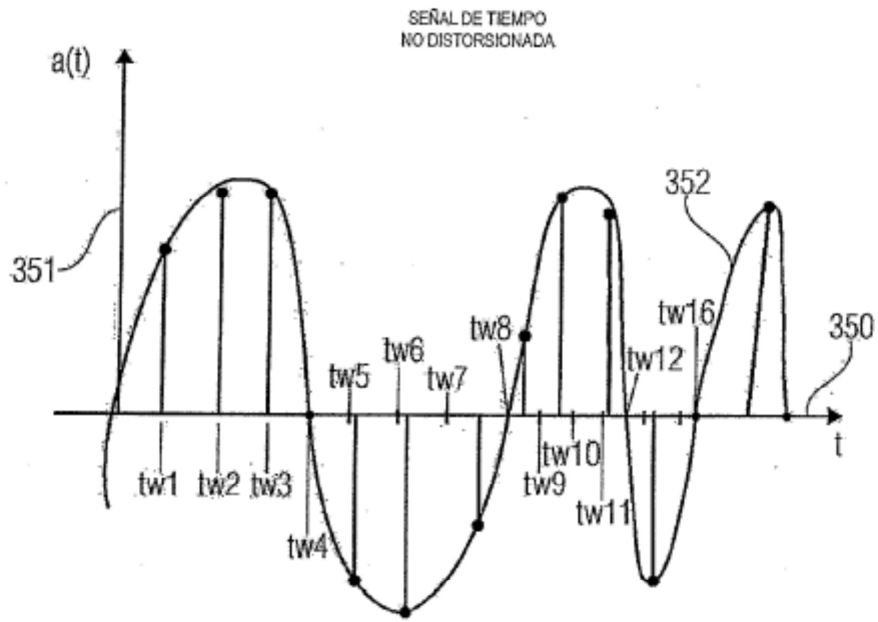


FIG 3G

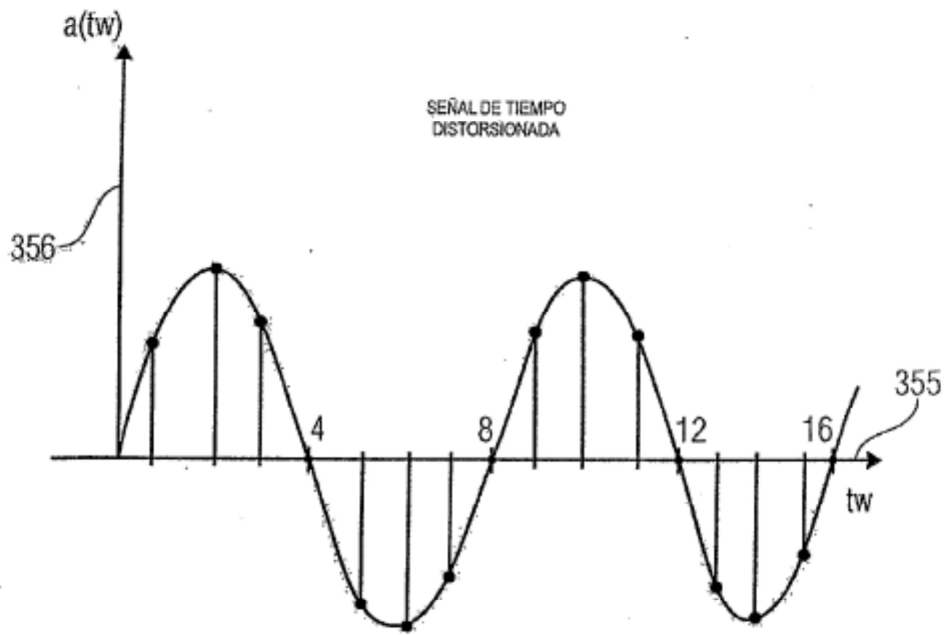


FIG 3H

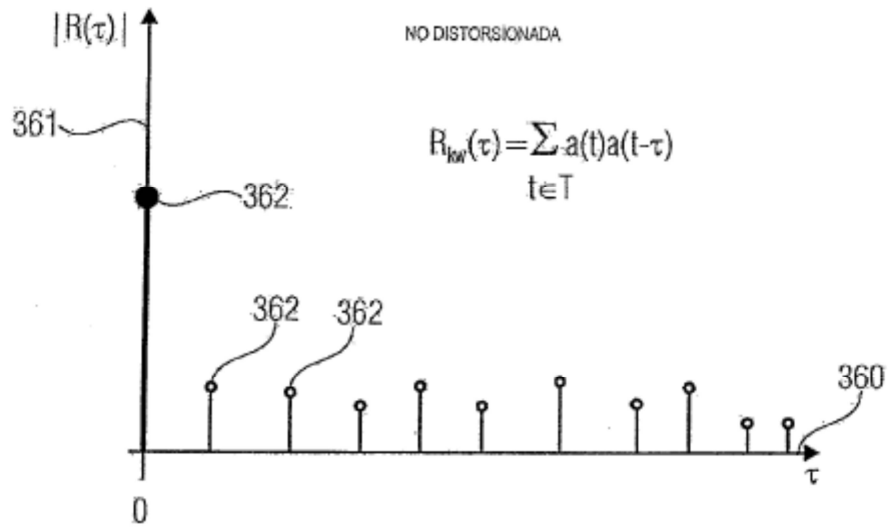


FIG 3I

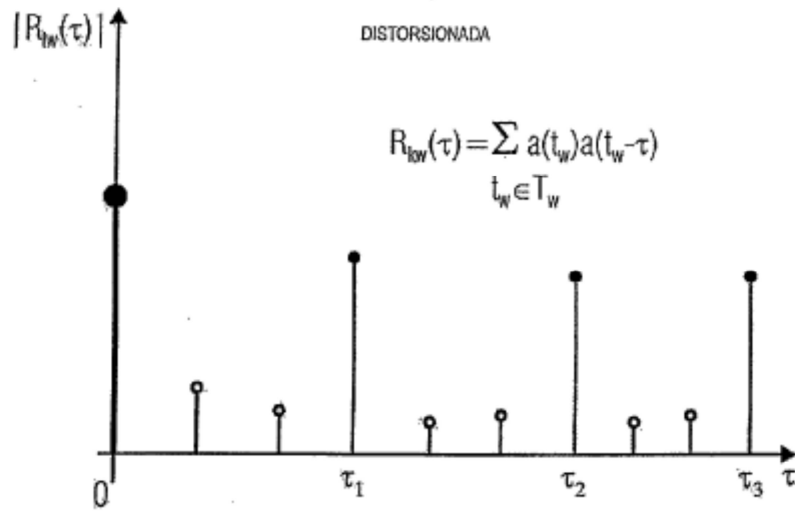


FIG 3J

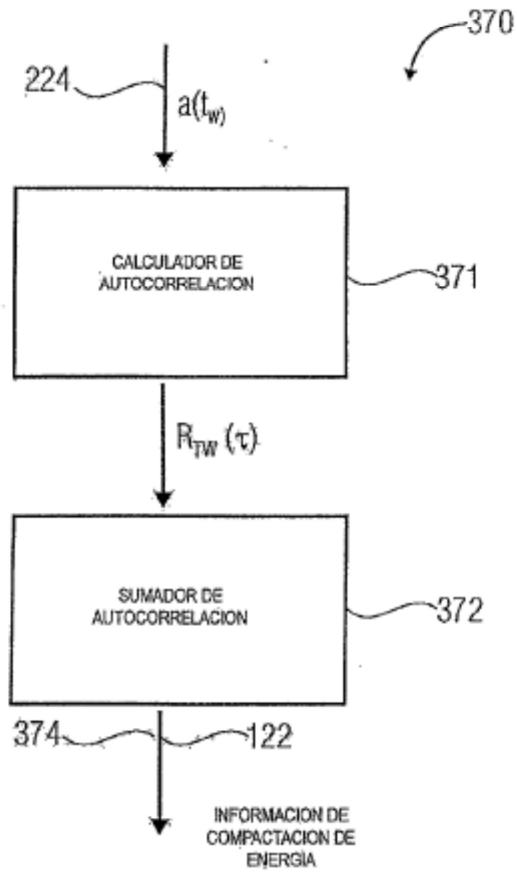


FIG 3K

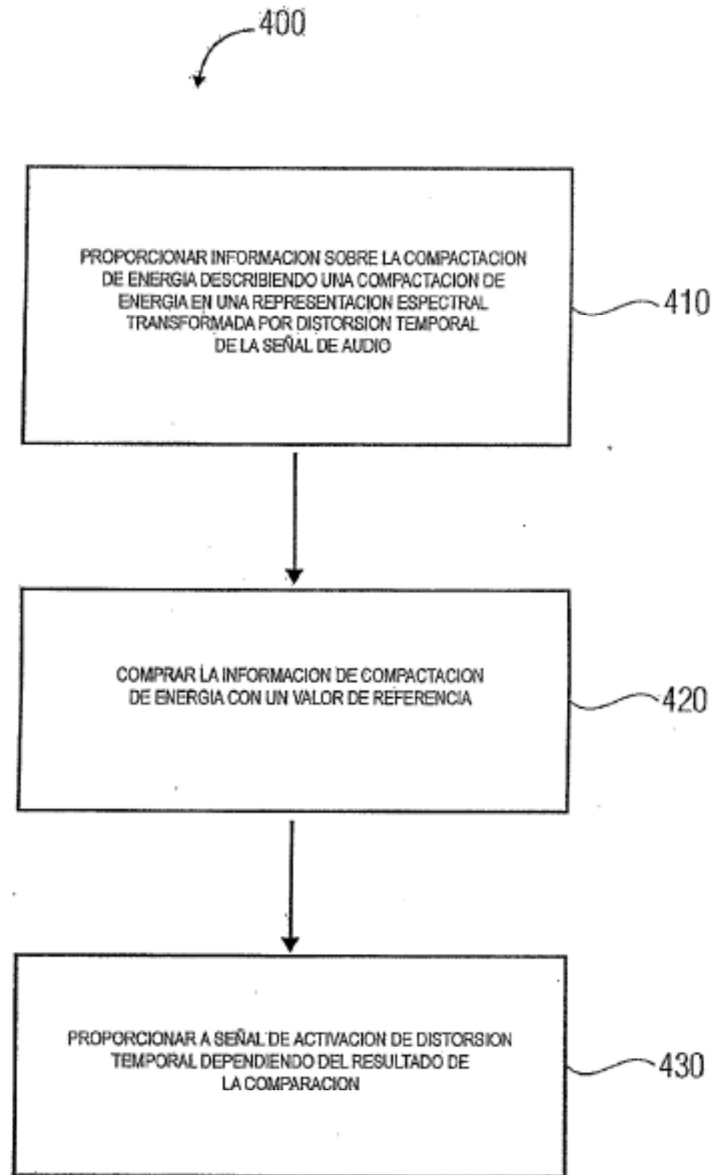


FIG 4A

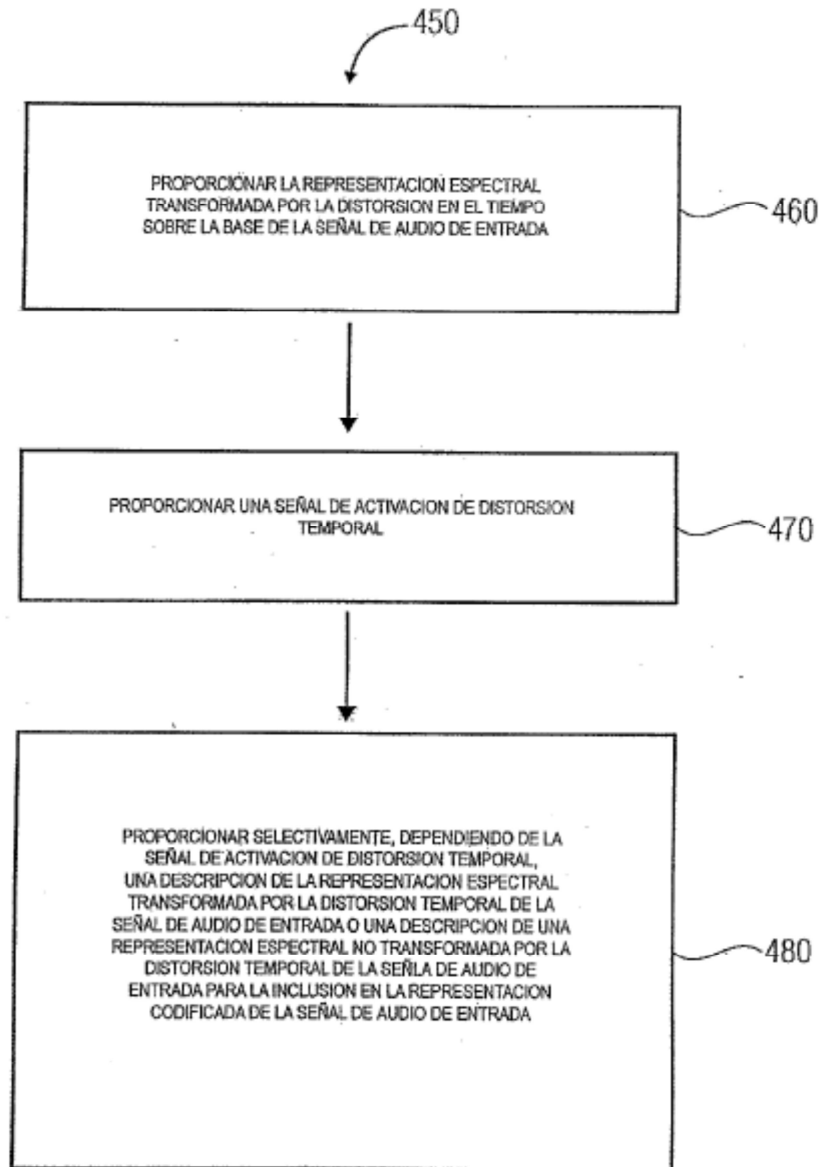


FIG 4B

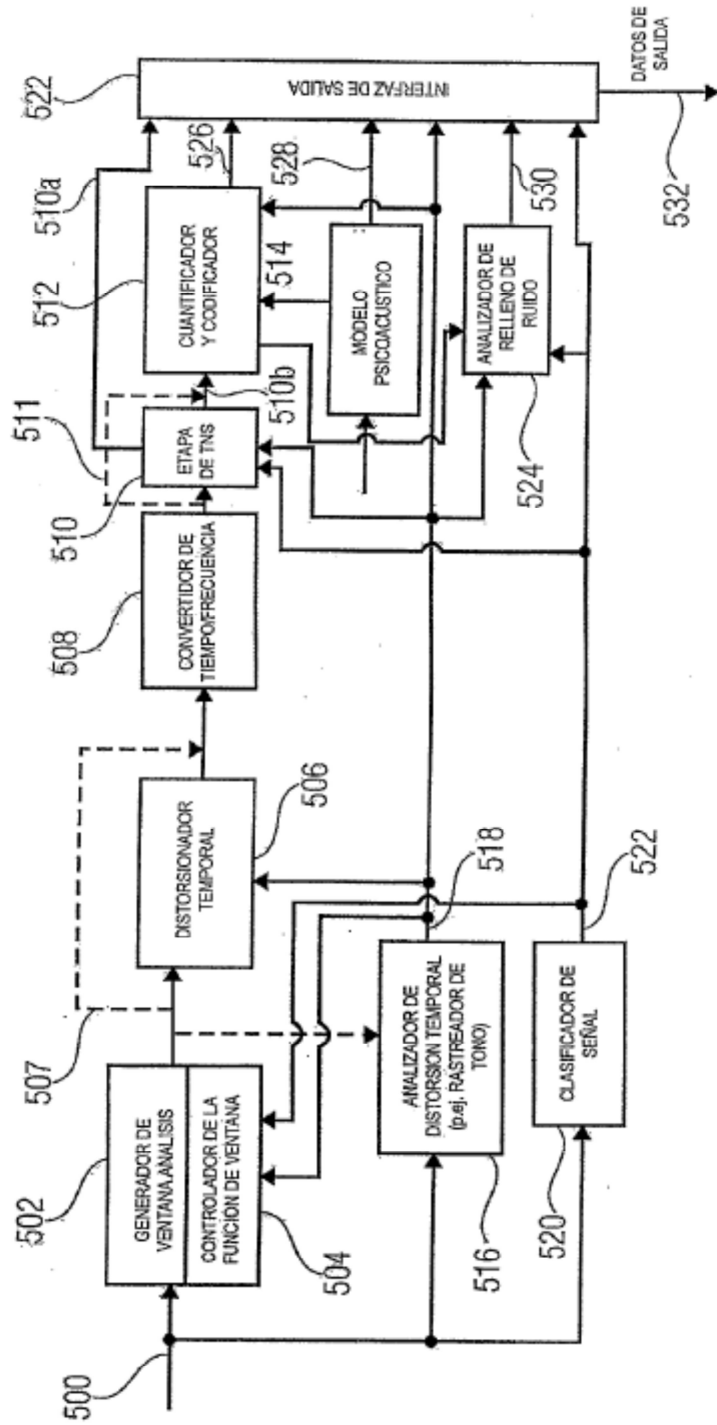


FIG 5A
(CODIFICADOR)

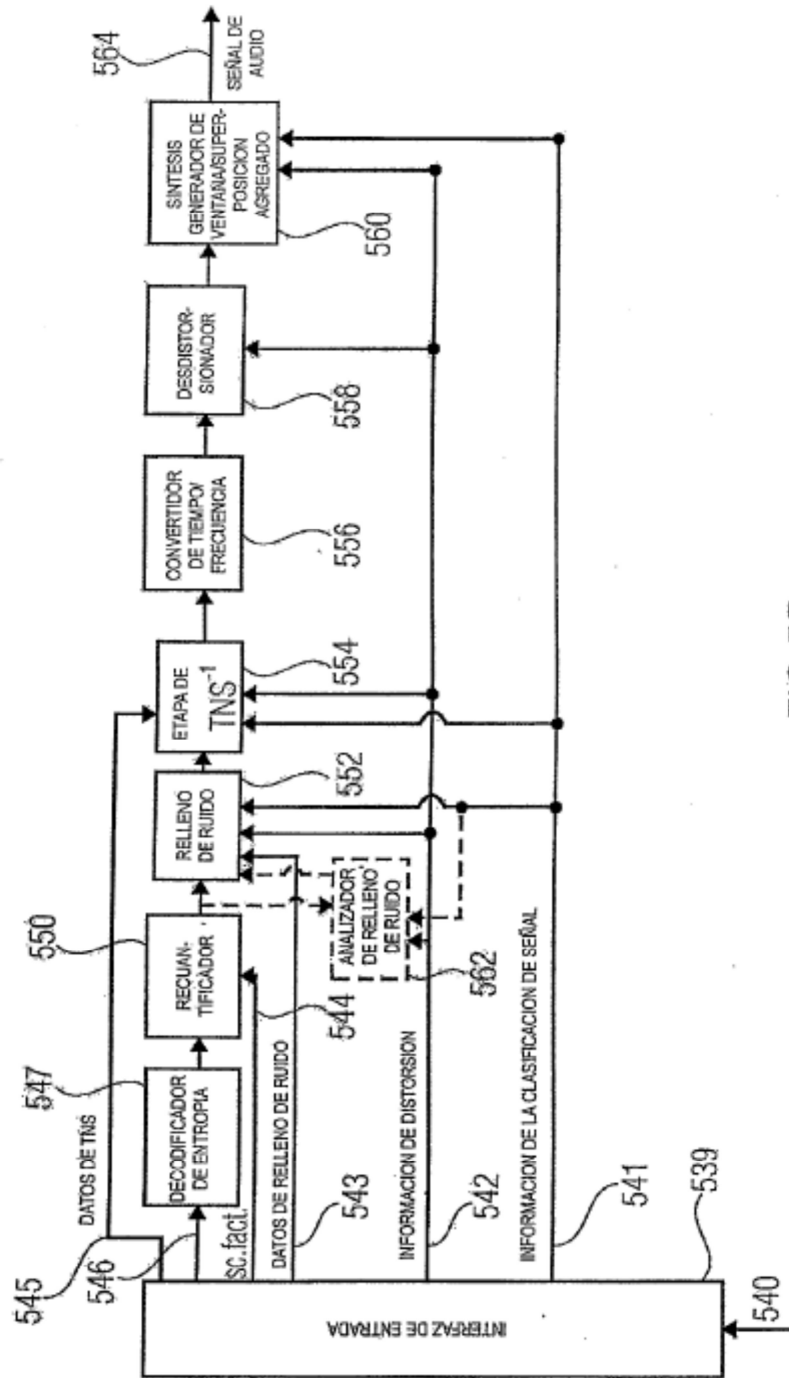


FIG 5B
(DECODIFICADOR)

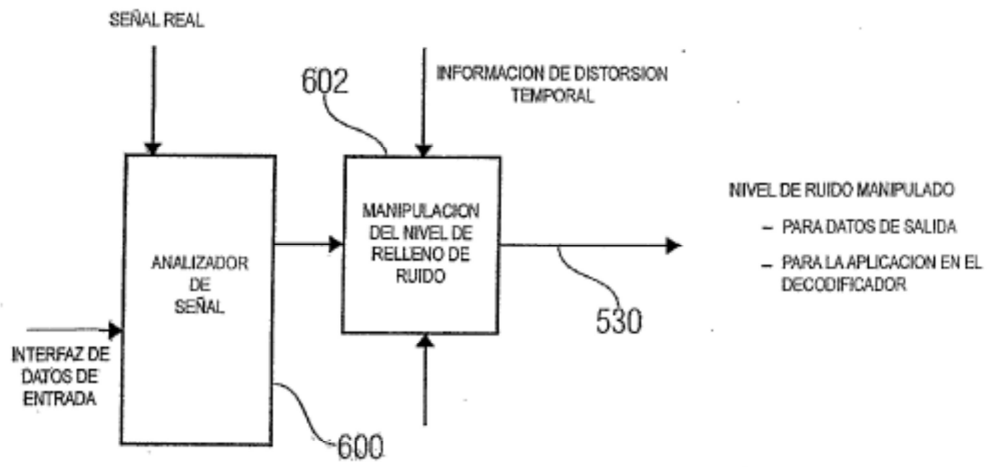


FIG 6A

CLASIFICACION DE EXPRESION		DISTORSION TEMPORAL		CLASIFICACION DE SEÑAL		NIVEL DE RELLENO DE RUIDO
V	UV	SI	NO	EXPRESION	NO EXPRESION	
			X		X	NORMAL
		X			X	BAJO
X		X		X		MUY BAJO / CERO
	X		X	X		NORMAL

FIG 6B

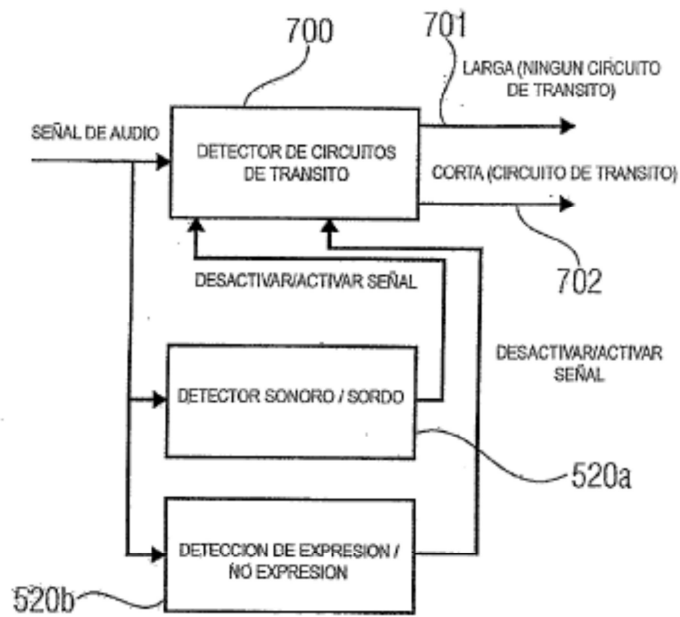


FIG 7A

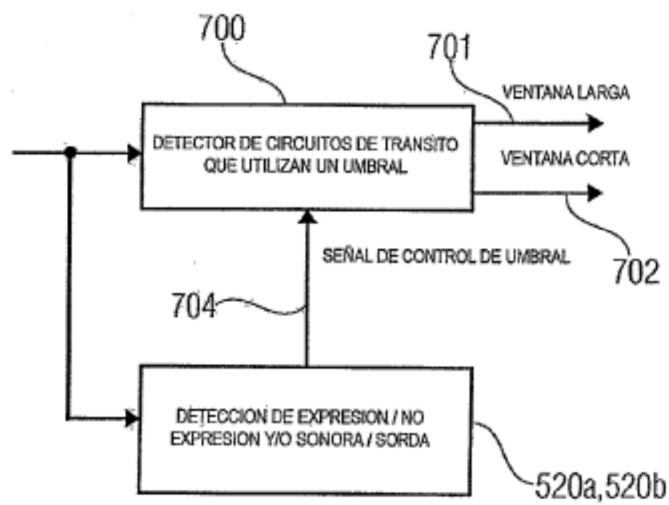


FIG 7B

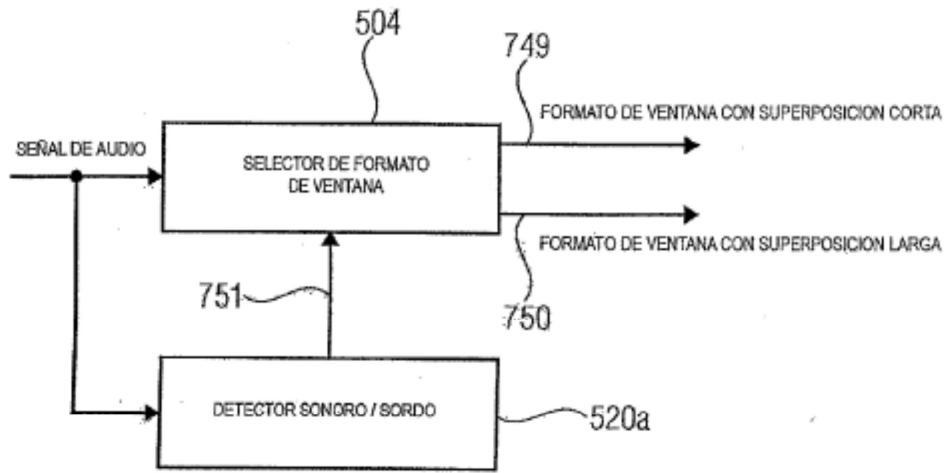


FIG 7C

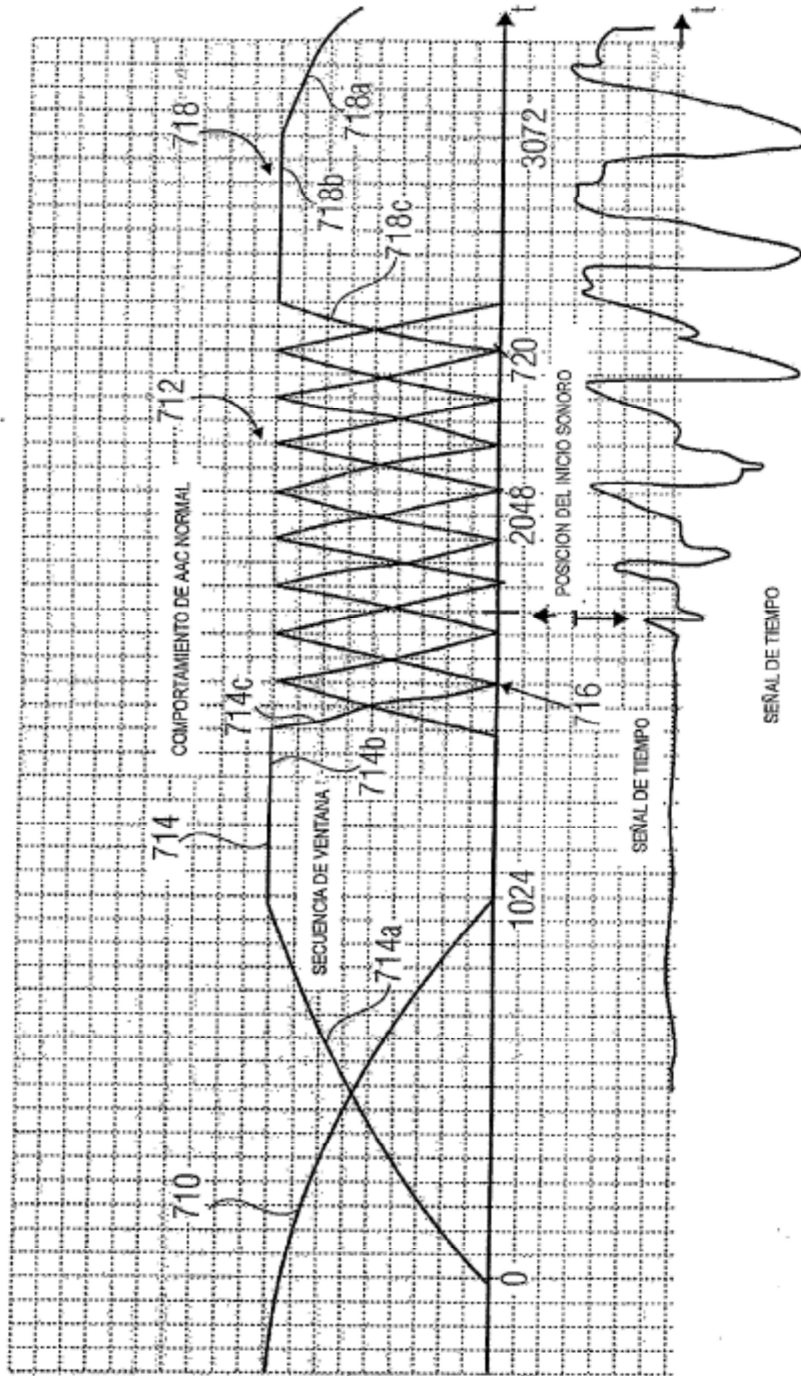


FIG 7D

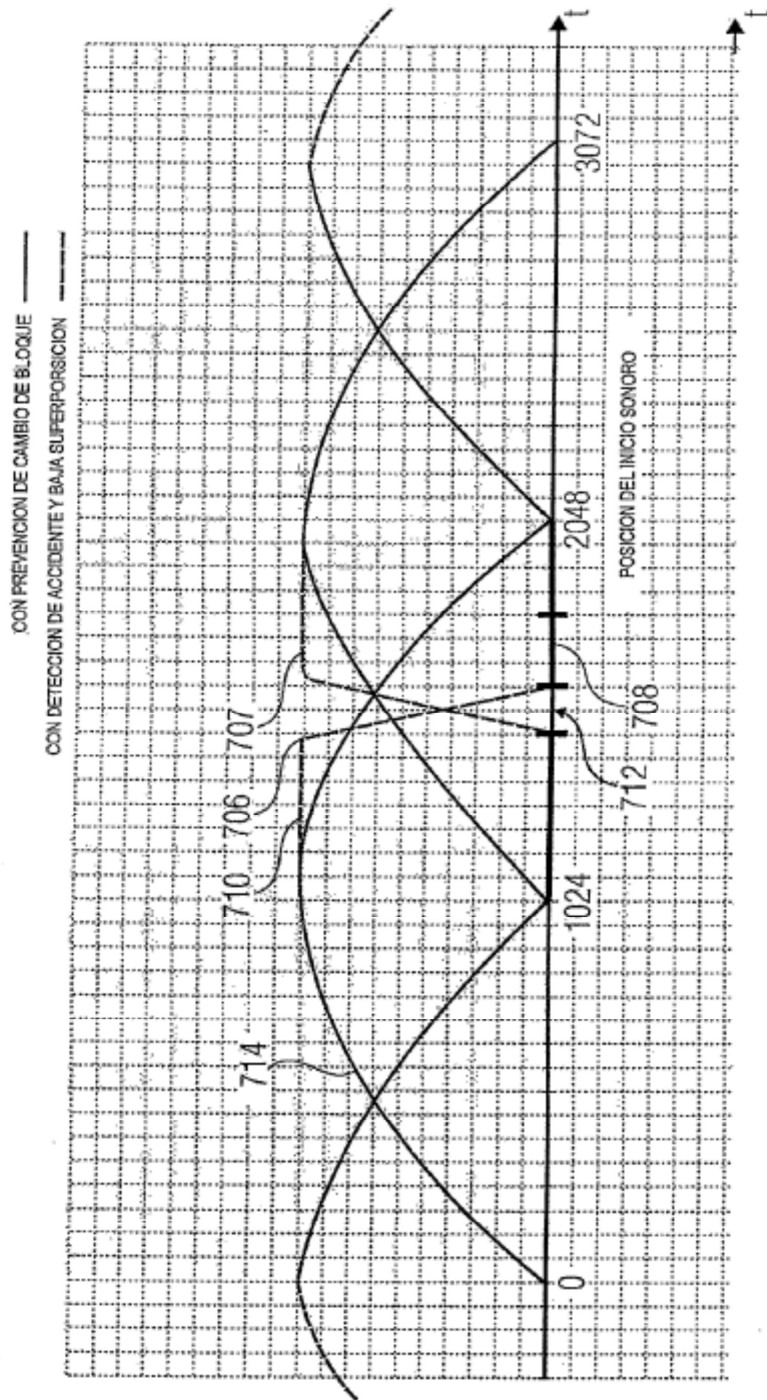


FIG 7E

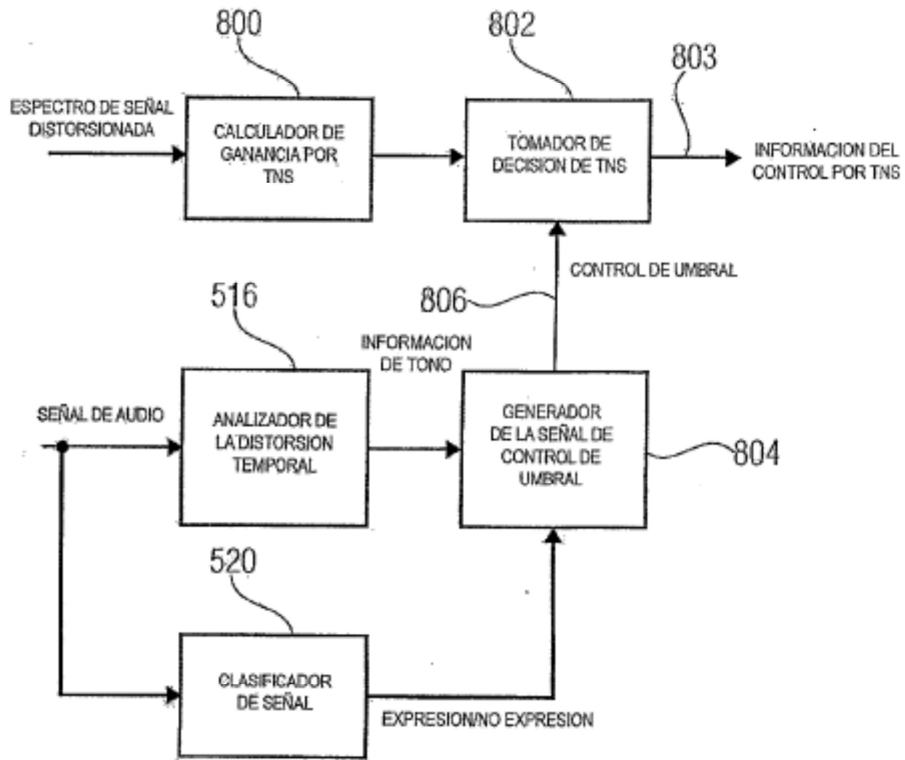


FIG 8A

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

FIG 8B

FIG 9A

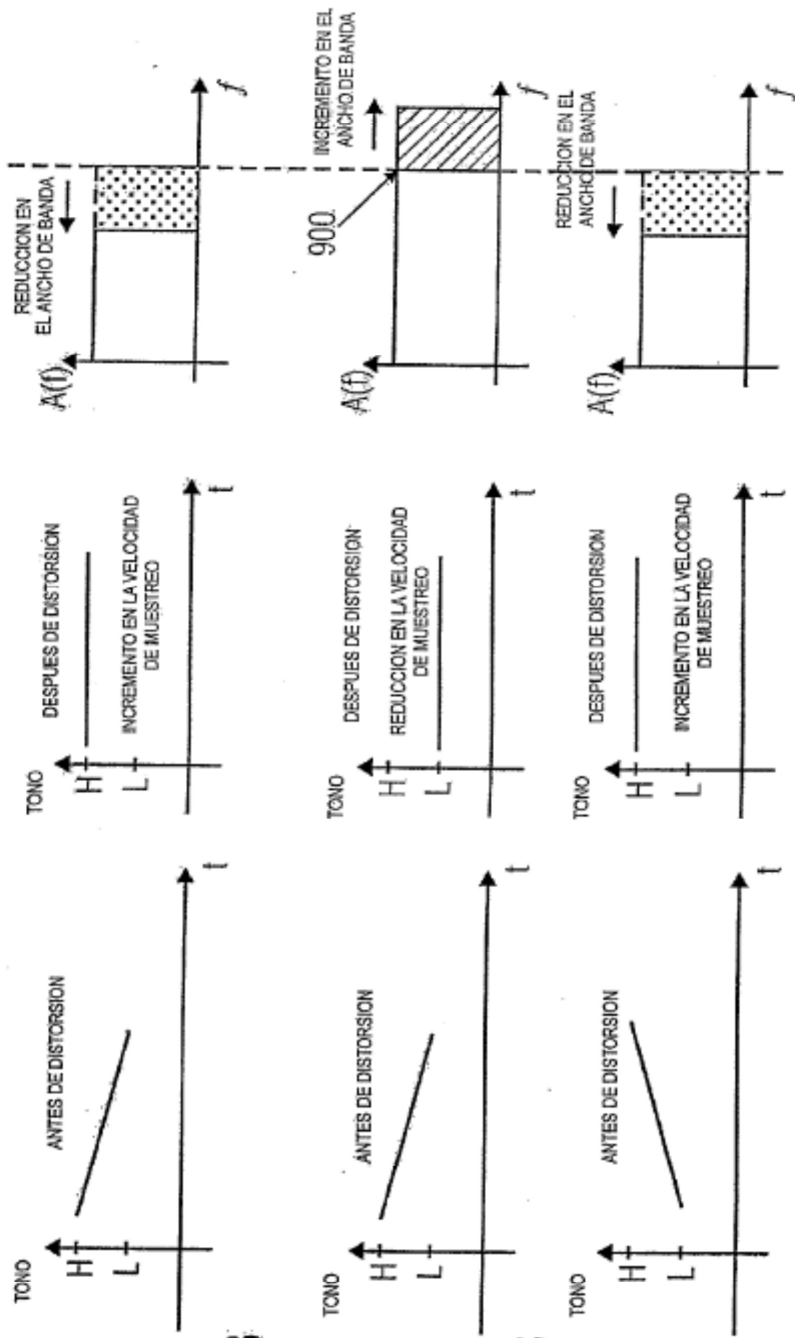
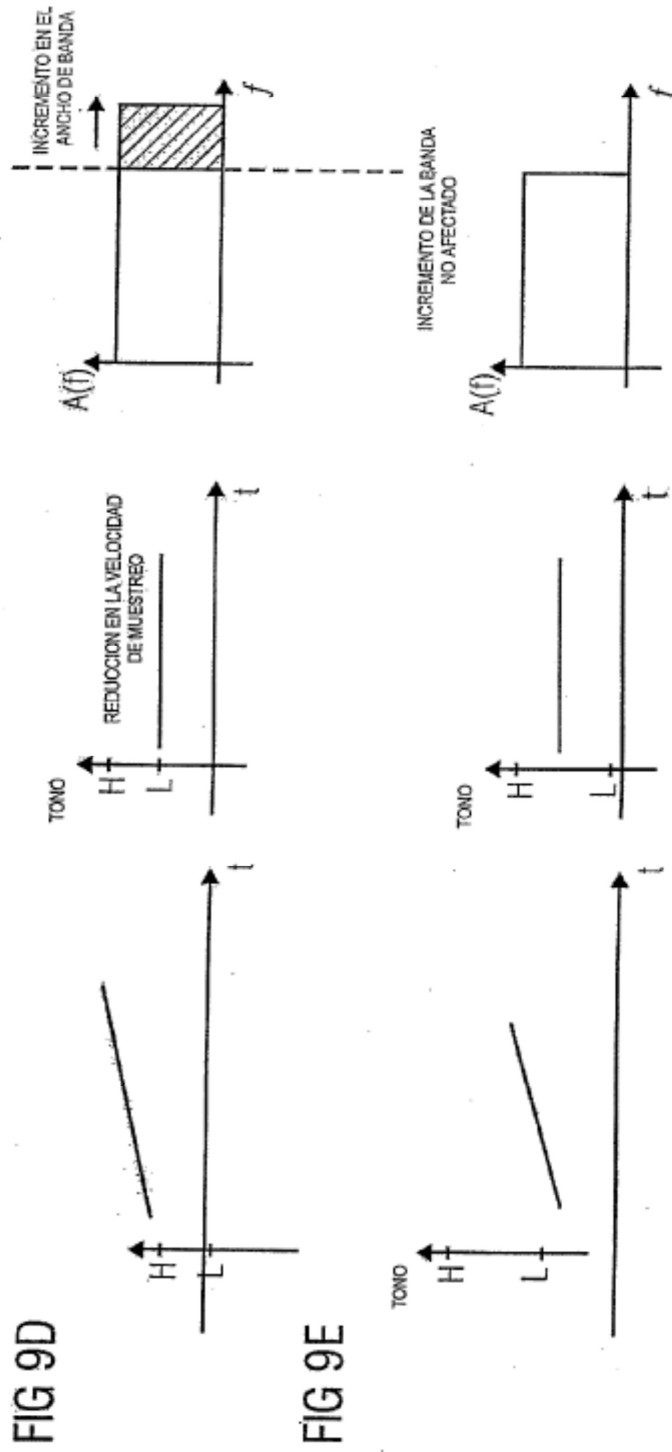


FIG 9B

FIG 9C



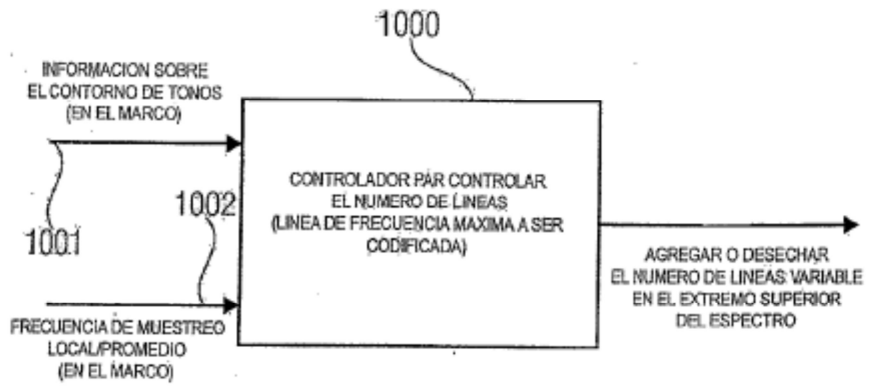


FIG 10A

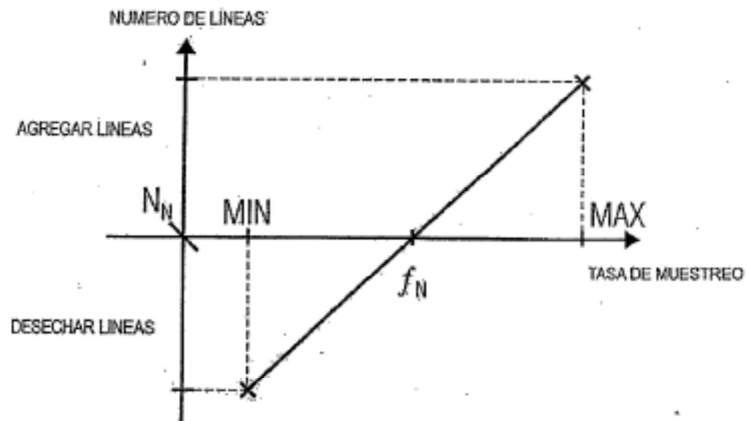


FIG 10B

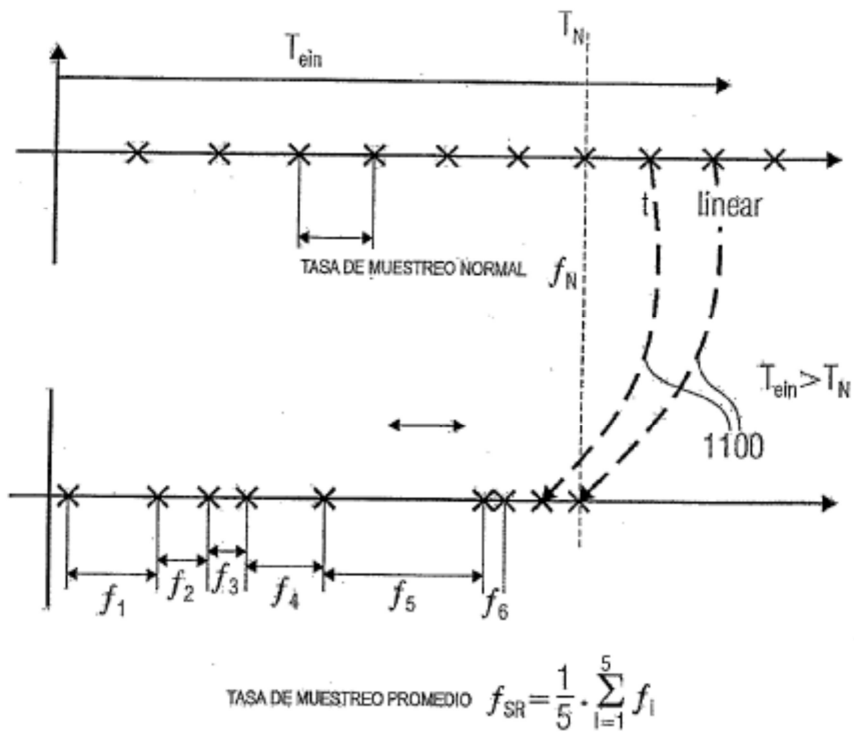


FIG 11

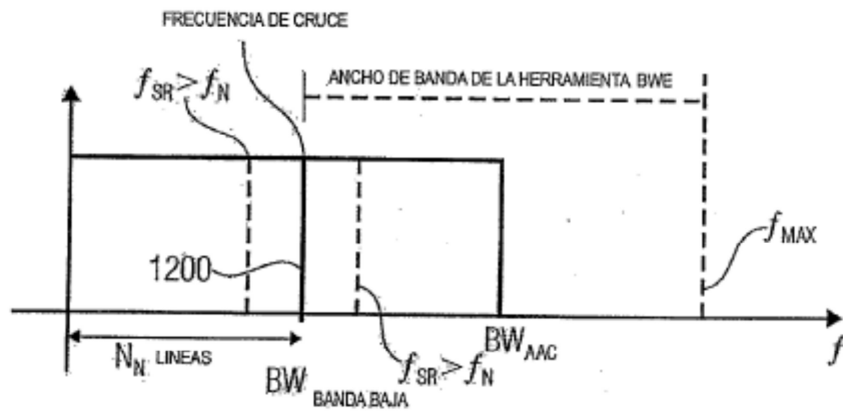


FIG 12A

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

N : NUMERO DE LINEAS SEGUN LO DETERMINADO
 POR EL PROCESADOR

T_{win} : LONGITUD DE MARCO NO DISTORSIONADO
 (DEPENDIENE DE LAS CARACTERISTICAS DE
 DISTORSION)

T_{win}	f_{SR}	Δf	N
$= T_N$	$= f_N$	$= \Delta f_N$	$= N_N$
$> T_N$	$> f_N$	$> \Delta f_N$	$< N_N$ ELIMINAR LINEAS
$< T_N$	$< f_N$	$< \Delta f_N$	$> N_N$ AGREGAR LINEAS

FIG 12B